A transposable element insertion is associated with an alternative life history strategy

Alyssa Woronik1,7*, Kalle Tunström1, Michael W. Perry2,8, Ramprasad Neethiraj1, Constanti Stefanescu3,4, Maria de la Paz Celorio-Mancera1, Oskar Brattström5, Jason Hill1,9, Philipp Lehmann1, Reijo Käkelä6 & Christopher W. Wheat1*

Tradeoffs affect resource allocation during development and result in fitness consequences that drive the evolution of life history strategies. Yet despite their importance, we know little about the mechanisms underlying life history tradeoffs. Many species of Colias butterflies exhibit an alternative life history strategy (ALHS) where females divert resources from wing pigment synthesis to reproductive and somatic development. Due to this reallocation, a wing color polymorphism is associated with the ALHS: either yellow/orange or white. Here we map the locus associated with this ALHS in Colias crocea to a transposable element insertion located downstream of the Colias homolog of BarH-1, a homeobox transcription factor. Using CRISPR/Cas9 gene editing, antibody staining, and electron microscopy we find white-specific expression of BarH-1 suppresses the formation of pigment granules in wing scales and gives rise to white wing color. Lipid and transcriptome analyses reveal physiological differences associated with the ALHS. Together, these findings characterize a mechanism for a female-limited ALHS.
A life history strategy is a complex pattern of co-evolved life history traits (e.g. number of offspring, size of offspring, and lifespan), that is fundamentally shaped by tradeoffs that arise because all fitness components cannot simultaneously be maximized. Therefore, finite resources are competitively allocated to one life history trait versus another within a single individual, and selection acts on these allocation patterns to optimize fitness. Evolutionary theory predicts that positive selection will remove variation from natural populations, as genotypes with the highest fitness go to fixation. However, across diverse taxa alternative life history strategies (ALHSSs) are maintained within populations at intermediate frequencies due to balancing selection. Life history theory was developed using methods, such as quantitative genetics, artificial selection, demography and modeling to gain significant insights into the causes and consequences of genetic and environmental variation on life history traits. Yet despite these advances, a key challenge that remains is to identify the proximate mechanisms underlying tradeoffs, especially for ecologically relevant tradeoffs that occur in natural populations. Here, we characterize one such mechanism underlying an ALHS in the butterfly *Colias crocea* (Pieridae, Lepidoptera) (Geoffroy, 1785).

*Colias* butterflies (the clouded sulphurs) are common throughout the Holarctic and can be found on every continent except Australia and Antarctica. In approximately a third of the nearly 90 species within the genus, females exhibit two alternative wing-color morphs: yellow or orange (depending on the species) and white5–8 (Fig. 1a). The wing color polymorphism arises because during pupal development the white morph, also known as Alba, reallocates larval derived resources from the synthesis of energetically expensive colored pigments to reproductive and somatic development. This tradeoff has been well characterized in *Colias crocea*, the Old World species that we focus upon in this work, via radio-labelled metabolite tracking in pupae, as well as in the New World species *Colias eurytheme* (Pieridae, Lepidoptera) (Boisduval, 1852) using ultraviolet spectrophotometry. As a result of the resource reallocation, Alba females exhibit faster pupal development, a larger fat body, and significantly more mature eggs at eclosion compared to orange females. However, despite these developmental advantages and the dominance of the Alba allele, the polymorphism is maintained by several abiotic and biotic factors. For example, males preferentially mate with orange females, as wing color is an important cue for mate recognition. This mating bias likely has significant fitness costs for Alba females because males transfer essential nutrients during mating, and multiply mated females have more offspring over their lifetime. The mating bias against Alba females is conserved within the genus *Colias*, as Alba females of the North American species *C. eurytheme* also exhibit fewer pigment granules within wing scales compared to orange females. Finally, we use lipid and transcriptome analyses to characterize physiological differences associated with the ALHS in *C. crocea* and find evidence that the fitness-related traits associated with the ALHS are also conserved between *C. crocea* and *C. eurytheme*. Together these findings characterize the mechanism underlying a female-limited ALHS.

**Results**

**Mapping the Alba locus.** Using a de novo reference genome for *C. crocea* that we generated via Illumina and PacBio sequencing, and three rounds of bulk segregant analyses (BSA) using whole-genome sequencing from a female and two male informative crosses for Alba, we mapped the Alba locus to an ~3.7 Mbp region. Then, with whole-genome re-sequencing data from 15 Alba and 15 orange females from diverse population backgrounds, a SNP association study fine mapped the Alba locus to a ~430 kb contig that fell within the ~3.7 Mbp locus identified using the BSA crosses. The majority of SNPs significantly associated with Alba (n = 70 of 72) were within or flanking a Jockey-like transposable element (TE) (Fig. 1c). We determined that the TE insertion was unique to the Alba morph in *C. crocea* by quantifying differences in read depth between morphs within and flanking the insertion (Supplementary Fig. 1) and assembling orange and Alba haplotypes for this region (Supplementary Fig. 2). We also validated the insertion by testing that reads for each morph mapped as expected across the two haplotypes (Supplementary Fig. 2 & 3) and used PCR to validate the presence or absence, respectively, of the insertion in 25 Alba and 57 orange wild-captured females (Supplementary Fig. 4). We also found no evidence of a TE insertion in the homologous region of other Lepidopteran genomes (*Bombyx mori* & *Heliconius melpomene*) (Supplementary Fig. 1).

**Functional investigation of the Alba locus.** The Alba-specific insertion was located ~30 kb upstream of a gene encoding a DEAD-box helicase, and ~6 kb downstream of the *Colias* homolog of *BarH-1*, a homeobox transcription factor. We use antibody staining to confirm Alba-specific expression of *BarH-1* in the scale building cells of pupal wings and use CRISPR/Cas9 gene editing to validate *BarH-1*’s functional role in the wing-color switch. We then use scanning electron microscopy to determine that *BarH-1* expression gives rise to white wing color by reducing the number of pigment granules within Alba wing scales. We find evidence that the Alba mechanism is likely conserved across the genus *Colias*, as Alba females of the North American species *C. eurytheme* also exhibit fewer pigment granules within wing scales compared to orange females. Finally, we use lipid and transcriptome analyses to characterize physiological differences associated with the ALHS in *C. crocea* and find evidence that the fitness-related traits associated with the ALHS are also conserved between *C. crocea* and *C. eurytheme*. Together these findings characterize the mechanism underlying a female-limited ALHS.
displayed no white/orange mosaic on the wing. These results indicate BarH-1 expression suppresses orange coloration in the wings. We also observed black and green mosaic coloring of eyes in KO males and females of both morphs, where green eyes are the wild-type color (Fig. 1e). These results indicate BarH-1 also plays a role in Colias eye development.

