EVOLUTION

Deep cis-regulatory homology of the butterfly wing pattern ground plan

Anyi Mazo-Vargas[1,2]*, Anna M. Langmüller[3], Alexis Wilder[2], Karin R. L. van der Burg[1], James J. Lewis[1,4], Philipp W. Messer[3], Linlin Zhang[1,5], Arnaud Martin[2], Robert D. Reed[1]*

Butterfly wing patterns derive from a deeply conserved developmental ground plan yet are diverse and evolve rapidly. It is poorly understood how gene regulatory architectures can accommodate both deep homology and adaptive change. To address this, we characterized the cis-regulatory evolution of the color pattern gene *WntA* in nymphalid butterflies. Comparative assay for transposase-accessible chromatin using sequencing (ATAC-seq) and in vivo deletions spanning 46 cis-regulatory elements across five species revealed deep homology of ground plan–determining sequences, except in monarch butterflies. Furthermore, noncoding deletions displayed both positive and negative regulatory effects that were often broad in nature. Our results provide little support for models predicting rapid enhancer turnover and suggest that deeply ancestral, multifunctional noncoding elements can underlie rapidly evolving trait systems.

Trait evolution frequently occurs through sequence divergence in noncoding regions of the genome that control gene expression (*1*). Few case studies, however, have characterized the history of regulatory systems that underlie rapidly evolving traits (*2*). In this work, we performed comparative chromatin analyses and regulatory knockouts of the butterfly wing pattern gene *WntA* to investigate how trait homology is reflected in regulatory sequences of a highly diverse, continually adapting character system. *WntA* encodes a signaling ligand that induces major color pattern elements of the butterfly wing pattern ground plan (*3*–*6*), and *WntA* noncoding variation underlies color pattern adaptation in multiple unrelated butterfly species (*4*, *7*). Thus, allelic variation at the *WntA* locus underlies pattern variation at microevolutionary scales yet also explains macroevolutionary aspects of pattern divergence.

To characterize the cis-regulatory architecture of the *WntA* [i.e., identities and locations of regulating cis-regulatory elements (CREs)], we first used Hi-C to infer the topologically associating domain (TAD) of the *WntA* locus in developing wings. Inside individual TADs, CREs and genes preferentially interact with each other (*8*). In imaginal discs of *Junonia coenia*, when *WntA* is expressed (*5*), we identified a TAD that spans *WntA* and its two intergenic regions (Fig. 1, A and B). The strongest CRE-to-promoter interactions occurred just upstream of the *WntA* promoter and across its lengthy first intron. These data, coupled with sequence association studies (*7*, *9*), led us to focus our functional screens for *WntA* CREs on these regions.

We performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) to profile chromatin accessibility in heads, forewings, and hindwings from the last larval instar of five nymphalids (Fig. 1C and fig. S1). By comparing head and wing profiles, we identified regions showing wing-specific activity. We next asked to what extent individual wing-specific CREs are conserved or are lineage specific. By overlapping the most conserved sequences (Fig. 1C) with the differentially accessible chromatin regions, we observed that 69 to 88% of wing-specific CREs in the *WntA* TAD were in areas with strong sequence conservation between the nymphaline, satyrine, and heliconiine subfamilies. The exception was the monarch butterfly, *Danaus plexippus*, for which 70.6% of the ATAC-seq peaks were in danaine-specific regions (fig. S2). Whereas the nymphaline and heliconiine datasets highlighted both orthologous and novel CREs, most wing-specific CREs were conserved within and between these nymphalid subfamilies. By contrast, the sister group to the rest of the nymphalids, the Danainae clade showed a largely lineage-specific repertoire of CREs (figs. S1 and S2), consistent with the divergent mode of *WntA* expression previously reported in monarchs (*6*).

