A macromutation eliminates colour patterning in captive butterflies

Candidate Journal: PLoS Genetics

Authors
Joseph J. Hanly*[,1,2], Luca Livraghi[2,3], Christa Heryanto[1], W. Owen McMillan[2], Chris D. Jiggins[3], Lawrence E. Gilbert[4], Arnaud Martin[1]

Affiliations
[1] Department of Biological Sciences, The George Washington University, Washington, DC, USA
[2] Smithsonian Tropical Research Institute, Gamboa, Panama
[3] Department of Zoology, Downing St, University of Cambridge, Cambridge, CB2 3EJ, UK
[4] Department of Integrative Biology, University of Texas, Austin. TX, USA
*: Corresponding author, email : joe.hanly@gmail.com

Funding
This work was supported by NSF awards IOS-2110534 to A.M. and IOS-2110532 to W.O.M., a GWU UFF research grant to A.M. and J.J.H., BBSRC grant BB/R007500 to C.D.J. and L.L., and past NSF grants for greenhouse facilities as well as the Worthington Foundation Endowment to L.E.G.

Abstract

Captive populations often harbor variation that is not present in the wild due to artificial selection. Recent efforts to map this variation have provided insights into the genetic and molecular basis of variation. *Heliconius* butterflies display a large array of pattern variants in the wild and the genetic basis of these patterns has been well-described. Here we sought to identify the genetic basis of an unusual pattern variant that is instead found in captivity, the *ivory* mutant, in which all scales on both the wings and body become white or yellow. Using a combination of autozygosity mapping and coverage analysis from 37 captive individuals, we identify a 78kb deletion at the *cortex* wing patterning locus as the *ivory* mutation. This deletion is undetected among 458 wild *Heliconius* genomes samples, and its dosage explains both homozygous and heterozygous *ivory* phenotypes found in captivity. The deletion spans a large 5' region of the *cortex* gene that includes a facultative 5'UTR exon detected in larval wing disk transcriptomes. CRISPR mutagenesis of this exon replicates the wing phenotypes from coding knock-outs of *cortex*, consistent with a functional role of *ivory*-deleted elements in establishing scale color fate. Population demographics reveal that the stock giving rise to the ivory mutant has a mixed origin from across the wild range of *H. melpomene*, and supports a scenario where the *ivory* mutation occurred after the introduction of *cortex* haplotypes from Ecuador. Homozygotes for the *ivory* deletion are inviable, joining 40 other examples of allelic variants that provide heterozygous advantage in animal populations under artificial selection by fanciers and breeders. Finally, our results

highlight the promise of autozygosity and association mapping for identifying the genetic basis of aberrant mutations in captive insect populations.

Introduction

Captive populations can harbour variation that is not observed in the wild. Aside from domestication for agricultural or commercial yield, animals and plants have often been selectively bred for 'beautiful', 'interesting' or 'unnatural' aesthetic traits, especially colour and pattern variants (1,2). In recent decades, the genetic basis of artificially selected variation has begun to be mapped in many organisms used in agriculture, floriculture and the pet trade, including the genetic mapping of a large number of genes responsible for flower colour variation (3,4), melanin-based coat colour in domestic mammals (1), plumage patterns and colours in birds (5,6), and chromatophore distribution in squamates and fishes (7). These variants can affect genes involved in the enzymatic production, transport and deposition of pigments like the melanin and carotenoid pathways (8,9), or can be caused by genes that affect signalling or cell type differentiation (as with chromatophore distribution in squamates(10)). The extensive set of artificial and domesticated genetic variants has been used repeatedly both for studies of genotype-to-phenotype relationship, and for understanding the genetics of disease states in humans (11).

Notably, several cases have been identified where a gene underlying an artificially selected variant is also responsible for natural variation in other populations or species. This is the case with *BCO2*, where an artificially selected protein coding variant in cattle affects milk colour (12), and a wild type regulatory variant in wall lizards causes changes to ventral scale colour (13). The gene Agouti has repeatedly been mapped in cases of pigment differences, both in cases of domestication [*e.g.* in the chicken (14), horse (15), and rabbit (16)], and in natural variation (*e.g.* in warblers (17), humans (18) and snowshoe hare (19)). This reuse of hotspot genes in wild and captive populations can even be observed at the within-species level; multiple separate variants in agouti have been linked to pigment variation in sheep - both in wild and captive populations, and caused by protein coding mutations, *cis*-regulatory mutations and copy number variation (20,21). As such, studies of domesticated variation can inform our understanding of natural variation too, both on macro- and micro-evolutionary scales

As the primary selective force in captive populations is the eye of the selector, one may often observe variation that is not seen in nature because it is not optimally fit (22–25). Fanciers and breeders have purposefully selected for variation that causes deleterious effects in combination with colour and pattern differences, applying a regime of selection that favours the phenotype of interest over fitness. This applies especially to distinctive colouration, which can be straightforward to observe and maintain in a captive stock. Examples include coat color variants for Merle dogs, where homozygotes for a retrotransposon insertion in the *SILV* gene have an increased risk of deafness and blindness (26,27), and in overo horses, where foals homozygous for a mutated Endothelin Receptor B (*EDNRB)* develop Lethal White Syndrome (28). "Lemon frost" geckos have been selected for their unique colour, but exhibit an increased risk of iridophoroma, the formation of tumours from iridophores (analogous to melanoma) (7). Such reduced fitness has also been identified in floriculture, where white

petunias with mutations to the *AN1* gene have deficient vacuole acidification and a weakened seed coat (29), though an artificially selected mutation to the same gene causing a similar effect in the morning glory does not appear to carry the same deleterious fitness effects (3). Whether they are the direct targets of selective breeding, or collateral effects of inbreeding and bottlenecks, deleterious mutations that would be purged by natural selection in the wild are common in captive populations (30).

The majority of work mapping artificially selected or domesticated variation in animals has taken place in vertebrates, with the notable exceptions of the economically important silkworm (31). *Heliconius* butterflies, a model system for the genetic study of color pattern adaptations in the wild (32,33), have also been maintained in captivity by butterfly breeders continuously since at least the mid-20th century, including at notable tourist attractions like Butterfly World in Florida. There, they have been selectively bred for hundreds of generations with special attention paid to pattern variations that deviate from phenotypes normally observed in the wild. Over an extended period of inbreeding, selection and occasional out-crossing, novel color variations have appeared in this captive population, including the 'Piano Keys' (PK) pattern, and a deleterious mutation that occurred within the PK genetic stock, here dubbed *ivory* (Fig 1).

The genetics of wing patterning have been extensively studied in *Heliconius* (34). Linkage and association mapping have pinpointed regulatory regions around a small toolkit of genes as being responsible for much of the pattern variation seen in wild populations, including the transcription factor *optix* (35–38), the signaling ligand *WntA* (39,40), and *cortex,* a *cdc20* homolog with a currently-unidentified function (41,42). We used whole genome resequencing, association mapping and autozygosity mapping to determine that the *ivory* mutation is caused by a large deletion at the patterning gene *cortex*, which likely occurred *de novo* in the captive population.

Methods

Stock history

The captive population of *Heliconius melpomene* was initially by JRG Turner and others from around Central and South America, and were transferred from his genetic research stocks at the University of Leeds to the former London Butterfly House (LBH) initiated by Clive Farrell 1981 as a multi-race hybrid population. They were then acquired from Tom Fox of LBH by R. Boender at the MetaScience Butterfly Farm in Florida about 1985. These hybrid stocks formed a core of Boender's inhouse hybrid *Heliconius* display stock when Butterfly World (BW) opened in 1988. Only one known introduction of wild-caught *H. melpomene* occurred to this stock, of *H. m. cythera* collected in the early 1980's by R. Boender and T. Emmel around Tinalandia, Ecuador. After this introduction, Boender began selection for a novel pattern that is termed "Piano Key" in this paper. The exact mix of races that were transferred from LBH to BW are not known, but contain pattern alleles that are characteristic of *H. melpomene* races of Suriname and the Guianas (LEG, personal obs).

Since the Piano Key phenotype was first noticed by R. Boender over 30 years ago, the Piano Key stock population has been maintained in separate compartments for around 400

generations at Butterfly World and has long been fixed for the PK phenotype (RB, personal communication). Thus, we referred to it as *H. melpomene* BWPK throughout. Within the last decade, novel mutant butterflies appeared with fewer melanic scales, initially termed "pale PK". These were subsequently separated to form a selected sub-population at BW. Soon after, virtually pure white butterflies incapable of flight began appearing in this selected stock, herein referred to as *ivory*. In 2015 "Pale PK" were sent to UT Austin and maintained by LEG in climate-controlled greenhouses (see S1 Table for phenotypes and collection dates) to investigate the genetics of this mutation. We now know that "Pale PK" phenotypes are heterozygous for ivory mutation and the pure white and flight-disabled phenotypes are homozygous for the ivory mutation (see below).

Imaging

Pinned specimens were imaged with a Nikon D5300 camera mounted with a Micro-NIKKOR 105mm f/2.8 lens. Scale phenotypes were imaged with a Keyence VHX-5000 microscope mounted with a VH-Z100R lens.

Short read DNA sequencing

DNA was extracted from thoracic tissues of 37 *H. m.* BWPK individuals using the Qiagen DNeasy Blood & Tissue Kit, RNAse-treated, and used to prepare a multiplexed sequencing library with the TruSeq PCR-free DNA protocol. Samples were sequenced on an Illumina NovaSeq S1 150bp PE run, yielding 13x mean coverage per individual. Sequencing reads are accessible via NCBI SRA under the project accession number PRJNA663300.

Genomic analyses

Samples were aligned to Hmel2.5 (43) retrieved from LepBase (44) with BWA-MEM using default parameters (45), and variants called using GATK v4.1 with tools HaplotypeCaller and GenotypeGVCFs using default parameters (46). Variant sites were accepted if they were biallelic and the quality (QUAL) value was ≥30. SNPs were phased with Beagle 4.1 (47). Phased SNP variants were used to perform principal component analysis (PCA) using the Eigensoft module SmartPCA(48). SNP association was carried out in PLINK v.1.9, with 1000 permutations(49). For phylogenetic analyses, we followed the TWISST pipeline(50); briefly, data were phased in Beagle 4.1 with a window size of 10,000 and overlap of 1000. Trees were built from windows of 50 SNPs, and then analysed with TWISST using the five groups "East", "West", "Ecuador", "Atlantic", "BWPK" (S2 Table).