We next investigated how the Alba color change manifests within wings. Butterfly wing color can arise either due to the absorption of light by pigments deposited within the scales, or by the scattering of light via regularly arranged nanostructures in the scales19. Colias butterflies have pteridine pigments. These pigments are synthesized within the wings and previous work using ultraviolet spectrophotometry in C. eurytheme found Alba females exhibit dramatic reductions in colored pteridine pigments compared to orange9,10,20,21. Studies on Drosophila eyes indicate pteridines are synthesized in pigment granules22,23 and pigment granules containing pteridines are concentrated within wing scales of Pierid butterflies24. However, whether morphs differed in wing scale morphology was unknown. To investigate wing morphology, we used scanning electron microscopy and found white scales from Alba individuals exhibited a dramatic and significant reduction in pigment granules, compared to orange scales (t5.97 = 2.93, p = 0.03, 95% confidence interval = −25.034, 144.967, Welch two sample t-test, n = 6 individuals, mean Alba scales = 66.9, mean orange scales = 126.9) (Fig. 3a, b). These results indicate the color change to white is caused by reduced pigment granule formation. Congruent with this interpretation, CRISPR KO Alba individuals exhibited significantly fewer pigment granules in scales from the white wild-type region compared to scales in orange BarH-1 KO regions (Fig. 3c) (t5.45 = 10.78, p < 0.001, 95% confidence interval = 57.10, 91.70, Welch two sample t-test, n = 10 wing scales within a single mosaic individual, mean white scales = 32.2, mean orange scales = 106.6). To further test whether reduction in pigment granule amount alone was sufficient for the orange to white color change, we chemically removed the pigment granules containing pteridines from the wing of an orange C. crocea female. This resulted in formerly orange regions turning white (Fig. 3d). Wings likely appear white after granule removal due to the scattering of light from the remaining non-lamellar

Fig. 1 Color variation in Colias crocea and the genetic mechanism of Alba. a Colias crocea male, orange female, and Alba female (left to right). b SNPs significantly associated with the Alba phenotype (red) within the ~3.7 Mbp Alba locus identified via three rounds of bulk segregant analysis. Contigs in this region shown as alternating dark and light blue. c The location of Alba-associated SNPs (red) on the ~430 kb outlier contig identified in the GWAS. Gene models for the DEAD-box helicase, the Jockey-like transposable element, and BarH-1 shown at the top of the panel. d Wings of a female with an Alba genotype following CRISPR/Cas9 mosaic knockout of BarH-1, wild-type regions are white, knockout regions are orange. Orange color is seen on the dorsal forewing (top) and hindwing (bottom). e BarH-1 mosaic knockout also leads to black regions in the eyes, wild-type regions are green. Source data are provided as a Source Data file. Panel d and e photo credit John Hallmén.
These results demonstrate that BarH-1’s suppression of pigment granule formation in Alba wing scales, results in the white color of Alba females in C. crocea. Thus, we propose the resource tradeoff between color and development arises due to a classic Y reallocation model, wherein limited resources are competitively allocated and increased investment in one trait results in a decreased investment to another. Within the energetically closed system of a developing pupa, reduced pigment granule formation would likely result in reduced pigment synthesis, which would in turn leave more resources free to be used for other developmental processes. Finally, we also observed scale building cells in black regions of both morphs express BarH-1 (Fig. 2a, d) and also lack pigment granules (Fig. 3a, b), but these scales appear black due to melanin deposition within the scale. These results suggest BarH-1 may also repress pigment granule formation within black scales.

The Alba mechanism is assumed to be conserved across Colias. Therefore, we wished to test whether Alba females from the New World species Colias eurytheme also exhibited fewer pigment granules than orange females. Indeed, we found orange C. eurytheme scales exhibited abundant pigment granules, while Alba scales almost lacked granules (Fig. 3e, f). These results demonstrate white wing color arises via the same morphological mechanism within Colias and corroborate previous assumptions that Alba is conserved across the genus.

Comparing morph physiology within and between Colias species. To validate that other aspects of the Alba/orange alternative life history strategy are conserved across the genus we tested whether one of the physiological tradeoffs of Alba reported for a New World species was also seen in C. crocea. In C. eurytheme, Alba females have larger fat bodies than orange females and the strength of the Alba advantage increased in cold temperatures. To compare abdominal lipid stores between morphs in C. crocea, we conducted high performance thin layer chromatography on 2-day-old adult females reared under two temperature treatments (Hot: 27 °C vs. Cold: 15 °C during pupal development). Adults were not allowed to feed before samples were taken, therefore these measurements reflect larval stores, where the putative energetic tradeoff should be more clearly visible. We found that recently eclosed adult Alba females had larger abdominal lipid stores than orange in both temperature treatments, though the difference was only significant in the cold treatment (cold: n = 32, mean Alba = 0.545, mean orange = 0.346, t_{29.12} = 3.42, P = 0.002, 95% confidence interval = 0.080, 0.318, hot: n = 25, mean Alba = 0.654, mean orange = 0.575, t_{22.71} = 0.67, P = 0.51, 95% confidence interval = −0.166, 0.324, Welch two sample t-test, two sided) (Fig. 4a). These results are consistent with previous reports from New World Colias species and indicate that the morph-specific tradeoff associated with the color change is also conserved across the genus.
upregulated within Alba abdomens (all p-values reported from the GSEA are the result of a Fisher’s Exact Test in the R package topGO using the weight01 algorithm, Supplementary Data 1). Additionally, in our differential expression analysis a gene encoding a triacylglycerol lipase was significantly upregulated within Alba abdomen tissue (log fold change [log FC] of 4.8) (Fig. 4b). Triacylglycerol composes more than 90% of the lipids stored in the fat body and during times of energy demand triacylglycerol lipases mobilize these stores27. For example, during oogenesis there is a massive shift in lipid distribution from the fat body to ovaries as lipids comprise 30–40% of the dry weight of insect oocytes27. Taken together these results suggest that, similar to C. eurytheme, Alba females of C. crocea may be benefitting from increased oogenesis compared to orange females. We also observe an enrichment of ‘defense response to Gram-positive bacterium’ (GO:0050830, 0.00027) for genes upregulated within Alba abdomens. Interestingly, previous work has suggested that Alba females may have increased sensitivity to viral infection19. Further investigation of potential morph-specific tradeoffs between wing color and immunity is of interest.

