

1 **A transposable element insertion is the switch between alternative life history strategies**

2

3 **Authors:** Alyssa Woronik<sup>\*1,2</sup>, Kalle Tunström<sup>1</sup>, Michael W. Perry<sup>2,3</sup>, Ramprasad Neethiraj<sup>1</sup>,  
4 Constanti Stefanescu<sup>4,5</sup>, Maria de la Paz Celorio-Mancera<sup>1</sup>, Oskar Brattström<sup>6</sup>, Jason Hill<sup>1,7</sup>,  
5 Philipp Lehmann<sup>1</sup>, Reijo Käkälä<sup>8</sup>, Christopher W. Wheat<sup>\*1</sup>

6

7 **Author affiliations:**

8 <sup>1</sup> Department of Zoology, Stockholm University, S106 91 Stockholm, Sweden

9 <sup>2</sup> Department of Biology, New York University, New York, New York 10003, USA

10 <sup>3</sup> Division of Biological Sciences, University of California San Diego, La Jolla, California 92093,  
11 USA

12 <sup>4</sup> Museum of Natural Sciences of Granollers, Granollers, Catalonia 08402, Spain

13 <sup>5</sup> CREAM, Cerdanyola del Valles, Catalonia 08193, Spain

14 <sup>6</sup> Department of Zoology, University of Cambridge, Cambridge CB23EJ, United Kingdom

15 <sup>7</sup> Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

16 <sup>8</sup> Helsinki University Lipidomics Unit, Helsinki Institute for Life Science (HiLIFE) and Molecular  
17 and Integrative Biosciences Research Programme, University of Helsinki, FI00014 Helsinki,  
18 Finland

19

20 **Correspondence to:** AW [alyssa.woronik@zoologi.su.se](mailto:alyssa.woronik@zoologi.su.se) and CWW [chris.wheat@zoologi.su.se](mailto:chris.wheat@zoologi.su.se)

21

22 Tradeoffs affect resource allocation during development and result in fitness consequences that  
23 drive the evolution of life history strategies. Yet despite their importance, we know little about  
24 the mechanisms underlying life history tradeoffs in wild populations. Many species of *Colias*  
25 butterflies exhibit an alternative life history strategy (ALHS) where females divert resources from  
26 wing pigment synthesis to reproductive and somatic development. Due to this reallocation, a  
27 wing color polymorphism is associated with the ALHS: individuals have either yellow/orange or  
28 white wings. Here we map the genetic basis of the ALHS switch in *Colias crocea* to a  
29 transposable element insertion downstream of the *Colias* homolog of *BarH-1*, a homeobox  
30 transcription factor. Using CRISPR/Cas9 gene editing, antibody staining, and electron  
31 microscopy we find morph-specific expression of *BarH-1* suppresses the formation of  
32 pigment granules in wing scales. Lipid and transcriptome analyses reveal physiological  
33 differences associated with the ALHS. These findings characterize a novel mechanism for a

34 female-limited ALHS and show that the switch arises via recruitment of a transcription factor  
35 previously known for its function in cell fate determination in pigment cells of the retina.

36

37 A life-history strategy is a complex pattern of co-evolved life history traits (e.g. number of  
38 offspring, size of offspring, and lifespan<sup>1</sup>), that is fundamentally shaped by tradeoffs that arise  
39 because all fitness components cannot simultaneously be maximized. Therefore, finite  
40 resources are competitively allocated to one life history trait versus another within a single  
41 individual, and selection acts on these allocation patterns to optimize fitness<sup>2</sup>. Evolutionary  
42 theory predicts that positive selection will remove variation from natural populations, as  
43 genotypes with the highest fitness go to fixation<sup>3</sup>. However, across diverse taxa alternative life  
44 history strategies (ALHSs) are maintained within populations at intermediate frequencies due to  
45 balancing selection<sup>4</sup>. Life history theory was developed using methods such as quantitative  
46 genetics, artificial selection, demography, and modeling to gain significant insights into the  
47 causes and consequences of genetic and environmental variation on life history traits. Yet  
48 despite these advances, a key challenge that remains is to identify the proximate mechanisms  
49 underlying tradeoffs, especially for ecologically relevant tradeoffs that occur in natural  
50 populations<sup>5</sup>. Here, we identify the mechanism underlying one such ALHS in the butterfly *Colias*  
51 *crocea* (Pieridae, Lepidoptera) (Geoffroy, 1785).

52

53 *Colias* butterflies (the “clouded sulphurs”) are common throughout the Holarctic and can be  
54 found on every continent except Australia and Antarctica<sup>6</sup>. In approximately a third of the nearly  
55 90 species within the genus, females exhibit two alternative wing-color morphs: yellow or  
56 orange (depending on the species) and white<sup>6,7</sup> (Fig. 1A). The wing color polymorphism arises  
57 because during pupation the white morph, also known as Alba, reallocates larval derived  
58 resources from the synthesis of energetically expensive colored pigments to reproductive and  
59 somatic development<sup>8</sup>. This tradeoff has been well characterized in *Colias crocea*, the Old  
60 World species that we focus upon in this work, via radio-labelled metabolite tracking in pupae<sup>9</sup>  
61 as well as in the New World species *Colias eurytheme*<sup>8</sup> (Pieridae, Lepidoptera) (Boisduval,  
62 1852) using ultraviolet spectrophotometry. As a result of the resource reallocation, Alba females  
63 have faster pupal development, a larger fat body, and significantly more mature eggs at  
64 eclosion compared to orange females<sup>10</sup>. However, despite these developmental advantages  
65 and the dominance of the Alba allele, the polymorphism is maintained by several abiotic and  
66 biotic factors<sup>10-14</sup>. For example, males preferentially mate with orange females, as wing color is  
67 an important cue for mate recognition<sup>10,12,13</sup>. This mating bias likely has significant fitness costs

68 for Alba females because males transfer essential nutrients during mating, and multiply mated  
69 females have more offspring over their lifetime<sup>15,16</sup>. The mating bias against Alba females is  
70 strongest in populations that frequently co-occur with other white Pierid butterfly species due to  
71 interference competition<sup>13</sup>. Also, Alba's development rate advantage is temperature dependent,  
72 with Alba females having faster development in cold temperatures<sup>10</sup>. Field studies confirm Alba  
73 frequency and fitness increases in species that inhabit cold and nutrient poor habitats, where  
74 the occurrence of other white Pierid butterflies is low. While in warm environments with nutrient  
75 rich host plants and a high co-occurrence of other white species, orange females exhibit  
76 increased fitness and frequency<sup>12-14</sup>. Previous work has also suggested Alba females have a  
77 higher sensitivity to viral infections<sup>9</sup>. In all *Colias* species where it has been investigated (n=6),  
78 the switch between the Alba or the orange strategy is controlled by a single, autosomal locus<sup>6</sup>.  
79 This fact, along with ancestral state reconstruction<sup>7</sup>, has led to the assumption that the Alba  
80 locus is conserved within the genus *Colias*, and potentially across the subfamily Coliadinae. Yet,  
81 despite over a century of research on various aspects of Alba biology the mechanism underlying  
82 this polymorphism remained unknown.

83

84 Using a *de novo* reference genome for *C. crocea* that we generated via Illumina and PacBio  
85 sequencing, and three rounds of bulk segregant analyses (BSA) using whole genome  
86 sequencing from a female and two male informative crosses for Alba, we mapped the Alba  
87 locus to a ~3.7 Mbp region (Supplementary Fig.1, & Supplementary Information). Then, with  
88 whole genome re-sequencing data from 15 Alba and 15 orange females from diverse population  
89 backgrounds, a SNP association study fine mapped the Alba locus to a ~430 kb contig that fell  
90 within the ~3.7 Mbp locus identified using the BSA crosses (Fig. 1B and Supplementary  
91 Information). The majority of SNPs significantly associated with Alba (n=70 of 72) were within or  
92 flanking a *Jockey-like* transposable element (TE) (Fig. 1C). We determined that the TE insertion  
93 was unique to the Alba morph in *C. crocea* by assembling orange and Alba haplotypes for this  
94 region, then quantifying differences in read depth between morphs within and flanking the  
95 insertion (Supplementary Information and Supplementary Figs. 2, 3, & 4). We then used PCR to  
96 validate the presence or absence, respectively, of the insertion in 25 Alba and 57 orange wild-  
97 caught females (Supplementary Fig. 7). We also found no evidence of a TE insertion in the  
98 homologous region of other butterfly genomes (*Danaus plexippus* & *Heliconius melpomene*)  
99 (Supplementary Fig. 2).

