

Current Biology

Aristaless Controls Butterfly Wing Color Variation Used in Mimicry and Mate Choice

Highlights

- Butterfly wing color maps to a putative cis-regulatory element of *aristaless1*
- Aristaless1 is differentially expressed between white and yellow wings
- CRISPR knockout of *aristaless1* turns white wings yellow
- Wing color has been shared among species via hybridization

Authors

Erica L. Westerman,
Nicholas W. VanKuren,
Darli Massardo, ..., Nathan Buerkle,
Stephanie E. Palmer,
Marcus R. Kronforst

Correspondence

mkronforst@uchicago.edu

In Brief

Westerman, VanKuren et al. show that butterfly wing color maps to a putative cis-regulatory element adjacent to two *aristaless* genes. The genes are differentially expressed between white and yellow wings and CRISPR knockout of *aristaless1* causes white wings to develop yellow. Both colors have been shared among species via hybridization.



Aristaless Controls Butterfly Wing Color Variation Used in Mimicry and Mate Choice

Erica L. Westerman,^{1,2,10} Nicholas W. VanKuren,^{1,10} Darli Massardo,¹ Ayşe Tenger-Trolander,¹ Wei Zhang,^{1,3} Ryan I. Hill,⁴ Michael Perry,⁵ Erick Bayala,¹ Kenneth Barr,^{1,6} Nicola Chamberlain,⁷ Tracy E. Douglas,⁸ Nathan Buerkle,⁹ Stephanie E. Palmer,⁹ and Marcus R. Kronforst^{1,11,*}

¹Department of Ecology & Evolution, University of Chicago, Chicago, IL 60637, USA

²Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

³State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

⁴Department of Biological Sciences, University of the Pacific, Stockton, CA 95211, USA

⁵Department of Biology, New York University, New York, NY 10003, USA

⁶Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL 60637, USA

⁷FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA

⁸Department of BioSciences, Rice University, Houston, TX 77251, USA

⁹Department of Organismal Biology & Anatomy, University of Chicago, Chicago, IL 60637, USA

¹⁰These authors contributed equally

¹¹Lead Contact

*Correspondence: mkronforst@uchicago.edu

<https://doi.org/10.1016/j.cub.2018.08.051>

SUMMARY

Neotropical *Heliconius* butterflies display a diversity of warningly colored wing patterns, which serve roles in both Müllerian mimicry and mate choice behavior. Wing pattern diversity in *Heliconius* is controlled by a small number of unlinked, Mendelian “switch” loci [1]. One of these, termed the *K* locus, switches between yellow and white color patterns, important mimicry signals as well as mating cues [2–4]. Furthermore, mate preference behavior is tightly linked to this locus [4]. *K* controls the distribution of white versus yellow scales on the wing, with a dominant white allele and a recessive yellow allele. Here, we combine fine-scale genetic mapping, genome-wide association studies, gene expression analyses, population and comparative genomics, and genome editing with CRISPR/Cas9 to characterize the molecular basis of the *K* locus in *Heliconius* and to infer its evolutionary history. We show that white versus yellow color variation in *Heliconius cydno* is due to alternate haplotypes at a putative cis-regulatory element (CRE) downstream of a tandem duplication of the homeodomain transcription factor *aristaless*. *Aristaless1* (*al1*) and *aristaless2* (*al2*) are differentially regulated between white and yellow wings throughout development with elevated expression of *al1* in developing white wings, suggesting a role in repressing pigmentation. Consistent with this, knockout of *al1* causes white wings to become yellow. The evolution of wing color in this group has been marked by retention of the ancestral yellow color in many lineages, a single origin of white coloration

in *H. cydno*, and subsequent introgression of white color from *H. cydno* into *H. melpomene*.

RESULTS

The diverse color patterns of neotropical *Heliconius* butterflies are well known for their role in Müllerian mimicry [5–9], or mimicry among mutually toxic species [10]. In *Heliconius*, many instances of mimicry involve the sharing of color patterns among different *Heliconius* species, all of which are chemically defended and unpalatable to predators [5, 11]. In addition to serving as signals to predators—both as warning coloration and in the context of mimicry—*Heliconius* wing patterns also play an essential role in intraspecific communication. In particular, a number of *Heliconius* species and subspecies have been shown to mate assortatively, and this is driven by male preference for females with conspecific color patterns [2–4, 12–14]. Wing patterns are highly variable within and between *Heliconius* species, but the genetic architecture of this variation is dominated by a small number of large-effect loci [15–17]. These largely unlinked, Mendelian loci switch color pattern elements on particular portions of the wing [18]. Three of the four largest effect switch loci have been positionally cloned and traced back to the actions of specific genes: *optix*, *WntA*, and *cortex* [19–21]. The fourth switch locus, called the *K* locus, switches light portions of the wing between white and yellow, but the molecular basis of this locus has not been identified.

Species and subspecies in the *Heliconius cydno* complex display a diversity of color patterns resulting primarily from variation in melanin patterning, as well as white versus yellow variation controlled by the *K* locus. This phenotypic variation exists because across the range of *H. cydno*, which spans a number of subspecies as well as closely related species such as *H. pachinus*, members of this clade are precise co-mimics of distantly related *Heliconius* species such as *H. sapho*,