We functionally assessed regions containing candidate *WntA* CREs using a CRISPR-Cas9 shotgun mosaic deletion approach (*10*), where we injected multiple single-guide RNAs (sgRNAs) tiled across open chromatin regions (Fig. 2A, fig. S3, and tables S1 and S2). This approach results in pattern mutant clones derived from a spectrum of deletions of different lengths and positions around candidate CREs. To identify regions that potentially play a role in establishing the nymphalid ground plan, we targeted wing-specific CREs conserved between *J. coenia* and *Vanessa cardui*—two species with ancestral-like *WntA*-induced color patterns (*5*, *6*). We observed that most of our deletions generated mutant clones affecting similar or overlapping wing color pattern elements (fig. S3 and data S1) (*11*) and also affected basal, central, and distal pattern elements across both wings (Fig. 2B and fig. S3). This high prevalence of overlapping phenotypic effects is consistent with the Hi-C data, which reveal physical interactions across multiple CREs and the *WntA* promoter (Fig. 1A) and support a model where color patterns are determined by a spatially distributed array of physically interacting noncoding sequences.

This conserved *WntA* regulatory architecture prompted us to investigate the role of recently evolved sequences in pattern formation. To test this, we deleted a region centered on *CRE 24*, which appears to be specific to *V. cardui*. *CRE 24* is not found in congenerics *Vanessa tameamea* or *Vanessa atalanta* (which diverged ~10 to 15 million years ago) (*12*) (fig. S4) or any other currently sequenced butterfly. Deletion of this region caused the reduction and/or loss of basal, central, and marginal pattern elements (figs. S1 and S5), thus demonstrating how even recently evolved noncoding sequences can be integrated into cis-regulatory networks.

Color pattern homologies of *Heliconius* butterflies are a long-standing question (*13*, *14*). *WntA* specifies melanic patches in this genus that may be derived from the nymphalid ground plan (*3*, *6*). We thus investigated to what degree ancestral versus *Heliconius*-specific CREs determine these patterns. ATAC-seq and comparative sequence analysis showed a large number of *WntA* CREs with deep sequence conservation between heliconiines and nymphalines (Fig. 1C and fig. S1), including CREs required for ground plan patterning in *J. coenia* and *V. cardui*. We generated deletions centered on five of these deeply conserved CREs in *Heliconius himera* (figs. S3C and S4). Notably, deletions spanning all five CREs, including on opposite ends of the first intron, had similar broad effects on melanic *Heliconius* wing patterns (Fig. 2D) (*11*). Deletions spanning two additional heliconiine-specific CREs revealed similar, overlapping phenotypes (*11*). We conclude that *Heliconius WntA* shares a conserved cis-regulatory architecture with nymphaline butterflies and that the highly derived mimicry-related color patterns of *Heliconius* appear to share deep regulatory homology with the nymphalid ground plan.

It has been speculated that the color patterns of basal heliconiines, which typically show fragmented black, brown, and silver spots, may represent an intermediate state bridging the ancestral ground plan with the *Heliconius* pattern archetype (*13*, *14*). We tested this in the

[1]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA. [2]Department of Biological Sciences, The George Washington University, Washington, DC, USA. [3]Department of Computational Biology, Cornell University, Ithaca, NY, USA. [4]Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA. [5]CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China.
*Corresponding author. Email: anyimv@gwu.edu (A.M.-V.); robertreed@cornell.edu (R.D.R.)