100

101 The Alba-specific insertion was located ~30 kb upstream of a gene encoding a DEAD-box  
102 helicase, and ~6kb downstream of the *Colias* homolog of *BarH-1*, a homeobox transcription  
103 factor (Fig. 1C). *BarH-1* was an intriguing find as it affects color via pigment granule  
104 development within eyes of *Drosophila melanogaster*<sup>17</sup>. To investigate BarH-1 expression in  
105 developing *C. crocea* wings, we used *in situ* hybridization of BarH-1 on wings from two day old  
106 pupae of orange and Alba females. We found the BarH-1 protein is expressed in scale building  
107 cells within the white wing regions in Alba females (Fig. 2B). We did not observe BarH-1 in scale  
108 building cells from orange areas of the wing in orange females (Fig, 2C). Interestingly however,  
109 we found BarH-1 is expressed in scale building cells within black regions for both morphs (Fig,  
110 2A&D). To validate the functional role of *BarH-1* in the Alba phenotype, we generated  
111 CRISPR/Cas9-mediated deletions within exons 1 and 2 using a mosaic knockout (KO)  
112 approach (Supplementary Information). *BarH-1* KO gave rise to a white/orange color mosaic on  
113 the dorsal side of the wings in females with an Alba genotype (i.e. TE insertion +) (Fig. 1D),  
114 while KO males and orange females displayed no white/orange mosaic on the wing. These  
115 results indicate *BarH-1* expression suppresses orange coloration in the wings. We also  
116 observed black and green mosaic coloring of eyes in KO males and females of both morphs,  
117 where green eyes are the wild type color (Fig. 1E). These results indicate *BarH-1* also plays a  
118 role in *Colias* eye development.

119  
120 We next investigated how the Alba color change manifests within wings. Butterfly wing color can  
121 arise either due to the absorption of light by pigments deposited within the scales, or by the  
122 scattering of light via regularly arranged nanostructures in the scales<sup>18</sup>. *Colias* butterflies have  
123 pteridine pigments. These pigments are synthesized within the wings and previous work using  
124 ultraviolet spectrophotometry in *C. eurytheme* found Alba females exhibit dramatic reductions in  
125 colored pteridine pigments compared to orange<sup>8,9</sup>. In insects, pteridines are synthesized in  
126 pigment granules and pigment granules are concentrated within wing scales of Pierid  
127 butterflies<sup>19,20</sup>. However, whether morphs differed in wing scale morphology was unknown. To  
128 investigate wing morphology, we used scanning electron microscopy and found white scales  
129 from Alba individuals exhibited a dramatic and significant reduction in pigment granules,  
130 compared to orange scales ( $t_{5,97} = 2.93$ ,  $p = 0.03$ ) (Fig 3 A&B). These results indicate the color  
131 change to white is caused by reduced pigment granule formation. Congruent with this  
132 interpretation, CRISPR KO Alba individuals exhibited significantly less pigment granules in  
133 scales from the white wild-type region compared to scales in orange *BarH-1* KO regions ( $t_{5,45} =$   
134 10.78,  $p < 0.001$ ) (Fig. 3C). To further test whether reduction in pigment granule amount alone

135 was sufficient for the orange to white color change, we chemically removed the pigment  
136 granules from the wing of an orange *C. crocea* female. This resulted in formerly orange regions  
137 turning white (Fig. 3D). Wings likely appear white after granule removal due to the scattering of  
138 light from the remaining non-lamellar nanostructures<sup>21</sup>. These results demonstrate that *BarH-1*  
139 suppresses pigment granule formation in wing scales, resulting in the white color of Alba  
140 females in *C. crocea*. Thus, we propose the resource tradeoff between color and development  
141 arises due to a classic Y reallocation model, wherein limited resources are competitively  
142 allocated and increased investment in one trait results in a decreased investment to another<sup>22</sup>.  
143 Within the energetically closed system of a developing pupa, reduced pigment granule  
144 formation would likely result in reduced pigment synthesis, which would in turn leave more  
145 resources free to be used for other developmental processes. Finally, we also observed scale  
146 building cells in black regions of both morphs express BarH-1 and also lack pigment granules  
147 (Fig 2 A&D and Fig 3 A&B), but these scales appear black due to melanin deposition within the  
148 scale<sup>18</sup>. These results suggest BarH-1 may also repress pigment granule formation within black  
149 scales.

150  
151 The Alba mechanism is assumed to be conserved across *Colias*. Therefore, we wished to test  
152 whether Alba females from the New World species *Colias eurytheme* also exhibited significantly  
153 less pigment granules than orange females. Indeed, we found orange *C. eurytheme* scales  
154 exhibited abundant pigment granules while Alba scales almost entirely lacked granules (Fig. 3  
155 E&F). These results demonstrate white wing color arises via the same morphological  
156 mechanism within *Colias* and corroborate previous assumptions that Alba is conserved across  
157 the genus. To further validate that other aspects of the Alba/orange alternative life history  
158 strategy are conserved across the genus we tested whether one of the physiological tradeoffs of  
159 Alba reported for New World species was also seen in *C. crocea*. In *C. eurytheme*, Alba females  
160 have larger fat bodies than orange females and the strength of the Alba advantage increased in  
161 cold temperatures<sup>10</sup>. To compare abdominal lipid stores between morphs in *C. crocea*, we  
162 conducted high performance thin layer chromatography on two day old adult females reared  
163 under two temperature treatments (Hot: 27°C vs. Cold: 15°C during pupal development). Adults  
164 were not allowed to feed before samples were taken, therefore these measurements reflect  
165 larval stores, where the putative energetic tradeoff should be more clearly visible. We found  
166 Alba females had larger abdominal lipid stores than orange in both temperature treatments,  
167 though the difference was only significant in the cold treatment (cold: n=32,  $t_{29,12} = 3.42$ ,  $P =$   
168 0.002, hot: n=25,  $t_{22,71} = 0.67$ ,  $P = 0.51$ ) (Fig. 4A). These results are consistent with previous

169 reports from New World *Colias* species and indicate that the morph-specific tradeoff associated  
170 with the color change is also conserved across the genus.

171  
172 We then investigated the transcriptome of pupal abdomen and wing tissue at the time of  
173 pteridine synthesis to identify genes that exhibited differential expression between morphs and  
174 therefore may play a role in the morph-specific differences in physiology that arise due to the  
175 resource tradeoff (Fig. 4 B&C, Supplementary Information, Supplementary Tables 5,6,&7). In *C.*  
176 *eurytheme* Alba females emerge from the pupa with significantly more mature eggs than orange  
177 females<sup>10</sup> and we find evidence that suggests similar dynamics are occurring in *C. crocea*. A  
178 gene set enrichment analysis (GSEA) revealed that ‘embryo development ending in birth or egg  
179 hatching’ (GO:0009792,  $p = 0.00072$ ), ‘proteasome-mediated ubiquitin-dependent protein  
180 catabolic\_process’ (GO:0043161,  $p = 0.00073$ ), and ‘proteolysis’ (GO:0006508,  $p = 0.00101$ )  
181 were within the top 5 terms enriched and upregulated within Alba abdomens (Supplementary  
182 Table 6). Additionally our differential expression analysis identified a that gene which encodes a  
183 triacylglycerol lipase was significantly upregulated within Alba abdomen tissue (log fold change  
184 [log FC] of 4.8) (Fig. 4B and Supplementary Table 5). Triacylglycerol composes more than 90%  
185 of the lipids stored in the fat body and during times of energy demand triacylglycerol lipases  
186 mobilize these stores<sup>23</sup>. For example, during embryogenesis there is a massive shift in lipid  
187 distribution from the fat body to ovaries as lipids comprise 30-40% of the dry weight of insect  
188 embryos<sup>23</sup>. Taken together these results suggest that, similar to *C. eurytheme*, Alba females of  
189 *C. crocea* may be benefitting from increased embryogenesis compared to orange females. We  
190 also observe an enrichment of ‘defense response to Gram-positive bacterium’ (GO:0050830,  
191 0.00027) for genes upregulated within Alba abdomens. Interestingly, previous work has  
192 suggested that Alba females may have enhanced sensitivity to viral infection<sup>9</sup>. Further  
193 investigation of potential morph-specific tradeoffs between wing color and immunity is of  
194 interest.

195  
196 For genes downregulated in Alba abdomens the GSEA revealed significant enrichment of  
197 ‘regulation of nucleoside metabolic process’ (GO:0009118,  $p$  value  $< 0.0001$ ) and ‘regulation of  
198 purine nucleotide catabolic process’ (GO:0033121,  $p$  value  $< 0.0001$ ) (Supplementary Table 6).  
199 *Colias* wings use purine precursors to synthesize pteridines<sup>24</sup>. Downregulation of these GO  
200 terms in Alba abdomens suggests that the decreased pteridine synthesis observed in Alba  
201 females<sup>8</sup>, which likely arises due to the decrease in pigment granules in the wings, leads to a  
202 decrease in purine precursors being shunted from the abdomen to the wings. Additionally,

203 consistent with previous reports of GTP reallocation from wings to other areas of development  
204 in Alba females<sup>8</sup> we also observed significant enrichment of 'positive regulation of GTPase  
205 activity' (GO:0043547, p value < 0.0001). Additionally, *RIM*, a Rab GTPase effector<sup>25</sup>, was one  
206 of the most highly differentially expressed (DE) genes in both tissues (logFC increase in Alba of  
207 3.4 in the abdomen and 5.1 in the wings) (Fig. 4 B&C and Supplementary Table 5). RIM acts as  
208 a molecular switch by converting guanosine diphosphate to guanosine triphosphate (GTP),  
209 thereby activating its associated Rab GTPase, which is in turn involved in synaptic vesicle  
210 exocytosis and secretory pathways<sup>26</sup>.