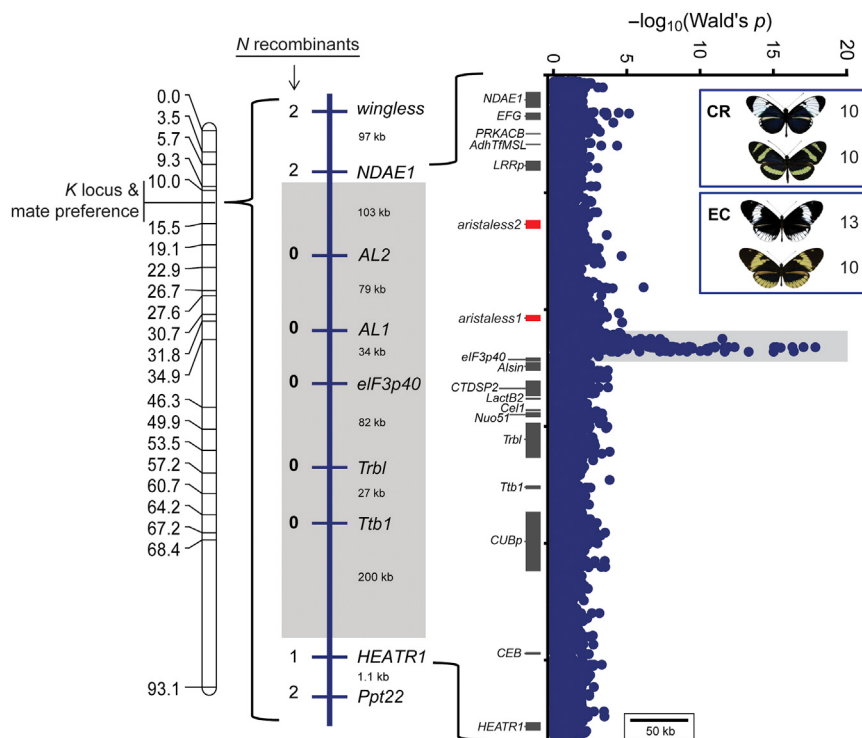


Figure 1. Genetic Mapping of the *Heliconius cydno* K Locus

Linkage and QTL mapping have shown that wing color and male color preference are genetically linked on chromosome 1 (left). Fine mapping yellow versus white wing color, using 127 F2 inter-cross hybrids between white *H. cydno galanthus* and yellow *H. cydno alithea* identified a 525 kbp zero-recombinant window (middle). This region, from *NDAE1* to *HEATR1*, spans 16 genes, including a tandem duplication of the developmental transcription factor *aristaless*. Genome-wide association mapping of wing color based on 43 whole-genome sequences (11,499,871 variants) identified a single, strong peak of association downstream of *aristaless1*, including three perfectly associated variants (right).

ated SNPs fell in a non-coding intergenic region 28 kbp downstream of *aristaless1* (*al1*) and *aristaless2* (*al2*), a Lepidoptera-specific tandem duplication of the developmental transcription factor *aristaless* [22]. *Aristaless* is a paired-like homeo-domain transcription factor with important roles in appendage patterning in *Drosophila* [23], and *al2* expression has

been shown to correlate with discal cell wing pattern elements in a variety of butterflies [22], making one or both of these genes good candidates for the *Heliconius* K locus. To explore the potential function of *al1* and *al2* in *Heliconius* wing coloration, we analyzed expression of these genes in developing wing discs of white and yellow *H. cydno* using quantitative real-time PCR (qPCR). Interestingly, we found largely non-overlapping expression patterns, with *al1* being highly expressed throughout development in white *H. cydno* (Figure 2A) and *al2* being expressed in yellow *H. cydno*, albeit at a lower level (Figure 2B). The white color of scales on *Heliconius* wings is structurally based, whereas yellow scales are produced by those same scales producing or importing a yellow pigment, the omochrome precursor 3-hydroxy-L-kynurenine (3-OHK) [24]. Because the white K locus allele, which is the absence of pigment, is dominant to the yellow allele, we hypothesized that the white allele may function as a dominant repressor of pigmentation. If *al1* is functionally responsible for switching wing color, we hypothesized that eliminating expression of *al1*, which is expressed at high levels during the development of white wings, should result in a yellow pigmented wing. To test this, we used CRISPR/Cas9 [25, 26] to knockout *al1* in white *H. cydno galanthus*, and we observed the effect in first-generation (G0) individuals that were somatic mosaics. In support of our hypothesis, *al1* knockout resulted in streaks of yellow clones spanning the white patch on the forewing (Figures 2C–2E). Experiments with white *H. cydno alithea* produced similar results (Figure S2). Our data suggest that the narrow genomic interval we mapped downstream of *al1* is a cis-regulatory element that leads to differential expression of the *aristaless* genes in developing butterflies, ultimately generating white or yellow wings.

H. eleuchia, and *H. hewitsoni*. Previously, we found that the K locus plays a dominant role in mediating assortative mating between *H. cydno galanthus* and *H. pachinus*, white and yellow sister taxa from Costa Rica (Figure S1A), because males use wing color (white versus yellow) as the primary cue in choosing females [4]. Furthermore, male mate preference is genetically linked to the K locus itself [4]. Our previous genetic and QTL mapping results showed that the Mendelian K locus and quantitative variation in male mate preference both map to the same end of chromosome 1 in the vicinity of *wingless*, a color patterning candidate gene. Further supporting the linkage between color and preference, we also found that white and yellow males from the polymorphic subspecies *H. cydno alithea* differ in their preference for white and yellow females; white males do not, on average, prefer a color, but yellow males significantly prefer yellow females [2].

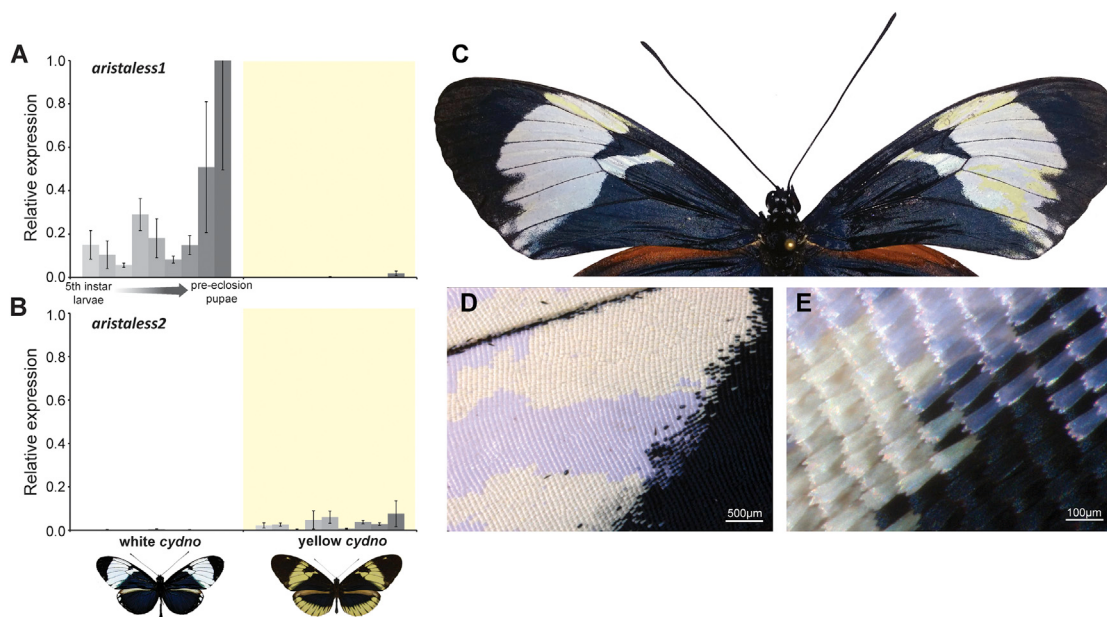


Figure 2. *Aristaless1* Controls *Heliconius* Wing Color

(A and B) Gene expression levels (mean $\pm$ SEM) of *aristaless1* (A) and *aristaless2* (B) differ between white and yellow winged butterflies over the course of wing disc development, with *aristaless1* upregulated in white *H. cydno*.