***De novo* assembly for indel breakpoints**

To find precise indel breakpoints, a subset of samples were *de novo* assembled with Velvet (51) using default parameters. The resulting contigs were searched with BLAST (52) for the genomic region including the deletion. Scaffolds that bridged the indel breakpoints were selected, and aligned to the Hmel2.5 reference genome with MAFFT (53).

CRISPR/Cas9 mutagenesis

We designed sgRNAs against the putative *H. erato cortex* promoter corresponding to N_{20}NGG sites within the distal promoter/5'UTR. In order to mutagenize the locus, we designed three sgRNAs (Table 1) using the "find CRISPR sites" algorithm within the Geneious software (https://www.geneious.com/). Guide specificity and off-target effects were assessed

by scoring against the *H. erato* reference genome. sgRNAs displaying low off-target scores were then synthesised commercially by Synthego, and mixed with Cas9 at a concentration of 500:500 ng/ul respectively. Embryonic injections were performed as previously described (42) within 1-3 hours after egg laying, after which larvae were allowed to develop on a diet of *P. biflora* until adult emergence.

Results

H. melpomene BWPK - artificially selected wing patterns in an insectary stock

The stock maintained at Butterfly World in Florida and at UT Austin is here named *H. melpomene* BWPK (see Methods - Stock history section). The insectary-bred line is polymorphic for several wing pattern elements found in wild populations, including the red pattern elements (Fig 1). Additionally, *H. m.* BWPK includes aberrant pattern features not observed in natural populations. First, in the pattern dubbed 'Piano Keys' (Fig 1A), white or yellow hindwing marginal elements extend along the veins towards the discal cell. While distal yellow hindwing pattern elements are observed in some wild pattern forms of *Heliconius*, including *H. melpomene cythera* from Ecuador and *H. cydno*, the extent of these marginal elements in *H. m. BWPK* is much greater than any of these, occurring over most of the area of the hindwing except for the intervein regions that would otherwise be taken by the red hindwing rays.

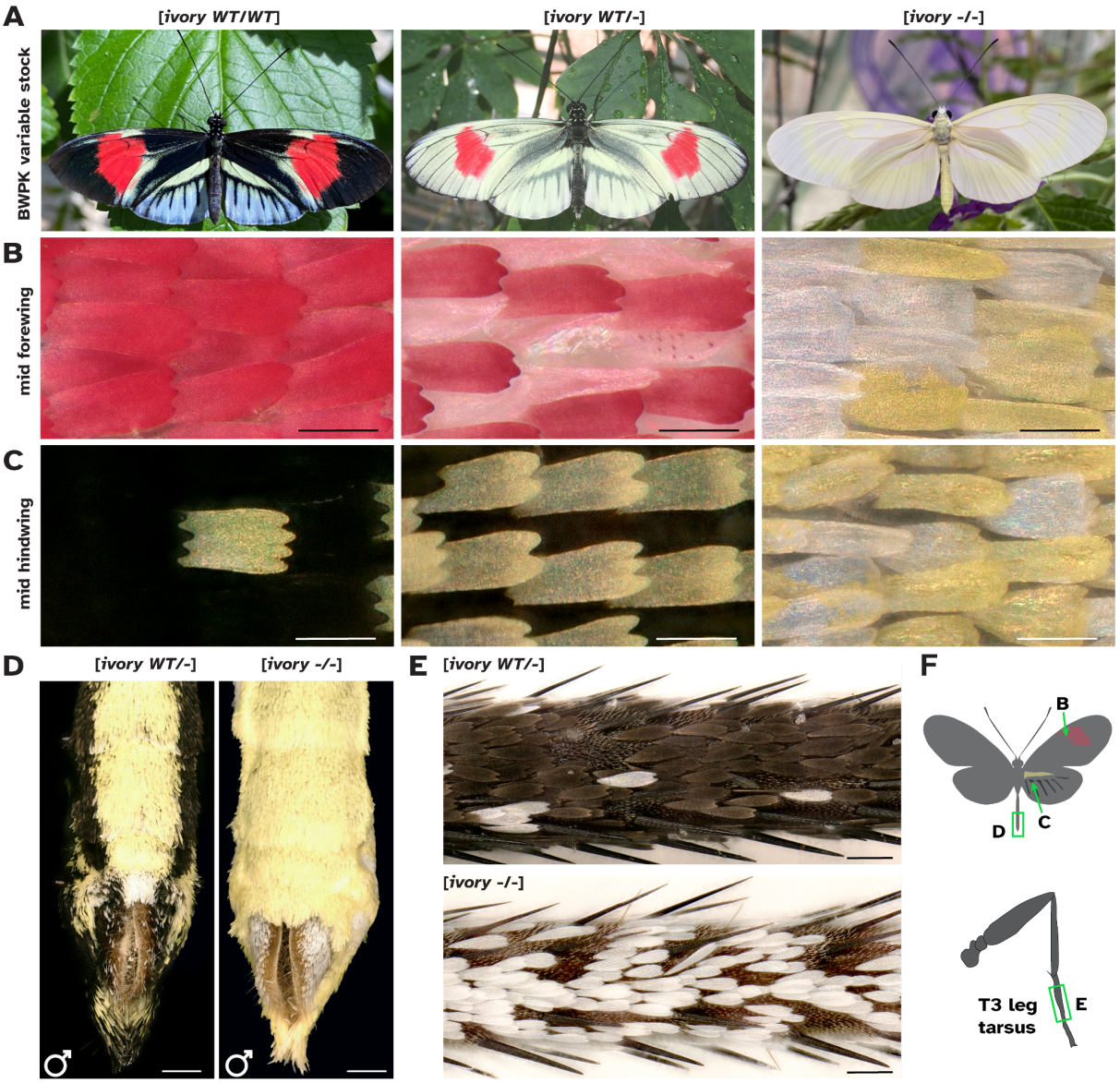


Fig 1. Phenotypes of *H. melpomene* BWPK butterflies. (A) Ivory phenotypes in the *H. melpomene* "Piano Keys" in the UT Austin stock (BWPK). The three color states (from left to right "Dark BWPK", "Pale BWPK", "ivory") depend on the allelic dosage of a co-dominant mutation. Ivory homozygotes are inviable and only found among offspring from two pale heterozygotes. **(B)** Magnified view of the forewing red band region. In the *[ivory WT/-]* state, abnormal scales have formed in the red region, with red pigment in granules and scales curled, while in [*ivory -/-*], all scales are yellow or white. **(C)** Magnified view of a central hindwing region that is black in [*ivory WT/WT*]. The [*ivory WT/-*] wing has all cover scales as yellow or white while all ground scales remain black. All scales are yellow or white in [*ivory -/-*]. **(D-E)** Complete replacement of melanic scales by yellow-white scales in the abdomen **(D)** and legs **(E)** from [*ivory -/-*] homozygotes. **(F)** Cartoon of image locations for B-E. Scale bars : **B-C** = 50 μm ; **D** = 500 μm ; **E** = 100 μm.

Occasional butterflies with more extensive regions of yellow or white scales were identified (Fig. 1A centre). Other than the wing pattern differences, both forms of butterflies exhibit typical feeding, mating, egg-laying and flight behaviours to other insectary-reared *Heliconius* (*e.g.* S1 Movie). However, the offspring of a mating of two pale *H. m.* BWPK will include completely white-winged and -bodied butterflies (Fig 1C). We dub these butterflies 'ivory' in reference to their emergence in the 'Piano Keys' stock, and inferred that the pale morph of

H.m. BWPK represents a heterozygous state for a mutation that causes the ivory phenotype in the homozygous state.

The ivory butterflies have melanin pigments in their body cuticle and eyes, indicating that their capacity to synthesize and deposit melanin has not been perturbed. However, all scales on the wings and body are white or yellow, with no black or red scales, with the exception of a 1 mm patch of red at the base of the hindwing in some individuals. This is in stark contrast to wild-type or BWPK butterflies which have black scales covering most of the body. Unlike the other BWPK butterflies, ivory homozygote butterflies do not fly and have not been observed to successfully mate or lay eggs. The wings appear to be of a normal strength and sturdiness, but flight is weak and will not occur when prompted by dropping (S1 Movie).

In order to investigate the genetic basis of these wing patterns, we sequenced 37 butterflies from the UT Austin *H. m.* BWPK stock to an average coverage of 13 x, including 11 Dark BWPKs, 9 Pale BWPKs and 10 ivory butterflies (Fig 1A).

Autozygosity mapping identifies associated SNPs

We used a combination of GWAS and patterns of SNP autozygosity to determine the locations of regions associated with Mendelian pattern variants (Fig 2). As proof of principle, we mapped the Dennis pattern element which was previoulsy described as a *cis*-regulatory element of the gene *optix*. GWAS for presence vs absence of Dennis gave a narrow association peak centred near *optix*, in a region previously identified as associated with the Dennis pattern element (38) (S1 Fig). We then examined the inheritance patterns of SNPs, filtering for SNPs where butterflies with no Dennis element were homozygous for the reference allele, and individuals with the Dennis element were either heterozygous or homozygous for an alternate allele. In total, 3430 SNPs matched this pattern of inheritance, with 3425 in a cluster centred on the gene *optix,* indicating that autozygosity mapping is appropriate for mapping pattern elements in this data set.

GWAS for ivory gave a broad association peak on chromosome 15 (Fig 2D). We expected dark morph BWPK butterflies to carry the reference allele, pale morphs to be heterozygous, and ivory morphs to be homozygous for a non-reference allele. We expected that at the causative locus, this allelic segregation pattern should be fixed in every individual. This was the case at just 131 SNPs in the whole genome, all within a 9 kb interval on chromosome 15, within the broad association peak (Fig 2D, blue marks). These SNPs sit within the first intron of the gene *cortex*. *Cortex* is a switch gene necessary for differentiation into black Type II and red Type III scales (41,42), and has previously been linked to melanic switches in many species, including in industrial melanism in the peppered moth (54,55).

We noted that in [*ivory WT/-*] butterflies, scales in red regions exhibited aberrant morphologies (Fig 2E-F). Granular clumps of red pigments could be observed in the body of some scales, whereas wild type scales are solidly-pigmented with no granularity. Additionally, we observed scales that were curled at the edges rather than flat. Similar scale phenotypes occur in the wings of *cortex* crispant butterflies (Fig 2E-F), thus supporting a role for *cortex* loss-of-function in the ivory colour phenotypes (42).