211  
212 Within wings, *BarH-1* was not differentially expressed between morphs at this stage, indicating  
213 that morph specific expression differences are temporal. However, we did observe genes  
214 downregulated in Alba wings were significantly enriched for 'xanthine dehydrogenase activity' (p  
215 = 0.02, GO:0004854) (Supplementary Table 7). Xanthine dehydrogenase is the enzyme that  
216 catalyzes the xanthopterin to leucopterin conversion during pteridine synthesis in *Colias*  
217 butterflies<sup>8</sup>. These results are consistent with previous studies in *C. eurytheme* that reported the  
218 level of xanthopterin in Alba wings was 7-8 fold less than in orange<sup>8</sup>. Additionally we observed  
219 enrichment of 'MAP kinase activity' (GO:0004709, p = 0.00109) in genes downregulated within  
220 Alba wings. In *Drosophila*, BarH-1 represses Decapentaplegic a morphogen that is homolog to  
221 TGFβ<sup>27</sup>. TGFβ can activate signalling cascades, including MAP kinase pathways<sup>28</sup>. Previous  
222 work in *Drosophila* has also suggested an interaction between Bar homeobox genes and  
223 Ras/MAP kinase signalling during eye development<sup>29</sup>. Future functional studies of candidate  
224 genes are needed to better understand their mechanistic roles in wing development and the  
225 tradeoffs associated with the ALHS.

226  
227 Here we identified the proximate mechanism underlying a female-limited ALHS in a natural  
228 population. Historically, the field of life history research has treated mechanistic details as a  
229 black box<sup>5</sup>, though recently several genetic mechanisms underlying ecologically relevant ALHSs  
230 have been identified, e.g. in the wall lizard<sup>30</sup>, ruff<sup>31,32</sup>, white throated sparrow<sup>33</sup>, and fire ant<sup>34</sup>.  
231 The majority of these studies found that supergenes, large loci that maintain many genes in tight  
232 linkage due to structural variation, gave rise to the alternative morphs<sup>31-35</sup>. Such findings  
233 established that structural variation facilitates the evolution of complex traits. However these  
234 genomic architectures make determining the specific contributions of individual genes to ALHSs  
235 difficult, though there have been significant advances made in the white throated sparrow<sup>33</sup>.  
236 Interestingly, our results, and recent work in the wall lizard<sup>30</sup>, found that ALHSs arose due to

237 changes in the regulatory region of either a single or two genes, respectively. This raises the  
238 question of how these regions give rise to the other fitness-related traits associated with the  
239 ALHS. We propose the Alba-associated physiological and developmental traits arise due to a  
240 classic Y reallocation model, where reduced pigment granule formation results in reduced  
241 pigment synthesis, which in turn leaves more resources free to be used for other developmental  
242 processes. Previous literature studying the Alba phenotype had been unable to determine  
243 whether allocation of resources from the fatbody, or pigment biosynthesis within the wing  
244 scales, was the basis of Alba. Here, the mosaicism of the BarH-1 KO documents the cell level  
245 autonomy of the Alba polymorphism, as abdomen level provisioning would affect all wing scales  
246 equally. Nevertheless, there may be other pleiotropic effects of the reallocation, such as the  
247 sensitivity of the Alba to viral disease, or of the TE insertion itself as it may affect BarH-1  
248 expression in tissues other than the wing.

249  
250 Previous work has shown that *BarH-1* plays a role in the morphogenesis of neurons, leg  
251 segments, and eyes in *Drosophila*<sup>36</sup>. Specifically, *BarH-1* expression is required for the  
252 formation of pigment granules and the deposition of red pteridine pigments in the *Drosophila*  
253 eye<sup>17</sup>. We find that *BarH-1* also plays a role in eye and wing color in *Colias* butterflies. However,  
254 as *BarH-1* expression represses the formation of pigment granules within *Colias* wings, we find  
255 it has a reversed function in *Drosophila* and *Colias*. This may be one of several examples where  
256 either whole or a part of a gene regulatory network that regulates eye development has been  
257 co-opted to give rise to a novel trait in the insect wing<sup>37</sup>. If so, future work could investigate what  
258 aspects of the network have been co-opted and how this lead to BarH-1's contrasting roles in  
259 morphogenesis.

260  
261 Additionally, recent work in the field of butterfly wing evolutionary-development has found that  
262 several genes are repeatedly involved in wing color variation across distantly related species.  
263 Such genes form a patterning "toolkit" (e.g. *optix*<sup>38</sup>, *WntA*<sup>39</sup>, and *cortex*<sup>40</sup>). *BarH-1* might serve  
264 as another toolkit gene for patterning wing color in butterflies beyond *Colias* as we found BarH-1  
265 expression in scale building and socket cells of developing wings in *Vanessa cardui*  
266 (Nymphalidae, Lepidoptera) (Linnaeus, 1758) pupae (Supplementary Fig. 10). However, the  
267 functional role of BarH-1 in *V. cardui* wings remains to be determined. BarH-1 may have a novel  
268 function within *V. cardui* wings. Alternatively, the function of BarH-1 as a repressor of pigment  
269 granule formation could be conserved, as *V. cardui* scales do not have pigment pigment  
270 granules<sup>41</sup>



271  
272 Under the latter assumption, we would expect that BarH-1 is not expressed in closely related  
273 Pierinae species that, despite appearing white, exhibit abundant pigment granules that are  
274 primarily filled with the UV-absorbing pteridine called leucopterin<sup>42</sup>. Future work investigating the  
275 evolutionary history of BarH-1's co-option to the wing and function in other species could shed  
276 light on how complex traits such as ALHSs evolve.

277

## 278 **Author Contributions**

279 AW conducted butterfly rearings and lab work, analysed the data, and wrote the manuscript with  
280 CWW and input from the coauthors. AW, MWP, KT, and CWW conducted the CRISPR/Cas9  
281 knockout experiment. AW and KT conducted the electron microscopy. MWP conducted  
282 antibody staining. RN and JH assisted with bioinformatics. PL and RK conducted HPTLC and  
283 AW and PL analyzed the data. AW, CS, CWW and OB conducted fieldwork. MC conducted lab  
284 work. CWW supervised the work at all stages.

285

## 286 **Acknowledgements**

287 We would like to thank Lovisa Wennerström, Elishia Harji, Jofre Carnicer, and Christina Hansen  
288 Wheat for help with fieldwork. We thank Marianne Ahlbom for assistance with the SEM. Finally  
289 we would like to thank Karin Kiontke, Christen Bossu, Naomi Keehnen, and Peter Pruesscher for  
290 helpful comments on the manuscript. We thank the Department of Zoology at Stockholm  
291 University, the Swedish Research Council 2012–3715, the Academy of Finland 131155, the  
292 Knut and Alice Wallenberg Foundation 2012.0058 and the Erik Philip-Sörensens foundation for  
293 funding.

294

## 295 **References**

- 296 1 Stearns, S. C. *The Evolution of Life Histories*. (Oxford University Press, 1992).  
297 2 Stearns, S. C. Trade-Offs in Life-History Evolution. *Funct Ecol* **3**, 259-268, doi:Doi  
298 10.2307/2389364 (1989).  
299 3 Fisher, R. A. *The Genetical Theory of Natural Selection*. (Oxford University Press,  
300 1930).  
301 4 Gross, M. R. Alternative reproductive strategies and tactics: Diversity within sexes (vol  
302 11, pg 92, 1996). *Trends Ecol Evol* **11**, 263-263 (1996).  
303 5 Flatt, T. & Heyland, A. *Mechanisms of Life History Evolution: The Genetics and*  
304 *Physiology of Life History Traits and Trade-Offs*. (Oxford University Press, 2011).  
305 6 Remington, C. L. The genetics of *Colias* (Lepidoptera). *Adv Genet* **6**, 403-450 (1954).  
306 7 Limeri, L. B. & Morehouse, N. I. The evolutionary history of the "alba" polymorphism in  
307 the butterfly subfamily Coliadinae (Lepidoptera: Pieridae). *Biol J Linn Soc* **117**, 716-724,  
308 doi:10.1111/bij.12697 (2016).

