



Functional genetic elements of a butterfly mimicry supergene

Nicholas W. VanKuren^{a,1} , Sofia I. Sheikh^a , Claire L. Fu^a , Darli Massardo^a , Namiko N. Service^{a,b} , Wei Lu^a, and Marcus R. Kronforst^{a,1}

Edited by Gene Robinson, University of Illinois at Urbana-Champaign Institute for Genomic Biology, Urbana, IL; received April 23, 2025; accepted September 9, 2025

Development requires the coordinated action of many genes across space and time, yet numerous species can develop discrete, alternate phenotypes. Such complex balanced polymorphisms are often controlled by supergenes: multiple tightly linked loci that function together to control development of a complex phenotype. Supergenes are widespread in nature. However, the evolution and functions of supergene alleles remain obscure because the identities of the functional loci, and the causative variation between them, remain essentially unknown. The *doublesex* supergene controls mimicry polymorphism in the swallowtail butterflies *Papilio polytes* and *Papilio alphenor*. Alternate alleles cause development of discrete mimetic or nonmimetic wing patterns. We found that the mimetic allele evolved by gaining six new *cis*-regulatory elements (CREs) and an inversion that locked those CREs together with *dsx* and the novel noncoding gene *U3X*. At least four of these new CREs are essential for *dsx* expression and mimetic pattern development. Genome-wide assays of DSX binding suggest that *dsx* controls mimetic pattern development by directly regulating the expression of both itself and a handful of unlinked genes. The *dsx* supergene thus contains multiple functional genetic elements, each required for the phenotype switch and linked together by an inversion, and likely exerts its effects on color pattern development through direct regulation of unlinked “modifier” genes. Our results therefore support classic theories of supergene evolution, but update those ideas to match what we have learned about gene regulation since their development over a half century ago.

autoregulation | polymorphism | sexual dimorphism | co-option | epistasis

Complex balanced polymorphisms are widespread in nature, ranging from sexual dimorphisms to alternative life history and mating strategies in many animals and plants. These fascinating systems have provided important examples of how natural and sexual selection drive the evolution and maintenance of phenotypic variation (1–5). Although complex balanced polymorphisms often involve a suite of behavioral and morphological traits, many are controlled by genetic variation in just a single genomic locus (6–8). In some cases, this locus is a single developmental transcription factor (9). More often, complex balanced polymorphisms are associated with large chromosomal regions of low recombination containing tens to hundreds of genes, frequently called “supergenes” (10–12).

Supergenes control a wide array of complex balanced polymorphisms including fire ant colony social structures (2, 5), bird mating morphs (3, 12), fish migration (10, 13), and most sexual dimorphisms (6). In each system, the switch between discrete phenotypes is controlled by alternate supergene alleles. Classic theory predicts that each allele contains multiple genes coadapted to control different aspects of the complex polymorphism (14). Reduced recombination between these genes, often aided by one or more inversions, prevents shuffling of coadapted alleles that could produce intermediate phenotypes with low fitness (15). Recent work in mimetic butterflies has extended the concept of a supergene to encompass multiple tightly linked functional elements, such as *cis*-regulatory elements, coadapted to control the phenotype switch (8, 16, 17).

Despite these clear theoretical expectations of what a supergene really is, the causative gene(s), let alone functional genetic differences between supergene alleles, have been difficult to identify. This is because classic supergenes often span multiple megabases, contain tens to hundreds of genes, and exhibit extremely high linkage disequilibrium that blocks traditional genetic mapping approaches to identify causative genetic variation (2, 10–12, 18, 19). Recent candidate approaches have identified a single gene each within the fire ant social supergene and the ruff mating morph supergene that are strongly associated with those polymorphisms (2, 20). These candidate genes—the pheromone binding protein *Gp-9* and the hormone synthesis enzyme *HSD17B*, respectively—could potentially coordinate phenotype development themselves; it is not known whether any other genes are required for the phenotype switch. It is therefore unknown 1) whether multiple coadapted alleles exist

Significance

Balanced polymorphisms are often controlled by genetic variation in just a single locus that switches between alternate phenotypes. Here, we establish that novel *cis*-regulatory elements (CREs) of the conserved transcription factor gene *doublesex* underlie the switch between mimetic and nonmimetic butterfly wing patterns. *dsx* and these novel CREs are locked together by an inversion, showing that this *dsx* “supergene” combines traditional concepts of multigene supergenes with our current understanding of the complexity of gene regulation. We identify a molecular basis for the functions of a historically important supergene and provide insight into how conserved genes can be co-opted into new roles in development.

Author affiliations: ^aDepartment of Ecology & Evolution, The University of Chicago, Chicago 60637, IL; and ^bCommittee on Development, Regeneration, and Stem Cell Biology, The University of Chicago, Chicago 60637, IL

Author contributions: N.W.V., S.I.S., and M.R.K. designed research; N.W.V., S.I.S., C.L.F., D.M., N.N.S., and W.L. performed research; N.W.V. analyzed data; and N.W.V. and M.R.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: nvankuren@uchicago.edu or mkrnforst@uchicago.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2509864122/-/DCSupplemental>.

Published October 8, 2025.

and 2) when and why recombination suppression evolved during the evolution of these balanced polymorphisms.

Here, we investigated the composition, function, and evolution of a supergene controlling mimicry polymorphism in *Papilio* swallowtail butterflies. Many palatable butterfly species have evolved wing color patterns resembling those of distantly related toxic species, thereby protecting mimics from visual predators (21). Negative frequency-dependent selection often results in mimicry polymorphisms. In some species, mimetic patterns are limited to one sex, usually females, resulting in complex balanced polymorphisms in which males develop a single nonmimetic color pattern while females can develop one or more discrete mimetic patterns (15). Based on crossing experiments in *P. polytes* (22), R. A. Fisher argued that mimicry polymorphisms evolved in two steps (23). First, an initial large-effect mutation switched on development of a novel, but imperfect, mimetic pattern. After this “switch gene” became established in the population, natural selection for mimicry caused the evolution of modifier genes that function epistatically with the switch gene to improve the novel mimetic pattern (23–26). Classic supergene theory is an extension of Fisher’s basic model to situations where the switch and modifier loci are tightly linked (14, 16, 26–29). Thus, if we can identify the functional genetic differences between switch gene alleles and the molecular basis of epistasis between them and their modifiers, then we may better understand the evolution of supergenes and the complex balanced polymorphisms they control.

Here, we dissect the functional genetic basis of supergene mimicry in *Papilio alphenor*, formerly considered a subspecies of *P. polytes* (30, 31). Female-limited mimicry polymorphism in *P. polytes* and *P. alphenor* is controlled by the *doublesex* supergene (32, 33). This supergene comprises an ancestral *h* allele and a novel, dominant *H* allele that is defined by a 150 kb inversion containing *dsx*—a key transcription factor gene that controls insect sexual differentiation—and the novel noncoding gene *untranslated three exons* (*U3X*; Fig. 1 *A* and *B*). Previous RNA interference (RNAi) experiments demonstrated that *dsx^H* completely controls the switch from nonmimetic to mimetic female patterns (33, 34). Although the DSX proteins encoded by the two alleles are identical within their dimerization domains and DNA binding domains, *dsx^H* exhibits a unique expression pattern in pupal mimetic female wings that is essential to switch on mimetic pattern development (Fig. 1 *B* and *C*) (32–35). The causes of allele-specific *dsx* expression patterns, and therefore the functional genetic basis of the mimicry switch, remained unknown.