For genes downregulated in Alba abdomens the GSEA revealed significant enrichment of ‘regulation of nucleoside metabolic process’ (GO:0009118, p-value < 0.0001) and ‘regulation of purine nucleotide catabolic process’ (GO:0033121, p-value < 0.0001) (Supplementary Data 2). Colias synthesize pteridines in their wings from purine precursors and Alba females exhibit dramatic reductions in colored pteridines compared to orange females20. Thus, downregulation of these GO terms in Alba abdomens may result from a decrease in purine precursors (guanosine triphosphate [GTP]) being shunted from the abdomen to the wings for pteridine synthesis. Additionally, consistent with previous reports of GTP reallocation from wings to other areas of development in Alba females9 we also observed significant enrichment for ‘positive regulation of GTPase activity’ (GO:0043547, p-value < 0.0001). Additionally, RIM, a Rab3 GTPase effector28, was one of the most highly differentially expressed (DE) genes in both tissues (logFC increase in Alba of 3.4 in the abdomen and 5.1 in the wings) (Fig. 4b, c). RIM is localized to the plasma membrane and forms a GTP-dependent complex between the membrane and vesicles to mediate calcium-regulated exocytosis29. If and how RIM plays a role in the Alba-associated GTP reallocation is unknown. However, RIM is known to be involved in exocytosis across diverse tissues and taxa. For example, in the neuronal synapse, RIM is involved in neurotransmitter release, a function that is evolutionarily conserved across mammals and insects28,30. While in the prothoracic gland of D. melanogaster, RIM plays an essential role in the release of the hormone ecdyson31. Ecdysone is largely known for its role in regulating the timing of molting and metamorphosis in insects32; however, it can also affect immunity33, longevity34 and reproduction, specifically ovarian maturation and oogenesis35. The physiological changes associated with ecdyson are congruent with the fitness traits associated with Alba, thus investigating whether morphs differ in ecdyson levels may be interesting. Additionally, RIM expression has not been previously reported in wings and investigating the role it may play in this tissue is of interest.

Within wings, BarH-1 was not differentially expressed at the time of pigment synthesis (~70% of pupal development, Supplementary Table 1), indicating that morph-specific expression differences are temporal (antibody staining of BarH-1 in pupal wings was conducted 48 h after pupation in 27 °C with 20 h of daylight, ~30% of pupal development). However, we did observe that genes downregulated in Alba wings were significantly enriched for ‘xanthine dehydrogenase activity’ (p = 0.02, GO:0004854) (Supplementary Data 3). Xanthine dehydrogenase is the enzyme that catalyzes the xanthopterin to leucopterin

We then investigated the transcriptome of pupal abdomen and wing tissue at the time of pteridine synthesis (>70% of pupal development, Supplementary Table 1) to identify genes that exhibited differential expression between morphs and therefore may play a role in the morph-specific differences in physiology that arise due to the resource tradeoff (Fig. 4b, c). In C. eurytheme Alba females emerge from the pupa with significantly more mature eggs than orange females31 and we find evidence that suggests similar dynamics are occurring in C. crocea. A gene set enrichment analysis (GSEA) revealed that ‘embryo development ending in birth or egg hatching’ (GO:0009792, p = 0.00072), ‘proteasome-mediated ubiquitin-dependent protein catabolic process’ (GO:0043161, p = 0.00073), and ‘protein degradation’ (GO:0006508, p = 0.00101) were within the top 5 terms enriched and

Fig. 3 Colias forewings and scanning electron microscopy of wing scale nanostructures. a C. crocea wild-type Alba female wing and wing scale structure. The top panel shows the scanning electron microscope (SEM) image of a black scale; pigment granules are absent. The bottom panel shows a white scale, exhibiting near absence of pigment granules. b Wing and wing scale structures of a wild-type orange C. crocea female. The top panel shows a black scale, pigment granules are absent. The bottom panel shows an orange scale with abundant pigment granules. c Wing and wing scales of a C. crocea female with an Alba genotype (i.e. transposable element insertion present) exhibiting CRISPR/Cas9 mosaic knockout of BarH-1. The top panel shows a wild-type white scale, where pigment granules are mostly absent. The bottom panel shows a scale in an orange BarH-1 KO region. It exhibits significantly more pigment granules than the white scales (t\textsubscript{45} = 10.78, p < 0.001, Welch two sample t-test, n = 10 scales in a single mosaic individual). d Wing and wing scale of an orange C. crocea female where pigment granules have been chemically removed from the distal half of the wing. The SEM image shows a scale from the white region with pigment granules completely missing. The white color of this wing section presumably results from light reflection off the remaining scale nanostructures. e Wing and wing scale structure of a C. eurytheme Alba female. Wing scales exhibit few pigment granules, similar to the phenotype observed in C. crocea. f Wing and wing scale structures of a C. eurytheme orange female. Orange scales show abundant pigment granules, again consistent with the orange phenotype observed in C. crocea. All scale bars are 2 µm.
conversion during pteridine synthesis in Colias butterflies\textsuperscript{9,21} and previous work in D. melanogaster found xanthine dehydrogenase is localized within type II pigment granules, which synthesize and store pteridines, within the eye\textsuperscript{22,23,36,37}. These results are consistent with previous studies in C. eurytheme that reported the level of xanthopterin in Alba wings was 7–8 fold less than in orange\textsuperscript{9}. Additionally we observed enrichment of ‘MAP kinase activity’ (GO:0004709, \( p = 0.00109 \)) in genes downregulated within Alba wings. In Drosophila, BarH-1 represses Decapentaplegic, a morphogen that is homolog to TGF\( \beta \)\textsuperscript{38}. TGF\( \beta \) can activate signalling cascades, including MAP kinase pathways\textsuperscript{39}. Previous work in Drosophila has also suggested an interaction between Bar homebox genes and Ras/MAP kinase signalling during eye development\textsuperscript{40}. Future functional studies of the above mentioned candidate genes are needed to better understand their mechanistic roles in morph-specific development and the trade-offs associated with the ALHS.

### Discussion

Here we characterize the proximate mechanisms underlying a female-limited ALHS in a natural population. Historically, the field of life history research has treated mechanistic details as a black box\textsuperscript{5}, though recently several genetic mechanisms underlying ecologically relevant ALHSs have been identified, e.g. in the wall lizard\textsuperscript{41}, ruff\textsuperscript{42,43}, white throated sparrow\textsuperscript{44} and fire ant\textsuperscript{45}. The majority of these studies found that supergenes, large loci that maintain many genes in tight linkage due to structural variation, gave rise to the alternative morphs\textsuperscript{42–46}. Such findings established that structural variation facilitates the evolution of complex traits. However these genomic architectures make determining the specific contributions of individual genes to ALHSs difficult, though there have been significant advances made in the white throated sparrow\textsuperscript{44}. In contrast, recent work in the wall lizard\textsuperscript{41}, found that ALHSs arose due to changes in the regulatory regions of two genes. Based on the intergenic location of the Alba-associated insertion and the Alba-specific expression of BarH-1, we predict that the insertion affects the regulation of BarH-1. However, this raises the question of how the locus gives rise to the other fitness-related traits associated with the ALHS. Our parsimonious hypothesis is that the Alba-associated physiological and developmental traits arise due to a classic Y location model, where reduced pigment granule formation results in reduced pigment synthesis, which in turn leaves more resources free to be used for other developmental processes within the energetically closed system of the developing pupa. However, other possibilities also exist. The other fitness-related traits may emerge due to the insertion affecting the expression of BarH-1 in other tissues, or surrounding genes could also have altered expression, either in the wings or other tissues. Alternatively, other mutations, located near the insertion, may cause these traits. Dissecting these details is an important avenue of ongoing research.