- 309 8 Watt, W. B. Adaptive Significance of Pigment Polymorphisms in Colias Butterflies .3.  
310 Progress in Study of Alba Variant. *Evolution* **27**, 537-548, doi:Doi 10.2307/2407188  
311 (1973).
- 312 9 Descimon, H. & Pennetier, J. L. Nitrogen-Metabolism in Colias-Croceus (Linne) and Its  
313 Alba Mutant (Lepidoptera, Pieridae). *J Insect Physiol* **35**, 881-885, doi:Doi  
314 10.1016/0022-1910(89)90104-2 (1989).
- 315 10 Graham, S. M., Watt, W. B. & Gall, L. F. Metabolic Resource-Allocation Vs Mating  
316 Attractiveness - Adaptive Pressures on the Alba Polymorphism of Colias Butterflies. *P*  
317 *Natl Acad Sci-Biol* **77**, 3615-3619, doi:DOI 10.1073/pnas.77.6.3615 (1980).
- 318 11 Woronik, A., Stefanescu, C., Kakela, R., Wheat, C. W. & Lehmann, P. Physiological  
319 differences between female limited, alternative life history strategies: The Alba  
320 phenotype in the butterfly Colias croceus. *J Insect Physiol* **107**, 257-264,  
321 doi:10.1016/j.jinsphys.2018.03.008 (2018).
- 322 12 Nielsen, M. G. & Watt, W. B. Behavioural fitness component effects of the alba  
323 polymorphism of Colias (Lepidoptera, Pieridae): resource and time budget analysis.  
324 *Funct Ecol* **12**, 149-158, doi:DOI 10.1046/j.1365-2435.1998.00167.x (1998).
- 325 13 Nielsen, M. G. & Watt, W. B. Interference competition and sexual selection promote  
326 polymorphism in Colias (Lepidoptera, Pieridae). *Funct Ecol* **14**, 718-730, doi:DOI  
327 10.1046/j.1365-2435.2000.00472.x (2000).
- 328 14 Hovanitz, W. The biology of Colias butterflies. II. Parallel geographic variation of  
329 dimorphic color phases in North American species. . *Wasmann Journal of Biology* **8**,  
330 197-219 (1950).
- 331 15 Boggs, C. L. & Watt, W. B. Population structure of pierid butterflies IV. Genetic and  
332 physiological investment in offspring by male Colias. *Oecologia* **50**, 320-324,  
333 doi:10.1007/BF00344970 (1981).
- 334 16 Wiklund, C., Karlsson, B. & Leimar, O. Sexual conflict and cooperation in butterfly  
335 reproduction: a comparative study of polyandry and female fitness. *Proc Biol Sci* **268**,  
336 1661-1667, doi:10.1098/rspb.2001.1719 (2001).
- 337 17 Higashijima, S. *et al.* Dual Bar homeo box genes of Drosophila required in two  
338 photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development.  
339 *Genes Dev* **6**, 50-60 (1992).
- 340 18 Nijhout, H. F. *The Development and Evolution of Butterfly Wing Patterns*. (Smithsonian  
341 Institution Press, 1991).
- 342 19 Mackenzie, S. M., Howells, A. J., Cox, G. B. & Ewart, G. D. Sub-cellular localisation of  
343 the white/scarlet ABC transporter to pigment granule membranes within the compound  
344 eye of Drosophila melanogaster. *Genetica* **108**, 239-252 (2000).
- 345 20 Morehouse, N. I., Vukusic, P. & Rutowski, R. Pterin pigment granules are responsible for  
346 both broadband light scattering and wavelength selective absorption in the wing scales  
347 of pierid butterflies. *P R Soc B* **274**, 359-366, doi:10.1098/rspb.2006.3730 (2007).
- 348 21 Rutowski, R. L., Macedonia, J. M., Morehouse, N. & Taylor-Taft, L. Pterin pigments  
349 amplify iridescent ultraviolet signal in males of the orange sulphur butterfly, Colias  
350 eurytheme. *P R Soc B* **272**, 2329-2335, doi:10.1098/rspb.2005.3216 (2005).
- 351 22 van Noordwijk, A. J. & de Jong, G. Acquisition and Allocation of Resources: Their  
352 Influence on Variation in Life History Tactics. *The American Naturalist* **128**, 137-142  
353 (1986).
- 354 23 Arrese, E. L. & Soulages, J. L. Insect Fat Body: Energy, Metabolism, and Regulation.  
355 *Annu Rev Entomol* **55**, 207-225, doi:10.1146/annurev-ento-112408-085356 (2010).
- 356 24 Watt, W. B. Pteridine biosynthesis in the butterfly Colias eurytheme. *J Biol Chem* **242**,  
357 565-572 (1967).
- 358 25 Pavlos, N. J. & Jahn, R. Distinct yet overlapping roles of Rab GTPases on synaptic  
359 vesicles. *Small GTPases* **2**, 77-81, doi:10.4161/sgtp.2.2.15201 (2011).

- 360 26 Stenmark, H. Rab GTPases as coordinators of vesicle traffic. *Nat Rev Mol Cell Biol* **10**,  
361 513-525, doi:10.1038/nrm2728 (2009).
- 362 27 Kang, J., Yeom, E., Lim, J. & Choi, K. W. Bar represses dPax2 and decapentaplegic to  
363 regulate cell fate and morphogenetic cell death in *Drosophila* eye. *PLoS One* **9**, e88171,  
364 doi:10.1371/journal.pone.0088171 (2014).
- 365 28 Derynck, R. & Zhang, Y. E. Smad-dependent and Smad-independent pathways in TGF-  
366 beta family signalling. *Nature* **425**, 577-584, doi:10.1038/nature02006 (2003).
- 367 29 Hayashi, T., Kojima, T. & Saigo, K. Specification of primary pigment cell and outer  
368 photoreceptor fates by BarH1 homeobox gene in the developing *Drosophila* eye. *Dev*  
369 *Biol* **200**, 131-145, doi:10.1006/dbio.1998.8959 (1998).
- 370 30 Andrade, P. *et al.* Regulatory changes in pterin and carotenoid genes underlie balanced  
371 color polymorphisms in the wall lizard. *Proc Natl Acad Sci U S A* **116**, 5633-5642,  
372 doi:10.1073/pnas.1820320116 (2019).
- 373 31 Lamichhaney, S. *et al.* Structural genomic changes underlie alternative reproductive  
374 strategies in the ruff (*Philomachus pugnax*). *Nat Genet* **48**, 84-88, doi:10.1038/ng.3430  
375 (2016).
- 376 32 Kupper, C. *et al.* A supergene determines highly divergent male reproductive morphs in  
377 the ruff. *Nat Genet* **48**, 79-83, doi:10.1038/ng.3443 (2016).
- 378 33 Horton, B. M. *et al.* Estrogen receptor alpha polymorphism in a species with alternative  
379 behavioral phenotypes. *Proc Natl Acad Sci U S A* **111**, 1443-1448,  
380 doi:10.1073/pnas.1317165111 (2014).
- 381 34 Wang, J. *et al.* A Y-like social chromosome causes alternative colony organization in fire  
382 ants. *Nature* **493**, 664-668, doi:10.1038/nature11832 (2013).
- 383 35 Schwander, T., Libbrecht, R. & Keller, L. Supergenes and complex phenotypes. *Curr*  
384 *Biol* **24**, R288-294, doi:10.1016/j.cub.2014.01.056 (2014).
- 385 36 Reig, G., Cabrejos, M. E. & Concha, M. L. Functions of BarH transcription factors during  
386 embryonic development. *Dev Biol* **302**, 367-375, doi:10.1016/j.ydbio.2006.10.008  
387 (2007).
- 388 37 Monteiro, A. Gene regulatory networks reused to build novel traits: co-option of an eye-  
389 related gene regulatory network in eye-like organs and red wing patches on insect wings  
390 is suggested by optix expression. *Bioessays* **34**, 181-186, doi:10.1002/bies.201100160  
391 (2012).
- 392 38 Zhang, L., Mazo-Vargas, A. & Reed, R. D. Single master regulatory gene coordinates  
393 the evolution and development of butterfly color and iridescence. *Proc Natl Acad Sci U S*  
394 *A* **114**, 10707-10712, doi:10.1073/pnas.1709058114 (2017).
- 395 39 Mazo-Vargas, A. *et al.* Macroevolutionary shifts of WntA function potentiate butterfly  
396 wing-pattern diversity. *Proc Natl Acad Sci USA* **114**, 10701-10706,  
397 doi:10.1073/pnas.1708149114 (2017).
- 398 40 Nadeau, N. J. *et al.* The gene cortex controls mimicry and crypsis in butterflies and  
399 moths. *Nature* **534**, 106+, doi:10.1038/nature17961 (2016).
- 400 41 Dinwiddie, A. *et al.* Dynamics of F-actin prefigure the structure of butterfly wing scales.  
401 *Developmental Biology* **392**, 404-418, doi:10.1016/j.ydbio.2014.06.005 (2014).
- 402 42 Makino, K., Satoh, K., Koike, M. & Ueno, N. Sex in *Pieris rapae* L. and the pteridin  
403 content of their wings. *Nature* **170**, 933-934 (1952).
- 404 43 Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively  
405 parallel sequence data. *Proc Natl Acad Sci U S A* **108**, 1513-1518,  
406 doi:10.1073/pnas.1017351108 (2011).
- 407 44 Chin, C. S. *et al.* Phased diploid genome assembly with single-molecule real-time  
408 sequencing. *Nat Methods* **13**, 1050-1054, doi:10.1038/nmeth.4035 (2016).
- 409 45 Wences, A. H. & Schatz, M. C. Metassembler: merging and optimizing de novo genome  
410 assemblies. *Genome Biol* **16**, 207, doi:10.1186/s13059-015-0764-4 (2015).