Recent work has also begun to identify potential modifiers of the *dsx* switch (34, 35, 37). RNA-seq and RNAi experiments have identified over a dozen genes required for mimetic pattern development, including genes that are linked to the *dsx* supergene such as *U3X*, *sir2*, and *prospero* (37) and unlinked genes like *engrailed* that are involved in Wnt and Hedgehog signaling (34). Although these studies clearly showed that these genes function downstream of *dsx^H* specifically in mimetic pattern development, the causes of epistasis between the switch gene and these potential modifiers remained unknown. The goals of this study were thus to identify the functional genetic elements of the *dsx* supergene, determine the causes of the functional differences between alleles, and identify potential molecular mechanisms of epistasis between this switch gene and its unlinked modifiers.

Results

The Mimetic Allele Gained Multiple Novel *dsx* CREs. Mimetic *dsx^H* has evolved a unique expression pattern in mimetic female wings that is essential to switch on mimetic pattern development, yet the

DSX proteins are nearly identical (Fig. 1 *B* and *C*). We therefore expected that the functional genetic differences between supergene alleles were regulatory. We first searched for *cis*-regulatory elements (CREs) controlling *dsx* expression in the developing wing using ATAC-seq (Fig. 1 *D* and *SI Appendix*, Figs. S1–S4 and Tables S1–S3). Peaks of ATAC-seq coverage indicate genome regions accessible to transcription factor binding, and therefore potential CREs (38). We found 28 ATAC peaks within the mimetic *H* allele in early pupal wings, when *dsx* expression spikes in mimetic females. Most peaks were in or near *dsx^H*, including peaks 0.47 kb and 3.9 kb upstream of the *dsx^H* promoter, 22 peaks within introns, and peaks within exons 1 and 6. Peaks were also found at the *U3X* and *U3X* promoters. The nonmimetic *h* allele contained 33 peaks in early pupal wings located in similar relative positions to those in the mimetic allele.

The different numbers of CREs in the two alleles suggested that CRE gain or loss may have played a role in this supergene’s evolution. We tested this idea by using BLAST to identify orthologous CRE sequences between the *h* and *H* alleles and three monomorphic outgroup species’ genomes (39). Despite enormous sequence divergence between the alleles, CRE sequences and synteny were highly conserved over 20 My of evolution (Fig. 1 *E* and *SI Appendix*, Fig. S5 and Tables S4 and S5). All CREs found in the *P. alphenor h* allele were also found in the *P. polytes h* allele and at least two outgroup species. However, six CREs were uniquely shared by the *P. alphenor* and *P. polytes H* alleles—one at the *U3X* promoter and five within or near *dsx*. This result strongly suggested that the derived *H* allele gained multiple novel *dsx* CREs. These *H*-specific CREs are likely involved in the unique spike of *dsx* expression in early pupal female wings because 1) all were accessible in early pupal wings but inaccessible in heads and mid-pupal wings, and 2) three of the six *H*-specific CREs, all within or near *dsx*, were significantly more accessible in mimetic females than in males in early pupal wings (*SI Appendix*, Figs. S3 and S4 and Tables S6 and S7).

It is unlikely that CREs outside the inversion regulate *dsx* expression. Hi-C experiments showed that each allele was contained within a single topologically associating domain (TAD), a 3D genome organization which largely defines the local regulatory region (Fig. 1 *F*) (40). The right TAD boundaries coincided with the right inversion breakpoint in both alleles, while the left TAD boundaries were 21.9 kb and 17.9 kb outside of the inversion, encompassing *dsx* and adjacent genes *sir-2*, *rad51*, and *nach*. The inversion and high divergence between the alleles thus caused few changes to the local chromatin structure. Four CREs, in or near *nach* and *sir2*, were found within the TADs but outside of the inversion. However, these peaks were found in all samples, and none were differentially accessible between mimetic and nonmimetic females. The mimetic *dsx* allele therefore gained six new CREs, all contained within the inversion.

Mimetic Allele-Specific CREs Are Essential for the Mimicry Switch.

We next tested whether these novel CREs were required for the mimicry switch. Knocking down *dsx* expression in mimetic *HH* or *Hh* females using RNA interference (RNAi) causes them to develop nonmimetic patterns (34). We expected to observe similar effects when we knocked out any CREs required for *dsx* expression in the mimetic wing. We therefore used CRISPR/Cas9 to individually delete target CREs in *Hh* eggs, then looked for G0 females with mosaic nonmimetic color patterns (Fig. 2). Injections targeting three conserved CREs, which we expected to be generally essential for *dsx* expression, yielded females with patches of nonmimetic pattern in an otherwise mimetic background (*SI Appendix*, Fig. S7 and Tables S8 and S9). Importantly, we also recovered mosaic females from injections targeting four of six *H*-specific CREs,