(C–E) Mosaic CRISPR/Cas9 knockout of *aristaless1* in white *H. cydno* generates yellow pigmented wing scales instead of wild-type white scales. See also Figure S2.

There are multiple, well-documented instances of adaptive introgression of wing pattern mimicry in *Heliconius* butterflies, many involving *H. melpomene* donating its color pattern to other species [27, 28]. It has been hypothesized that yellow-winged subspecies of *H. cydno*, including yellow *H. cydno alithea* and *H. pachinus*, may have independently acquired their wing color from sympatric *H. melpomene* [29]. We tested this by scanning the *K* locus interval for evidence of introgression between yellow winged *H. cydno* and yellow winged *H. melpomene* using the *D*-statistic [30], but our results revealed no evidence of yellow haplotype introgression (Figure 3A). In contrast, we did find a strong *D*-statistic signature (2.6–2.62 Mbp: $D = 0.421 \pm 0.113$, $Z = 3.735$, $p = 0.0002$) between white *H. cydno* and the white winged *H. melpomene cythera*, indicative of introgression of the white haplotype (Figure 3A). This signal was centered directly on the putative CRE downstream of *al1*. This region highlighted by the *D*-statistic also showed a peak in the related f_d statistic (2.6–2.62 Mbp: $f_d = 0.376 \pm 0.101$, $Z = 3.726$, $p = 0.0002$), as well as reduced DNA sequence divergence, d_{xy} , between white *H. cydno* and white *H. melpomene* (Figure 3A). These signatures at the putative CRE stood out relative to chromosome 1 as a whole (Figure S1B), and the focal 20 kbp window was significantly different from windows spanning chromosome 1 (D : $U = 2051$, $p = 2.94 \times 10^{-7}$, f_d : $U = 1980$, $p = 1.98 \times 10^{-7}$, d_{xy} : $U = 3438$, $p = 0.0002$). Furthermore, comparison of a phylogenetic tree inferred using SNPs in the *K* locus (Figure 3B) to a tree inferred from 35.6 million genome-wide SNPs (Figure 3C) revealed instances of discordance indicative of introgression in the color-determining region: the *K* locus phylogeny grouped samples by wing color, while the genome-wide phylogeny grouped samples by taxon. The introgression signal was specific to the

K locus, as trees based on intervals adjacent to, but excluding *K*, also grouped samples by taxon instead of color (Figure S1C). The grouping of all white *H. melpomene* *K* locus haplotypes with white *H. cydno* haplotypes suggests that the white haplotype originated in *H. cydno* and was subsequently transferred to *H. melpomene*. Introgression appears to have been recent and spatially restricted because the *K* locus tree clustered white *H. melpomene cythera* samples with white *H. cydno alithea*, both of which co-occur in western Ecuador. Similarly, the small cluster of yellow *H. melpomene* samples that group with yellow *H. cydno* suggests that a minority of the yellow variation present in *H. melpomene* was acquired from *H. cydno*. Again, this appears to have been recent introgression because the *K* locus tree clustered these two *H. melpomene* samples, which were collected from the same site in Costa Rica, with yellow *H. pachinus* from Costa Rica.

Introgression at the *K* locus appears to involve movement of alleles from *H. cydno* to *H. melpomene*. Furthermore, these results suggest that the *H. cydno* lineage was ancestrally yellow and that the current widespread yellow coloration in the group is due to segregating ancestral variation. Sometime after the origin of the *H. cydno* clade, there was a new innovation, the origin of white wing coloration, and this was subsequently exported to *H. melpomene* via introgression. This evolutionary inference allowed us to further dissect potential regulation of the *K* locus. If white is the derived phenotype, we might expect to see evidence of novel functional variants associated with the white haplotype in the location of the putative CRE. Indeed, we found white-associated SNPs in the location of the putative CRE that resulted in novel binding sites for transcription factors *Mitf*, *biniou*, *vismay*, and *aristaless* itself (Figure S3). However, the

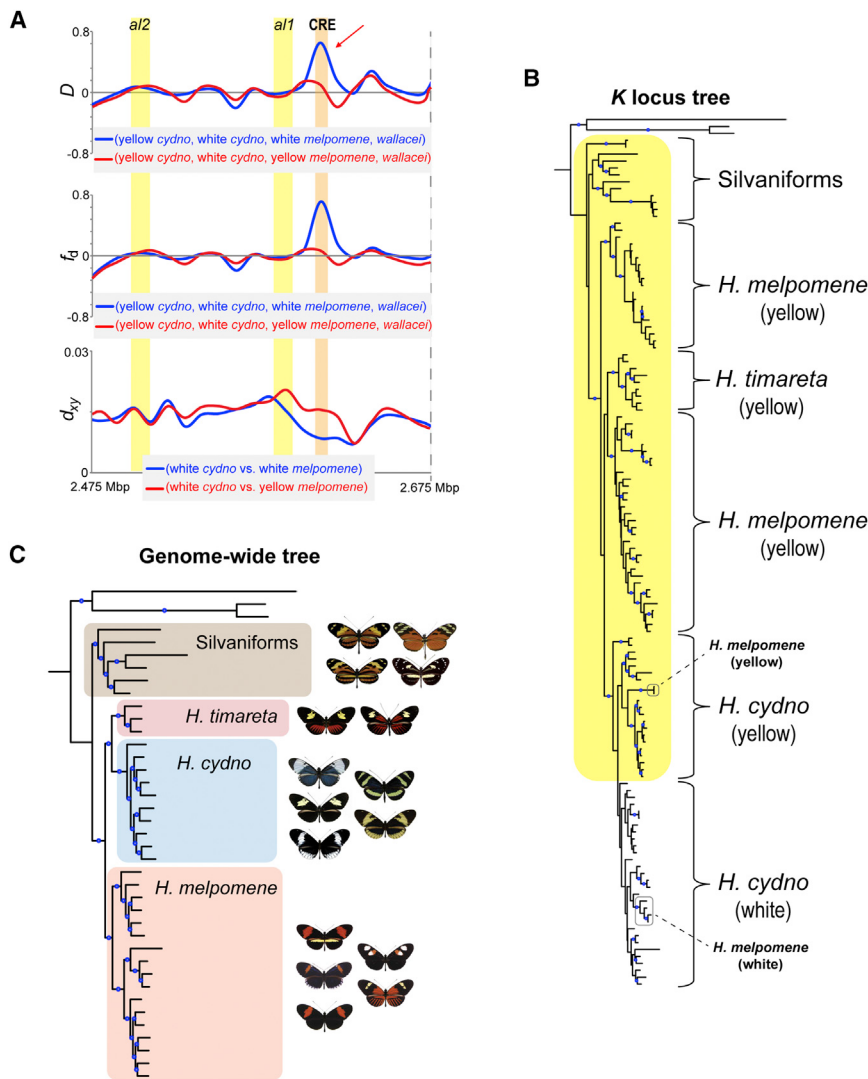


Figure 3. Evolution of Wing Color across *Heliconius* Butterflies

(A) Sliding-window analyses of Patterson's D -statistic, f_d , and d_{xy} identified peaks of biased allele sharing between white *H. cydno* and white *H. melpomene* at the putative CRE downstream of *al1*.