Previous work has shown that BarH-1 plays a role in the morphogenesis of neurons, leg segments, and eyes in Drosophila\textsuperscript{47}. Specifically, BarH-1 expression is required for the formation of pigment granules and red pteridine pigments in the Drosophila eye\textsuperscript{48}. We find that BarH-1 also plays a role in eye and wing color in Colias butterflies. However, as BarH-1 expression represses the formation of pigment granules within Colias wings, we find it has a reversed function in Drosophila and Colias. This may be one of several examples where either whole or a part of a

---

**Fig. 4 Physiological differences between female morphs of C. crocea.** a The mass corrected total neutral lipid content within abdomens for female morphs in two temperature treatments. White boxes are Alba females, while orange boxes are orange females. Alba females, on average, have larger neutral lipid stores in their abdomens than orange females. However there is an interaction between morph and temperature as the difference is only significant in the cold treatment (cold: \( t_{29.12} = 3.42, P = 0.002, n = 32 \) abdomens hot: \( n = 25 \) abdomens, \( t_{22.71} = 0.67, P = 0.51 \)). Boxplot center line, median; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range from the hinge; points, outliers. b Volcano plot to visualize gene expression differences between female morphs in pupal abdominal tissue. Each point is a gene. Genes not significantly differentially expressed between morphs are grey, while differentially expressed genes are blue. The black square is the triacylglycerol lipase and the black triangle is RIM. The X-axis is the log of the fold change (FC), positive log(FC) indicates the gene is upregulated in Alba individuals. c Volcano plot to visualize gene expression differences between female morphs in pupal wing tissue. Color coding, shapes, and axes are the same as in panel B. Source data are provided as a Source Data file.
gene regulatory network that regulates eye development has been co-opted to give rise to a novel trait in the insect wing. If so, future work could investigate what aspects of the network have been co-opted and how this lead to BarH-1’s contrasting roles in morphogenesis.

Additionally, recent work in the field of butterfly wing evolutionary-development has found that several genes are repeatedly involved in wing color variation across distantly related species. Such genes (e.g. optix, WntA, and cortex) form a patterning toolkit. BarH-1 might serve as another toolkit gene for patterning wing color in butterflies beyond Colias as we found BarH-1 expression in scale building and socket cells of developing wings in Vanessa cardui pupae (Nymphalidae, Lepidoptera) (Linnaeus, 1758) (Supplementary Fig. 5). However, the functional role of BarH-1 in V. cardui wings remains to be determined.

BarH-1 may have a novel function within V. cardui wings. Alternatively, the function of BarH-1 as a repressor of pigment granule formation could be conserved, as V. cardui scales do not have pigment granules. Under the latter assumption, we would expect that BarH-1 is not expressed in closely related Pierinae species that, despite appearing white, exhibit abundant pigment granules that are primarily filled with the UV-absorbing pteridine called leucopterin. Future work investigating the evolutionary history of BarH-1’s co-option to the wing and function in other species could shed light on how complex traits such as ALHSSs evolve.

Methods

Ethical approval. This study is compliant with all relevant ethical regulations for animal testing and research. Ethical approval is not required for experiments involving Lepidoptera in Sweden, thus no approval was sought.

Genome assembly. An orange female and male carrying Alba (offspring from wild-caught butterflies, Catalonia, Spain) were mated in the lab. DNA from an Alba female offspring of this cross was extracted using a salting-out method. DNA quality and quantity were assessed using a Nanodrop 8000 spectrophotometer (Thermo Scientific) and a Qubit 2.0 fluorometer (Invitrogen). A 190 bp insert size paired-end library (101 bp reads) was prepared (TruSeq PCR-free) and sequenced on an Illumina Hiseq 4000 at the Beijing Genomics Institute (Shenzhen, China). A Nextera mate-pair library with a 3 kb insert size was prepared and paired-end library (101 bp reads) was prepared (TruSeq PCR-free) and Illumina sequencing (101 bp PE HiSeq2500), at the Beijing Genomics Institute (Shenzhen, China). The same read cleaning, mapping and SNP calling pipeline used on the female informative cross was applied to this dataset. Resulting SNP sites from the F1 pools were filtered in R, for a read depth ≥20 and ≤50, a biallelic state, and a minimum minor allele frequency of 3. The mother’s mpileup was analyzed as described for the female informative cross and SNP sites were filtered for a read depth ≥20 and ≤50. A SNP site was considered a male informative cross I Alba SNP when it met the following expectations: (1) homozygous in the orange mother, (2) homozygous in the orange pool, (3) the allele frequency difference in the Alba pool compared to the orange pool was 0.45–0.55.

Male Informative Cross II. A male carrying Alba mated an orange female in the lab. DNA was extracted using a combination of salting-out and phenol chloroform (Supplementary Methods), then was evaluated and prepared as described above for 26 Alba and 28 orange female offspring. This resulted in two DNA pools. Library preparation (TruSeq PCR-free) and Illumina sequencing (151 bp paired-end reads with 350 bp insert, Hiseq-X), was performed at Science for Life Laboratory (Stockholm, Sweden). The same read cleaning, mapping and SNP calling pipeline used on the female informative and male informative I crosses was applied, except that there was no mother sequenced for the second male informative cross. The resulting output table for the pools was filtered in R, for sites with a read depth ≥20 and ≤50, a biallelic state, and a minor allele frequency of 3. For each pool, we calculated the allele frequency difference in the Alba pool compared to the orange pool and were filtered for a read depth ≥20 and ≤50. A SNP site was considered a male informative cross II Alba SNP if (1) the SNP site was homozygous in the orange pool, (2) the allele frequency difference in the Alba pool compared to the orange pool was 0.45–0.55.

A contig was considered Alba-associated if it had ≥3 Alba SNPs in all three crosses. Nineteen Alba-associated contigs were identified. They totaled ~3.7 Mbp and are considered the Alba BSA locus (Supplementary Table 2).