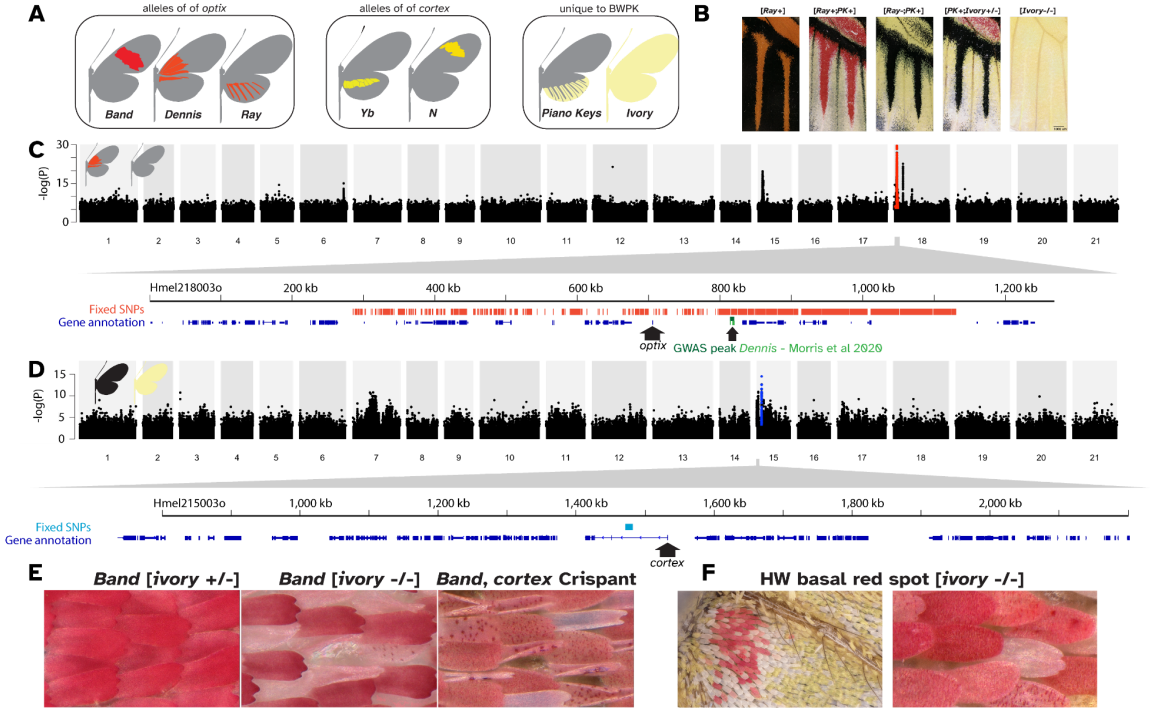


Fig 2: Genome wide association and autozygosity mapping in *H. melpomene* BWPK. (A) Cartoons of the allelic variation present in the *H. m.* BWPK stock. **(B)** the hindwing veins, showing the composite effects of different alleles depicted in panel A. **(C)** GWAS for the Dennis element; Manhattan plot of Wald test P-values, with fixed SNPs in red, and a magnified annotation of the region around the gene *optix*. **(D)** GWAS for ivory, with fixed SNPs in blue and a magnified annotation of the *cortex* region, with the point of the arrow at the annotated TSS. **(E)** comparison of red scales in [*ivory WT/WT*], [*ivory -/-*], and *cortex* crispant wings. **(F)** Some ivory butterflies have a small red dot at the base of the ventral hindwing, the only scales on the entire butterfly that are not white or yellow. They share the atypical phenotype of red scales from *cortex* crispants (**E**).

Ivory butterflies carry a large deletion including the *cortex* promoter

Many phenotypes involve structural genomic variation, both under artificial selection (56), and in natural adaptation (57,58). As such, we checked read depth across this region in all our sequenced samples. Immediately adjacent to the block of fixed SNPs identified by autozygosity mapping, we found a large region where read depth in [*ivory -/-*] butterflies dropped to zero, while in [*ivory WT/-*] it dropped by half. This indicated that a large deletion spans the first exon and promoter of *cortex* (Fig 3A).

To determine the precise breakpoints of this deletion, we *de novo* assembled short read data for all [*ivory WT/WT*] and [*ivory -/-*] individuals, and BLASTed the resulting contigs against the *cortex* region. This allowed us to recover individual contigs from some individuals that bridged the two ends of the deletion, giving a precise breakpoint at positions Hmel2150003o:1494902-1573177, with a length of 78,275 bp relative to the reference genome (Fig 3B).

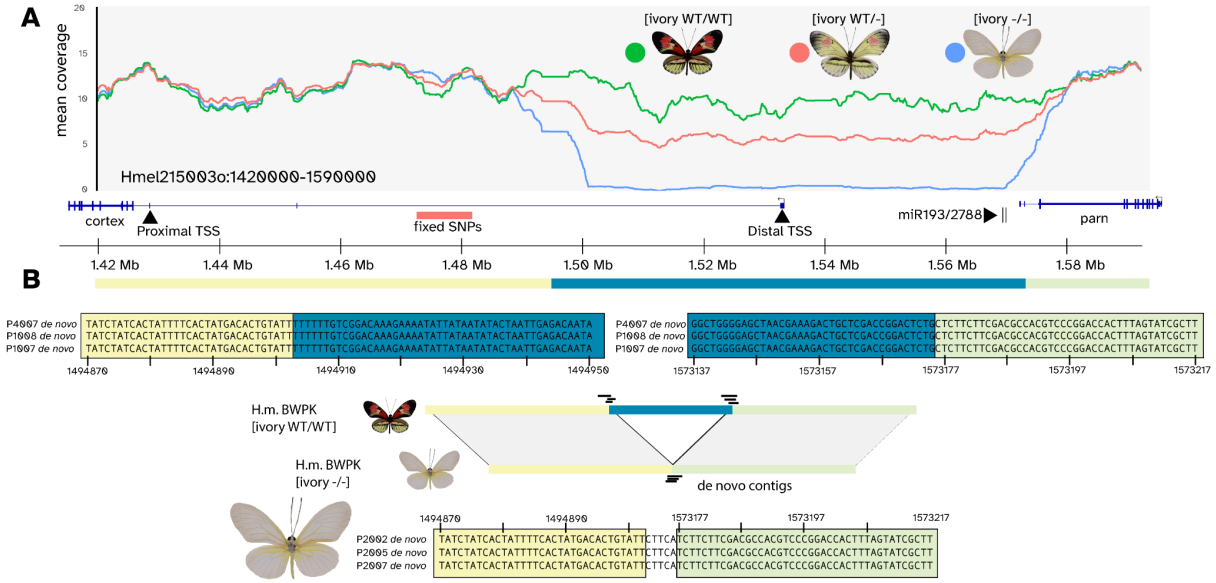


Fig 3: ivory is associated with a large deletion in the gene *cortex* (A) mean coverage plot for the region around *cortex*, with [*ivory WT/WT*] in green, [*ivory WT/-*] in red, and [*ivory -/-*] in blue. **(B)** Alignments of *de novo* scaffolds, indicating the precise breakpoints of the *ivory* deletion. Colours indicate the orientation of sequences relative to the reference scaffold depicted in panel A.

The *ivory* deletion includes one of the two *cortex* transcription start sites

The *ivory* deletion contains the only annotated transcription start site (TSS) and promoter for *cortex*. While butterflies carrying this mutation can complete development, even with effects in scale development across the whole organism and the inability to fly, coding knockouts of *cortex* (including clonal G_0 crispants) have been observed to have high levels of embryonic lethality, with very few crispant clones observed (42).

The ability of *ivory* mutants to survive in spite of the loss of the promoter, as well as the presence of a long first intron in *cortex*, led us to hypothesise that there may be a second, alternate TSS. In order to determine if this is the case, we mapped RNAseq from *H. melpomene* for a variety of samples including larval and pupal wings (59), adult ovaries (43), and embryos, adult head and adult abdomen (60), and looked at splicing and alignment of

the 5' end of the transcript. No expression was detected in adult head or abdomen, but in embryos, ovaries, and pupal wings, transcription initiates at position Hmel2150003o:1424680 in the third annotated exon, which is adjacent to the coding sequence (the 'proximal promoter') (Fig 4A, S2-3 Figs). Expression in larval wings initiates at the annotated TSS (the 'distal promoter'), and the exon with the proximal promoter is skipped. In pupal wings, transcription again initiated at the proximal promoter. This indicates that *cortex* has two alternative TSSs, a proximal promoter used in multiple tissues and a distal promoter specific to the larval wing. Only the distal promoter is included in the *ivory* deletion.

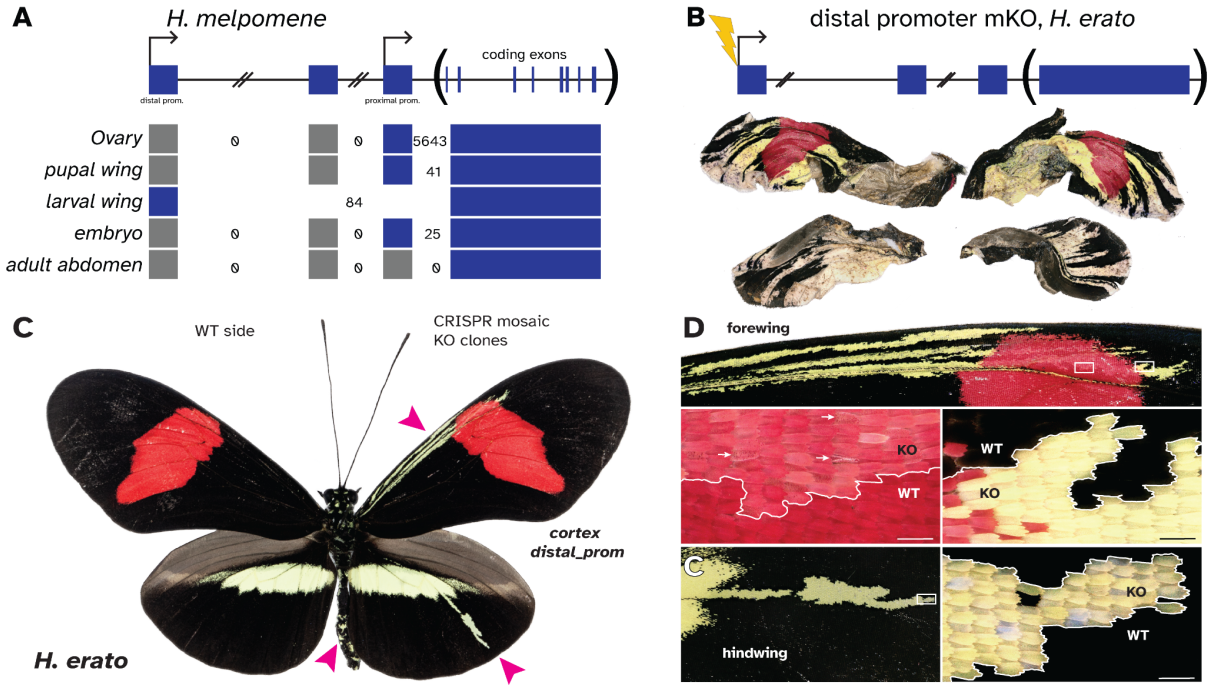


Fig 4: Knockout of the distal promoter phenocopies *cortex* protein coding knockouts. (A) mapping of RNAseq reads at *cortex* suggests that there are two promoters and transcription start sites. Numbers of intron-spanning reads are indicated (see Sashimi plots in Figs S2-3). **(B-C)** Knockout of the distal promoter in *H. erato* causes a transformation of black scales to white/yellow in a phenocopy of *cortex* protein coding knockouts. B shows a butterfly with extensive clones but that emerged poorly, indicating pleiotropic effects not observed in the *ivory* mutants, while C shows much smaller clones similar to those reported by Livraghi *et al.* 2021, with clone positions indicated by pink arrows. **(D)** High magnification images of mutant clones. Scale bars: 100 µm.