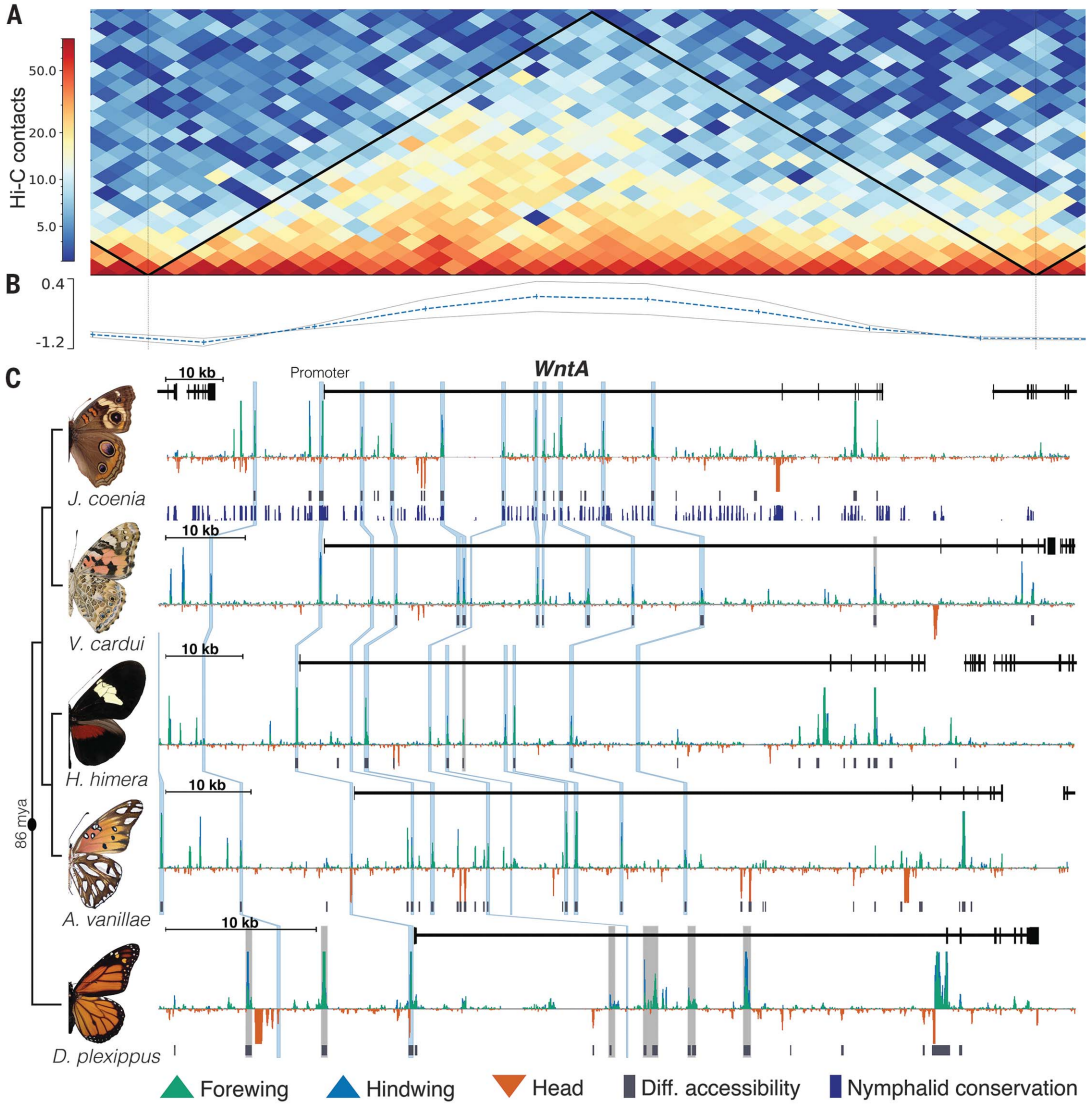

Fig. 1. Deeply conserved chromatin landscape of the *WntA* regulatory region. (**A**) Hi-C reveals abundant chromatin interactions across the upstream and first-intron regions of *WntA*. Color intensity corresponds to the contact frequency per bin. (**B**) TAD separation score. Black lines in (A) depict TAD boundaries as predicted by the TAD separation score. (**C**) Chromatin accessibility tracks (ATAC-seq) for each species sampled in this study, with phylogenetic relationships shown. Orthologous CREs assessed in this study are in blue shadows connecting different species, whereas lineage-specific elements are shown in gray; see figs. S1 and S4 for details. mya, million years ago; Diff., differential.

basal heliconiine *Agraulis vanillae* by producing deletions spanning the same regulatory regions tested in *H. himera* (Fig. 2D) and several additional heliconiine-specific CRE regions (fig. S4). Again, we found similar results—deletions of regulatory sequences across very different regions of the first intron had overlapping effects distributed across all *WntA*-induced color patterns (data S1) (*11*). This supports models that heliconiine color patterns evolved through simplification of the nymphalid ground plan and suggests that this process occurred partly through the tinkering of an ancestral cis-regulatory apparatus.