- 411 46 Woronik, A. & Wheat, C. W. Advances in finding Alba: the locus affecting life history and  
412 color polymorphism in a *Colias* butterfly. *J Evol Biol* **30**, 26-39, doi:10.1111/jeb.12967  
413 (2017).
- 414 47 Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate  
415 read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791,  
416 doi:10.1093/bioinformatics/btt468 (2013).
- 417 48 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,  
418 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 419 49 Kofler, R., Pandey, R. V. & Schlotterer, C. PoPoolation2: identifying differentiation  
420 between populations using sequencing of pooled DNA samples (Pool-Seq).  
421 *Bioinformatics* **27**, 3435-3436, doi:10.1093/bioinformatics/btr589 (2011).
- 422 50 R: A language and environment for statistical computing. (R Foundation for Statistical  
423 Computing, Vienna, Austria, 2019).
- 424 51 Kofler, R. *et al.* PoPoolation: a toolbox for population genetic analysis of next generation  
425 sequencing data from pooled individuals. *PLoS One* **6**, e15925,  
426 doi:10.1371/journal.pone.0015925 (2011).
- 427 52 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158,  
428 doi:10.1093/bioinformatics/btr330 (2011).
- 429 53 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based  
430 linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).
- 431 54 Perry, M. *et al.* Molecular logic behind the three-way stochastic choices that expand  
432 butterfly colour vision. *Nature* **535**, 280-284, doi:10.1038/nature18616 (2016).
- 433 55 Zaharia, M. *et al.* Faster and More Accurate Sequence Alignment with SNAP. *arxiv*,  
434 doi:arXiv:1111.5572 (2011).
- 435 56 Folch, J., Lees, M. & Sloane Stanley, G. H. A simple method for the isolation and  
436 purification of total lipides from animal tissues. *J Biol Chem* **226**, 497-509 (1957).
- 437 57 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a  
438 reference genome. *Nat Biotechnol* **29**, 644-652, doi:10.1038/nbt.1883 (2011).
- 439 58 Gilbert, D. in *7th annual arthropod genomics symposium* (Notre Dame, 2013).  
440 59 csvkit (2016).
- 441 60 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for  
442 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-  
443 140, doi:10.1093/bioinformatics/btp616 (2010).
- 444 61 Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology  
445 Assignment by eggNOG-Mapper. *Mol Biol Evol* **34**, 2115-2122,  
446 doi:10.1093/molbev/msx148 (2017).
- 447 62 topGO: Enrichment Analysis for Gene Ontology v. R package version 2.34.0 (2018).  
448

## 449 **Methods**

450 For detailed methods, including all bioinformatic commands, please see the supplementary  
451 information.

452 **Data availability:** SRA reference numbers for the genome and sequencing data will be included  
453 upon acceptance.

454 **Genome assembly:** An orange female and male carrying Alba (offspring from wild caught  
455 butterflies, Catalonia, Spain) were mated in the lab. DNA from an Alba female offspring of this  
456 cross was extracted. Quality and quantity were assessed using a Nanodrop 8000  
457 spectrophotometer (Thermo Scientific) and a Qubit 2.0 fluorometer (dsDNA BR, Invitrogen). A

458 180 insert size paired end library (101bp reads) was prepared (TruSeq PCR free) and  
459 sequenced on an Illumina HiSeq 4000 at the Beijing Genomics Institute (Shenzhen, China). A  
460 Nextera mate-pair library with a 3 kb insert size was prepared and sequenced on an Illumina  
461 HiSeq 2500 (125bp reads) at the Science for Life Laboratory (Stockholm, Sweden). Raw data  
462 was cleaned and high quality reads were used as input for the AllPaths-LG (v. 50960)<sup>43</sup>  
463 assembly pipeline. High molecular weight DNA was extracted from two more Alba females from  
464 the above mentioned cross (i.e full siblings). Equal amounts of DNA from each individual were  
465 pooled sent to the Science for Life Laboratory (Stockholm, Sweden) for PacBio sequencing on  
466 24 SMRT cells (~17GB of data was produced). A Falcon (v. 0.4.2)<sup>44</sup> assembly was generated by  
467 the Science for Life Laboratory. We then used Metassembler (v. 1.5)<sup>45</sup> to merge our AllPathsLG  
468 and Falcon assemblies, using the AllPathsLG assembly as the primary assembly.

469 **Bulk segregant analyses (BSA):** The female informative cross data and mapping protocol  
470 described in Woronik and Wheat, 2017<sup>46</sup> was applied to the high quality reference genome to  
471 identify the contigs that made up the Alba chromosome. Male Informative Cross (MIC) I: DNA  
472 was extracted from a wild caught orange mother (Catalonia Spain) and 26 of her Alba and 24 of  
473 her orange female offspring. DNA quality and quantity of each individual was assessed via a  
474 Nanodrop 8000 spectrophotometer (Thermo Scientific, MA, USA) and a Qubit 2.0 Fluorometer  
475 (dsDNA BR; Invitrogen, Carlsbad, CA, USA) before pooling equal amounts of high-quality DNA  
476 from Alba and orange offspring into two pools, respectively. The library preparation (TruSeq  
477 PCR-free) and Illumina sequencing (101 bp PE HiSeq2500), was performed at the Beijing  
478 Genomics Institute (Shenzhen, China). Raw reads were cleaned and then mapped to the  
479 reference genome using NextGenMap v0.4.10 (-i 0.09)<sup>47</sup>. SAMTOOLS v1.2<sup>48</sup> was used to filter  
480 (view -f 3 -q 20), sort and index the bam files and generate mpileup files for the two pools and  
481 the orange mother. Popoolation2<sup>49</sup> were used to calculate the allele frequency difference  
482 between Alba and orange pools. SNP sites were filtered in R<sup>50</sup>, for a read depth  $\geq 30$  and  $\leq 300$ ,  
483 a bi-allelic state, and a minimum minor allele frequency of 3. The orange mother mpileup was  
484 similarly analyzed using Popoolation<sup>51</sup> (read depth  $\geq 5$  and  $\leq 30$ ); but the major and minor allele  
485 frequencies were calculated in R<sup>50</sup> by dividing the major and minor allele count by the read  
486 depth at each site respectively. A SNP site was considered a MIC I Alba SNP when it met the  
487 following expectations: 1) homozygous in the orange mother, 2) homozygous in the orange  
488 pool, 3) the allele frequency difference in the Alba pool compared to the orange was 0.45-0.55.  
489 MIC II: A male carrying Alba mated an orange female in the lab at Stockholm University. DNA  
490 was prepared as described above for 26 Alba and 28 orange female offspring resulting in two  
491 DNA pools. Library preparation (TruSeq PCR-free) and Illumina sequencing (150 bp paired-end  
492 reads with 350bp insert, HiSeqX), was performed at Science for Life Laboratory (Stockholm,  
493 Sweden). The same mapping and SNP calling pipeline used on the MIC I was applied. A site  
494 was considered an Alba SNP if 1) it was homozygous in the orange pool and 2) the allele  
495 frequency difference in the Alba pool compared to the orange was 0.45-0.55. A contig was  
496 considered Alba associated if it had  $\geq 3$  Alba SNPs in all crosses. Nineteen Alba associated  
497 contig were identified. They total ~3.7Mbp and are considered the Alba BSA locus.

498 **Genome wide association study:** DNA for genome re-sequencing was extracted from 15 Alba  
499 and 15 orange females from diverse population backgrounds (Catalonia, Spain and Capri, Italy).  
500 High quality DNA was prepared using Illumina TruSeq and sequenced at the Science for Life  
501 Laboratory (Stockholm, Sweden) (150 bp paired-end reads HiSeqX). Cleaned reads were  
502 mapped to the annotated reference genome using NextGenMap v0.4.10 (-i 0.6 -X 2000)<sup>47</sup>. Bam  
503 files were filtered and sorted using SAMTOOLS v1.2 (view -f 3 -q 20)<sup>48</sup>. A VCF file was  
504 generated using SAMTOOLS v1.2 (-t DP -t SP -Q 15)<sup>48</sup> and bcftools v.1.2 (-Ov -m)<sup>48</sup>. VCFtools  
505 (v0.1.13)<sup>52</sup> was used to call SNP sites with no more than 50% missing data, an average read  
506 depth between 15-50 across individuals, and a minimum SNP quality of 30. An association

507 analysis was performed with PLINK (v1.07)<sup>53</sup> and a Benjamini & Hochberg step-up FDR control  
508 was applied. SNPs with FDR <0.05 were considered Alba SNPs. We conducted this analysis  
509 both genome wide and only within the BSA locus. Both analyses fine mapped the Alba locus to  
510 the same genomic region.

511 **Antibody Generation and Staining:** A Rabbit-anti-Bar antibody was generated against the full  
512 length sequence of the *Vanessa cardui* Bar homolog. Protein was generated by GenScript  
513 (Piscataway, NJ) and purified to >80% purity. DNA sequences to produce this protein were  
514 codon-optimized for bacterial expression and made via gene synthesis. GenScript injected  
515 resultant protein into host animals, collected serum for testing, and affinity purified the product  
516 using additional target protein bound to a column. Antibody staining was performed as  
517 described previously for *Drosophila* and butterfly tissues<sup>54</sup>. Staged pupal wings and retinas were  
518 dissected and fixed 48 hours post-pupation. The Rabbit-anti-Bar antibody was used at 1:100,  
519 followed by secondary antibody staining with AlexFluor-555-anti-Rabbit secondaries at 1:500  
520 and counterstaining with DAPI. Images were captured using standard confocal microscopy on a  
521 Leica SP5.