Mutagenesis of the distal promoter phenocopies *cortex* null effects

We reasoned that if the loss of the distal promoter causes the ivory phenotype, knocking out the promoter with CRISPR/Cas9 should phenocopy the effects of the *ivory* deletion. We generated G_0 mosaic knockouts (crispants) of the promoter in the related species *Heliconius erato* (Fig 4B-D). Resulting butterflies had extensive yellow/white clones on the wings, however several failed to emerge from the pupa, and a low survival rate and penetrance were observed (Table 1), suggesting that the few wing crispants obtained are rare mosaic escapers of a loss-of-function experiment with lethal effects (61). Thus while the pleiotropic effects of deleting the TSS were reminiscent of the previously reported *cortex* protein coding knockouts, which also exhibited low survival and penetrance (42), this experiment did not fully recapitulate the ability of the *ivory* deletion allele to generate large wing clones. This

difference in pleiotropy may be due to the use of *H. erato* in the CRISPR assay, and further experiments are needed to test the functionality of deleted elements within a *H. melpomene* stock. However, we conclude from these experiments that the disruption of 5' distal elements of *cortex* are necessary for normal color scale patterning, consistent with a causal role of the 78 kb deletion in underlying the ivory phenotypes.

Table 1. CRISPR mutagenesis of the *cortex* distal promoter/5'UTR in *H. erato*

Batch	sgRNA target	Injected	Pupae	Adults	Crispants	Penetrance
1	*sgProm1 :* 5'AGCGGTAACTTTATCGCGAT	512	42	37	3	0.6 %
2	*sgProm2:* 5'ATGGGATATGTTTTAATAGT *sgProm3:* 5'AACCTGTCACATCAGTACAG	695	78	73	1	0.1 %

Demographic origins

The *H. m.* BWPK stock was kept in insectaries for 30 years (c. 400 generations) before producing the *ivory* mutation. The precise history and ancestry of the stock is not recorded, but we do know that the original stock was generated with a mix of butterflies from multiple locations. after the addition of butterflies from the vicinity of Tinalandia, Ecuador in the 1980s. The first pale PK (*i.e.* heterozygotes for *ivory*) appeared in the PK culture at Butterfly World after 2013, and were shared with L.E.G. for study in 2014-15.

Whole genome PCA clustering of a geographic spread of samples of *H. melpomene*, as well as sister taxa *H. cydno* and *H. timareta*, replicates the geographic and species clustering previously described by Martin *et al.* (62), with the addition of a distinct "Atlantic" group within *H. melpomene* (S1 Table). *H.m.* BWPK cluster together near *H. melpomene* from the west of the Andes. Similarly, PCA of just the *cortex* locus, using many more individuals generated by selective sequencing by Moest et al (63) (S3 Table), also places BWPK closest to western *H. melpomene*.

Given the mixed ancestry of *H. m* BWPK, and the observation that ivory patterns emerged after introduction of butterflies from Ecuador, we examined heterogeneity in genome-wide ancestry by building phylogenetic trees in windows of 50 SNPs across the genome and using topology weighting to determine the closest neighbour of *H.m* BWPK at all genomic positions. Of the 15 possible topologies, the most frequent topology places BWPK as an outgroup to the four other *H. melpomene* taxa. The second most common topology groups BWPK with *H. melpomene* from western Ecuador (*H. m vulcanus* and *H. m. plesseni*), and the third most common places BWPK with *H. melpomene* from the East of the Andes. In contrast, sliding windows across the *cortex* locus (including the selective sequencing data) show the most common tree topology here places BWPK sister to the western Ecuador taxon. This result is supportive of *H.m.* BWPK having an admixed genome, but with the allele at *cortex* having originated in Ecuador. Importantly, the *ivory* deletion was not detected in any of 458 wild individuals.

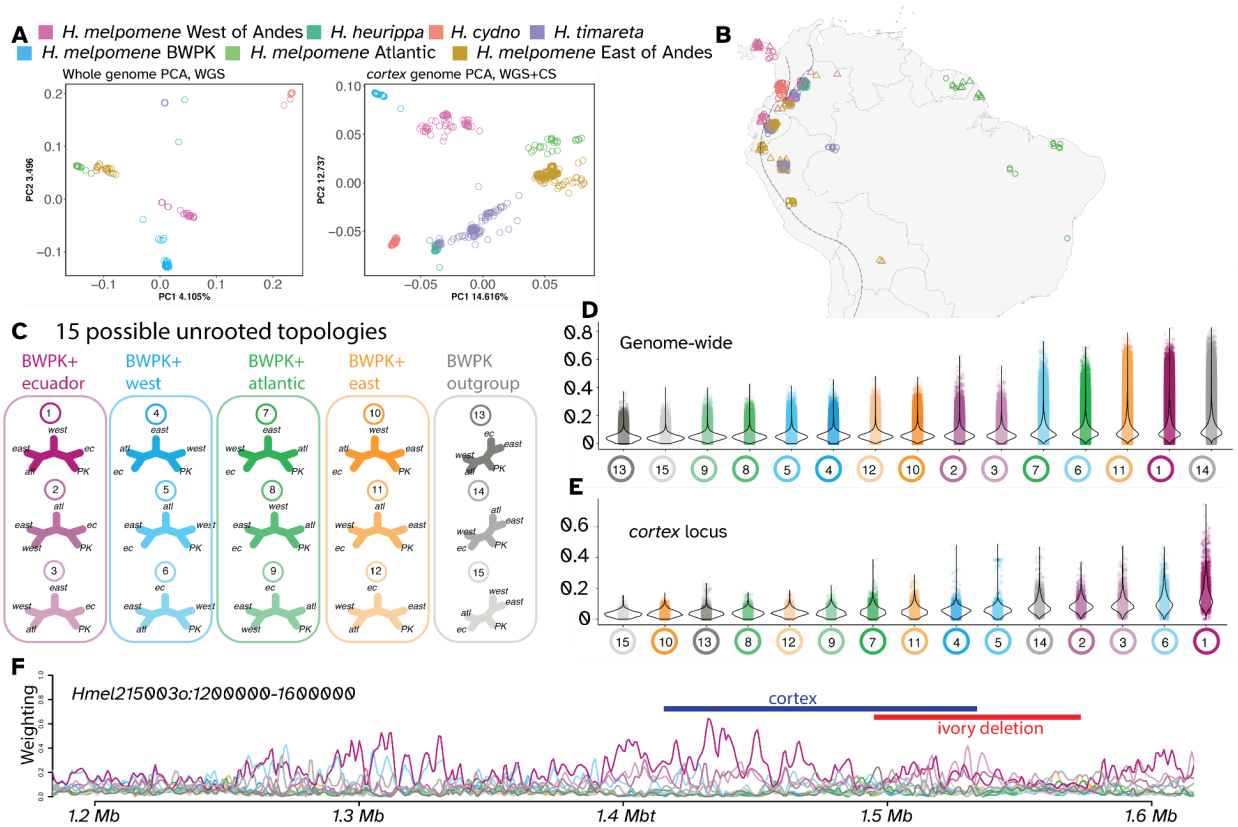


Figure 5: Phylogenetic clustering and weighting implies a mixed origin for *H. melpomene* BWPK. Clustering by PCA from both whole genome data (A) and from the *cortex* locus using additional selective sequencing data (B) show similar geographic clustering to that previously reported by Martin *et al.* (62), C), with the addition of separation between *H. melpomene* from East of the Andes and *H. melpomene* from French Guiana, Suriname and Brazil (here termed "Atlantic"). *H. m.* BWPK form a distinct cluster, closest to the West cluster of *H. melpomene*. Topology weighting was performed using TWISST (50), with five defined groups (West of Andes, Ecuador, East of Andes, Atlantic, and BWPK) (C). (D) Genome-wide, the most common topologies were 14 (BWPK as outgroup), 11 (BWPK clustered with East of Andes) and 1 (BWPK clustered with Ecuador). In contrast to this in (E), TWISST results at the *cortex* locus showed that topology 1 was most common, supporting an Ecuadorian origin for the *cortex* variants in *H. m. BWPK*.

Discussion

Using whole genome sequencing of a set of related individuals, we were able to map the *ivory* mutation to a large deletion that removed the distal promoter of the wing patterning gene *cortex*. Cortex was first identified as a switch between melanic and non-melanic patterns in both *Heliconius* and the peppered moth *Biston betularia*, and has later been mapped as a switch between melanic and non-melanic scale types in a number of other species, including three other geometrid moths (55), and *Papilio* butterflies (64), and also has a role in wing pattern polyphenism in *J. coenia* (65). The list of genes that have been identified as the targets of selection in butterfly wing pattern variation includes transcription factors like Optix and Bab (66,67), signalling pathway components like WntA (68), or terminal effector genes like Yellow (69), all of which have molecular functions that support a role in developmental patterning or pigment synthesis (34). Cortex, on the other hand, has no clear functional association with colour pattern or cell differentiation, and yet has been mapped in a very broad phylogenetic spread of species, suggesting its function may be ancestral in the

Lepidoptera. The gene does not appear to have a role in *Drosophila* wing development or patterning (41), limiting our ability to make inferences about its molecular function and interactions. In identifying this large deletion, this study provides novel insight into the genetics of this hotspot locus, and highlights the potential utility of looking at insectary mutants to understand butterfly wing development.