We next examined *D. plexippus*, the monarch butterfly—an exemplar of the nymphalid subfamily Danainae—to investigate how deeply the cis-regulatory architecture described above is conserved in nymphalids. In monarchs, *WntA* shows distinctive vein-associated expression patterns, and *WntA* knockouts cause the loss of these patterns (*6*). These patterns are highly derived and challenging to homologize with the nymphalid ground plan (*6*). Overall, the noncoding region of the monarch *WntA* locus shares relatively little sequence similarity with those of other nymphalids and shows a reduced number of ATAC-seq peaks (Fig. 1C), most of which are in danaine-specific genomic sequences (figs. S2 and S4). Although there are a few orthologous CREs, including the *WntA* promoter, most show no identifiable sequence similarity with other nymphalids, which suggests that they are independently derived or that their sequences are so divergent that orthology is difficult to ascertain (*15*, *16*). To test the wing patterning function of monarch CREs, we generated mosaic deletions centered on six danaine-specific CREs and one ancestrally conserved CRE. Again, we found that even distantly spaced regions had similar effects on *WntA*-induced color patterns (Fig. 2E and data S1). Thus, although many nymphalid wing patterns appear to derive from a deeply conserved regulatory architecture, there are also cases where divergent regulatory sequences underlie lineage-specific patterns.

Previous work has shown that *WntA* knockouts result in a highly specific loss of *WntA*-expressing color patterns (*3*, *6*). Our deletions that phenocopy *WntA* coding knockouts (Fig. 2) validate the enhancer-like function of these noncoding regions. However, we were surprised to observe expansions of *WntA*-expressing color patterns in several mosaic deletion experiments (data S1) (*11*). These expansions are phenocopied by heparin injections, which enhance *WntA* signaling during color pattern formation (*5*, *6*, *17*). Because deletion-induced color pattern expansion accurately replicates *WntA* gain-of-function effects, we speculate that some regulatory deletions had a positive impact on *WntA* transcription. These dual gain- and loss-of-function effects are well illustrated in *A. vanillae*, in which *WntA* is expressed in a subset of silver and black spots (Fig. 3A). When

heparin is injected, the *WntA*-expressing silver spots expand, whereas *WntA*-negative spots melanize or disappear. *WntA* coding knockouts present the opposite results—*WntA*-expressing silver spots disappear, whereas *WntA*-negative spots extend (*6*). Both expansion and reduction effects were observed in many deletion clones (Fig. 3B) across all species (fig. S6 and data S1). Our results show that *WntA* regulatory sequences encode both positive and negative regulatory instructions for color pattern formation.

The color pattern expansion and contraction phenotypes described above are present in mosaic individuals that bear deletion alleles