522 **CRISPR/Cas9 knockouts:** The guide-RNA (gRNA) sequences were generated using the  
523 protocol described in Perry et al. 2016. Viable Cas9 target-sites were located by manually  
524 looking for PAM-sites (NGG) in the exon region of *BarH-1*. Uniqueness of the target regions was  
525 confirmed using a NCBI nucleotide blast (ver. 2.5.0+ using blastn-short flag and filtering for an  
526 e-value of 0.01) against the *C. crocea* reference genome. gRNA constructs were ordered from  
527 Integrative DNA Technologies (Coralville, Iowa, USA) as DNA (gBlocks). Full gRNA constructs  
528 had the following configuration: an M13F region, a spacer sequence, a T7-promotor sequence,  
529 the Target specific sequence, a Cas9 binding sequence, and finally a P505 sequence. Upon  
530 delivery, gBlocks were amplified using PCR to generate single-stranded guide RNA (sgRNA).  
531 For each gBlock, four 50ul reactions were conducted using the M13f and P505 primers and  
532 Platinum Taq (Invitrogen cat. 10966-034). The four reactions were then combined and purified  
533 in a Qiagen Minelute spin column (cat. 28004, Venlo, Netherlands). The resulting template was  
534 transcribed using the Lucigen AmpliScribe T7-flash Transcription Kit from Epicentre/Illumina  
535 (cat. ASF3507, Madison, WI, USA) followed by purification via ammonium acetate precipitation.  
536 Products were resuspended with Qiagen buffer EB, concentrations were quantified by Qubit and  
537 further diluted to 1000 ng/μl. They were then mixed with Cas9-NLS protein (PNA Bio, Newbury  
538 Park, CA, USA) and diluted to a final concentration of 125-250 ng/μl. *C. crocea* females (n > 40)  
539 from Aiguamolls de l'Empordà, Spain were captured and kept in morph-specific flight cages in  
540 the lab at Stockholm University where they oviposited on alfalfa (*Medicago sativa*). Eggs were  
541 collected between 1-7h post-laying and sterilized in 7% benzalkonium chloride for ~5 minutes  
542 before injection. Injections were either at a concentration of 125 or 250 ng/ul and conducted  
543 using a M-152 Narishige micromanipulator (Narishige International Limited, London, UK) with a  
544 50 ml glass needle syringe, with injection pressure applied by hand via a syringe fitting.

545 **CRISPR/Cas9 validation:** To validate the mutation, Cas9 cut sites were PCR-amplified and a  
546 ~370bp region, centered on the intended cut site were sequenced using Illumina MiSeq 300bp  
547 paired-end sequencing. Primers were designed using Primer3  
548 ([http://biotools.umassmed.edu/bioapps/primer3\\_www.cgi](http://biotools.umassmed.edu/bioapps/primer3_www.cgi)). DNA was isolated from KO-  
549 individuals using KingFisher Cell and Tissue DNA Kit from ThermoFisher Scientific (N11997)  
550 and the robotic Kingfisher Duo Prime purification system. DNA quality and quantity were  
551 assessed via a Nanodrop 8000 spectrophotometer (Thermo Scientific, MA, USA) and a Qubit  
552 2.0 Fluorometer (dsDNA BR; Invitrogen, Carlsbad, CA, USA). Aliquots were then taken and  
553 diluted to 1ng/ul before amplifying the region over the cleavage-site. Sequences were amplified  
554 and ligated with Illumina adapter and indexes in a two-step process following the protocol

555 provided by Science for Life Laboratories (Stockholm, Sweden) and Illumina. First, we amplified  
556 the ~370bp long sequence around the cut sites and attach the first Illumina adapter, onto which  
557 we later attach Illumina handles and Index using a second round of PCR (Accustart II PCR  
558 Supermix from Quanta Bio [Beverly, MA, USA], settings 94C x 2 min followed by 40 cycles of 94  
559 C x 30 sec + 60 C x 15 sec + 68 C x 1 min followed by 68 C x 5 min). PCR products were  
560 purified using Qiagen Qiaquick (Cat. 28104). Concentration and quality of the product were  
561 assessed via Nanodrop and gel electrophoresis. DNA was diluted to ~0.5ng/ul and then the  
562 unique double indices were attached by the second round of PCR (same protocol as above).  
563 The final PCR products were purified again using Qiaquick spin columns and concentration and  
564 size was assessed using Qubit fluorometer and gel electrophoresis. All samples were then  
565 mixed at equal molarity and sent for sequencing at Science for Life Laboratories (Stockholm,  
566 Sweden). Sequences were aligned to their respective fragments (area surrounding cut site)  
567 using SNAP (ver. 1.0beta18)<sup>55</sup>, identical reads were clustered using the collapse utility in Fastx-  
568 Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Sequences containing deletions were extracted  
569 and the most abundant sequences containing deletions were selected for confirmation of  
570 deletion in the expected region.

571 **Electron Microscopy:** To quantify pigment granule differences between Alba and orange  
572 individuals pieces of the forewing were mounted on aluminum pin stubs (6mm length) with the  
573 dorsal side upwards. Samples were coated in gold for 80 seconds using an Agar sputter coater  
574 and imaged under 5 kV acceleration voltage, high vacuum, and ETD detection using a scanning  
575 electron microscope (Quanta Feg 650, FEI, Hillsboro, Oregon, USA). To quantify pigment  
576 granules within the photos we selected images from the same magnification and randomly  
577 placed three 4  $\mu\text{m}^2$  squares on the images. We counted the number of pigment granules within  
578 each square and took the average, then conducted a two sample t-test in R. To quantify  
579 pigment granule differences between KO and wild type regions in our CRISPR KO mosaic  
580 individuals, a biopsy hole punch a 2mm in diameter circle was used to cut out one piece mostly  
581 containing white scales and one piece with mostly orange scales. These pieces were first  
582 photographed using a Leica EZ4HD stereo microscope in order to allow us to confirm the color  
583 of each scale once they were covered with gold sputter. Five white and five orange scales were  
584 then selected and the granules from a 4 $\mu\text{m}^2$  square were counted from each of those scales  
585 and a two sample t-test was then conducted in R.

586 **Lipid Analysis:** Wild caught *C. crocea* Alba females (Catalonia, Spain) oviposited in the lab.  
587 Eggs were moved into individual rearing cups and split between two temperature treatments  
588 (hot: 27°C and 16 hour day length during larval and pupal development, cold: reared at 22°C  
589 with a 16 hour day length during larval development and 15°C with a 16 hour day length during  
590 pupal development). Once pupated, individuals were checked a minimum of every 12 hours.  
591 Upon eclosion adults were stored at 4°C until the next day to provide time for meconium  
592 excretion. Butterflies were not allowed to feed before dissection. Body weight was taken using a  
593 Sauter RE1614 scale before dissection. Total lipids were extracted using the Folch method<sup>56</sup>  
594 according to the procedures outlined in Woronik et. al. 2018<sup>11</sup>. HPTLC was conducted as  
595 described in Woronik et. al. 2018<sup>11</sup>. In brief, 5  $\mu\text{l}$  of the sample lipid extract was applied on a  
596 silica plate with a Camag Automatic TLC Sampler 4 (Camag, Muttenz, Switzerland). After the  
597 silica plate developed it was scanned with a Camag TLC plate scanner 3 at 254 nm using a  
598 deuterium lamp with a slit dimension of 6 × 0.45 mm and analyzed with the Win-CATS 1.1.3.0  
599 software. Peaks representing the four major neutral lipid classes (diacylglycerols,  
600 triacylglycerols, cholesterol and cholesterol esters) were identified by comparing their retention  
601 times against known standards. Then the peak areas were integrated and the amount of lipid  
602 within each class was calculated using the formula:  $\text{pmol}_{\text{sample}} = (\text{Area}_{\text{sample}} / \text{Area}_{\text{standard}}) \times$   
603  $\text{pmol}_{\text{standard}}$ . The total lipid content (nmol per abdomen) was calculated as a sum of pmol

604 contents of all neutral lipid classes. For the statistical analyses this value was regressed against  
605 abdomen weight and standardized residuals (i.e. mass-corrected storage lipid amount) and  
606 were subsequently used as dependent variable.

607 **Transcriptome assembly and differential expression analysis:** Offspring from a wild caught  
608 Alba female from Catalonia, Spain were reared at Stockholm University. When larvae reached  
609 the fifth instar they were checked at least every six hours and the pupation time of each  
610 individual was recorded. Tissue was collected between 82% and 92% of pupal development.  
611 Pupae were dissected in PBS solution, and the abdomen and wings were flash frozen in liquid  
612 nitrogen and stored at -80 °C. RNA was extracted from the abdomen and wing tissues using  
613 Trizol. RNA quality and quantity was assessed using a Nanodrop 8000 spectrophotometer  
614 (Thermo Scientific) and an Experion electrophoresis machine using the manufacturer protocol  
615 (Bio-Rad, Hercules, CA). Library preparation (Strand-specific TruSeq RNA libraries using poly-A  
616 selection) and sequencing (101 bp PE HiSeq2500 - high output mode) was performed at the  
617 Science for Life Laboratories (Stockholm, Sweden). In total 16 libraries were sequenced (4  
618 orange and 4 Alba individuals - wings and abdomen from each individual). Raw data was  
619 cleaned and reads from all libraries were used in a de novo transcriptome assembly (Trinity  
620 version trinityrnaseq\_r2013\_08\_14 with default parameters)<sup>57</sup>. To reduce the redundancy  
621 among contigs and produce a biologically valid transcript set, the tr2aacds pipeline from the  
622 EvidentialGene software package<sup>58</sup> was run on the raw Trinity assembly. The sixteen RNA-Seq  
623 libraries were mapped to the resulting transcriptome using NextGenMap v0.4.10 (-i 0.09)<sup>47</sup>.  
624 SAMTOOLS v1.2<sup>48</sup> was then used to filter (view -f 3 -q 20), sort and index the sixteen bam files.  
625 SAMTOOLS v1.2<sup>48</sup> idxstats was then used to calculate the read counts per gene for each of the  
626 sorted bam files. These counts were then joined in a CSV file using an in-house pipeline and  
627 csvjoin<sup>59</sup>. A differential expression analysis was conducted in EdgeR<sup>60</sup>. A Benjamini Hochberg  
628 correction was applied to the raw p values to correct for false discovery rate and differentially  
629 expressed genes were called (adjusted p value <0.05). eggNOG-mapper (v.1)<sup>61</sup> was used with  
630 default settings to functionally annotate the transcriptome. The R package topGo<sup>62</sup> was used to  
631 conduct a gene set enrichment analysis on genes that exhibited > 1 or < -1 log fold change in  
632 the differential expression analysis.