The loss of all black and red scales in the *ivory* mutant supports a model in which white/yellow scales are the 'default' state during scale cell differentiation, with scale cell precursors failing to differentiate into red or black cell types. As such, when comparing pattern homologies between butterflies, *Heliconius* white/yellow regions can be considered as the 'background', and black and red regions can be considered as 'pattern elements' layered on top of this background. This model was initially proposed based on observation and ultrastructure and pigmentation of scales (70), as well as on the analysis of pattern homologies of the Heliconiini (71).

Structural variation and mutations

Large deletions have repeatedly been observed in cases of both natural variation and domestication, with 29 deletions affecting regulatory regions which are larger than 1 kb listed on GepheBase (72). Recurrent deletions of a *pitx1* enhancer in sticklebacks, up to 8 kb in length, have caused convergent pelvic reduction (73), and the deletion of a 60.7 kb regulatory region at the *AR* gene in humans, which removes enhancers present in chimpanzee and conserved in other mammals, leads to loss of penile spines (74). Similarly large deletions have also been confirmed in the genetic basis of domestication phenotypes, including a 44 kb deletion, also at *pitx1*, causing feathered feet in chickens (75), and a 141 kb deletion upstream of *agouti/ASIP* in Japanese quail (76).

The large deletion that causes ivory is a form of structural variation. Other structural variants affecting this locus have been found in wild populations of *Heliconius*; multiple inversions have formed a wing pattern supergene in *H. numata*, and an independent inversion, similar in size and position, has been found in multiple species of the *erato* clade (77,78). It is possible that this genomic region will prove to be prone to a higher-than-average level of structural variation, or, conversely, that the region is more tolerant to structural variation than other parts of the genome.

Regulatory consequences of the *ivory* mutation

Large deletions have the potential to remove a large section of *cis*-regulatory sequence, as well as transcribed sequences. The *ivory* deletion causes the loss of one of two promoters. As many as 40% of developmentally expressed genes in *Drosophila* have two promoters which can cause distinct regulatory programs (79), and multiple promoters are also common in human genes (80). This both increases the complexity of gene regulatory interactions and increases the number of transcript isoforms per gene - for example, an alternate promoter in the *Drosophila* gene *Zfh1* creates an isoform that has a shorter 5' UTR which is missing miRNA seed sites on its 5'UTR, permitting differential degradation of the mRNA by *miR-8* (81). We determined that the *ivory* deletion contains one annotated promoter of *cortex*, but that in some contexts during development, another transcription start site is used, much

closer to the translation start site. Alternate promoter usage is likely to be common and widespread in animals.

We expect that the use of two separate, context-specific promoters at *cortex* will create a hierarchical regulatory organisation, where CREs used in different tissues or at different times can loop to different promoters. Additionally, transcripts from each of the two promoters contain a different arrangement of 5' non-coding exons, giving them different UTRs, possibly leading to differential translational regulation or degradation of the mRNA.

The *ivory* deletion does not include protein-coding sequence, which led us to hypothesise that by removing just one promoter, *cortex* would still be expressed in tissues that use the alternate promoter, and that therefore *ivory* mutants would bypass the pleiotropic, highly lethal effects observed in CRISPR experiments that target coding exons of *cortex* (42). However, crispants with deletions in the distal promoter in *H. erato* did not appear to be free of these pleiotropic effects, and exhibited high lethality and comparatively small clone sizes, dissimilar to the *H. melpomene* [*ivory -/-*] mutants. This may be due to inherent differences in the regulatory function of the distal TSS between these two species, or to the existence of other functional sequences within the deletion which might be necessary for generating the pale phenotypes. Specifically, the 78-kb deletion removes part of the 3'UTR of the neighbouring gene *parn* as well as two miRNAs (82), and it is also possible that a large deletion affects other cortex cis-regulatory elemnts or affect 3D chromatin interactions on a large portion of the chromosome. Additionally, the *ivory* butterflies carry a set of fixed SNPs adjacent to the deletion, which could contribute to the phenotype. In summary, the complete loss of black and red scales caused by the *ivory* deletion and by CRISPR-Cas9 mutagenesis in two *Heliconius* species indicates that either the distal promoter or another functional sequence within the 5' non-coding region of *cortex* are necessary for black and red scale development.

Heterozygote advantage of deleterious mutations in captivity

The *cortex* mutant phenotype has been artificially selected by a breeder in the heterozygous state, when unusually widespread yellow and white patterns emerged in the stock. The heterozygote carriers of the *ivory* mutation ([*ivory WT/-*]) do not appear to have any reduction in fitness or vigor. In contrast, homozygous *ivory* butterflies of either sex are unable to fly, limiting their ability to feed and reproduce – we have not observed any successful matings involving an [*ivory -/-*] butterfly. Those recessive effects make the mutant allele unfit for breeding in the homozygous state, resulting in a form of heterozygote advantage where mutant heterozygotes are selectively bred in spite of inviable or undesirable effects in homozygotes(83,84). We compiled similar situations from the literature using a database of gene-to-phenotype relationships (72), and found 38 analogous gene-to-trait relationships spanning domesticated mammals and birds (Table 2). Large structural variations (SVs) such as the *ivory* deletion accounted for 9 out of 46 derived alleles in this dataset, suggesting that the recessive deleterious effects of macromutations occasionally provide heterozygous states of interest to artificial selection. Of note, 17 out the 38 cases of captive heterozygote advantage involve selection for depigmentation traits, highlighting the trend among breeders and fanciers to select for conspicuous variants (Table 2). Finally, cases of heterozygous advantage also occur in the wild, such as in ruff birds where a large inversion haplotype

provides male coloration and behavioral traits that are maintained by sexual selection in spite of being recessive lethal (85). In wild populations of the polymorphic butterfly *Heliconius numata*, *cortex* itself is situated in the center of an inversion polymorphism under balancing selection (86). Further work will be needed to determine the sub-gene level elements that are within the *ivory* deletion and mediate homozygous inviability.

Table 2: Known cases of heterozygote advantage in animal breeding

Gene	Species	Heterozygote advantage (HA) trait for breeding	Detrimental recessive effect	Nb of HA alleles	Mutation types
ACAN	Cattle	Short stature	Lethal	1 1	Frameshift (4-bp ins) New start codon (SNP)
ACAN	Horse	Short stature	Lethal	1 1	Loss of 6 a.a. (18-bp del) Frameshift (1-bp del)
ALX1	Cat	Brachycephaly	Frontonasal dysplasia	1	Loss of 4 a.a. (12-bp del)
BBS9	Pig	Fast growth	Lethal (misexpressed *BMPER*)	1	Large SV (212-kb del)
BMP15	Sheep	Increased female fecundity	Female infertility	5 2 1 1 1	Missense (SNP) Nonsense (SNP) Frameshift (1-bp ins) Frameshift (17-bp del) Cis-regulatory (SNP)
cortex	*Heliconius*	White patterns	Flightless	1	Large SV (78-kb del)
EDNRB	Horse	White patterns	Lethal	1	Missense (SNP)
FGF3/4/19 cluster	Dog	Dorsal hair ridge	Dermoid sinuses	1	Large SV (133-kb duplication)
FGFR3	Sheep	Increased bone length	Chondrodysplasia	1	Missense (SNP)
FOXI3	Dog	Hairlessness	Ectodermal dysplasia	1	Frameshift (7-bp duplication)
GDF9	Sheep	Increased female fecundity	Female infertility	5	Missense (SNP)
HMGA2	Rabbit	Short stature	Lethal	1	Large SV (12.1-kb del)
HOXC10	Chicken	Head crest	Cerebral Hernia - likely epistatic	1	Large SV (197-bp duplication)
IHH	Chicken	Short legs	Lethal	1	Large SV (11.9-kb del)
KIT	Camel	White patterns	Lethal	1	Frameshift (1-bp del)
KIT	Dog	White patterns	Lethal	1	Frameshift (1-bp ins)
KIT	Donkey	White patterns	Lethal - not proven	1 1	Missense (SNP) Splice Site mutation (SNP)
KIT	Fox	White patterns	Lethal	1 1	Missense (SNP) Splice Site mutation (SNP)
KIT	Horse	White patterns	Lethal	3 2 2 1 1	Frameshift (1-bp, 1-bp, 4-bp del) Large SV (1.3-kb, 1.9-kb del) Nonsense (SNP) In-Frame Del (54-bp) Splice Site mutation (SNP)
KIT	Pig	White patterns	Lethal - not proven	1	Splice Site mutation (SNP)
MITF	Buffalo	White patterns	Microphtalmia (inferred from mice)	1 1	Nonsense (SNP) Splice Site mutation (SNP)
MITF	Dog	White patterns	Deafness (risk increase*)	1	Cis-regulatory (195-bp ins)

MITF	Horse	White patterns	Microphtalmia (inferred from mice)	2 1 1	Large SV (8.7-kb, 63-kb del) Frameshift (4-bp del) Missense (SNP)
MITF	Quail	White patterns	Small size and slow growth	1	Frameshift (2-bp del)
Mlana	Pigeon	White patterns	Eye defects	4	CNV including 3 other genes
MNR2 locus	Chicken	Head comb	Male infertility	1	Large SV (7.4-Mb inversion)
MRC2	Cattle	High muscle	Musculoskeletal defects	1 1	Frameshift (2-bp del) Missense (SNP)
*MSTN***	Dog	Racing performance	Non-conform appearance	1	Frameshift (2-bp del)
PAX3	Horse	White patterns	Lethal	2	Missense (SNP)
PMEL17	Dog	White patterns	Auditory and ocular defects	6	Splice Site mutation (TE insertions)
PMEL17	Horse	White patterns	Ocular defects	1	Missense (SNP)
RNASEH2Ba locus	Cattle	Milk yield (causal gene unknown)	Lethal (due to *RNASEH2Ba* KO)	1	Large SV (660-kb deletion)
RYR1	Pig	Meat yield and quality	Hyperthermia	1	Missense (SNP)
PRLR / SPEF2	Pig	Increased female fecundity	Sperm defects (due to *SPEF2* KO)	1	Cis-regulatory (TE insertion)
SPINT1	Gecko	White patterns	Metastases	1	Unknown
STX17	Horse	White patterns	Melanoma	1	Cis-regulatory (4.6-kb duplication)
TBXT (Brachyury)	Cat	Short or no tail	Lethal	4	Frameshift (1-bp del; 14-bp ins)
TBXT (Brachyury)	Dog	Short or no tail	Lethal	1	Missense (SNP)
TRPM1	Horse	White patterns	Ocular defects	1	Cis-regulatory (TE insertion)
TRPV4	Cat	Short ears	Chondrodysplasia	1	Missense (SNP)

* See Gephebase (www.gephebase.org) for references. The corresponding entries can be retrieved using a search for "@HeterozygoteAdvantage".
** recommendations to outcross are recent in the affected stock.
*** *MSTN* loss-of-function alleles exist in other bred species. Homozygotes sometimes requires assisted birthing, but are viable and artificially selected.