Genome wide association study. DNA for genome re-sequencing was extracted from 15 Alba and 15 orange females from diverse population backgrounds (Spain, and Capri, Italy) using a salting-out protocol. Library preparation was conducted using Illumina TruSeq and sequencing was conducted at the Science for Life Laboratory (Stockholm, Sweden) (150 bp paired-end reads Hiseq-X). Raw reads were filtered and trimmed as described in the genome assembly section. Cleaned reads were mapped to the annotated reference genome using NextGenMap. Bam files were filtered and sorted using SAMTOOLS. A VCF file was generated using SAMTOOLS and bcftools. Read depth per site was calculated using VCFtools. VCFtools was used to call SNP sites with minor allele frequency >50%. Missing data, an average read depth between 15-50 across individuals, and a minimum SNP quality of 30. An analysis association was performed with PLINK and a Benjamini & Hochberg step-up FDR control was applied. SNPs with FDR < 0.05 were considered Alba SNPs. We conducted this analysis both genome wide and only within the BSA locus. Both analyses fine mapped the Alba locus to the same genomic region.

Validating the Alba insertion. Synteny, or gene order, is highly conserved within Lepidoptera. Thus, to validate that the contig carrying the Alba locus (C. crocea contig 12) was properly assembled we compared gene order among homologous

1% agarose gel. Equal amounts of high-quality DNA from the 21 F1 Alba individuals was combined into a single pool, and the same was conducted for the 21 F1 orange individuals. The two mixed individual pools of DNA and DNA from the Alba mother underwent library preparation (TruSeq PCR-free) and Illumina sequencing (101 bp PE Hiseq2500), at the Beijing Genomics Institute (Shenzhen, China). Raw reads were filtered and trimmed as described in the genome assembly section. Cleaned reads were mapped to the C. crocea reference genome using NextGenMap. SAMTOOL54 was used to filter, sort, and index the bam files and generate mpileup files for the two pools and the Alba mother. Insertions and deletions were identified and masked using Popoolation26 and Popoolation66, respectively. Popoolation was used to convert the F1 mpileup files to a sync files and calculate the allele frequency difference between Alba and orange pools. Resulting SNP sites from the F1 pools were filtered in R, for a read depth ≥20 and ≤50 and a biallelic state. For the Alba mother, the major and minor allele frequencies were calculated in R by dividing the major and minor allele count in the mpileup file by the read depth at each site. Heterozygous SNPs in the mother were further filtered for a read depth ≥15 and ≤50. Based on the frequency of Alba to orange females in the F1 we determined the Alba mother was heterozygous for Alba (Aa) and had mated a male also heterozygous for Alba (Aa). Thus, a SNP was considered Alba-associated in the female informative cross if it met the following criteria: (1) the site was heterozygous in the Alba mother (i.e. allele frequency between 0.4 and 0.6). (2) The same SNP site was homozygous in the F1 orange pool (i.e. major allele at 100% frequency). (3) The allele frequency difference between the F1 orange pool and the F1 Alba pool was between 0.4 and 0.8, (4) The nucleotide change at the SNP site was the same in both the F1 Alba pool and the Alba mother datasets.

Male Informative Cross I: DNA was extracted from a wild-caught orange mother (Catalonia, Spain) and 26 of her Alba and 24 of her orange female offspring using a salting-out method (Supplementary Methods). DNA quality and quantity of each individual was assessed as in the female informative cross, before pooling equal amounts of high-quality DNA from Alba and orange offspring into two pools, respectively. Library preparation (TruSeq PCR-free) and Illumina sequencing (101 bp PE Hiseq2500), was performed on the two pools and the orange mother at the Beijing Genomics Institute (Shenzhen, China). The same read cleaning, mapping and SNP calling pipeline used on the female informative cross was applied to this dataset. Resulting SNP sites from the F1 pools were filtered in R, for a read depth ≥20 and ≤50, a biallelic state, and a minimum minor allele frequency of 3. The mother’s mpileup was analyzed as described for the female informative cross and SNP sites were filtered for a read depth ≥20 and ≤50. A SNP site was considered a male informative cross I Alba SNP when it met the following expectations: (1) homozygous in the orange mother, (2) homozygous in the orange pool, (3) the allele frequency difference in the Alba pool compared to the orange pool was 0.45–0.55.

Male Informative Cross II: A male carrying Alba mated an orange female in the lab. DNA was extracted using a combination of salting-out and phenol chloroform (Supplementary Methods), then was evaluated and prepared as described above for 26 Alba and 28 orange female offspring. This resulted in two DNA pools. Library preparation (TruSeq PCR-free) and Illumina sequencing (151 bp paired-end reads with 350 bp insert, Hiseq-X), was performed at Science for Life Laboratory (Stockholm, Sweden). The same read cleaning, mapping and SNP calling pipeline used on the female informative and male informative I crosses was applied, except that there was no mother sequenced for the second male informative cross. The resulting output table for the pools was filtered in R, for sites with a read depth ≥20 and ≤50, a biallelic state, and a minor allele frequency of 3. For each pool, we calculated the allele frequency difference in the Alba pool compared to the orange pool and were filtered for a read depth ≥20 and ≤50. A SNP site was considered a male informative cross II Alba SNP if (1) the SNP site was homozygous in the orange pool, (2) the allele frequency difference in the Alba pool compared to the orange pool was 0.45–0.55.

A contig was considered Alba-associated if it had ≥3 Alba SNPs in all three crosses. Nineteen Alba-associated contigs were identified. They totaled ~3.7 Mbp and are considered the Alba BSA locus (Supplementary Table 2).