(B) A gene genealogy based on 30 kbp encompassing the putative CRE reveals evidence of ancestral yellow variation across *Heliconius*, as well as instances of putative introgression from *H. cydno* into *H. melpomene*. The close affiliation of *H. timareta* and *H. melpomene* on the *K* locus tree may also indicate introgression but that remains less certain because of the highly variable position of *H. timareta* across gene genealogies (Figure S1C). Internal nodes with 100% bootstrap support are marked with a blue circle.

(C) Phylogeny of the *melpomene/cydno/silvaniform* clade of *Heliconius* inferred based on > 35 million genome-wide SNPs. See also Figure S1.

Wallace first proposed the concept of evolution by natural selection [33]. Both Henry Walter Bates and Fritz Müller built their concepts of mimicry, respectively known as Batesian and Müllerian mimicry, around their observations of Amazonian butterflies [10, 34]. *Heliconius* butterflies were much discussed in those early days of mimicry theory and have continued to fuel our understanding of adaptation and diversification ever since. Because *Heliconius* color patterns also contribute to assortative mating and reproductive isolation [13], they also provide a window into the process of speciation—in particular, the mechanisms by which divergent ecological selection may facilitate speciation [35]. Recently,

general pattern was of potentially greater regulation of genes on the yellow haplotype, with transcription factors generally having higher binding affinities for yellow haplotype sequences than white haplotype sequences across the *K* locus interval (Figure S3). Furthermore, white-associated alleles at two highly associated SNPs caused disruption of putative *Mitf* and *vismay* binding sites (Figure S3). The potential involvement of *Mitf* here suggests intriguing parallels between *Heliconius* wing coloration and the developmental genetics of stripe patterning in rodents, where an *aristaless*-like gene, *Alx3*, interacts with *Mitf* to switch between light and dark hairs [31]. A single copy of *Mitf* is present in *Heliconius* (HMELO06056 in *H. melpomene*, release 2, and HEL_005372 in *H. erato lativitta*) and is expressed in developing wings [32], but additional data will be required to compare the color patterning pathways.

DISCUSSION

Wing pattern mimicry in butterflies has served as an important example of adaptation since Charles Darwin and Alfred Russell

the molecular basis of *Heliconius* color pattern variation has been revealed, with critical observations connecting different Mendelian switch loci to the effects of the genes *optix*, *WntA*, and *cortex*. Here, we have shown that a central component of mimicry in *Heliconius cydno* and related species, what has historically been called the *K* locus, is controlled by the gene *aristaless1*. The white versus yellow wing color that is controlled by *aristaless1* serves as both a mimicry signal as well as a mating cue, making this discovery relevant to understanding the genetics of both adaptation and speciation in the clade.

Our findings raise important questions that have yet to be resolved. First, we still do not understand the regulation and developmental biology of this color switch. Our results indicate that *al1* contributes to the development of distinct wing colors but is *al2* involved as well? Previous work indicates that *al2* expression may play a widespread role in patterning discal cell elements on the wings of butterflies [22]. It is unclear to what extent those observations intersect with our characterization of the *K* locus here. Even within *Heliconius*, we do not know whether *al1* controls color pattern variation in other subclades,

such as the pupal mating species *H. eleuchia*, *H. sapho*, and *H. hewitsoni*, which are variably yellow and white and are the co-mimics of the *H. cydno* clade. Finally, we have previously shown that mate preference is also genetically linked to the *K* locus [4], and it remains to be seen whether this color associated region, or a different linked region, underlies mate preference variation. Future work, combining behavior, genomics, and functional genetics will help to clarify these important, outstanding questions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
 - Genetic mapping
 - Genome-wide association mapping
 - qRT-PCR
 - CRISPR/Cas9 knockouts
 - Evolutionary analyses
 - Transcription factor binding site analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and three tables and can be found with this article online at <https://doi.org/10.1016/j.cub.2018.08.051>.

ACKNOWLEDGMENTS

We thank Sumitha Nallu and Carlos Sahagun for assistance in the lab and greenhouse. We also thank Nancy Greig, Durrell Kapan, and Larry Gilbert for help and guidance with various aspects of this project. This project was funded by a Chicago Biomedical Consortium Postdoctoral Research Grant to E.L.W., Big Ideas Generator funding from the University of Chicago and Alfred P. Sloan Foundation funding to S.E.P., NIH grant P50GM068763 to Harvard's FAS Center for Systems Biology, the Pew Biomedical Scholars program, University of Chicago's Neubauer funds, NSF grant IOS-1452648 and NIH grant GM108626 to M.R.K.

AUTHOR CONTRIBUTIONS

E.W. and N.V.K. sequenced butterfly genomes, performed GWAS data analysis, and contributed to writing the manuscript; D.M. generated CRISPR/Cas9 knockouts and contributed to writing the manuscript; A.T.T. performed *al1* & *al2* qPCR and contributed to fine mapping; W.Z. performed phylogenetic and evolutionary analyses, assisted with GWAS analyses, and contributed to writing the manuscript; R.I.H. performed fine mapping; M.P. assisted with CRISPR/Cas9 experiments; E.B. and K.B. performed transcription factor binding site analyses; N.C. collected samples for qRT-PCR and assisted with fine mapping; T.E.D. generated butterfly crosses; N.B. and S.P. planned research and assisted with data analysis; M.R.K. planned and directed the research, performed crosses, collected tissue, assisted with data analysis, and contributed to writing the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 19, 2018