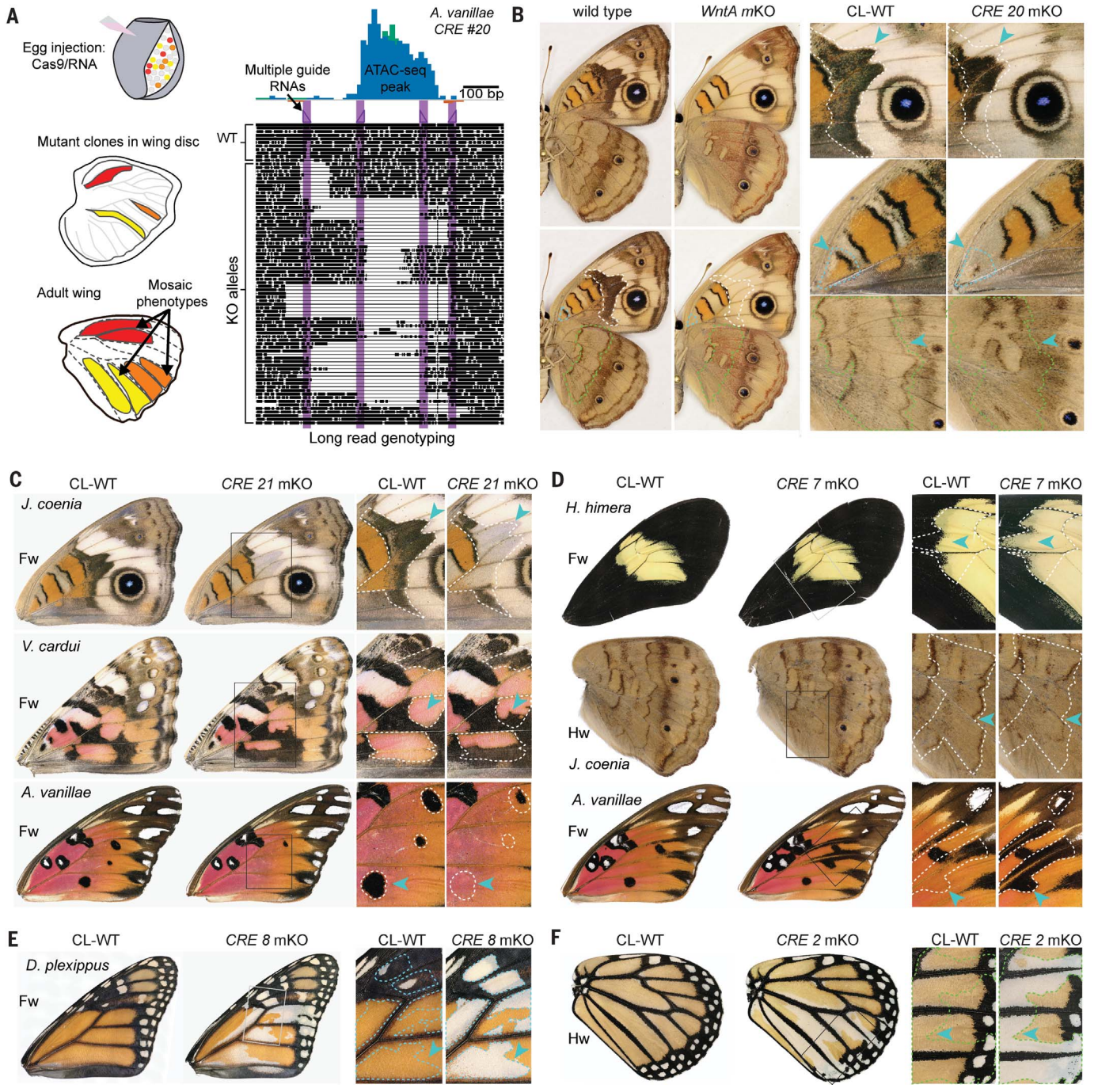

Fig. 2. In vivo mosaic deletions of *WntA* CREs reveal evolutionarily conserved wing pattern development functions. (**A**) Shotgun deletion generates butterflies that are mosaic for different deletion lengths. bp, base pair; WT, wild type. (**B**) A *J. coenia WntA* null mutant shows loss of *WntA*-expressing color pattern elements (left). This effect is phenocopied by the *CRE 20* mosaic knockout (mKO). (**C**) *CRE 21* ortholog mKOs across nymphaline (*J. coenia* and *V. cardui*) and heliconiine (*A. vanillae*) species illustrate deep evolutionary conservation of wing pattern ground plan CREs. (**D**) *CRE 7* ortholog mKOs across nymphaline (*V. cardui*) and heliconiine (*H. himera* and *A. vanillae*) species suggest that the highly divergent *Heliconius* wing patterns share a deep regulatory architecture with the nymphalid ground plan. (**E** and **F**) *D. plexippus* lineage-specific wing pattern CREs illustrated by *CRE 8* (E) and *CRE 2* (F) mKOs. CL-WT refers to contralateral wings with mostly or completely wild-type color patterns from the same individuals as the pictured mKO phenotypes. Cyan arrowheads point to asymmetric color patterns. Fw, forewing; Hw, hindwing.