Bulge segregant analyses (BSA). Female Informative Cross: This data were previously published in Woronick and Wheat, 2017 to identify the chromosome carrying the Alba locus. Methods in brief are as follows. DNA was extracted from a wild-caught Alba mother (Catalonia, Spain) and 21 of her Alba and 21 of her orange female offspring using a salting-out method (Supplementary Methods). DNA quality and quantity for each individual was assessed via a Nanodrop 8000 spectrophotometer (Thermo Scientific, MA, USA), a Qubit 2.0 Fluorometer (dsDNA BR; Invitrogen, Carlsbad, CA, USA), and by running extracted DNA on a
regions in Bombyx mori (chromosome 15) and Helicobasid melpomene (scaffold Hmel211089) by doing a tblastn search against Kaikobase v.3.2.2, default settings) and BLAST-n (blastn), respectively, that were annotated to 
C. croce contig 12 (Supplementary Fig 1A&B). Next, an analysis of the read depth using the 15 Alba and 15 orange re-sequencing datasets mapped to our high-quality reference genome indicated that the locus was an Alba-specific insertion (Supplementary Fig 1C). Within this predicted insertion, MESPA annotated a T7-like promoter sequence, an orange nucleotide insertion. These nucleotides lacked a TE insertion in this region as we assembled the orange haplotype by performing a de novo genome assembly on the wild-caught, orange mother of male larvae. The primer amplification reads from 12 of the 15 Alba individuals, could properly map across the predicted insertion site on the orange haplotype. We surmised that the reads from the Alba individuals that could map properly on the orange haplotype likely arose from an orange allele and that these individuals were heterozygous for Alba (An). To test this we looked whether these reads could map to Alba haplotype. Due to the reads already identified via SNP annotation the reads did not properly map to the Alba haplotype, lending support to our hypothesis (Supplementary Fig 2). To further test that these individuals were heterozygous for Alba, we compared the read depth within the insertion to the read depth of a nearby conserved genomic region. We found a reduction of the read depth to be half the read depth within the insertion. This confirmed the heterozygous nature of these individuals, lending further support to our prediction that they were heterozygous for Alba (Supplementary Fig 3). Finally, we validated the insertion using PCR. We developed primers that spanned the insertion (i.e. one primer sat within the insertion the other right outside) (Alba F: TGGGAgCTGTTAGTATGCT, Alba R: TGCAATTTGGCTGCAAGAAAGCTG). Primers amplifying a region of Cytochrome C were used as a positive control (F: GGATCACCTGATATAGCATTCCC, R: CCTGGTAAAATTAAAATATAA). To test this, staged pupal wings were dissected and fixed 48 h post-pupation. Pupal wings were fixed using 4% paraformaldehyde in 1X PBS, then rinsed twice with 1X PBS, then again for 30 min on a shaker at room temperature. For blocking, PBS was replaced with 200 ul of 5% normal goat serum in PBST and incubated for 30 min on a shaker at room temperature. The blocking solution was then replaced with primary antibody solution and incubated overnight at 4°C. The primary antibody solution was then replaced with PBST, rinsed three times, and was washed on a shaker at room temperature for at least one hour with 3–4 additional solution changes. Wash solution was replaced with secondary antibody solution and incubated on a shaker either overnight at 4°C or 2 h at room temperature. The antibody solution was then washed with PBST and rinsed three times. Washing in PBST continued for at least 1 h on a shaker at room temperature. The Rabbit-anti-Bar antibody was used at 1:100, followed by secondary antibody staining with Alexa Fluor 555-anti-Rabbit secondary antibody (ThermoFisher, Waltham, MA, USA), Cat #A11031. After 1:500 and counterstaining with DAPI at 1ug/ml in 1X PBS. Images were captured using standard confocal microscopy on a Leica SP5.

CRISPR/Cas9 knockouts. The guide-RNA (gRNA) sequences were generated by first manually looking for PAM-sites (NGG) in the exon region of BarH-1. Preference was given to sites in the first half of the gene and near the ends of the exons (Supplementary Table 3). Uniqueness of the target regions was confirmed using a NCBI nucleotide blast (ver. 2.5.0 + using blastn-short flag and filtering for an e-value of 0.01) against the C. croce reference genome. gRNA constructs were ordered from Integrated DNA Technologies (Corvallis, Iowa, USA) as DNA (gBlocks). Full gRNA constructs had the following configuration: a M13F region, a spacer sequence, a T7-promoter sequence, the target specific sequence, a Cas9 binding site, and finally a P505 sequence. The M13F sequence was amplified using PCR. For each gBlock, forty 50 ul reactions were conducted using the M13F and P505 primers (P505: AAAAAAAGACCGCAGCTGGTGCGCG, M13F: GTAAGAACGGCGCCGACGAT and Taq polymerase (Invitrogen cat. 10966-034). The four reactions were then combined and purified in a Qiagen MinElute spin column (cat. 28004, Venlo, Netherlands). The resulting template was transcribed using the Lucigen AmpliScribe T7-flash Transcription Kit from Epicentre/illumina (cat. ASF3507, Madison, WI, USA) followed by purification via ammonium acetate precipitation. Products were resuspended with QiaGen buffer EB, concentrations were quantified by Qubit and further diluted to 1000 ng/μl. They were then mixed with Cas9-NLS microinjection protein (PNA Bio, Norbury Park, CA, USA), diluted to a final concentration of 125–250 ng/μl. C. croce females (n = 40) from Aquamaillons de l’Empordá, Spain were captured and kept in morph-specific flight cages in the lab at Stockholm University where they oviposited on alfalfa (Medicago sativa). Eggs were collected between 1–2 h-post laying, removed from the leaf with a paintbrush, and sterilized in 7% benzalkonium chloride for ~5 min. Eggs were then rinsed in water and dried before being attached to a glass slide. Eggs were held in place on the slide using double-sided tape. Needle tips were broken using forceps and front-loaded with 0.8–1 μl of the construct (gRNA mixed with Cas9). Injections were either at a concentration of 125 or 250 ng/μl and conducted using a M-152 Narishige micromanipulator (Narishige International Limited, London, UK) with a 50 μl glass needle syringe, with injection pressure applied by hand via a syringe fitting. For more details regarding this protocol see Perry et. al. 2016. Slides with injected eggs were placed in petri dishes together with a damp paper towel placed in a sealed plastic container with another piece of damp paper towel in order to maintain high humidity until the eggs hatch. Freshly hatched larvae were transferred to fresh M. sativa plants using a paintbrush and kept in a climate room at 27°C with 20 h of daylight. Upon eclosion, adults were visually expected for knockout phenotypes. To see results from various gblocks see Supplementary Table 4.

CRISPR/Cas9 validation. To validate the mutation, Cas9 cut sites were PCR amplified and a ~370 bp region centered on the intended cut site was sequenced using Sanger sequencing. Two lane MiSeq 300 bp paired-end sequencing was designed using Primer3. DNA was isolated from KO-individuals using KingFisher Cell and Tissue DNA Kit from ThermoFisher Scientific (N1997) and the robotic KingFisher Duo Purification system. DNA quality and quantity was assessed via a Nanodrop 8000 spectrophotometer (Thermo Scientific, MA, USA) and a Qubit 2.0 Fluorometer (DNA RR; Invitrogen, Carlsbad, CA, USA). Aliquots were then taken and diluted to 1 ng/μl before ampling the region over the cleavage-site. Sequences were amplified and ligated with Illumina adapter and indexed in a two-step process following the protocol provided by Science for Life Laboratories (Stockholm, Sweden) and illumina. First, we amplified the ~370 bp long sequence around cut sites and attached Illumina adapter and indexed using a second round of PCR (Accustart II PCR Supermix from Quanta Bio [Beverly, MA, USA], settings 94°C x 2 min followed by 40 cycles of 94°C x 30 sec + 60°C x 15 sec + 72°C x 1 min followed by 68°C x 5 min). PCR products were purified using Qiagen Qiaquick (Cat. 28104). Concentration and quality of the product were assessed via NanoDrop and gel electrophoresis. DNA was diluted to -0.5 ng/μl and then the unique double indices were attached by the second round of PCR (same protocol as above). The final PCR products were purified again using Qiaquick spin columns and concentration and size was assessed using Qubit fluorometer and gel electrophoresis. All samples were then mixed at equal molarity and sent for sequencing at Science for Life Laboratories (Stockholm, Sweden). Sequences were aligned to their respective fragments (area surrounding cut site) using SNAP (ver. 1.0beta18), identical reads were clustered using the collapse utility in Fastx-Toolkit. Sequences containing deletions were extracted and the most abundant sequences containing deletions were selected for confirmation of deletion in the expected region (Supplementary Fig 6).