Conclusion

By using autozygosity mapping and association on a comparatively small pool of individuals, we were able to identify a structural mutation involved with butterfly wing pattern. This approach may prove fruitful in other studies of butterfly wing patterning or in the identification of *de novo* mutants, especially if combined with recent developments in whole genome sequencing from dried museum specimens (87,88). This could allow the mapping of other cases like *Hindsight* in *Junonia coenia*, or *pseudozorro* in *Parnassius apollo* (89,90). The identification and characterisation of spontaneous mutants has provided very valuable insights into the genetics of development in model organisms. Further whole genome sequencing of deleterious mutations in butterflies will help to identify parts of the genome that are functionally required for normal development, which will assist in a more complete understanding of their evolution and development.

Data availability

Whole genome sequences are accessible on the NCBI SRA under the Bioproject accession PRJNA610063.

Acknowledgements

We thank Ron Boender of Butterfly World, FL for generously providing living specimens for study by the Gilbert Lab at UT Austin. Thanks to greenhouse technicians T. Freiburger, N. Fogel, I. Terry, and E. Reese for keeping plant and butterfly cultures healthy in the Austin facility during the study, and G. Julian at the University of Cambridge for expert stock maintenance of *H. erato*. We also thank the GWU HPC team for computing infrastructure (91), and S.M. Van Belleghem for comments on an earlier draft.

References

1. Cieslak M, Reissmann M, Hofreiter M, Ludwig A. Colours of domestication. Biol Rev. 2011;86(4):885–99.
2. Driscoll CA, Macdonald DW, O'Brien SJ. From wild animals to domestic pets, an evolutionary view of domestication. Proc Natl Acad Sci. 2009 Jun 16;106(Supplement 1):9971–8.
3. Park K-I, Ishikawa N, Morita Y, Choi J-D, Hoshino A, Iida S. A bHLH regulatory gene in the common morning glory, Ipomoea purpurea, controls anthocyanin biosynthesis in flowers, proanthocyanidin and phytomelanin pigmentation in seeds, and seed trichome formation. Plant J. 2007;49(4):641–54.
4. Giovannini A, Laura M, Nesi B, Savona M, Cardi T. Genes and genome editing tools for breeding desirable phenotypes in ornamentals. Plant Cell Rep. 2021 Mar 1;40(3):461–78.
5. Price-Waldman R, Stoddard MC. Avian Coloration Genetics: Recent Advances and Emerging Questions. J Hered. 2021 Jul 1;112(5):395–416.
6. Domyan ET, Shapiro MD. Pigeonetics takes flight: Evolution, development, and genetics of intraspecific variation. Dev Biol. 2017 Jul 15;427(2):241–50.
7. Guo L, Bloom J, Sykes S, Huang E, Kashif Z, Pham E, et al. Genetics of white color and iridophoroma in "Lemon Frost" leopard geckos. PLOS Genet. 2021 Jun 24;17(6):e1009580.
8. Mundy NI, Stapley J, Bennison C, Tucker R, Twyman H, Kim K-W, et al. Red Carotenoid Coloration in the Zebra Finch Is Controlled by a Cytochrome P450 Gene Cluster. Curr Biol. 2016 Jun 6;26(11):1435–40.
9. Courtier-Orgogozo V, Martin A. The coding loci of evolution and domestication: current knowledge and implications for bio-inspired genome editing. Dickinson MH, Vosshall LB, Dow JAT, editors. J Exp Biol [Internet]. 2020 Feb 1 [cited 2021 Sep 15];223(Suppl_1). Available from: https://doi.org/10.1242/jeb.208934
10. Kuriyama T, Murakami A, Brandley M, Hasegawa M. Blue, Black, and Stripes: Evolution and Development of Color Production and Pattern Formation in Lizards and Snakes. Front Ecol Evol [Internet]. 2020 [cited 2021 Jul 13];8. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2020.00232/full
11. Andersson L. Domestic animals as models for biomedical research. Ups J Med Sci. 2016;121(1):1–11.
12. Berry SD, Davis SR, Beattie EM, Thomas NL, Burrett AK, Ward HE, et al. Mutation in bovine beta-carotene oxygenase 2 affects milk color. Genetics. 2009 Jul;182(3):923–6.
13. Andrade P, Pinho C, Pérez I de Lanuza G, Afonso S, Brejcha J, Rubin C-J, et al. Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. Proc Natl Acad Sci U S A. 2019 Mar 19;116(12):5633–42.
14. Yu S, Wang G, Liao J. Association of a novel SNP in the ASIP gene with skin color in black-bone chicken. Anim Genet. 2019 Jun;50(3):283–6.
15. Rieder S, Taourit S, Mariat D, Langlois B, Guérin G. Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). Mamm Genome Off J Int Mamm Genome Soc. 2001 Jun;12(6):450–5.
16. Fontanesi L, Forestier L, Allain D, Scotti E, Beretti F, Deretz-Picoulet S, et al. Characterization of the rabbit agouti signaling protein (ASIP) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. Genomics. 2010 Mar;95(3):166–75.
17. Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, et al. Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. Curr

Biol CB. 2016 Sep 12;26(17):2313–8.
18. Bonilla C, Boxill L-A, Donald SAM, Williams T, Sylvester N, Parra EJ, et al. The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. Hum Genet. 2005 Apr;116(5):402–6.
19. Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJR, et al. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science. 2018 Jun 22;360(6395):1355–8.
20. Gratten J, Pilkington JG, Brown EA, Beraldi D, Pemberton JM, Slate J. The genetic basis of recessive self-colour pattern in a wild sheep population. Heredity. 2010 Feb;104(2):206–14.
21. Norris BJ, Whan VA. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Res. 2008 Aug;18(8):1282–93.
22. Marsden CD, Vecchyo DO-D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc Natl Acad Sci. 2016 Jan 5;113(1):152–7.
23. Moyers BT, Morrell PL, McKay JK. Genetic Costs of Domestication and Improvement. J Hered. 2018 Feb 14;109(2):103–16.
24. Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet. 2004 Mar;5(3):202–12.
25. Makino T, Rubin C-J, Carneiro M, Axelsson E, Andersson L, Webster MT. Elevated Proportions of Deleterious Genetic Variation in Domestic Animals and Plants. Genome Biol Evol. 2018 Jan 1;10(1):276–90.
26. Strain GM, Clark LA, Wahl JM, Turner AE, Murphy KE. Prevalence of Deafness in Dogs Heterozygous or Homozygous for the Merle Allele. J Vet Intern Med. 2009;23(2):282–6.
27. Langevin M, Synkova H, Jancuskova T, Pekova S. Merle phenotypes in dogs – SILV SINE insertions from Mc to Mh. PLOS ONE. 2018 Sep 20;13(9):e0198536.
28. Metallinos DL, Bowling AT, Rine J. A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung Disease. Mamm Genome. 1998 Jun 1;9(6):426–31.
29. Spelt C, Quattrocchio F, Mol J, Koes R. ANTHOCYANIN1 of Petunia Controls Pigment Synthesis, Vacuolar pH, and Seed Coat Development by Genetically Distinct Mechanisms. Plant Cell. 2002 Sep 1;14(9):2121–35.
30. Bosse M, Megens H-J, Derks MFL, Cara ÁMR de, Groenen MAM. Deleterious alleles in the context of domestication, inbreeding, and selection. Evol Appl. 2019;12(1):6–17.
31. Futahashi R, Osanai-Futahashi M. Pigments in Insects. In: Hashimoto H, Goda M, Futahashi R, Kelsh R, Akiyama T, editors. Pigments, Pigment Cells and Pigment Patterns [Internet]. Singapore: Springer; 2021 [cited 2021 Oct 8]. p. 3–43. Available from: https://doi.org/10.1007/978-981-16-1490-3_1
32. Jiggins CD, Wallbank RWR, Hanly JJ. Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns? Philos Trans R Soc B Biol Sci. 2017 Feb 5;372(1713):20150485.
33. Van Belleghem SM, Lewis JJ, Rivera ES, Papa R. Heliconius butterflies: a window into the evolution and development of diversity. Curr Opin Genet Dev. 2021 Aug 1;69:72–81.
34. McMillan WO, Livraghi L, Concha C, Hanly JJ. From Patterning Genes to Process: Unraveling the Gene Regulatory Networks That Pattern Heliconius Wings. Front Ecol Evol [Internet]. 2020 [cited 2021 Jul 13];8. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2020.00221/full
35. Zhang L, Mazo-Vargas A, Reed RD. Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. Proc Natl Acad Sci. 2017 Oct 3;114(40):10707–12.

36. Lewis JJ, Geltman RC, Pollak PC, Rondem KE, Belleghem SMV, Hubisz MJ, et al. Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. Proc Natl Acad Sci. 2019 Nov 26;116(48):24174–83.
37. Wallbank RWR, Baxter SW, Pardo-Diaz C, Hanly JJ, Martin SH, Mallet J, et al. Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. PLOS Biol. 2016 Jan 15;14(1):e1002353.
38. Morris J, Hanly JJ, Martin SH, Van Belleghem SM, Salazar C, Jiggins CD, et al. Deep Convergence, Shared Ancestry, and Evolutionary Novelty in the Genetic Architecture of Heliconius Mimicry. Genetics. 2020 Nov 1;216(3):765–80.
39. Mazo-Vargas A, Concha C, Livraghi L, Massardo D, Wallbank RWR, Zhang L, et al. Macroevolutionary shifts of WntA function potentiate butterfly wing-pattern diversity. Proc Natl Acad Sci. 2017 Oct 3;114(40):10701–6.
40. Van Belleghem SM, Rastas P, Papanicolaou A, Martin SH, Arias CF, Supple MA, et al. Complex modular architecture around a simple toolkit of wing pattern genes. Nat Ecol Evol. 2017 Jan 30;1(3):1–12.
41. Nadeau NJ, Pardo-Diaz C, Whibley A, Supple MA, Saenko SV, Wallbank RWR, et al. The gene *cortex* controls mimicry and crypsis in butterflies and moths. Nature. 2016 Jun;534(7605):106–10.
42. Livraghi L, Hanly JJ, Belleghem SMV, Montejo-Kovacevich G, Heijden ESM van der, Loh LS, et al. Cortex cis-regulatory switches establish scale colour identity and pattern diversity in Heliconius. bioRxiv. 2021 Mar 17;2020.05.26.116533.
43. Pinharanda A, Rousselle M, Martin SH, Hanly JJ, Davey JW, Kumar S, et al. Sexually dimorphic gene expression and transcriptome evolution provide mixed evidence for a fast-Z effect in Heliconius. J Evol Biol. 2019;32(3):194–204.
44. Challis RJ, Kumar S, Dasmahapatra KK, Jiggins CD, Blaxter M. Lepbase: the Lepidopteran genome database. bioRxiv. 2016 Jun 6;056994.
45. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio [Internet]. 2013 May 26 [cited 2021 Jul 13]; Available from: http://arxiv.org/abs/1303.3997
46. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep 1;20(9):1297–303.
47. Browning SR, Browning BL. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. Am J Hum Genet. 2007 Nov 1;81(5):1084–97.
48. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904–9.
49. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.
50. Martin SH, Van Belleghem SM. Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. Genetics. 2017 May 1;206(1):429–38.
51. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008 May;18(5):821–9.
52. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10(1):421.
53. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul 15;30(14):3059–66.
54. van’t Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The

industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016 Jun;534(7605):102–5.
55. Van't Hof AE, Reynolds LA, Yung CJ, Cook LM, Saccheri IJ. Genetic convergence of industrial melanism in three geometrid moths. Biol Lett. 2019 Oct 31;15(10):20190582.
56. Bruders R, Hollebeke HV, Osborne EJ, Kronenberg Z, Maclary E, Yandell M, et al. A copy number variant is associated with a spectrum of pigmentation patterns in the rock pigeon (Columba livia). PLOS Genet. 2020 May 20;16(5):e1008274.
57. Junker J, Rick JA, McIntyre PB, Kimirei I, Sweke EA, Mosille JB, et al. Structural genomic variation leads to genetic differentiation in Lake Tanganyika's sardines. Mol Ecol. 2020;29(17):3277–98.
58. Mérot C, Oomen RA, Tigano A, Wellenreuther M. A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. Trends Ecol Evol. 2020 Jul 1;35(7):561–72.
59. Hanly JJ, Wallbank RWR, McMillan WO, Jiggins CD. Conservation and flexibility in the gene regulatory landscape of heliconiine butterfly wings. EvoDevo. 2019 Jul 16;10(1):15.
60. Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 2012 Jul;487(7405):94–8.
61. Livraghi L, Martin A, Gibbs M, Braak N, Arif S, Breuker CJ. Chapter Three - CRISPR/Cas9 as the Key to Unlocking the Secrets of Butterfly Wing Pattern Development and Its Evolution. In: ffrench-Constant RH, editor. Advances in Insect Physiology [Internet]. Academic Press; 2018 [cited 2021 Sep 21]. p. 85–115. (Butterfly Wing Patterns and Mimicry; vol. 54). Available from: https://www.sciencedirect.com/science/article/pii/S0065280617300346
62. Martin SH, Davey JW, Salazar C, Jiggins CD. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLOS Biol. 2019 Feb 7;17(2):e2006288.
63. Moest M, Belleghem SMV, James JE, Salazar C, Martin SH, Barker SL, et al. Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. PLOS Biol. 2020 Feb 6;18(2):e3000597.
64. VanKuren NW, Massardo D, Nallu S, Kronforst MR. Butterfly Mimicry Polymorphisms Highlight Phylogenetic Limits of Gene Reuse in the Evolution of Diverse Adaptations. Mol Biol Evol. 2019 Dec 1;36(12):2842–53.
65. van der Burg KRL, Lewis JJ, Brack BJ, Fandino RA, Mazo-Vargas A, Reed RD. Genomic architecture of a genetically assimilated seasonal color pattern. Science. 2020 Nov 6;370(6517):721–5.
66. Reed RD, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, et al. optix Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. Science. 2011 Aug 26;333(6046):1137–41.
67. Ficarrotta V, Hanly JJ, Loh LS, Francescutti CM, Ren A, Tunström K, et al. A genetic switch for male UV-iridescence in an incipient species pair of sulphur butterflies. bioRxiv. 2021 May 23;2021.05.21.445125.
68. Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, Halder G, et al. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc Natl Acad Sci. 2012 Jul 31;109(31):12632–7.
69. Martin SH, Singh KS, Gordon IJ, Omufwoko KS, Collins S, Warren IA, et al. Whole-chromosome hitchhiking driven by a male-killing endosymbiont. PLOS Biol. 2020 Feb 27;18(2):e3000610.
70. Gilbert LE, Forrest HS, Schultz TD, Harvey DJ. Correlations of ultrastructure and pigmentation suggest how genes control development of wing scales of Heliconius

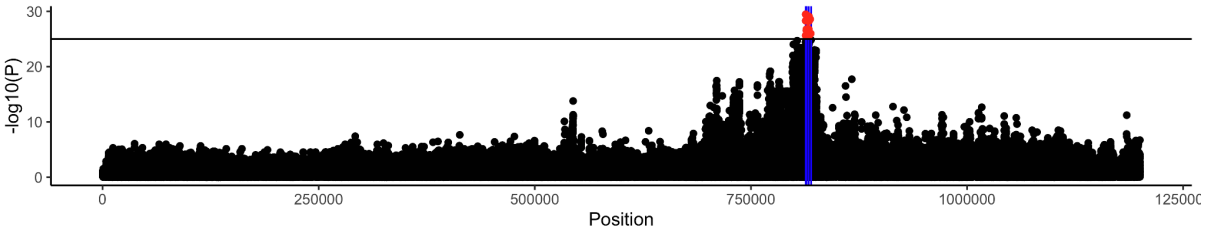
butterflies. J Res Lepidoptera USA [Internet]. 1987 [cited 2019 Oct 29]; Available from: http://agris.fao.org/agris-search/search.do?recordID=US8909729
71. Nijhout HF, Wray GA. Homologies in the colour patterns of the genus Heliconius (Lepidoptera: Nymphalidae). Biol J Linn Soc. 1988 Apr;33(4):345–65.
72. Courtier-Orgogozo V, Arnoult L, Prigent SR, Wiltgen S, Martin A. Gephebase, a database of genotype-phenotype relationships for natural and domesticated variation in Eukaryotes. Nucleic Acids Res. 2020 Jan 8;48(D1):D696–703.
73. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. Science. 2010 Jan 15;327(5963):302–5.
74. McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature. 2011 Mar 10;471(7337):216–9.
75. Domyan ET, Kronenberg Z, Infante CR, Vickrey AI, Stringham SA, Bruders R, et al. Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. eLife. 2016 Mar 15;5:e12115.
76. Nadeau NJ, Minvielle F, Ito S, Inoue-Murayama M, Gourichon D, Follett SA, et al. Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian ASIP (agouti) gene. Genetics. 2008 Feb;178(2):777–86.
77. Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature. 2011 Sep;477(7363):203–6.
78. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, et al. Genomic architecture and introgression shape a butterfly radiation. Science. 2019 Nov 1;366(6465):594–9.
79. Bhardwaj V, Semplicio G, Erdogdu NU, Manke T, Akhtar A. MAPCap allows high-resolution detection and differential expression analysis of transcription start sites. Nat Commun. 2019 Dec;10(1):3219.
80. Singer GA, Wu J, Yan P, Plass C, Huang TH, Davuluri RV. Genome-wide analysis of alternative promoters of human genes using a custom promoter tiling array. BMC Genomics. 2008 Jul 25;9(1):349.
81. Boukhatmi H, Bray S. A population of adult satellite-like cells in Drosophila is maintained through a switch in RNA-isoforms. eLife. 2018 Apr 9;7:e35954.
82. Surridge AK, Lopez-Gomollon S, Moxon S, Maroja LS, Rathjen T, Nadeau NJ, et al. Characterisation and expression of microRNAs in developing wings of the neotropical butterfly Heliconius melpomene. BMC Genomics. 2011 Jan 26;12(1):62.
83. Gemmell NJ, Slate J. Heterozygote Advantage for Fecundity. PLOS ONE. 2006 Dec 27;1(1):e125.
84. Hedrick PW. Heterozygote advantage: the effect of artificial selection in livestock and pets. J Hered. 2015 Apr;106(2):141–54.
85. Küpper C, Stocks M, Risse JE, Dos Remedios N, Farrell LL, McRae SB, et al. A supergene determines highly divergent male reproductive morphs in the ruff. Nat Genet. 2016 Jan;48(1):79–83.
86. Maisonneuve L, Chouteau M, Joron M, Llaurens V. Evolution and genetic architecture of disassortative mating at a locus under heterozygote advantage. Evolution. 2021;75(1):149–65.
87. Cong Q, Shen J, Zhang J, Li W, Kinch LN, Calhoun JV, et al. Genomics Reveals the Origins of Historical Specimens. Mol Biol Evol. 2021 May 1;38(5):2166–76.
88. Grewe F, Kronforst MR, Pierce NE, Moreau CS. Museum genomics reveals the Xerces blue butterfly (Glaucopsyche xerces) was a distinct species driven to extinction. Biol Lett. 17(7):20210123.

89. Weatherbee SD, Frederik Nijhout H, Grunert LW, Halder G, Galant R, Selegue J, et al. Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. Curr Biol. 1999 Feb 11;9(3):109–15.
90. Pierrat V, Descimon H. A new wing pattern mutant in the Apollo butterfly, *Parnassius apollo* (L. 1758) (Lepidoptera: Papilionidae). Ann Société Entomol Fr NS. 2011 Jan;47(3–4):293–302.
91. MacLachlan G, Hurlburt J, Suarez M, Wong KL, Burke W, Lewis T, et al. Building a Shared Resource HPC Center Across University Schools and Institutes: A Case Study. ArXiv200313629 Cs [Internet]. 2020 Apr 2 [cited 2021 Jul 13]; Available from: http://arxiv.org/abs/2003.13629