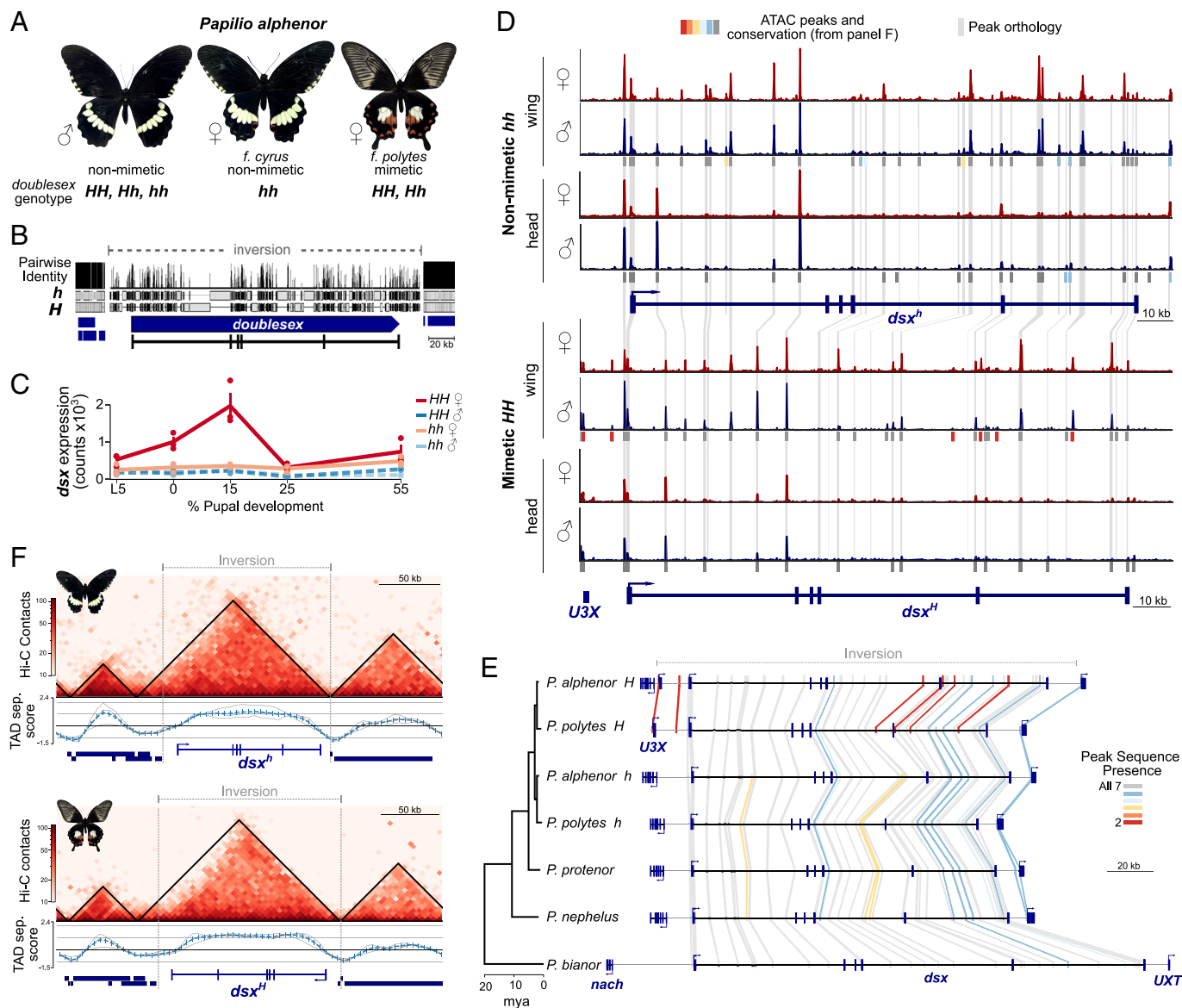


Fig. 1. Regulatory architecture of the *doublesex* supergene. (A) Female-limited mimicry polymorphism in *Papilio alphenor*. (B) Pairwise sequence alignment of the *dsx* supergene alleles and flanking 20 kb. The mimetic H allele is uninverted for alignment and display. (C) *dsx* hindwing expression across development, from (34). (D) Normalized ATAC-seq coverage tracks from early pupal wings and heads within the inversion. All tracks show the range [0 to 450]. Peaks of coverage indicate open chromatin and potential CREs. Statistical peak calls are shown below the x-axis for each tissue and are colored according to conservation. (E) Orthologous peak sequences between the alleles are connected by shaded lines. *SI Appendix, Figs. S1–S4 and Table S1* contain additional information. (F) Orthology and synteny between *P. alphenor* peaks and outgroup alleles. Orthology was determined by BLASTing *P. alphenor* peak sequences to the genome regions bounded by *nach* and *UXT* in each species. The *P. alphenor* and *P. polytes* H alleles are uninverted for display. Peak conservation (i.e. the number of alleles in which the peak sequence is present) is denoted by color, and lines connect orthologous regions. The phylogeny is modified from refs. 30 and 36. See also *Materials and methods* and *SI Appendix, Tables S2–S7*. (F) Hi-C contact heatmaps, topologically associating domains (TADs; black triangles), and TAD separation scores near the inversion in early pupal female hindwings. Gene models are shown along the x axes. Low TAD separation scores indicate boundaries between adjacent TADs. *SI Appendix, Fig. S6* contains additional Hi-C results.

strongly suggesting that these CREs are each essential for the mimicry switch (Fig. 2). Knockouts affected all mimetic pattern elements, including forewing stripes, scale color in light patches, and the size and color of submarginal hindwing spots (Fig. 2). *U3X* promoter knockouts yielded no phenotype, contrary to RNAi results in *P. polytes* (37). Long-read whole genome sequencing of three mosaic butterflies strongly supported the conclusion that their phenotypes were caused by knockout of the target CRE, rather than disruption of the entire gene or multiple nearby CREs (Fig. 2G and *SI Appendix, Fig. S8*). H-specific CREs are unlikely to be essential for *dsx* expression because both the mimetic and nonmimetic *dsx* alleles control sexual differentiation equally well. Instead, these novel CREs likely mediate the unique spike of *dsx*

expression that is required for mimetic pattern development (Fig. 1C).

Together, our results show that the functional genetic elements of the *dsx* supergene are novel *dsx* cis-regulatory elements. The mimetic supergene allele gained at least four novel *dsx* CREs spread across 150 kb that function together to switch on *dsx* expression in early pupal female wings, triggering development of the novel mimetic wing pattern.

***dsx* Is Autoregulated.** Gene expression depends on both CREs and the combinations of transcription factor proteins that specifically bind to those CREs and drive transcription. We next sought to identify transcription factors (TFs) that could drive

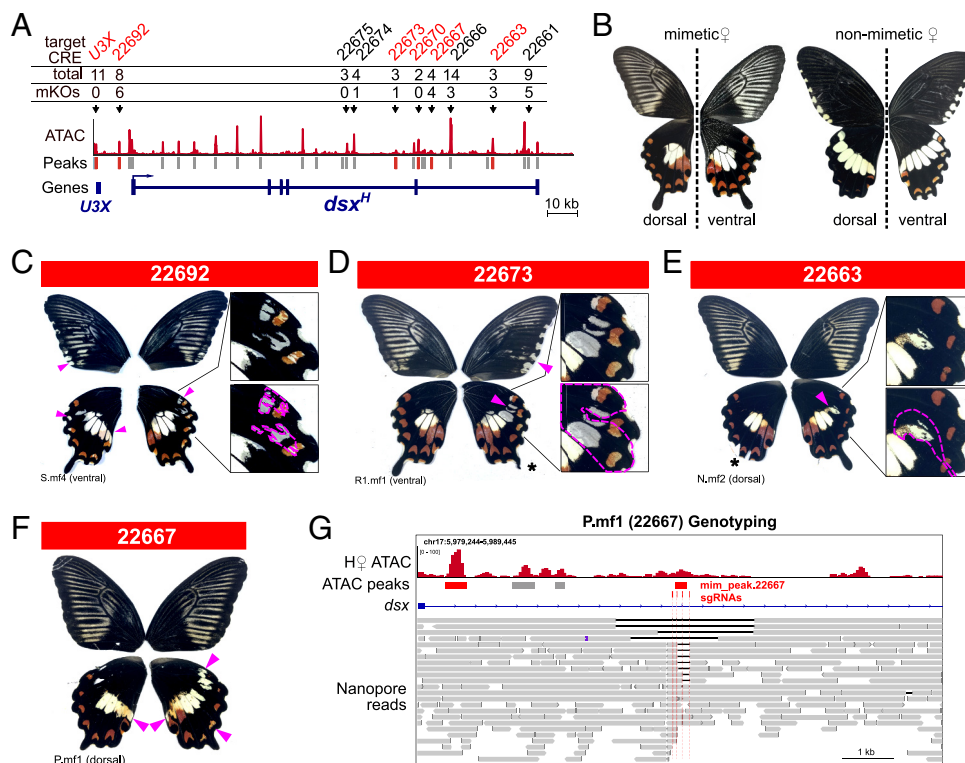


Fig. 2. Multiple *H*-specific CREs are required for mimetic pattern development. (A) Schematic of the mimetic allele indicating the ATAC peaks (CREs) that we targeted for deletion. Each candidate CRE was targeted for deletion by injecting 2 to 4 sgRNAs and Cas9 into fresh, heterozygous *Hh* eggs (SI Appendix, Tables S8 and S9). We targeted all six *H*-specific CREs (red) and four conserved CREs (black) individually. The total numbers of mimetic females (total) and mosaic knockout (mKO) females recovered are shown for each target CRE above the schematic. CRE orthology is from Fig. 1. (B) Expected dorsal and ventral wild-type female color patterns. mKO females were identified by the appearance of nonmimetic color patterns in otherwise mimetic wings (indicated by pink arrows, or outlined with pink dashed lines in the following panels). (C–F) Examples of mKO females recovered from injections targeting *H*-specific CREs 22692 (C), 22673 (D), 22663 (E), or 22667 (F). Sample identifiers and surfaces are shown below images. mKOs of additional, conserved CREs can be found in SI Appendix, Fig. S7. *: Tail damage, not developmental defect. (G) Whole-genome long-read sequencing from the individual shown in (F), demonstrating the range of deletions recovered in a single individual. Black lines within read alignments indicate deletions relative to the reference. Note the absence of deletions affecting CDS or CREs beyond the target CRE 22667. Indels and variation < 5 bp long are not highlighted. Normalized ATAC-seq data from an early pupal mimetic female wing are shown, with peak calls below. *dsx* exon 5 and intron 5 are shown peaks. Additional mKO females can be found in SI Appendix, Fig. S7.

allele-specific *dsx* expression in the *P. alphenor* wing, particularly through novel *H*-specific CREs. TFs that directly control *dsx* expression are unknown in any organism, but ChIP-seq experiments in *Drosophila* showed that over 230 different proteins bound to *dsx* CREs in whole adults. DSX itself was bound to the most *dsx* CREs of any protein (25 CREs), suggesting that it could play an important role regulating its own transcription (41). To identify CREs bound by DSX in *P. alphenor* wings, we assayed genome-wide DSX binding using CUT&RUN in early- and mid-pupal wings (Fig. 3 and SI Appendix, Figs. S9–S11 and Table S10). Similar to ChIP-seq, peaks of CUT&RUN coverage indicate regions of the genome bound by the target protein (42). We found 10,318 DSX peaks genome-wide among all samples. Most DSX peaks overlapped ATAC peaks (88.2%), consistent with the idea that CUT&RUN identified DSX binding to active CREs. Importantly, the strongest DSX peaks in each sample were found in *dsx* CREs (Fig. 3 A and B). These observations matched the *Drosophila* data and strongly support the idea that *dsx* is autoregulated in the developing butterfly wing.