**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 3 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.

**Fig. 2. *BarH-1* is expressed in white but not orange regions of the wing in *C. crocea* females.** DAPI (nuclei, left, blue) and *BarH-1* antibody (right, red) staining of pupal wings. Large nuclei are in scale building cells, small nuclei are in epithelial cells. The right part of the panel shows the approximate location of the stained area and the scales in this region in an adult wing. Scale bars are 2 $\mu$ m. (A) Staining of the forewing of an Alba female in an area at the black margin (top) and a white area (bottom). *BarH-1* is expressed in melanic as well as white Alba scale building cells. (B) Antibody staining of the forewing of an orange female. *BarH-1* is not expressed in these scale building cells. (C) Antibody staining of the hindwing of an orange female. *BarH-1* is heterogeneously expressed in the scale building cells within this region. This staining pattern presumably corresponds to the variation in scale color, with melanic scale building cells expressing *BarH-1* but orange lacking expression.

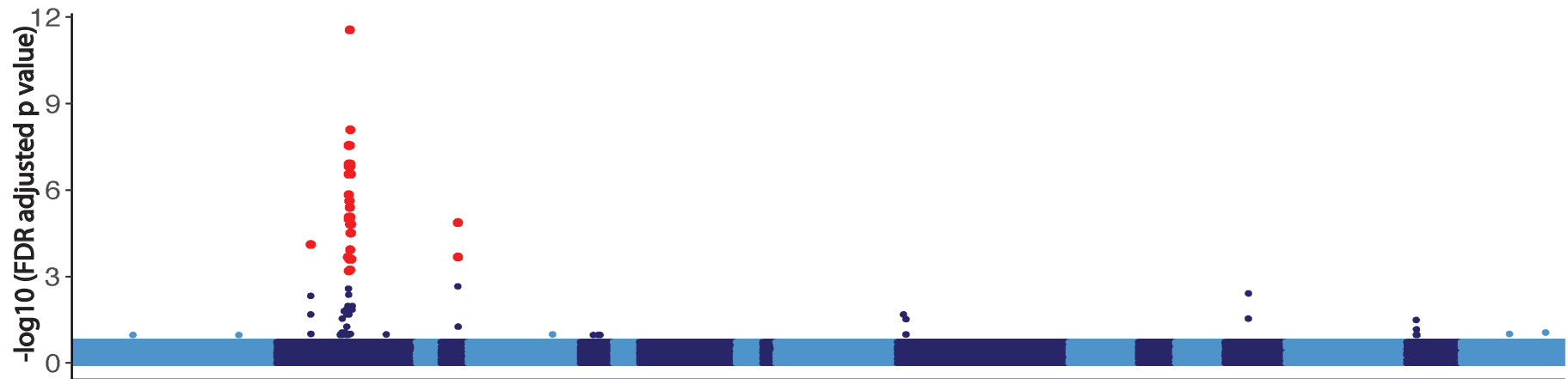
**Fig. 3. *Colias* forewings and scanning electron microscope (SEM) images of their wing scale nanostructures.** (A) *C. crocea* Alba female wing and wing scale structure. The top panel shows the SEM image of a black scale; pigment granules are absent. The bottom panel shows a white scale, exhibiting 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 genetically Alba female exhibiting CRISPR/Cas9 mosaic knockout of *BarH-1*. The top panel shows the 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. (D) Wing and wing scales 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 lack 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*.

**Fig. 4. Physiological differences between female morphs of *C. crocea*.** A) The mass corrected total neutral lipid content for female morphs in two temperature treatments. Alba females, on average, have larger neutral lipid stores than orange females. However there is an interaction between morph and temperature as the difference is only significant in the cold treatment. Error bars are the standard error (cold: n=32,  $t_{29,12} = 3.42$ ,  $P = 0.002$ , hot: n=25,  $t_{22,71} = 0.67$ ,  $P = 0.51$ ). 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 plots to visualize gene expression differences between female morphs in pupal wing tissue. Color coding, shapes, and axes are the same as above.