Pigment granule removal from wings. This protocol was published in Rutowski et. al. 2005. In brief, to remove the pterins, and consequently the pigment granules, from the wings of orange Colias females, wings were removed from dead females and thorough wetted with 70% isopropyl alcohol. The wing was then dipped in 1% NH4O solution made with reignet grade water. Wings were then placed on a paper towel under a glass slide to dry.

Electron microscopy. To quantify pigment granule differences between wild-type Alba and orange individuals pieces of the forewing were mounted on aluminum pin stubs (2 mm length) with the dorsal side upwards. Samples were then placed in gold for 80 sec using an Agar spatter coater and imaged under 5 kV acceleration voltage, high vacuum, and ETD detection using a scanning electron microscope (Quanta Feg 650, FEI, Hillsboro, Oregon, USA). The same protocol was used for the wing where pigment granules were chemically removed. To quantify pigment granules for each individual (three Alba and three orange) we selected images from the same magnification and randomly placed three 4 μm2 squares on the scale image. We counted the number of pigment granules within each square and took the average (Supplementary Table 5), then conducted a Welch two sample t-test in R. For additional SEM images that document the variation in pigment granule density of wild-type Alba and orange see Supplementary Figure S3.
**Lipid analysis.** Wild caught C. crocea Alba females (Catalonia, Spain) were dissected in the lab on M. sativa. Eggs were moved into individual rearing cups and randomly split between two temperature treatments (hot: 27 °C and 16 h day length during larval and pupal development, cold: reared at 22 °C with a 16 h day length during larval development and 15 °C with a 16 h day length during pupal development). Once pupated, individuals were checked a minimum of every 12 h. Upon eclosion adults were stored at 4 °C until the next day to provide time for meconium excretion. Butterflies were not allowed to feed before dissection. Body weight was taken using a Sauter RE1614 scale before dissection. Total lipids were extracted using the Folch method. In brief, frozen butterfly abdomens were first homogenized in 1 ml aliquot of chloroform:methanol (2:1,92. The tissue water estimate included) in Eppendorf tubes using a Tissuelyser Bead Homogenizer (Qiagen, Hilden, Germany), and then the lysate was transferred to 10 ml glass tubes and 4.5 ml of chloroform:methanol (2:1) was added. The chloroform and aqueous phases separated when 0.9 ml water was added, and after mixing and centrifugation steps the lower chloroform phase with the lipids was recovered. The remaining aqueous phase was re-extracted with the theoretical lower phase, and the recovered chloroform phases from the two extraction steps were combined, evaporated under nitrogen and dissolved into 750 μl of chloroform:methanol (1:2).

This sample solution (stored in deactivated glass vials for a maximum of three days at −20 °C) was used for high performance thin layer chromatography (HPTLC) analysis. The HPTLC was conducted as follows: Five microliters of the sample lipids extract was applied on a silica plate with a Camag Automatic TLC Sampler 4 (Camag, Muttenz, Switzerland). After the silica plate developed it was scanned with a Camag TLC plate scanner at 254 nm using a deuterium lamp with a slit dimension of 6 x 0.45 mm and analyzed with the WinCATS 1.1.3.0 software. Peaks representing the four major neutral lipid classes (diacylglycerols, triacylglycerols, cholesterol and cholesterol esters) were identified by comparing their retention times against known standards. Then the peak areas were integrated and the amount of lipid within each class was calculated using the formula: pmol sample = (Area sample × Area standard) / pmol standard. The total lipid content (nmol per abdomen) was calculated as a sum of pmol contents of all neutral lipid classes. For the statistical analyses this value was regressed against abdomen weight and standardized residuals (i.e. mass-corrected storage lipid amount). For more details regarding methods see Woronik et al. 20184.2.

**Transcriptome assembly, differential expression, and gene set enrichment analysis.** Offspring from a wild-caught Alba female from Catalonia, Spain were reared at Stockholm University. When larvae reached the fifth instar they were checked for differentiation every second and the pupation time of each individual was recorded. Tissue was collected between 82% and 92% of pupal development (Supplementary Table 7). Pupae were dissected in 1xPBS solution, and the abdomen and wings were flash frozen in liquid nitrogen and stored at −80 °C. RNA was extracted from the abdomen and wing tissues using Trizol. RNA quality and quantity was assessed using a NanoDrop spectrophotometer and a NanoPhotometer (Thermo Scientific) with a期待 (electrophoresis machine using the manufacturer protocol (Bio-Rad, Hercules, CA). Library preparation (Strand-specific TrueSeq RNA libraries using poly-A selection) and sequencing (101 bp PE HiSeq2500—high output mode) was performed at the Science for Life Laboratories (Stockholm, Sweden). In total 16 libraries were sequenced (4 orange and 4 Alba individuals—wings and abdomen from each individual). Raw reads were adapter filtered and trimmed using the BBmap software package (Bushnell B. sourceforge.net/projects/bbmap/). Cleaned reads from all libraries were used for a de novo transcriptome assembly using Trinity4.3, 4. To reduce the redundancy among contigs and produce a biologically valid transcript set, the raw reads pipeline from the Evidential software package was run on the raw Trinity assembly. The tr2aacds pipeline utilizes the raw Trinity assembly. The tr2aacds pipeline utilizes

**Code availability** Software versions and parameters can be found in the Supplementary Methods.

Received: 29 June 2019; Accepted: 15 November 2019; Published online: 17 December 2019