Patterns of DSX binding were significantly different between supergene alleles. DSX was bound to five CREs in the nonmimetic *h* allele in early pupal female wings. DSX was bound to these five orthologous CREs in the mimetic *H* allele, but also to an additional 15 conserved CREs and five of the six *H*-specific CREs (Fig. 3 C and D). Despite their recent origins, DSX binding sites in *H*-specific CREs were just as strong as DSX binding sites in conserved CREs. Log-odds probabilities, which measure how

similar DSX binding site sequences are to the consensus motif, were not significantly different between genome-wide DSX peaks, conserved peaks, or *H*-specific peaks (Fig. 3 E and F; all Welch's *t* test *P*-values > 0.10). In fact, *H*-specific CREs contain two of the top three strongest DSX binding sites in the *H* allele (SI Appendix, Table S11). Our ATAC, CRISPR, and CUT&RUN results strongly suggest that the mimetic supergene allele gained multiple new *dsx* CREs that mediate autoregulation and are essential for *dsx* expression in the mimetic wing.

Temporal dynamics of DSX binding in the supergene also support this conclusion. DSX is highly and uniformly expressed in the early pupal mimetic wing, but by mid-pupal development becomes restricted to regions of the wing that become white (34). By mid-pupal development, only one *H*-specific CRE was accessible and was not bound by DSX (SI Appendix, Fig. S3) (34). Instead, 13 conserved CREs were differentially bound by DSX between females and males at this stage (SI Appendix, Fig. S10). These results further suggest that DSX binding in *H*-specific CREs is required in early pupae to trigger the mimicry switch, while differential use of conserved CREs may refine DSX expression later in development.

Gene regulation depends on combinatorial binding of TFs to multiple CREs, and it is important to note that DSX is almost certainly not the only TF required for *dsx* regulation. We searched for other TFs potentially involved in *dsx* regulation by looking for known TF binding site sequences enriched in *dsx* CREs (SI Appendix, Fig. S12) (43). *dsx* CREs were most significantly

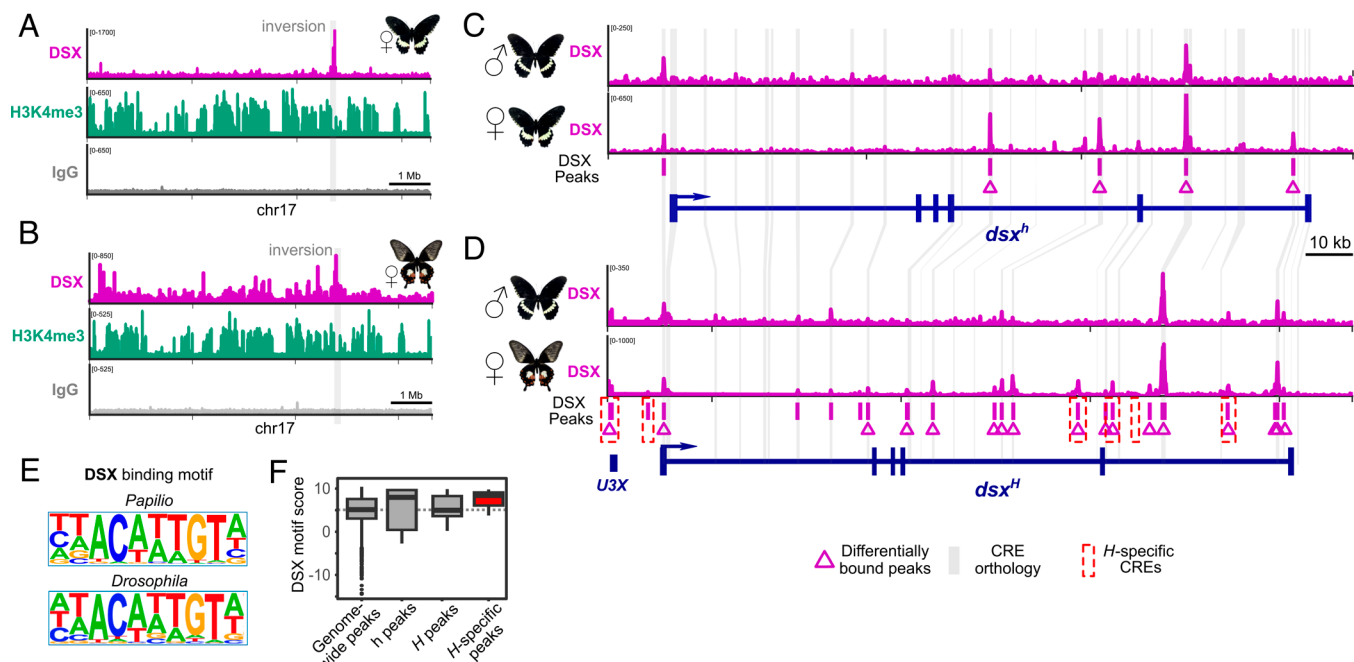


Fig. 3. The mimetic allele gained multiple novel CREs bound by DSX itself. (A and B) Normalized CUT&RUN coverage tracks for DSX, the active promoter/enhancer histone mark H3K4 trimethylation (H3K4me3), and negative control immunoglobulin G (IgG) in early pupal nonmimetic (A) and mimetic (B) female wings. Coverage across the whole of chromosome 17 is shown. Peaks of coverage indicate regions of the genome bound by the target proteins. The shaded region indicates the location of the *dsx* inversion. (C and D) DSX CUT&RUN coverage tracks in the nonmimetic *h* allele (C) and mimetic *H* allele (D) in early pupal wings. Statistically enriched DSX peaks are shown below the coverage tracks for each allele. Peaks that are significantly differentially bound by DSX between males and females are indicated by triangles. All differentially bound peaks are more strongly bound in females than in males (SI Appendix, Tables S11 and S12). Gray lines connect orthologous CREs (from Fig. 1). (E) The consensus DSX binding site motifs in *Papilio* and *Drosophila*. Letter heights are proportional to the base frequencies at each position in the binding site. See SI Appendix, Fig. S13. (F) Boxplot of the strength of DSX binding site sequences in different classes of DSX CUT&RUN peaks, measured using log-odds scores. Higher values indicate better matches between the peak sequence and the consensus motif (panel E). Scores are shown for sites in all peaks genome-wide (all peaks); peaks conserved between the *h* and *H* alleles (*h* peaks and *H* peaks); or in *H*-specific peaks. No pairwise comparisons were significantly different (all Welch's *t* test *P*-values >0.1).

enriched with binding sites for the Hox TFs Paired ($P = 1e-13$), Caudal ($P = 1e-11$), and Extradenticle ($P = 1e-10$), and DSX ($P = 1e-9$). Future work on these TFs may shed additional light onto the upstream factors required for the mimicry polymorphism.