Revised: July 24, 2018

Accepted: August 22, 2018

Published: October 25, 2018

REFERENCES

1. Kronforst, M.R., and Papa, R. (2015). The functional basis of wing patterning in *Heliconius* butterflies: the molecules behind mimicry. *Genetics* 200, 1–19.
2. Chamberlain, N.L., Hill, R.I., Kapan, D.D., Gilbert, L.E., and Kronforst, M.R. (2009). Polymorphic butterfly reveals the missing link in ecological speciation. *Science* 326, 847–850.
3. Kronforst, M.R., Young, L.G., and Gilbert, L.E. (2007). Reinforcement of mate preference among hybridizing *Heliconius* butterflies. *J. Evol. Biol.* 20, 278–285.
4. Kronforst, M.R., Young, L.G., Kapan, D.D., McNeely, C., O'Neill, R.J., and Gilbert, L.E. (2006). Linkage of butterfly mate preference and wing color preference cue at the genomic location of *wingless*. *Proc. Natl. Acad. Sci. USA* 103, 6575–6580.
5. Brown, K.S. (1981). The biology of *Heliconius* and related genera. *Annu. Rev. Entomol.* 26, 427–456.
6. Eltringham, H. (1916). On specific and mimetic relationships in the genus *Heliconius*. *Trans. Entomol. Soc. Lond.* 1916, 101–148.
7. Kapan, D.D. (2001). Three-butterfly system provides a field test of mullerian mimicry. *Nature* 409, 338–340.
8. Turner, J.R.G. (1976). Adaptive radiation and convergence in subdivisions of the butterfly genus *Heliconius* (Lepidoptera: Nymphalidae). *Zool. J. Linn. Soc.* 58, 297–308.
9. Turner, J.R.G. (1981). Adaptation and evolution in *Heliconius*: A defense of Neo Darwinism. *Annu. Rev. Ecol. Syst.* 12, 99–121.
10. Müller, F. (1879). *Ituna* and *Thyridia*; a remarkable case of mimicry in butterflies. *Trans. Entomol. Soc. Lond.* 1879, xx–xxix.
11. Engler-Chaouat, H.S., and Gilbert, L.E. (2007). *De novo* synthesis vs. sequestration: negatively correlated metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. *J. Chem. Ecol.* 33, 25–42.
12. Jiggins, C.D., Estrada, C., and Rodrigues, A. (2004). Mimicry and the evolution of premating isolation in *Heliconius melpomene* Linnaeus. *J. Evol. Biol.* 17, 680–691.
13. Jiggins, C.D., Naisbit, R.E., Coe, R.L., and Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. *Nature* 411, 302–305.
14. Merrill, R.M., Gompert, Z., Dembeck, L.M., Kronforst, M.R., McMillan, W.O., and Jiggins, C.D. (2011). Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution* 65, 1489–1500.
15. Nijhout, H.F., Wray, G., and Gilbert, L.E. (1990). An analysis of the phenotypic effects of certain color pattern genes in *Heliconius* (Lepidoptera: Nymphalidae). *Biol. J. Linn. Soc. Lond.* 40, 357–372.
16. Sheppard, P.M., Turner, J.R.G., Brown, K.S., Benson, W.W., and C, S.M. (1985). Genetics and the evolution of Mullerian mimicry in *Heliconius* butterflies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 308, 433–613.
17. Turner, J.R.G. (1971). The genetics of some polymorphic forms of the butterflies *Heliconius melpomene* (Linnaeus) and *H. erato* (Linnaeus). II. The hybridization of subspecies from Surinam and Trinidad. *Zoologica* 56, 125–157.
18. Naisbit, R.E., Jiggins, C.D., and Mallet, J. (2003). Mimicry: developmental genes that contribute to speciation. *Evol. Dev.* 5, 269–280.
19. Martin, A., Papa, R., Nadeau, N.J., Hill, R.I., Counterman, B.A., Halder, G., Jiggins, C.D., Kronforst, M.R., Long, A.D., McMillan, W.O., and Reed, R.D. (2012). Diversification of complex butterfly wing patterns by repeated regulatory evolution of a *Wnt* ligand. *Proc. Natl. Acad. Sci. USA* 109, 12632–12637.
20. Nadeau, N.J., Pardo-Diaz, C., Whibley, A., Supple, M.A., Saenko, S.V., Wallbank, R.W., Wu, G.C., Maroja, L., Ferguson, L., Hanly, J.J., et al.