of different lengths (Fig. 2A). The mosaic nature of the shotgun mutations, however, limits our ability to make precise mechanistic conclusions about the functions encoded within individual CREs. Therefore, to link specific color pattern effects to specific mutant alleles, we in-crossed *V. cardui* G_0 crispants to generate F_1's and then in-crossed further to get F_2 germline mutants bearing deletions in *CRE 23* (Fig. 3, E and F, and fig. S7). We confirmed heterozygous and compound inheritance of deletions within this single CRE (fig. S7), which caused different color pattern changes exemplified by specific gain-of-function effects in dorsal and ventral melanic subelements on the forewings (Fig. 3, E and F). This allelic series shows that small changes in a single CRE are enough to cause localized phenotypic

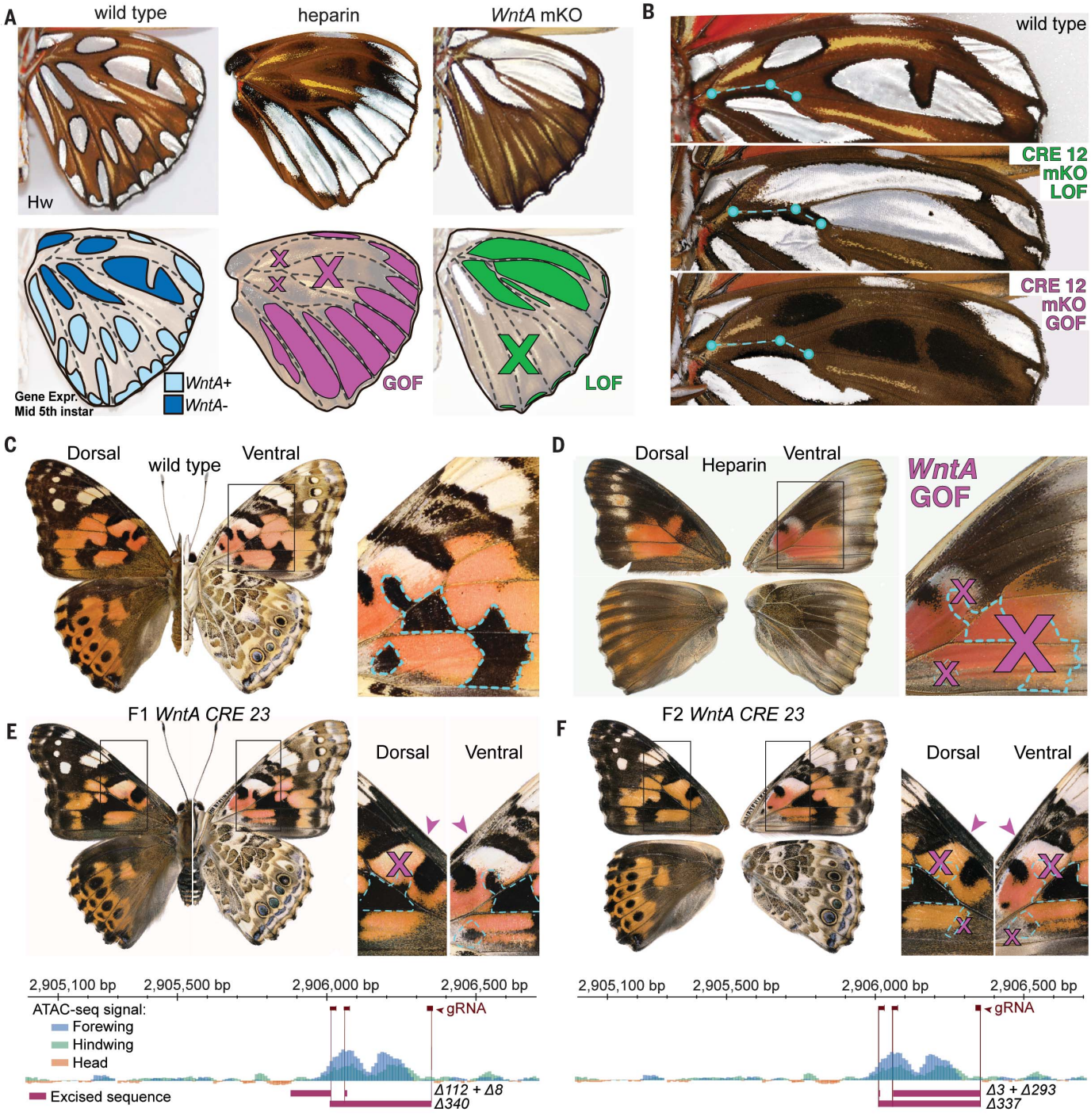

Fig. 3. Positive and negative regulatory activity is a characteristic of color pattern regulation. (**A**) Heparin injections [gain-of-function (GOF), magenta] and *WntA* knockouts [loss-of-function (LOF), green] in *A. vanillae* highlight the effects of experimental manipulations of the *WntA* signaling axis. Expr., expression. (**B**) Mosaic shotgun deletions of *WntA CRE 12* in *A. vanillae* variably result in expansion and contraction of the anterior hindwing silver spots, consistent with a *WntA* LOF and GOF, respectively. (**C**) Wild-type *V. cardui* butterfly with closeup of ventral middle forewing region. (**D**) Heparin injections in *V. cardui* illustrate the *WntA* GOF phenotype. (**E** and **F**) F_1 (E) and F_2 (F) *WntA CRE 23* deletion in *V. cardui*. Each individual represents a different combination of deletion alleles (fig. S7). Cyan dots and dashed annotations show wing landmarks. The "X" marks indicate an absence of pattern with respect to the wild-type phenotype. Arrowheads point to the extension of melanic pigmentation. gRNA, guide RNA.

changes (Fig. 3E) and illustrates the role of negative, silencer-like regulatory directives during patterning.

In testing the wing patterning functions of many orthologous and lineage-specific regulatory regions across five different butterfly species, we made several discoveries that reshape our understanding of the regulatory architecture of morphological evolution. First, in contrast to traditional models, most noncoding deletions that we studied acted globally across forewings and hindwings, often without restriction to any specific *WntA*-expressing color pattern elements. These broad effects suggest an unexpected regulatory fragility to wing patterning. Similar sensitivity to perturbation was also observed for the *Heliconius* color