DSX Directly Regulates Unlinked Modifier Expression. Mimicry polymorphisms are predicted to be controlled by both switch genes and epistatic modifier genes (23, 28). Epistatic modifiers, also known as “specific modifiers” (6, 15), function with the mimicry allele to control development of the mimetic pattern, but have little or no effect on nonmimetic pattern development. Although classic, multigene supergenes may evolve when modifiers are linked to the switch gene, most modifiers are expected to be unlinked to the mimicry locus (15, 26, 28, 44). Multiple genes are known to function downstream of *dsx*^H in *P. polytes* and *P. alphenor*, including components of canonical Wnt signaling and the transcription factors *rotund* and *engrailed* (3435, 37). Consistent with their roles as epistatic modifiers of the *dsx* switch, these genes are differentially expressed between mimetic and nonmimetic female wings and, importantly, RNAi of these genes disrupted mimetic patterns but had no effect on male or nonmimetic female patterns (34, 37). Our DSX CUT&RUN data immediately suggested that DSX could directly regulate the expression of these and other unlinked modifiers, providing a molecular basis for epistasis.

We tested this idea by intersecting the set of genes that were differentially expressed (DE) between mimetic and nonmimetic females with our DSX CUT&RUN results (SI Appendix, Figs. S14 and S15 and Dataset S1) (34). We found 266 DE genes that were also differentially bound by DSX in early pupal wings, strongly suggesting that these are direct targets of DSX regulation. These

potential modifiers include TFs known to be involved in the mimicry switch: *rotund*, *engrailed*, and *prospero* (34, 37); the butterfly color patterning TF *aristaleless-1* (45); the T-box TF *midline*; and at least eight other DNA-binding proteins (Fig. 4A and Dataset S1). All of these potential modifiers, except *pros*, are unlinked to *dsx*. Overall, DSX binding was strongly correlated with lower gene expression in early pupal wings (Pearson's $r = -0.286$, $P = 0.002$), consistent with its primary role as a repressor in *Drosophila* sexual differentiation (46).

We confirmed the roles of *al-1* and *mid* in the mimicry switch using RNAi. *mid* RNAi caused the boundary between medial red and white patterns to shift distally, and disrupted patterns of blue scales (Fig. 4D). *al-1* RNAi caused a more dramatic shift in the red/white boundary and a near-complete loss of blue scales (Fig. 4G). Consistent with their roles as epistatic modifiers of the mimicry switch, we observed minimal effects of *al-1* or *mid* RNAi on nonmimetic color patterns (SI Appendix, Fig. S16). These results and existing RNAi of *en*, *pros*, and *rn* confirmed the roles of these five direct targets in the mimicry switch (Fig. 4) (34, 37). These results provide strong evidence that the mimetic allele is epistatic to unlinked modifier genes because it directly binds to modifier CREs. Future work that specifically disrupts DSX binding sites would directly test this link.

Temporal Dynamics of the Mimicry Switch. *dsx*^H switches on mimetic pattern development in the early pupal wing, but becomes decoupled from this process as development continues. We know this because antibody stains showed that DSX^H expression 1) never fully prefigured the adult pattern and 2) by mid-pupal development becomes restricted to regions of the wing