- (2016). The gene *cortex* controls mimicry and crypsis in butterflies and moths. *Nature* 534, 106–110.
21. Reed, R.D., Papa, R., Martin, A., Hines, H.M., Counterman, B.A., Pardo-Diaz, C., Jiggins, C.D., Chamberlain, N.L., Kronforst, M.R., Chen, R., et al. (2011). *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333, 1137–1141.
 22. Martin, A., and Reed, R.D. (2010). *Wingless* and *aristaless2* define a developmental ground plan for moth and butterfly wing pattern evolution. *Mol. Biol. Evol.* 27, 2864–2878.
 23. Campbell, G., Weaver, T., and Tomlinson, A. (1993). Axis specification in the developing *Drosophila* appendage: the role of *wingless*, *decapentaplegic*, and the homeobox gene *aristaless*. *Cell* 74, 1113–1123.
 24. Gilbert, L.E., Forrest, H.S., Schultz, T.D., and Harvey, D.J. (1988). Correlations of ultrastructure and pigmentation suggest how genes control development of wing scales in *Heliconius* butterflies. *J. Res. Lepid.* 26, 141–160.
 25. Li, X., Fan, D., Zhang, W., Liu, G., Zhang, L., Zhao, L., Fang, X., Chen, L., Dong, Y., Chen, Y., et al. (2015). Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nat. Commun.* 6, 8212.
 26. Mazo-Vargas, A., Concha, C., Livraghi, L., Massardo, D., Wallbank, R.W.R., Zhang, L., Papador, J.D., Martinez-Najera, D., Jiggins, C.D., Kronforst, M.R., et al. (2017). Macroevolutionary shifts of *WntA* function potentiate butterfly wing-pattern diversity. *Proc. Natl. Acad. Sci. USA* 114, 10701–10706.
 27. Heliconius Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487, 94–98.
 28. Zhang, W., Dasmahapatra, K.K., Mallet, J., Moreira, G.R., and Kronforst, M.R. (2016). Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biol.* 17, 25.
 29. Gilbert, L.E. (2003). Adaptive novelty through introgression in *Heliconius* wing patterns: evidence for a shared genetic “tool box” from synthetic hybrid zones and a theory of diversification. In *Ecology and Evolution Taking Flight: Butterflies as Model Systems*, C.L. B. W.B. W. and P.R. E. eds. (Chicago: University of Chicago Press), pp. 281–318.
 30. Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252.
 31. Mallarino, R., Henegar, C., Mirasiera, M., Manceau, M., Schradin, C., Vallejo, M., Beronja, S., Barsh, G.S., and Hoekstra, H.E. (2016). Developmental mechanisms of stripe patterns in rodents. *Nature* 539, 518–523.
 32. Lewis, J.J., van der Burg, K.R.L., Mazo-Vargas, A., and Reed, R.D. (2016). ChIP-Seq-annotated *Heliconius erato* genome highlights patterns of cis-regulatory evolution in Lepidoptera. *Cell Rep.* 16, 2855–2863.
 33. Darwin, C., and Wallace, A.R. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zool. J. Linn. Soc.* 3, 46–50.
 34. Bates, H.W. (1862). Contributions to an insect fauna of the Amazon Valley. *Lepidoptera: Heliconidae*. *Trans. Linn. Soc. Lond.* 23, 495–566.
 35. Jiggins, C.D. (2008). Ecological speciation in mimetic butterflies. *Bioscience* 58, 541–548.
 36. Gallant, J.R., Imhoff, V.E., Martin, A., Savage, W.K., Chamberlain, N.L., Pote, B.L., Peterson, C., Smith, G.E., Evans, B., Reed, R.D., et al. (2014). Ancient homology underlies adaptive mimetic diversity across butterflies. *Nat. Commun.* 5, 4817.
 37. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
 38. Davey, J.W., Chouteau, M., Barker, S.L., Maroja, L., Baxter, S.W., Simpson, F., Joron, M., Mallet, J., Dasmahapatra, K.K., and Jiggins, C.D. (2016). Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3* 6, 695–708.
 39. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
 40. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
 41. Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339.
 42. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824.
 43. Perry, M., Kinoshita, M., Saldi, G., Huo, L., Arikawa, K., and Desplan, C. (2016). Molecular logic behind the three-way stochastic choices that expand butterfly colour vision. *Nature* 535, 280–284.
 44. Bassett, A., and Liu, J.L. (2014). CRISPR/Cas9 mediated genome engineering in *Drosophila*. *Methods* 69, 128–136.
 45. Kronforst, M.R., Hansen, M.E., Crawford, N.G., Gallant, J.R., Zhang, W., Kulathinal, R.J., Kapan, D.D., and Mullen, S.P. (2013). Hybridization reveals the evolving genomic architecture of speciation. *Cell Rep.* 5, 666–677.
 46. Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., and Jiggins, C.D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* 23, 1817–1828.
 47. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
 48. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
 49. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
 50. Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44 (W1), W242–5.
 51. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). A draft sequence of the Neandertal genome. *Science* 328, 710–722.
 52. Martin, S.H., Davey, J.W., and Jiggins, C.D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32, 244–257.
 53. Smith, J., and Kronforst, M.R. (2013). Do *Heliconius* butterfly species exchange mimicry alleles? *Biol. Lett.* 9, 20130503.
 54. Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M.A., Nowell, R.W., Mallet, J., Dasmahapatra, K.K., and Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. *Curr. Biol.* 28, 1839–1845.e3.
 55. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. *Genome Biol.* 8, R24.
 56. Shazman, S., Lee, H., Socol, Y., Mann, R.S., and Honig, B. (2014). OnTheFly: a database of *Drosophila melanogaster* transcription factors and their binding sites. *Nucleic Acids Res.* 42, D167–D171.
 57. Zhu, L.J., Christensen, R.G., Kazemian, M., Hull, C.J., Enameh, M.S., Basciotta, M.D., Brasefield, J.A., Zhu, C., Asriyan, Y., Lapointe, D.S., et al. (2011). FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* 39, D111–D117.
 58. Berg, O.G., and von Hippel, P.H. (1988). Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.* 200, 709–723.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Benzalkonium chloride	Sigma-Aldrich	Cat#63249
Cas9 protein	PNA Bio	Cat#CP01-200
Critical Commercial Assays		
TOPO TA vector	Invitrogen	Cat#K460001
Megascript T7 Kit	Ambion	Cat#ASF3257
Phusion polymerase	New England Biolabs,	Cat#M0561S
iScript cDNA Synthesis Kit	BioRad	Cat#1708890
DNeasy Blood and Tissue Kit	QIAGEN	Cat#69506
Nextera DNA Library Prep Kit	Illumina	Cat#FC-121-1030
Nextera XT Index Kit	Illumina	Cat#FC-131-1001
KAPA Hyper Prep Kit	Roche	Cat#KK8504
KAPA Single-Indexed Adaptor Kit	Roche	Cat#KK8700
Agencourt AMPure XP	Beckman Coulter	Cat#A63881
Genomic DNA Clean & Concentrator	ZYMO Research	Cat#D4010
Deposited Data		
NGS data	This paper	GenBank: PRJNA485723
Oligonucleotides		
See Table S3 for oligonucleotides	NA	NA
Software and Algorithms		
Trimmomatic Version 0.36	[37]	http://www.usadellab.org/cms/?page=trimmomatic
Bowtie2 v2.2.9	[47]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Picard v2.8.1	Broad Institute	https://broadinstitute.github.io/picard/
RealignerTargetCreator and IndelRealigner in GATK v3.4	[40]	https://software.broadinstitute.org/gatk/
UnifiedGenotyper in GATK v3.4	[48]	https://software.broadinstitute.org/gatk/
Haplotype Caller (GATK v3.4)	Broad Institute	https://software.broadinstitute.org/gatk/
RAXML	[49]	https://sco.h-its.org/exelixis/web/software/raxml/index.html
iTOL	[50]	https://itol.embl.de/
delly2	[41]	https://github.com/dellytools/delly
gemma 0.94	[42]	www.xzlab.org/software.html
Geneious v9.1.3	https://www.geneious.com/	N/A
Other		
FlyCRISPR tools	http://flycrispr.molbio.wisc.edu/tools	N/A
borosilicate needle	Sutter Instruments, Novato, CA, USA	Cat#BF100-50-10

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Marcus R. Kronforst (mkronforst@uchicago.edu).

METHOD DETAILS

Genetic mapping

We crossed a white *H. cydno galanthus* female with a yellow *H. cydno alithea* male to generate F1 hybrids and then individually paired five F1 males with yellow *H. cydno alithea* females to generate F2 backcross broods. Across the five families, we raised a total of 127 adult F2 butterflies (50 yellow, 77 white). We PCR amplified and Sanger sequenced nine markers in the *K* locus region in parents and

offspring to identify and score segregating variation, particularly heterozygous sites in F1 fathers that could be traced in F2 offspring. Primers are in Table S3. We detected a total of seven recombinants representing four unique recombination events, two on either side of the 525 kbp zero-recombinant interval (Figure 1).