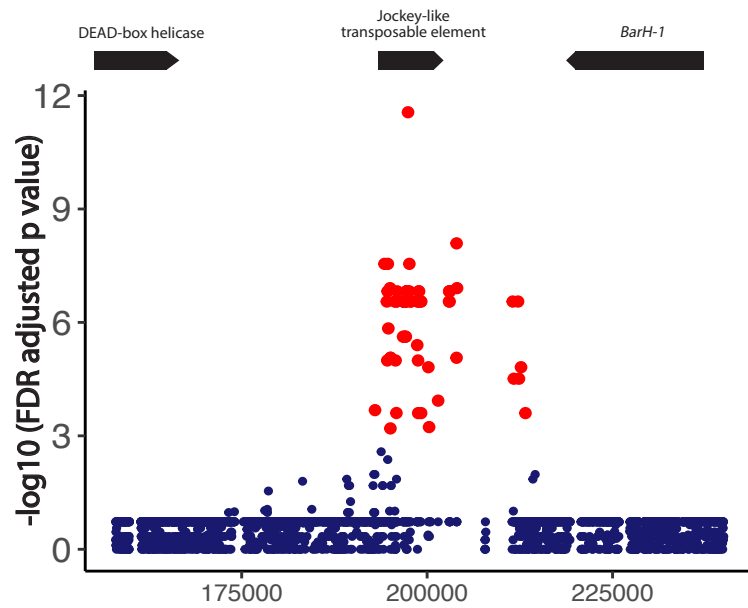
A



B



C



D



E



Figure 1

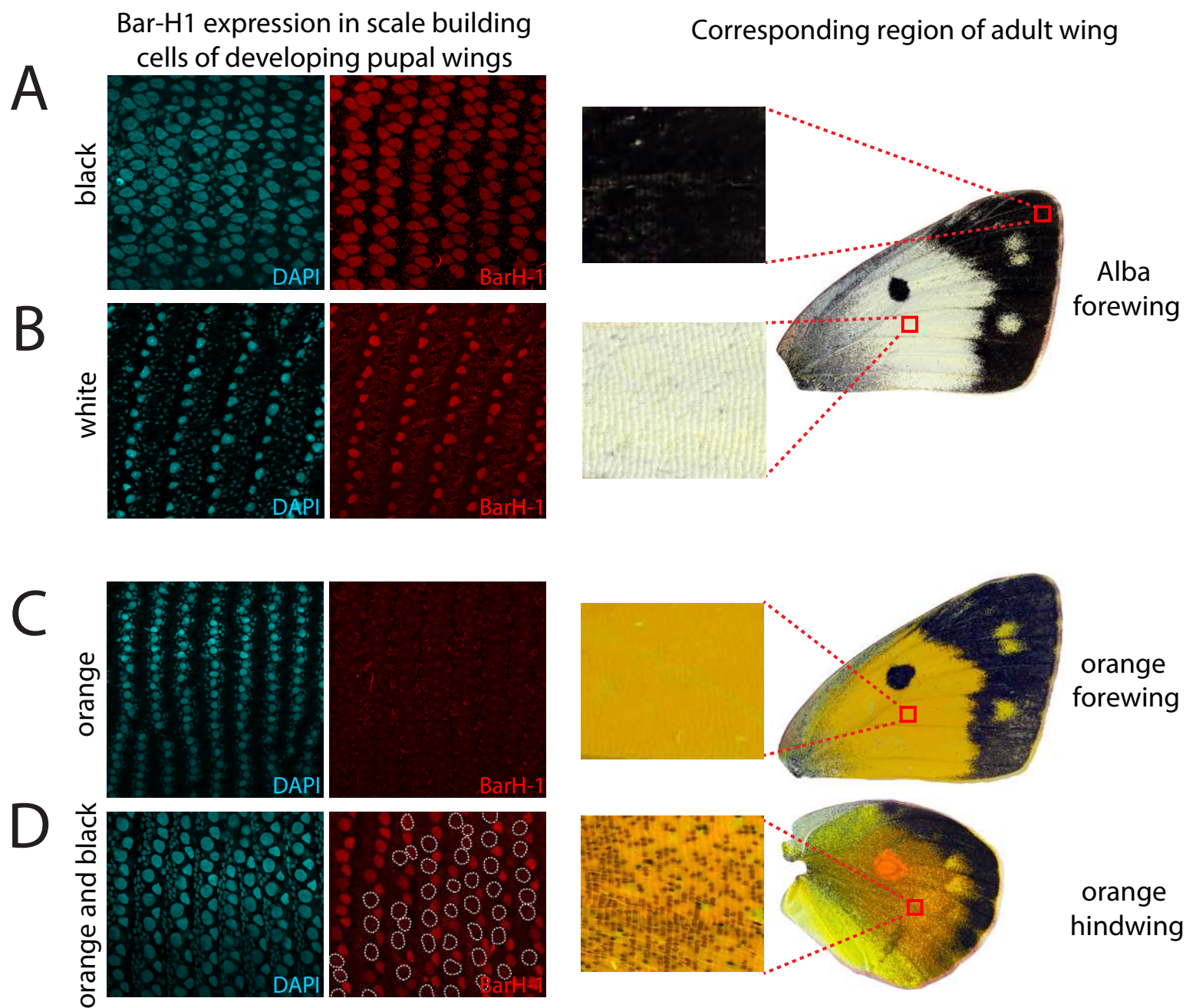


Figure 2

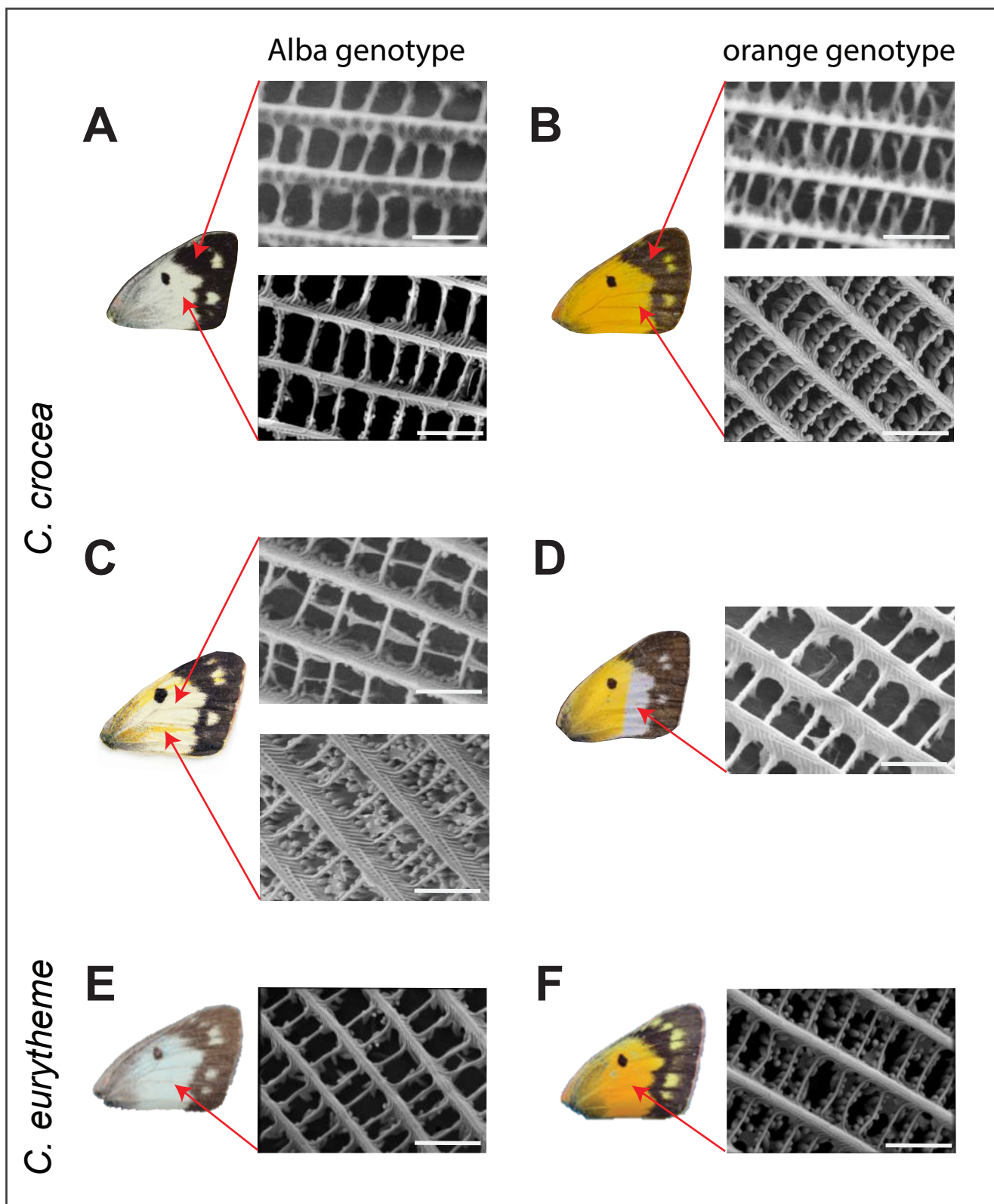


Figure 3

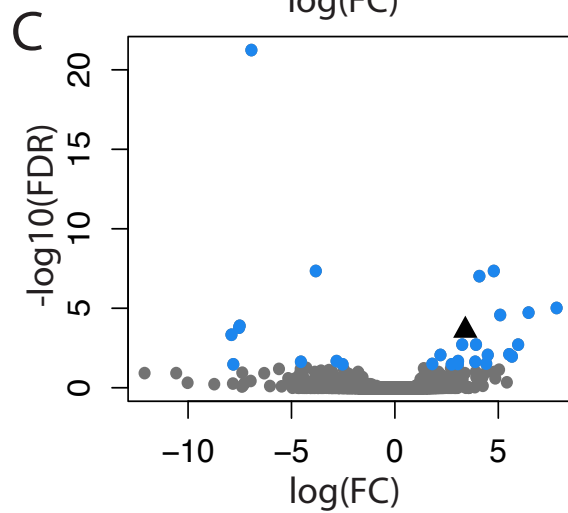
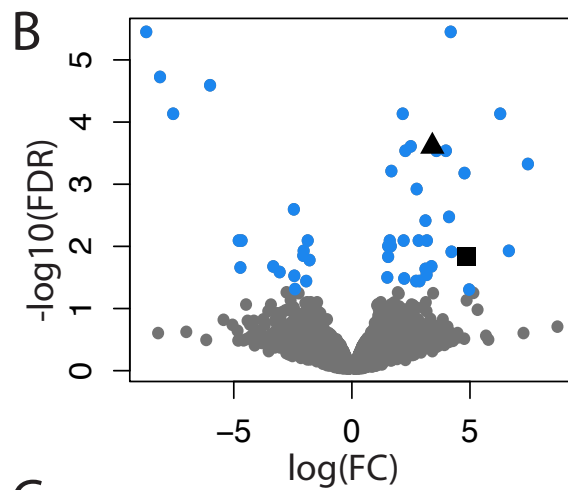
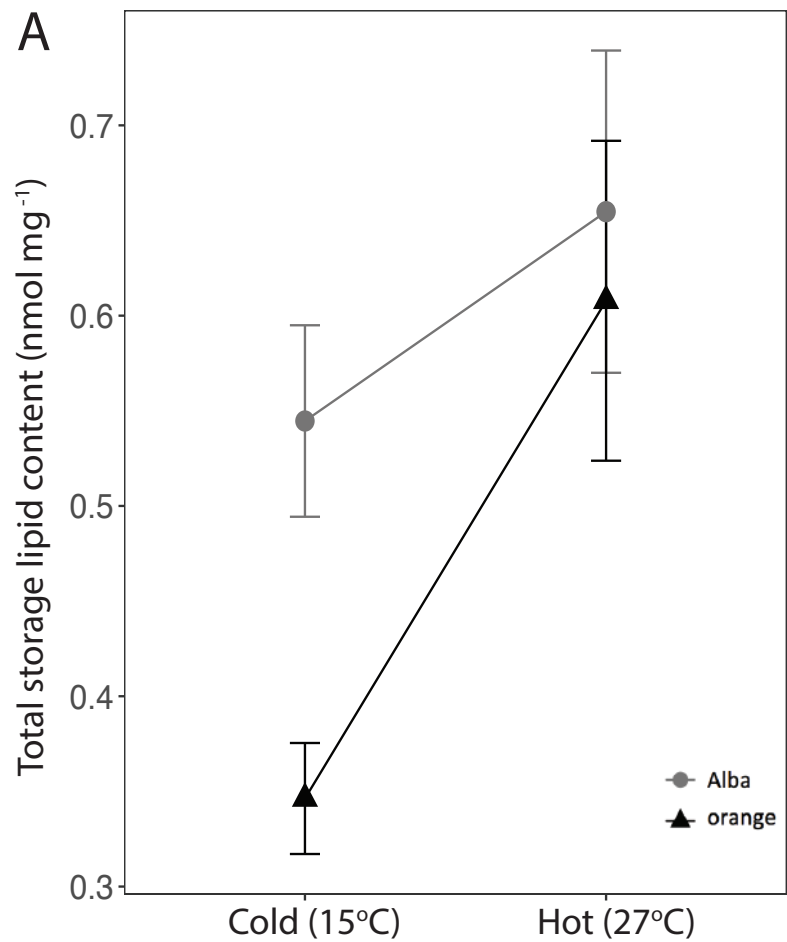


Figure 4