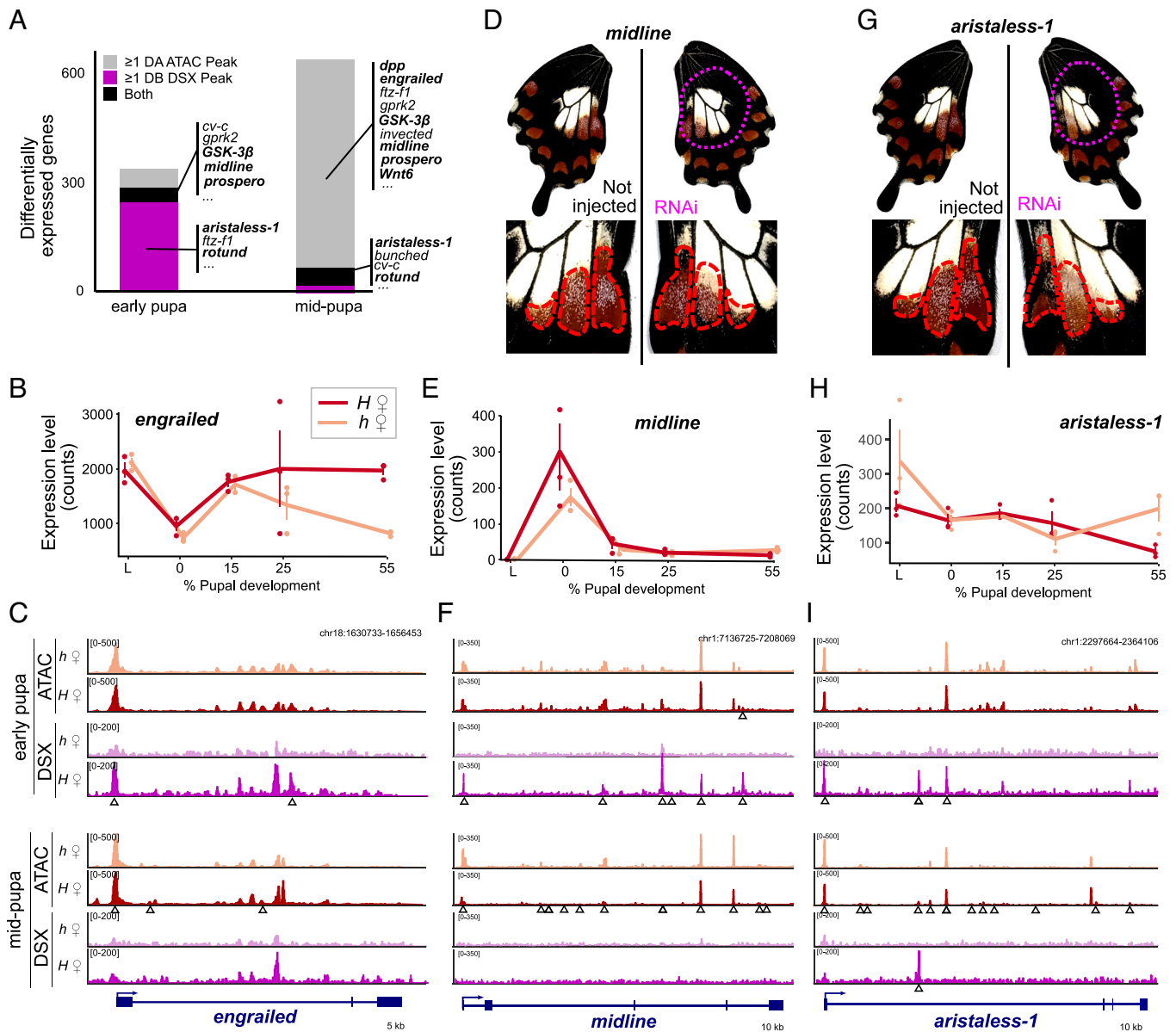


Fig. 4. Acute and long-term consequences of DSX binding in early pupal wings. (A) Stacked barchart of the numbers of differentially expressed genes between mimetic and nonmimetic females that also contain at least one differentially bound (DB) DSX peak and/or at least one differentially accessible ATAC peak. Select genes known to be involved in wing development are listed. Genes with experimental evidence for their involvement in *Papilio* mimicry are bolded (34, 37). (B) Expression pattern of *engrailed*, a known downstream target of DSX in the mimicry switch (34). (C) Normalized ATAC and DSX CUT&RUN coverage tracks from nonmimetic *h* and mimetic *H* females near *engrailed* in early (Top) and mid-pupal (Bottom) development. Differentially accessible ATAC peaks or differentially bound DSX peaks between mimetic and nonmimetic females are indicated by Δ underneath the pairs of tracks being compared. Gene models are shown with blue boxes (exons) and lines (introns) below. (D–F) Functional and functional genomic analysis of the transcription factor *midline*. (D) *midline* RNAi in a mimetic female. Short interfering RNAs targeting *midline* were electroporated into the ventral surface of the left wing following previous work (34, 47). RNAi-induced pattern changes are inferred from comparison to the *Right* (not injected) wing pattern. Pink dashes outline the area that was electroporated. Red dashed lines in the insets indicate the observed and expected boundaries between white and red patches. Note the distal shift of the white/red boundary and disruption of blue scale patterning in the RNAi wing. (E) *midline* expression. (F) Same as C, but for *midline*. (G–I) Functional and functional genomic analysis of *aristaless-1*. (G) *aristaless-1* RNAi in a mimetic female. Note the distal movement of the white/red boundary and nearly complete loss of blue scales in the RNAi wing. (H) *aristaless-1* expression. (I) Same as C, but for *aristaless-1*. Additional RNAi can be found in *SI Appendix*, Fig. S16. Peak and RNA-seq dataset can be found in *Dataset S1*.

which will become white (34). Temporal patterns of differential DSX binding and differential CRE accessibility further support the idea that DSX quickly becomes decoupled from mimetic pattern development (Fig. 4 and *SI Appendix*, Fig. S15). In early pupal development, 3640 (35.3%) DSX peaks (spread among 2187 genes) were differentially bound between mimetic and nonmimetic female wings. However, by mid-pupal development, only 557 (5.4%) DSX peaks were differentially bound, with just 46 peaks among 41 DE genes, including *rn* and *al-1*. This loose relationship between DSX binding and differential expression is

consistent with work in *Drosophila* that showed DSX binds many targets genome-wide, but affects expression of only a small number of those genes, presumably due to the availability of appropriate cofactors (46).

In contrast to DSX binding, 1.7% of CREs were differentially accessible between early pupal wings, while 27.7% were differentially accessible in mid-pupal wings (*SI Appendix*, Figs. S14 and S15). These changes were associated with differential expression independent of DSX binding because 5.6% of DE genes contained at least one differentially accessible CRE in early pupal wings, but

40.8% of DE genes contained at least one differentially accessible CRE in mid-pupal wings (Fig. 4) (34). DSX thus appears to directly regulate few genes early in development whose effects are propagated to later stages despite being decoupled from DSX binding itself.

Discussion

Structure and Evolution of a Supergene. Complex balanced polymorphisms require a simple genetic architecture to switch between alternate phenotypes. The basic concept of a supergene as a complex of coadapted genes was developed over a half-century ago yet we still lack much of the functional data necessary to say whether or not this concept is accurate (15). The key hurdle to testing these concepts has been identification of the functional genetic elements of supergenes. Until now, only the *Primula* S-locus has been demonstrated to contain multiple functional elements—genes in that case—that each contribute to the complex balanced polymorphism (48). Here, we show that the *dsx* supergene combines elements from classic models of supergene evolution and our current understanding of gene regulation: The *dsx* supergene contains multiple discrete functional elements within a single regulatory gene, each essential for *dsx* expression in the mimetic wing, linked together by an inversion. Recent authors have labeled these “multisite” supergenes to contrast them with classic, multigene, supergenes such as the *Primula* S-locus (6, 8, 16, 17).

How did the *dsx* supergene evolve? Although the supergene’s genomic structure is clear, its evolution remains murky because the *dsx* inversion and all six *H*-specific CREs were present in the last common ancestor of *P. polytes* and *P. alphenor* ~1.5 Mya (Fig. 1) (36). We hypothesize that this supergene originated via the gain of a novel CRE(s) that drove a spike of *dsx* expression in the early pupal wing that initiated mimetic pattern development. Subsequent gain of additional CREs may have helped refine the novel allele’s expression pattern across development, and the mimetic wing pattern in turn (33, 34). A key requirement for the evolution of supergenes is that these subsequent mutations are only beneficial when combined with the initial mutation—i.e. that they are conditionally advantageous. Importantly, our CRISPR/Cas9 experiments showed that at least four of the five novel *dsx* CREs are conditionally advantageous: Knocking out any one of these CREs completely breaks the mimicry switch (Fig. 2). Selection for mimicry would have then favored maintenance of an inversion that suppressed recombination between epistatic CREs along the 150 kb *dsx* region because linkage disequilibrium decays rapidly in butterflies, down to equilibrium within ~10 kb (49). Combinatorial CRE knockouts, or potentially knock-in of mimetic CREs into the nonmimetic allele, could help reconstruct the stepwise evolution of this supergene.

The evolution of butterfly wing patterns often depends on quantitative shifts of color pattern boundaries rather than the appearance of new pattern elements (50). It remains unclear whether a simple spike of *dsx* expression would alone be sufficient to induce development of a rudimentary mimetic pattern—step one in Fisher’s model of mimicry evolution (23). However, our observations that DSX directly regulates expression of genes that are known to control color pattern development in other butterflies, such as *engrailed* and *aristaleless-1*, suggest that the mimetic pattern likely evolved through modification of existing wing regulatory connections and color patterning programs (34, 50). DSX was bound to these genes in both mimetic and nonmimetic female wings, supporting the idea that those regulatory connections existed before mimicry evolved.

Co-Option and Evolution of Switch Genes. Complex balanced polymorphisms depend on alleles at a single locus that switch between alternate phenotypes. Switch genes may therefore be expected to be regulatory genes (TFs, signaling ligands, or receptors) that widely affect development. In addition to *dsx*, the Hox genes *engrailed* and *invected* are associated with the *Papilio dardanus* mimicry polymorphism, and *GLO2* controls anther position in heterostyly polymorphisms. Our work provides two insights into how such conserved genes can be co-opted into new roles like switch genes. First, we showed that *dsx* co-option occurred through the gain of new CREs. These CREs cannot be essential for its ancestral, conserved function in development of sexual dimorphisms. If the function of these novel CREs is limited to the developing wing—i.e. they do not have deleterious pleiotropic effects—then they are free to optimize their functions in mimetic pattern development. Our ATAC-seq data from developing heads provide some support for this scenario, as no *H*-specific CREs were accessible in heads. Although mutations in or loss of conserved CREs may have still played a role in the mimicry switch, the primary cause of its evolution appears to be the gain of novel regulation.

Second, we found strong evidence that *dsx* transcription is directly regulated by the DSX TF (i.e. autoregulation) (51). This does not imply that DSX is the only TF that regulates *dsx* expression in the developing wing, only that *dsx* expression depends on DSX. Future work that disrupts individual DSX binding sites within *H*-specific CREs, rather than entire CREs, would specifically test the autoregulation hypothesis. We propose that positive autoregulation could provide a simple route to switch gene evolution and, importantly, dominance. New alleles are often incompletely dominant and produce intermediate phenotypes in heterozygotes (23). Yet balanced polymorphisms depend on proper development of discrete, alternate phenotypes. Natural selection for mimicry, in particular, favors the rapid evolution of dominance at the switch gene because it exposes the new allele to selection and eliminates intermediate, poor mimics (23, 24). Gaining positive autoregulation could immediately cause new switch gene alleles to become dominant because it would boost the new allele’s expression over the ancestral allele. These effects would be limited to the novel allele, allowing further independent evolution of the alternate phenotypes. Autoregulation may also avoid deleterious pleiotropic effects, such as those from ectopic expression, because expression continues to be limited to tissues or stages where the gene is already expressed. Autoregulation could play an identical role in the evolution of classic supergenes, where TFs or other regulatory genes control their own and/or nearby gene expression. Divergence between alleles, in CREs or protein-coding sequences, and subsequent recombination suppression would then refine the supergene alleles’ functions according to classic models (15).