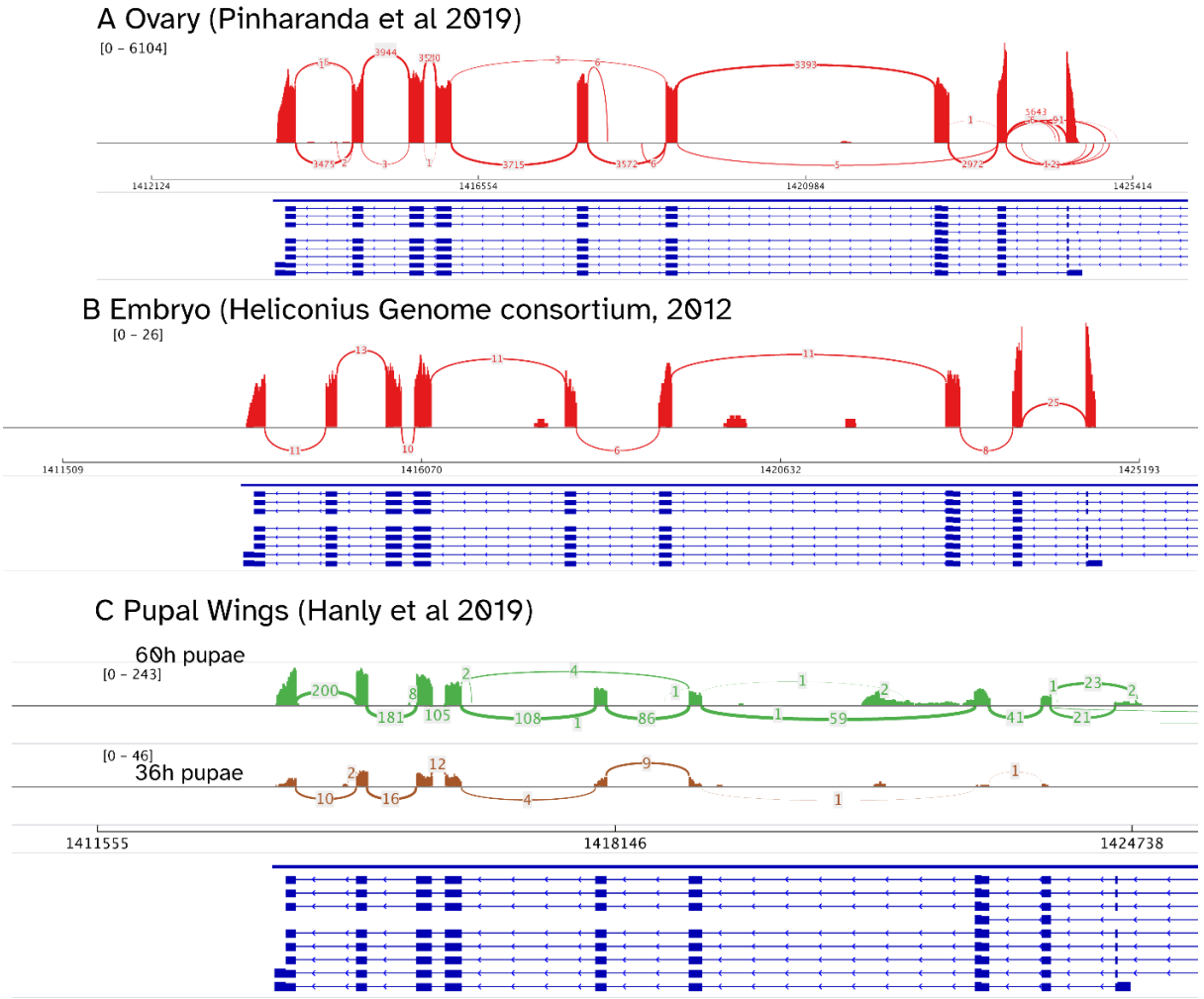
S1 Table : Samples and accessions.

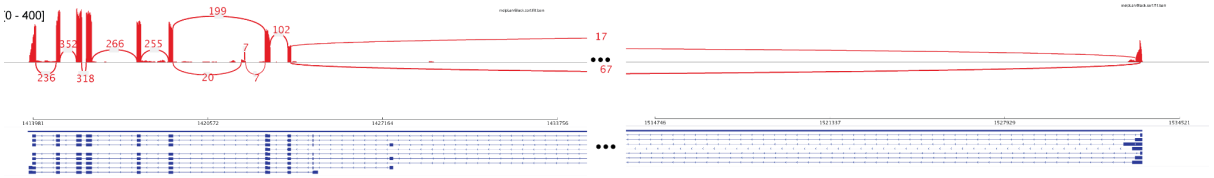
Accession	BioSample Accession	Sample Name	Phenotype	Inferred Genotype	Date Collected	Coverage	%GC
SRR11243376	SAMN14276602	P1001	Piano Keys, dark	*B/-, d/d, r/r, [ivory WT/WT]*	30 Oct 2014	13.01	32.53
SRR11243375	SAMN14276603	P1002	Piano Keys, dark	*B/-, d/d, r/r, [ivory WT/WT]*	30 Oct 2014	14.94	32.97
SRR11243364	SAMN14276604	P1006	Piano Keys, dark	*B/-, d/d, r/r, [ivory WT/WT]*	23 Feb 2014	16.18	32.52
SRR11243353	SAMN14276605	P1007	Piano Keys, dark	*B/-, D/-, r/r, [ivory WT/WT]*	19 Feb 2013	14.93	32.38
SRR11243344	SAMN14276606	P1008	Piano Keys, dark	*B/-, D/d, r/r, [ivory WT/WT]*	19 Feb 2013	13.42	32.83
SRR11243343	SAMN14276607	P1009	*cydno galanthus*		19 Nov 2019	16.14	32.41
SRR11243342	SAMN14276608	P1010	*cydno x* pale PK		19 Nov 2019	14.07	32.37
SRR11243341	SAMN14276609	P1011	F1 pale PK x *cydno*		6 Feb 2019	16.14	32.5
SRR11243340	SAMN14276610	P1012	F1 pale PK x *cydno*		2 Feb 2019	16.62	32.6
SRR11243339	SAMN14276611	P1013	BC of F1 x *cydno*		4 Feb 2019	12.49	32.66
SRR11243374	SAMN14276612	P1014	BC of F1 x *cydno*		4 Feb 2019	14.77	32.19
SRR11243373	SAMN14276613	P1015	BC of F1 x *cydno*		19 Feb 2019	15.93	32.44
SRR11243372	SAMN14276614	P2001	ivory	*[ivory -/-]*	21 Jan 2019	15.37	32.24
SRR11243371	SAMN14276615	P2002	ivory	*[ivory -/-]*	21 Jan 2019	11.69	32.48
SRR11243370	SAMN14276616	P2003	ivory	*[ivory -/-]*	6 Jan 2019	8.95	32.45
SRR11243369	SAMN14276617	P2004	ivory	*[ivory -/-]*	30 Sep 2019	16.11	32.49
SRR11243368	SAMN14276618	P2005	ivory	*[ivory -/-]*	23 Sep 2019	16.19	32.74
SRR11243367	SAMN14276619	P2006	ivory	*[ivory -/-]*	10 Feb 2019	11.47	33.03
SRR11243366	SAMN14276620	P2007	ivory	*[ivory -/-]*	21 Jan 2019	15.44	32.77
SRR11243365	SAMN14276621	P2008	ivory	*[ivory -/-]*	1 Oct 2019	17.40	32.6
SRR11243363	SAMN14276622	P2009	ivory	*[ivory -/-]*	24 Jan 2019	12.99	32.64
SRR11243362	SAMN14276623	P2010	Piano Keys, pale	*B/-, d/d, r/r, [ivory WT/-]*	24 Jan 2019	13.25	32.78
SRR11243361	SAMN14276624	P2011	ivory	*[ivory -/-]*	20 Jun 2019	16.63	32.32
SRR11243360	SAMN14276625	P3005	Piano Keys, dark	*B/-, d/d, r/r, [ivory WT/WT]*	19 Nov 2019	15.92	32.71
SRR11243359	SAMN14276626	P3006	Piano Keys, dark	*B/-, D/-, R/- [ivory WT/WT]*	19 Nov 2019	14.44	32.71
SRR11243358	SAMN14276627	P3007	Piano Keys, dark	*B/-, D/-, R/- [ivory WT/WT]*	16 Dec 2018	16.45	32.43
SRR11243357	SAMN14276628	P3016	Piano Keys, dark	*B/-, d/d, r/r, [ivory WT/WT]*	27 Nov 2018	9.35	32.58
SRR11243356	SAMN14276629	P3017	Piano Keys, dark	*B/-, d/d, r/r, [ivory WT/WT]*	14 Dec 2018	19.50	32.4
SRR11243355	SAMN14276630	P3019	Piano Keys, dark	*B/-, d/d, r/r, [ivory WT/WT]*	27 Jan 2019	15.40	32.41
SRR11243354	SAMN14276631	P4001	Piano Keys, dark	*B/-, d/d, r/r, [ivory WT/-]*	19 Feb 2019	13.24	32.89
SRR11243352	SAMN14276632	P4002	Piano Keys, pale	*B/-, D/-, R/-, [ivory WT/-]*	19 Nov 2019	13.85	32.62
SRR11243351	SAMN14276633	P4003	Piano Keys, pale	*B/-, d/d, r/r, [ivory WT/-]*	5 Feb 2019	13.20	32.48
SRR11243350	SAMN14276634	P4004	Piano Keys, pale	*B/-, d/d, r/r, [ivory WT/-]*	7 Dec 2019	15.82	32.87
SRR11243349	SAMN14276635	P4005	Piano Keys, pale	*B/-, d/d, r/r, [ivory WT/-]*	19 Nov 2019	17.01	32.68
SRR11243348	SAMN14276636	P4006	Piano Keys, pale	*B/-, D/-, R/-, [ivory WT/-]*	19 Nov 2019	18.17	32.63
SRR11243347	SAMN14276637	P4007	Piano Keys, pale	*B/-, d/d, r/r, [ivory WT/-]*	26 Jan 2019	9.74	32.77
SRR11243346	SAMN14276638	P4008	Piano Keys, pale	*B/-, d/d, r/r, [ivory WT/-]*	19 Dec 2018	11.18	33.58
SRR11243345	SAMN14276639	P4009	Piano Keys, pale	*B/-, d/d, r/r, [ivory WT/-]*	19 Dec 2019	13.96	32.63

S1 Figure: Dennis GWAS zoomed on Hmel218003o, showing the tip of the association peak (-log10P > 25) aligning precisely with the Dennis element designated by Morris et al (2020).



S2 Figure: Sashimi plot of *H. melpomene* RNA splicing at *cortex*. Ovary from Pinharanda et al 2019 (A), embryo from Dasmahapatra et al 2012 (B), pupal wings from Hanly et al 2019 (C), with 60h pupal wings above and 36h pupal wings below. Annotations of splice variants from *cortex* from the Hmel2.5 annotation are included, running in the antiparallel direction. Note that while expression of cortex is at least 20-fold higher in ovary compared to embryo or wing, no expression or splicing to the distal TSS was detected.





S3Figure: Sashimi plot of *H. melpomene* RNA splicing at the full cortex annotation from larval wings, from Hanly et al 2019. The ellipsis indicates an 81kb interval. Here, the TSS at the distal promoter is used, and the proximal promoter and TSS are not transcribed.