pattern gene *optix* (*10*) and may indicate that these loci require the assembly of clusters of CREs into transcriptional hubs to mediate gene expression, consistent with emerging superenhancer models (*18*–*20*). Second, many deeply conserved wing pattern CREs are shared between nymphalid butterflies. These conserved elements control homologous ground plan components as well as divergent color patterns in species that evolved derived modes of *WntA* expression. We propose that this deep conservation reflects an ancestral regulatory homology that underlies the nymphalid ground plan. Third, noncoding regions have the capacity to both promote and suppress *WntA* color patterns. Overall, the combination of deep conservation and dense functionality of *WntA* regulatory sequences suggests a mode of evolution marked less by the gain and loss of pattern-specific enhancers and more by nuanced modification of an array of deeply ancestral, multifunctional ground plan sequences.

REFERENCES AND NOTES

1. M. Rebeiz, M. Tsiantis, *Curr. Opin. Genet. Dev.* **45**, 115–123 (2017).
2. A. Monteiro, M. D. Gupta, *Curr. Top. Dev. Biol.* **119**, 205–226 (2016).
3. C. Concha *et al.*, *Curr. Biol.* **29**, 3996–4009.e4 (2019).
4. A. Martin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12632–12637 (2012).
5. A. Martin, R. D. Reed, *Dev. Biol.* **395**, 367–378 (2014).
6. A. Mazo-Vargas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10701–10706 (2017).
7. J. R. Gallant *et al.*, *Nat. Commun.* **5**, 4817–4827 (2014).
8. J. R. Dixon *et al.*, *Nature* **485**, 376–380 (2012).
9. S. M. Van Belleghem *et al.*, *Nat. Ecol. Evol.* **1**, 0052 (2017).
10. J. J. Lewis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24174–24183 (2019).
11. A. Mazo-Vargas, R. D. Reed, Deep cis-regulatory homology of the butterfly wing pattern groundplan, dataset, Dryad (2022); https://doi.org/10.5061/dryad.8cz8w9gpr.
12. J. Zhang, Q. Cong, J. Shen, P. A. Opler, N. V. Grishin, bioRxiv 829887 [Preprint] (2019).
13. H. F. Nijhout, *The Development and Evolution of Butterfly Wing Patterns* (Smithsonian Institution Press, 1991).
14. H. F. Nijhout, G. A. Wray, *Biol. J. Linn. Soc. Lond.* **33**, 345–365 (1988).
15. M. Kazemian *et al.*, *Genome Biol. Evol.* **6**, 2301–2320 (2014).
16. L. Minnoye *et al.*, *Genome Res.* **30**, 1815–1834 (2020).
17. A. Sourakov, L. T. Shirai, *Trop. Lepid. Res.* **30**, 4–19 (2020).
18. F. Grosveld, J. van Staalduinen, R. Stadhouders, *Annu. Rev. Genomics Hum. Genet.* **22**, 127–146 (2021).
19. E. Z. Kvon, R. Waymack, M. Gad, Z. Wunderlich, *Nat. Rev. Genet.* **22**, 324–336 (2021).
20. A. Tsai, M. R. Alves, J. Crocker, *eLife* **8**, e45325 (2019).