Molecular Basis of Epistasis Between Switch Genes and Modifiers. Mimicry polymorphisms depend on both the switch gene and the epistatic modifiers it controls (23). By definition, epistatic modifiers, sometimes called specific modifiers, are any genes required for the balanced polymorphism that respond differently to alternate switch gene alleles (15). Classic supergenes only evolve when switches and their modifiers are already tightly linked and, indeed, many modifiers are not linked to butterfly mimicry supergenes at all. Crosses in *Papilio dardanus* and *P. polytes* exemplify this phenomenon (31, 44, 52). Within-population crosses always yielded offspring with good mimetic patterns and complete dominance while between-population crosses yielded intermediate patterns and incomplete dominance.

That is, segregation revealed that each population evolved unlinked epistatic modifiers despite using the same switch (28, 44). These interactions rapidly evolve. For example, all *P. polytes* develop hindwing tails, but only mimetic *P. alphenor* develop hindwing tails. Thus, *P. alphenor* carries genetic variation at one or more unlinked tail loci that is epistatic with the mimetic *dsx* allele (31). The same is true for the size of the mimetic white patch and forewing color pattern (31). Furthermore, genes that are involved in *P. polytes* mimicry, such as *U3X*, do not appear to be involved in *P. alphenor* (Figs. 2 and 3) (37). We suspect that these results are explained by mutations in modifier CREs that are bound by the switch gene itself, and therefore only revealed in the presence of the dominant switch gene allele. Refinement of the mimetic pattern may occur through *cis*-regulatory divergence in modifier genes, particularly in CREs that are directly bound by the switch gene. Disrupting *DSX* binding sites in putative modifiers such as *en*, *mid*, and *al-1* would directly test this hypothesis.

Finally, epistasis between two genes can be caused by myriad mechanisms, from their participation in a protein complex to direct transcriptional regulation of one gene by another (53). Our results provide a counterpoint to recent work in ruffs (20). In these birds, development of alternative male morphs is strongly associated with the epistatic effects of circulating testosterone. Testosterone levels, in turn, are associated with one of the ~100 genes in the supergene, *HSD17B2*. Unlike direct regulation (i.e. the switch gene product itself binds to its downstream effectors to regulate their transcription), indirect regulation results from intra- or intercellular signaling cascades that culminate in differential gene expression. Indirect regulation also appears to underlie widespread differential expression between fire ant colony social organizations (54). Whether supergenes often control epistatic modifiers via direct regulation by switch genes, indirect regulation by circulating hormones, or other indirect mechanisms will be important for understanding the general mechanisms by which complex balanced polymorphisms evolve and function.

Materials and Methods

Additional details and information on experimental procedures can be found in the [SI Appendix, Extended materials and methods](#).

Butterflies. *Papilio alphenor* genotyping, care, developmental staging, and RNAi were performed following (34).

Genome Assembly and Annotation. High molecular weight genomic DNA was extracted from individual adult female thoraxes using the QIAgen (USA) Blood & Tissue Kit, then used to construct long-read sequencing libraries using the Oxford Nanopore Technologies (USA) LSK-110 kit. Each library was sequenced

on one R9.4.3 flow cell on a miniON Mk1b. Reads were then assembled using Flye (55) then polished using Medaka. Assemblies were annotated using our previous pipeline (34). See [SI Appendix, Extended materials and methods and Tables S13 and S14](#).

Hi-C. Hi-C libraries were constructed using hindwing tissue from *HH* or *hh* males or females and the Dovetail Genomics' (USA) Omni-C kit following the manufacturer's instructions. Hindwings from five individuals were pooled for each library. Libraries were pooled and sequenced PE150 on a NovaSeq 6000 S1 flow cell, then processed using the Juicer v1.6 pipeline (56). Data were used to scaffold the mimetic *P. alphenor* assembly and to identify TADs using hicExplorer (57).

ATAC-Seq. ATAC-seq was performed on hindwings or clean brain and retina from individual pupae following published protocols (38, 58). Samples were collected from *Hh* or *hh* males and females at two developmental stages (15 and 35% pupal development) in triplicate. Libraries were sequenced PE50 on an Illumina NovaSeqX for an average of 30M read pairs per sample. After mapping and quality control, we identified peaks using F-seq2. Orthologous ATAC peaks within the *dsx* region were identified using BLAST. See [SI Appendix, Extended materials and methods and Table S1](#).

CRISPR/Cas9-Mediated CRE Knockouts. We targeted individual CREs for knockout by injecting Cas9 protein and 2 to 4 sgRNAs flanking the CRE of interest into *Hh* eggs within 3 h of being laid. Surviving larvae were allowed to grow and emerge as adults before screening. Adults with large mutant clones were chosen for deep, long-read sequencing essentially following the Nanopore protocol described above. See [SI Appendix, Supplementary text and Tables S8 and S9](#).

CUT&RUN. CUT&RUN was performed following (42). Samples were collected from individual homozygous *HH* or *hh* males and females at two developmental stages (15 and 35% pupal development) in triplicate. We pulled down *DSX*, *H3K4me3*, and IgG from each sample (a single pair of hindwings). Libraries were sequenced PE50 on a NovaSeq 6000 to ~10M read pairs per sample. CUT&RUN peaks were identified using MACS3 and IgG as the control track (59). Peak annotation and motif identification were performed using HOMER 4.11 (43). See [SI Appendix, Supplementary text and Table S10](#).

RNA-Seq Reanalysis. RNA-seq data from ref. 34 were analyzed using the new reference genome and annotation following the pipeline described in that paper.

Data, Materials, and Software Availability. Illumina ATAC and CUT&RUN sequencing data are publicly available in the National Center for Biotechnology Information (NCBI) BioProject [PRJNA1062051](#) (60). Genome assemblies, annotations, R projects, and full analysis results are publicly available in Dryad accession [10.5061/dryad.tx95x6b75](#) (61). Previously published data from ref. 34 available in NCBI BioProject [PRJNA882073](#) were used in this work (62). Anti-*DSX* antibody is available from the authors upon request. All other data are included in the manuscript and/or [supporting information](#).

ACKNOWLEDGMENTS. We thank the University of Chicago greenhouse staff and the University of Chicago Functional Genomics Facility (RRID:SCR_019196) for research support and the University of Chicago's Center for Research Informatics for computational support. We also thank three anonymous reviewers for their constructive criticisms. This work was supported by NIH R35 GM131828 to M.R.K.