Genome-wide association mapping

We performed GWA for color using 43 previously published *H. cydno galanthus* (N = 10), *H. pacheus* (N = 10), and yellow (N = 10) and white (N = 13) *H. cydno alithea* whole genome sequencing datasets [36]. Sample details are in Table S1. We trimmed adapters and low-quality regions from raw reads using Trimmomatic 0.36 [37] before mapping to the *H. melpomene* (release 2) reference genome [38] using BWA MEM v0.7.12 [39] with default settings except the -M flag to mark secondary alignments. We marked duplicate reads using PicardTools v1.92, then called SNPs and indels using the Genome Analysis ToolKit's [GATK, v3.7-0 [40]] HaplotypeCaller and GenotypeGVCFs using default parameters except expected heterozygosity (-hets) was set to 0.005. We filtered out sites with overall quality less than 1000 and sites that exhibited strong supporting read biases to generate a final SNP/indel callset. We jointly called inversions (300 bp – 3 Mb), deletions (50 bp – 3 Mb), tandem duplications (300 bp – 3 Mb), and insertions using delly 0.7.6 [41]. We included only high-quality (≥ 3 supporting read pairs), nucleotide-resolution calls in our final call set. Filtered SNP, indel, and SV calls were merged before subsequent analyses. GWA was performed using a univariate linear mixed model in GEMMA 0.94 [42]. We excluded variant sites with > 20% missing genotypes or minor allele frequency < 0.05. We analyzed p values from site-wise Wald tests.

qRT-PCR

We compared *al1* and *al2* gene expression between white *H. cydno galanthus* and yellow *H. cydno alithea* using qRT-PCR. We assessed the relative expression of genes at developmental time points spanning 5th instar larvae to late pupal stages in forewing tissue. We extracted RNA using Trizol and then synthesized cDNA using BioRad's iscript cDNA synthesis kit. For each target, we designed gene specific primers and checked for high (> 90%) primer efficiency with standard curves. Primers are in Table S3. We also assessed expression of a control gene, *elongation factor 1 alpha* (*ef1 α*), to normalize expression levels of our targets. Reactions were run on an ABI 7500 fast HT machine using the ABI sybr green 2X master mix and ABI MicroAmp Fast Optical 96 well plates. We quantified gene expression levels using the $2^{-\Delta\Delta CT}$ method.

CRISPR/Cas9 knockouts

We reared *H. cydno galanthus* and *H. cydno alithea* in greenhouse insectaries at the University of Chicago. Larvae were fed on fresh host-plant material (*Passiflora*) under a light regime of 16 h light and 8 h dark at 26–27°C, 60%–80% humidity. We used two or three single guide-RNAs (sgRNAs) to generate long deletions or frameshifts (Figure S2) [26, 43]. sgRNAs were identified using FlyCRISPR tools (<http://flycrispr.molbio.wisc.edu/tools>) to search genomic regions for GGN₁₈NGG or N₂₀NGG sequences. The specificity of candidate sgRNA sequences was assessed using BLAST to confirm there were not multiple binding sites. Target sequences are in Table S3. A sgRNA template was generated by PCR amplification with a forward primer encoding a T7 polymerase-binding site and a sgRNA target site using Phusion polymerase (New England Biolabs, Ipswich, MA, USA), and a reverse primer encoding the remainder of the sgRNA sequence [43]. *In vitro* transcription was conducted using Megascript T7 Kit (Ambion, Waltham, MA, USA) and purified by phenol–chloroform extraction and isopropanol precipitation [44].

To retrieve eggs for injection, we offered host-plants to female butterflies to lay eggs for a period of 1–3 hr. The eggs were collected and then washed for 120 s in 7.5% benzalkonium chloride (Sigma-Aldrich), rinsed in water, and dried by air ventilation, in order to soften the chorion. Eggs were arranged on a double-sided adhesive tape glued to a glass slide with the micropyle facing up to allow the caterpillars to break the eggshell. Injection mixtures containing sgRNAs and recombinant Cas9 protein (PNA Bio) were injected using a 0.5-mm borosilicate needle (Sutter Instruments, Novato, CA, USA). The concentration of Cas9 and sgRNAs varied between 125 ng/ μ l–250 ng/ μ l and 83 ng/ μ l–125 ng/ μ l, respectively. After injection, embryos were placed into humid Petri dishes until hatching (~4 days) and moved to an incubator at 26–27°C and ~80% humidity to develop. Hatchlings were collected via paintbrush and transferred to the host plant. Adults eclosed approximately 10 days post-pupation. After emerging, the adults were frozen for genotyping and pinned.

For genotyping, we isolated DNA from tissue showing mutant phenotypes, or single legs, using a phenol–chloroform protocol, and PCR amplified a section of *al1* flanking the sgRNAs target region (Primers are in Table S3.). We gel-purified PCR products, subcloned into a TOPO TA vector (Invitrogen, Carlsbad, CA, USA), and then sequenced them with dye terminator technology. The sequence data were analyzed with Geneious v9.1.3 software.

Evolutionary analyses

We performed whole genome re-sequencing for 9 adult butterflies and combined these with 117 whole genome re-sequencing datasets from NCBI (PRJNA226620 and PRJNA308754) [28, 45] and ENA (ERP002440 and PRJEB8011) [20, 46]. The raw reads were processed with Trimmomatic Version 0.36 [37] and the high quality reads were aligned to the *H. melpomene* reference genome [38] using Bowtie2 v2.2.9 with the parameter -very-sensitive-local [46]. Picard v2.8.1 (<https://broadinstitute.github.io/picard/>) was used to remove PCR duplicates. RealignerTargetCreator and IndelRealigner in GATK v3.4 was used to realign indels [40] and UnifiedGenotyper was used to call genotypes [48]. SNPs with good quality (Qual > 30) were used in the downstream analyses. Sample details are listed in Table S2.

We performed a genome-wide maximum-likelihood phylogenetic analysis by extracting polymorphic genotype calls (approximately 35.64 million SNPs) from 39 individuals (one sample per species, subspecies, morph) with good quality, converting them into PHYLIP format and constructing a genome-wide tree using RAxML [49] with GTRGAMMA model and 100 bootstrap replicates (Figure 3C, Figure S1A). We also extracted genotype calls from 126 individuals and constructed a phylogenetic tree for 30 kb regions in the same manner (Figure 3B, Figure S1C). The tree images were visualized using iTOL [50].