ACKNOWLEDGMENTS

We thank J. Sanders for experimental and analysis input in addition to building a laboratory in the basement of his house and bringing laboratory equipment that allowed us to perform Nanopore genotyping at the beginning of the COVID-19 pandemic. We thank A. Picard Hastings and A. Agrawal for sharing monarch stocks; J. Hanly, S. Van Belleghem, and R. Papa for comments and experimental input; C. Peng, J. Kwan, B. Brack, R. Geltman, and C. Robertson for assistance rearing caterpillars; L. Livraghi and the personnel from genomics core facilities at Cornell University, Duke University, and the University of Maryland–Baltimore for DNA sequencing; and A. Fungtammasan at DNANexus for preassembling PacBio genomes. **Funding:** This study was supported by National Science Foundation grant DGE-1650441 (to A.M.-V.; Graduate Research Fellowship Program); National Science Foundation grants IOS-1656514, IOS-1753559, and IOS-2128164 (to R.D.R.); and National Science Foundation grant IOS-1656553 (to A.M.). **Authors contributions:** Conceptualization: A.M.-V. and R.D.R. Investigation: A.M.-V. Experiments: A.M.-V. and A.W. Writing – original draft: A.M.-V. and R.D.R. Writing – review & editing: A.M.-V. and R.D.R. Simulation analysis: A.M.L. and P.W.M. Resources: K.R.L.v.d.B., L.Z., J.J.L., A.M., and R.D.R. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Sequence data from ATAC-seq and Hi-C sequencing that support the findings of this study have been deposited in the National Center for Biotechnology Information BioProject PRJNA695303 under BioSamples accession nos. SAMN17611226 to SAMN17611270 and SRX8598136. Additionally, all butterfly wings showing mutations were scanned and deposited in a Dryad repository (*11*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi9407
Materials and Methods
Figs. S1 to S11
Tables S1 to S7
References (*21*–*47*)
MDAR Reproducibility Checklist
Data S1 and S2

View/request a protocol for this paper from *Bio-protocol*.

Submitted 17 April 2021; accepted 7 September 2022
10.1126/science.abi9407

Deep cis-regulatory homology of the butterfly wing pattern ground plan

Anyi Mazo-VargasAnna M. LangmüllerAlexis WilderKarin R. L. van der BurgJames J. LewisPhilipp W. MesserLinlin ZhangArnaud MartinRobert D. Reed

Science, 378 (6617), • DOI: 10.1126/science.abi9407

The butterfly's grand ground plan

In the 1920s, biologists proposed that butterfly wing pattern diversity evolved as variations of a ground plan of pattern elements that vary in color, shape, and position between different species. Mazo-Vargas *et al*. found that major aspects of this ground plan are determined by an ancient array of deeply conserved noncoding DNA sequences (see the Perspective by Espeland and Podsiadlowski). These regulatory sequences can have both positive and negative effects, and nuanced interactions between noncoding regions sculpt wing patterns. Deep homology of complex, rapidly evolving traits can thus be reflected in noncoding genomic sequences. —LMZ and DJ