**References**

1. Stearns, S. C. *The Evolution of Life Histories.* (Oxford University Press, 1992).
2. Stearns, S. C. Trade-offs in life-history evolution. *Funct. Ecol.* 3, 259–268 (1989).
3. Fisher, R. A. *The Genetic Theory of Natural Selection.* (Oxford University Press, 1930).
4. Gross, M. R. Alternative reproductive strategies and tactics: Diversity within sexes (vol 11, pg 92, 1996). *Trends Ecol. Evol.* 11, 263–268 (1996).
5. Platt, T. & Heyland, A. *Mechanisms of Life History Evolution: The Genetics and Physiology of Life History Traits and Trade-Offs.* (Oxford University Press, 2011).
6. Remington, C. L. *The genetics of Colias (Lepidoptera).* *Adv. Genet.* 6, 403–450 (1954).
7. Limeri, L. B. & Morehouse, N. I. The evolutionary history of the ‘alba’ polymorphism in the butterfly subfamily Coliadinae (Lepidoptera: Pieridae). *Biol. J. Linn. Soc.* 117, 716–724 (2016).
8. Gerould, H. H. The inheritance of polymorphism and sex in *Colias philodice.* *Am. Naturalist* 45, 257–283 (1911).
9. Watt, W. B. Adaptive significance of pigment polymorphisms in *Colias Butterflies.* *J. Prog. in Study of Alba Variant.* Evolution 27, 537–548 (1973).
10. Deschamps, H. & Bemettier, J. L. Nitrogen-Metabolism in the *Colias croceus* (Linne) and Its Alba Mutant (Lepidoptera, Pieridae). *J. Insect Physiol.* 35, 881–885 (1989).
11. Graham, S. M., Watt, W. B. & Gall, L. F. Metabolic resource-allocation vs mating attractiveness - adaptive pressures on the Alba polymorphism of *Colias Butterflies.* *Nat. Acad. Sci. Biol.* 77, 3615–3619 (1980).
12. Woronik, A., Stefanscu, C., Kakela, R., Wheat, C. W. & Lehmahn, P. Physiological differences between female limited, alternative life history strategies: The Alba phenotype in the butterfly *Colias croceus.* *J. Insect Physiol.* 107, 257–264 (2018).
13. Nielsen, M. G. & Watt, W. B. Behavioural fitness component effects of the alba polymorphism of *Colias* (Lepidoptera, Pieridae): resource and time budget analysis. *Funct. Ecol.* 12, 149–158 (1998).
14. Nielsen, M. G. & Watt, W. B. Interference competition and sexual selection promote polymorphism in *Colias Butterflies.* *Lepidoptera, Pieridae*. *Funct. Ecol.* 14, 718–730 (2000).
15. Hovanitz, W. The biology of *Colias butterlies.* II. Parallel geographic variation of dimorphic color phases in North American species. *Wasmann J. Biol.* 8, 197–219 (1950).
16. Boggs, C. L. & Watt, W. B. Population structure of pierid butterflies IV. Genetic and physiological investment in offspring by male *Colias. Oecologia* 50, 320–324 (1981).
17. Wijklund, C., Karlsson, B. & Leimar, O. Sexual conflict and cooperation in butterfly reproduction: a comparative study of polyandry and female fitness. *Proc. Biol. Sci.* 268, 1661–1667 (2001).
18. Higashijima, S. et al. Dual Bar homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. *Genes Dev.* 6, 50–60 (1992).
19. Niljohøt, H. The Development and Evolution of Butterfly Wing Patterns. (Smithsonian Institution Press, 1991).
20. Watt, W. B. Pigment biosynthesis in the butterfly *Colias eurytheme.* *J. Biol. Chem.* 242, 565–572 (1967).
21. Arrese, E. L. & Soulages, J. L. Insect fat body: energy, metabolism, and precursors of Drosophila melanogaster—iso-lization and characterization for synthesis of sepiapterin.

22. Montell, I., Rasmuson, A., Rasmuson, B. & Holmgren, P. Uptake and incorporation in pteridines of externally supplied GTP in normal and pigmentoocytes of Drosophila melanogaster.

23. Morehouse, N. I., Vukusic, P. & Rutowski, R. Pterin pigment granules are responsible for both broadband light scattering and wavelength selective absorption in the wing scales of Pieris butterflies.

24. Simon, A. F., Shih, C., Mack, A. & Benzer, S. Steroid control of longevity in Drosophila melanogaster.

25. Knott, L. & Whitfield, M. J. The rosy locus in Drosophila melanogaster.

26. Watt, W. B. Xanthine Dehydrogenase and Pteridine Metabolism in insects: from stem cells to ovarian follicle formation.

27. Geppert, M. & Sudhof, T. C. RAB3 and synaptotagmin: the yin and yang of neurotransmitter release.

28. Reaume, A. G., Knecht, D. A. & Chovnick, A. The rosy locus in Drosophila melanogaster.

29. Kupper, C. et al. A supergene determines highly divergent male reproductive strategies in the ruff (Philomachus pugnax).

30. Kline, K., Satoh, K., Naka, M. & Ueno, N. Sex in Pieris rapae L. and the supergene content of their wings.

31. Aljanabi, S. M. & Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques.

32. Nadeau, N. J. et al. The gene cortex controls mimicry and crypsis in butterflies and moths.

33. Neidwied, A. et al. Dynamics of F-actin prefigure the structure of butterfly wing scales.

34. Makino, K., Satoh, K., Koike, M. & Ueno, N. Sex in Pieris rapae L. and the supergene content of their wings.

35. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics.

36. Schwander, T., Libbrecht, R. & Keller, L. Supergenes and complex phenotypes.

37. Almean, A., Sampaio, M. & Moya, L. and the supergene.

38. Neethiraj, R., Hornett, E. A., Hill, J. A. & Wheat, C. W. Investigating the genomic basis of discrete phenotypes using a Pool-Seq-only approach: New insights into the genetics underlying color variation in diverse taxa.

39. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019.

40. Woronik, A. & Wheat, C. W. Advances in finding Alba: the locus affecting life history and color polymorphism in a Colias butterfly.

41. Monteiro, A. Gene regulatory networks reused to build novel traits: co-option.

42. Kline, K., Satoh, K., Koike, M. & Ueno, N. Sex in Pieris rapae L. and the supergene content of their wings.

43. Nadeau, N. J. et al. The gene cortex controls mimicry and crypsis in butterflies and moths.

44. Neidwied, A. et al. Dynamics of F-actin prefigure the structure of butterfly wing scales.

45. Makino, K., Satoh, K., Koike, M. & Ueno, N. Sex in Pieris rapae L. and the supergene content of their wings.
Acknowledgements
We thank Karin Kiontke, Christen Bossu, Naomi Keehnen, and Peter Pruisscher for helpful comments on the manuscript. We thank Lovisa Wennerström, Elishia Harji, Jofre Carnicer, and Christina Hansen Wheat for help with fieldwork. We thank Marianne Ahlbom for assistance with the SEM. We thank the Department of Zoology at Stockholm University, the Swedish Research Council 2012-3715, the Academy of Finland 131155, the Knut and Alice Wallenberg Foundation 2012.0058 and the Erik Philip-Sörensens foundation for funding. Sequencing was performed at SciLifeLab, National Genomics Infrastructure, Stockholm and Uppsala, Sweden. Open access funding provided by Stockholm University.

Author contributions
A.W. conducted butterfly rearings and lab work, analysed the data, and wrote the manuscript with C.W.W. and input from the coauthors. A.W., M.W.P., K.T., and C.W.W. conducted the CRISPR/Cas9 knockout experiment. A.W. and K.T. conducted the electron microscopy. M.W.P. conducted antibody staining. R.N. and J.H. assisted with bioinformatics. P.L. and R.K. conducted HPTLC and A.W. and P.L. analyzed the data. A.W., C.S., C.W.W. and O.B. conducted fieldwork. M.C. conducted lab work. C.W.W. supervised the work at all stages.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-13596-2.