- M. Joron *et al.*, A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol.* **4**, 1831–1840 (2006).
- J. Wang *et al.*, A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).
- C. Küpper *et al.*, A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016).
- J. Li *et al.*, Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nat. Plants* **2**, 16188 (2016).
- C. Martínez-Ruiz *et al.*, Genomic architecture and evolutionary antagonism drive allelic expression bias in the social supergene of red fire ants. *Elife* **9**, e55862 (2020).
- D. Charlesworth, The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol. Appl.* **9**, 74–90 (2016).
- T. Schwander, R. Libbrecht, L. Keller, Supergenes and complex phenotypes. *Curr. Biol.* **24**, R288–R294 (2014).
- J. Gutiérrez-Valencia, P. W. Hughes, E. L. Berdan, T. Slotte, The genomic architecture and evolutionary fates of supergenes. *Genome Biol. Evol.* **13**, evab057 (2021).
- A. Yassin *et al.*, The *pdm3* locus is a hotspot for recurrent evolution of female-limited color dimorphism in *Drosophila*. *Curr. Biol.* **26**, 2412–2422 (2016).
- M. Jamsandekar *et al.*, The origin and maintenance of supergenes contributing to ecological adaptation in Atlantic herring. *Nat. Commun.* **15**, 9136 (2024).
- M. Joron *et al.*, Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
- E. M. Tuttle *et al.*, Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344–350 (2016).
- D. E. Pearse *et al.*, Sex-dependent dominance maintains migration supergene in rainbow trout. *Nat. Ecol. Evol.* **3**, 1731–1742 (2019).
- K. Mather, The genetical architecture of heterostyly in *Primula sinensis*. *Evolution* **4**, 340–352 (1950).
- D. Charlesworth, B. Charlesworth, Theoretical genetics of batesian mimicry II. Evolution of supergenes. *J. Theor. Biol.* **55**, 305–324 (1975).
- M. J. Thompson, C. D. Jiggins, Supergenes and their role in evolution. *Heredity (Edinb.)* **113**, 1–8 (2014).

17. T. Booker, R. W. Ness, D. Charlesworth, Molecular evolution: Breakthroughs and mysteries in Batesian mimicry. *Curr. Biol.* **25**, R506–R508 (2015).
18. S. Lamichaney *et al.*, Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2016).
19. M. Matschner *et al.*, Supergene origin and maintenance in Atlantic cod. *Nat. Ecol. Evol.* **6**, 469–481 (2022).
20. J. L. Loveland *et al.*, A single gene orchestrates androgen variation underlying male mating morphs in ruffs. *Science* **387**, 406–412 (2025).
21. H. W. Bates, Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconidae. *Trans. Linn. Soc. Lond.* **23**, 495–566 (1862).
22. J. C. F. Fryer, An investigation by pedigree breeding into the polymorphism of *Papilio polytes*. *Lond. B Biol. Sci.* **204**, 227–254 (1914).
23. R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, ed. 1, 1930).
24. D. Charlesworth, B. Charlesworth, Theoretical genetics of Batesian mimicry III: Evolution of dominance. *J. Theor. Biol.* **55**, 325–337 (1975).
25. C. A. Clarke, P. M. Sheppard, The evolution of dominance under disruptive selection. *Heredity (Edinb.)* **14**, 73–87 (1960).
26. K. Mather, Polymorphism as an outcome of disruptive selection. *Evolution* **9**, 52 (1955).
27. P. M. Sheppard, Polymorphism, linkage and the blood groups. *Am. Nat.* **87**, 283–294 (1953).
28. C. A. Clarke, P. M. Sheppard, Super-genes and mimicry. *Heredity* **14**, 175–185 (1960).
29. E. B. Ford, The genetics of polymorphism in the Lepidoptera. *Adv. Genet.* **5**, 43–87 (1953).
30. F. L. Condamine *et al.*, A comprehensive phylogeny and revised taxonomy illuminate the origin and diversification of the global radiation of *Papilio* (Lepidoptera: Papilionidae). *Mol. Phylogenet. Evol.* **183**, 107758 (2023).
31. C. A. Clarke, P. M. Sheppard, The genetics of the mimetic butterfly *Papilio polytes*. *Lond. B Biol. Sci.* **263**, 431–458 (1972).
32. K. Kunte *et al.*, *Doublesex* is a mimicry supergene. *Nature* **507**, 229–232 (2014).
33. H. Nishikawa *et al.*, A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat. Genet.* **47**, 405–409 (2015).
34. VanKuren N. W. *et al.*, Acute and long-term consequences of Co-opted doublesex on the development of mimetic butterfly color patterns. *Mol. Biol. Evol.* **40**, msad196 (2023).
35. T. Iijima, S. Yoda, H. Fujiwara, The mimetic wing pattern of *Papilio polytes* butterflies is regulated by a doublesex-orchestrated gene network. *Commun. Biol.* **2**, 257 (2019).
36. W. Zhang, E. Westerman, E. Nitzany, S. Palmer, M. R. Kronforst, Tracing the origin and evolution of supergene mimicry in butterflies. *Nat. Commun.* **8**, 1269 (2017).
37. S. Komata *et al.*, Functional unit of supergene in female-limited Batesian mimicry of *Papilio polytes*. *Genetics* **223**, iyac177 (2022). [10.1093/genetics/iyac177](https://doi.org/10.1093/genetics/iyac177).
38. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213 (2013).
39. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
40. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
41. N. Nègre *et al.*, A cis-regulatory map of the *Drosophila* genome. *Nature* **471**, 527–531 (2011).
42. M. P. Meers, T. D. Bryson, J. G. Henikoff, S. Henikoff, Improved CUT&RUN chromatin profiling tools. *Elife* **8**, e46314 (2019).
43. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
44. C. A. Clarke, P. M. Sheppard, The evolution of mimicry in the butterfly *Papilio dardanus*. *Heredity (Edinb.)* **14**, 163–173 (1960).
45. E. L. Westerman *et al.*, *Aristaless* controls butterfly wing color variation used in mimicry and mate choice. *Curr. Biol.* **28**, 3469–3474 (2018).
46. E. Clough *et al.*, Sex- and tissue-specific functions of *Drosophila* Doublesex transcription factor target genes. *Dev. Cell* **31**, 761–773 (2014).
47. T. Ando, H. Fujiwara, Electroporation-mediated somatic transgenesis for rapid functional analysis in insects. *Development* **140**, 454–458 (2013).
48. C. N. Huu, B. Keller, E. Conti, C. Kappel, M. Lenhard, Supergene evolution via stepwise duplications and neofunctionalization of a floral-organ identity gene. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23148–23157 (2020).
49. S. H. Martin *et al.*, Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* **23**, 1817–1828 (2013).
50. H. F. Nijhout, Developmental perspectives on evolution of butterfly mimicry. *Bioscience* **44**, 148–157 (1994).
51. S. T. Crews, J. C. Pearson, Transcriptional autoregulation in development. *Curr. Biol.* **19**, R241–R246 (2009).
52. M. J. T. N. Timmermans, A. Srivathsan, S. Collins, R. Meier, A. P. Vogler, Mimicry diversification in *Papilio dardanus* via a genomic inversion in the regulatory region of *engrailed-invested*. *Proc. R. Soc. B Biol. Sci.* **287**, 20200443 (2020).
53. B. Lehner, Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27**, 323–331 (2011).
54. B. M. Jones *et al.*, The fire ant social chromosome exerts a major influence on genome regulation. *Mol. Biol. Evol.* **42**, msaf112 (2025).
55. B. Freire, S. Ladra, J. R. Parama, Memory-efficient assembly using Flye. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **19**, 3564–3577 (2022).
56. O. Dudchenko *et al.*, De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
57. F. Ramírez *et al.*, High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
58. J. J. Lewis, R. D. Reed, Genome-wide regulatory adaptation shapes population-level genomic landscapes in *Heliconius*. *Mol. Biol. Evol.* **36**, 159–173 (2019).
59. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137–R137 (2008).
60. N. W. VanKuren, M. R. Kronforst, "Functional genomic basis of the Papilio alphenor doublesex supergene." NCBI BioProject. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1062051>. Deposited 7 January 2024.
61. N. W. VanKuren, "Supergene evolution via gain of autoregulation." Dryad. <https://datadryad.org/dataset/doi:10.5061/dryad.tx95x6b75>. Deposited 22 March 2025.
62. N. W. VanKuren, M. R. Kronforst, "Papilio polytes (common Mormon): Differential gene expression underlying the Papilio polytes alphenor mimicry switch." NCBI BioProject. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA882073>. Deposited 19 September 2022.