We applied Patterson's D -statistic [30, 51] to identify potential introgression signatures around the K locus. We used *H. wallacei* as an outgroup taxon and set the ingroup taxa as (yellow *H. c. alithea*, white *H. c. alithea*, *H. m. cythera*) and (yellow *H. c. alithea*, white *H. c. alithea*, *H. m. rosina*). We have multiple individuals per taxon, so we used the frequency of the derived allele instead of binary counts of fixed ABBA and BABA counts to detect gene flow. The D -statistic was calculated as

$$D(P_1, P_2, P_3, O) = \frac{\sum_{i=1}^n \left[(1 - \hat{P}_{i1}) \hat{P}_{i2} \hat{P}_{i3} (1 - \hat{P}_{i4}) - \hat{P}_{i1} (1 - \hat{P}_{i2}) \hat{P}_{i3} (1 - \hat{P}_{i4}) \right]}{\sum_{i=1}^n \left[(1 - \hat{P}_{i1}) \hat{P}_{i2} \hat{P}_{i3} (1 - \hat{P}_{i4}) + \hat{P}_{i1} (1 - \hat{P}_{i2}) \hat{P}_{i3} (1 - \hat{P}_{i4}) \right]} \quad (1)$$

where P_1, P_2, P_3 and P_4 refer to four taxa and $\hat{P}_{ij}$ refers to the SNP frequency in the corresponding population [30]. We also calculated the related statistic, f_d , using the following equation:

$$\hat{f} = \frac{S(P_1, P_2, P_3, O)}{S(P_1, P_D, P_D, O)} \quad (2)$$

with P_1, P_2, P_3 and O as the four taxa of the comparison [52]. P_D was the higher frequency of the derived allele from either P_2 or P_3 .

For each 20 kbp window, the standard error was calculated using a moving block bootstrap approach according to Zhang et al. [28]. Then a two tailed z-test was performed to determine if the standard error for each D and f_d was significantly different from zero, indicative of potential gene flow.

We also calculated mean pairwise sequence divergence (d_{xy}) for two comparisons, white *H. cydno* versus white *H. melpomene cythera* and white *H. cydno* versus yellow *H. melpomene rosina*, around the K locus as well as chromosome 1 as a whole. We calculated mean d_{xy} using as

$$d_{xy} = \frac{1}{n} \sum_{i=1}^n \hat{p}_{ix} (1 - \hat{p}_{iy}) + \hat{p}_{iy} (1 - \hat{p}_{ix}) \quad (3)$$

where $\hat{P}$ refers to the reference allele frequency in the corresponding population [53].

We calculated D , f_d , and d_{xy} across chromosome 1 in non-overlapping windows of 20 kbp, including windows with a minimum of 1,000 SNPs (Figure S1B), following ref [54]. For a more detailed look at the K locus (Figure 3A), we calculated D , f_d , and d_{xy} in non-overlapping windows of 10 kbp, including windows with a minimum of 500 SNPs. We compared K locus values of D , f_d , and d_{xy} to windows spanning chromosome 1 using the Mann-Whitney U test.

Transcription factor binding site analysis

We tested if SNPs in the putative *aristaless* *cis*-regulatory element differentially affect transcription factor (TF) binding between yellow and white haplotypes using two related approaches. First, we generated 21 bp sequences for the top 10 color-associated SNPs by concatenating the 10 bp flanking the SNP and the "yellow" *H. pachinus* allele (fixed in *H. pachinus* and yellow *H. c. alithea*, but absent in *H. c. galanthus*) or "white" (*H. c. galanthus*) allele. We used TomTom [55] to manually search for TFBS sequences in the white and yellow sequences using TF position weight matrices (PWMs) from the *Drosophila melanogaster* OnTheFly database [56]. We kept for further study six TFs 1) that showed a presence/absence difference between the two sequences, 2) whose binding site encompassed the SNP site, and 3) had a strong PWM score. In our second approach, we analyzed the binding affinity of these six TFs at all SNP sites spanning an 11 kbp interval that included the putative CRE. We used position count matrices collected from the Fly Factor Survey database [57] and calculated relative affinity following ref [58]. Where there were multiple matrices for a factor in the database, we selected the matrix with the highest number of counts per nucleotide. To avoid bias in matrices with few counts, we added a pseudocount of 1 to all matrices (distributed evenly across nucleotides). We generated 21 bp sequences, as before, for all SNPs in the putative CRE. We then calculate the PWM score, S_i , of each sequence by calculating

$$S_i = \sum_j \ln \frac{f(b,j)}{0.25} \quad (4)$$

where $f(b,j)$ is the frequency of base b at position j in the PWM and 0.25 is background probability of observing that nucleotide at that position. The relative binding affinity, K_i , of a sequence is then given by

$$K_i = e^{S_i - S_m} \quad (5)$$

where S_m is the score of the consensus sequence for a PWM. We fix the parameter λ to 1 [58]. This gives the predicted binding affinity of a sequence in units of K_0 , the affinity of a consensus binding site for that factor.

QUANTIFICATION AND STATISTICAL ANALYSIS

We performed genome-wide association mapping of wing color using 43 previously published *H. cydno galanthus* (N = 10), *H. pachinus* (N = 10), and yellow (N = 10) and white (N = 13) *H. cydno alithea* whole genome sequencing datasets [36]. GWA was performed using a univariate linear mixed model in GEMMA 0.94 [42]. We excluded variant sites with > 20% missing genotypes or minor allele frequency < 0.05. We analyzed p values from site-wise Wald tests. We compared *al1* and *al2* gene expression between white *H. cydno galanthus* and yellow *H. cydno alithea* using qRT-PCR. We also assessed expression of a control gene, *elongation factor 1 alpha* (*ef1α*), to normalize expression levels of our targets. We quantified gene expression levels using the $2^{-\Delta\Delta CT}$ method. Values displayed in Figures 2A and 2B are scaled means $\pm$ sem, based on three biological replicates of each color/developmental stage. To study evolution of the *K* locus, we calculated *D* and *f_d* statistics using the following taxa (yellow *H. c. alithea*, white *H. c. alithea*, *H. m. cythera*, *H. wallacei*) and (yellow *H. c. alithea*, white *H. c. alithea*, *H. m. rosina*, *H. wallacei*). These calculations used data from four samples of each ingroup taxon, and one sample of *H. wallacei*. The four *H. m. rosina* samples designated as the yellow *melpomene* group did not include the two samples subsequently shown to cluster with yellow *cydno* on the *K* locus gene genealogy. A two tailed z-test was performed to determine if the standard error for each *D* and *f_d* was significantly different from zero. We also calculated *d_{xy}* between white *H. cydno* and white *H. melpomene cythera* and between white *H. cydno* and yellow *H. melpomene rosina*, around the *K* locus as well as chromosome 1 as a whole. These calculations used data from four samples of each taxon. We compared *K* locus values of *D*, *f_d*, and *d_{xy}* to windows spanning chromosome 1 using the Mann-Whitney *U* test.

DATA AND SOFTWARE AVAILABILITY

The accession number for the sequence data reported in this paper is GenBank: PRJNA485723.