Visual mate preference evolution during butterfly speciation is linked to neural processing genes

Matteo Rossi[1,2], Alexander E. Hausmann[1], Timothy J. Thurman[2,3], Stephen H. Montgomery[4], Riccardo Papa[2,5,6], Chris D. Jiggins[2,7], W. Owen McMillan[2] & Richard M. Merrill[1,2*]

1. Division of Evolutionary Biology, LMU, Munich, Germany; 2. Smithsonian Tropical Research Institute, Panama City, Panama; 3. Division of Biological Sciences, University of Montana, USA; 4. School of Biological Sciences, University of Bristol, Bristol, UK; 5. Department of Biology, University of Puerto Rico, San Juan, Puerto Rico; 6. Molecular Sciences and Research Center, University of Puerto Rico, San Juan, Puerto Rico; 7. Department of Zoology, University of Cambridge, Cambridge UK.

[*]corresponding authors**:** rossi@bio.lmu.de & merrill@bio.lmu.de

Many animal species remain separate not because they fail to produce viable hybrids, but because their individuals "choose" not to mate. However, we still know very little of the genetic mechanisms underlying changes in these mate preference behaviours. *Heliconius* butterflies display bright warning patterns, which they also use to recognize conspecifics. Here, we couple QTL for divergence in visual preference behaviours with population genomic and gene expression analyses of neural tissue (central brain, optic lobes and ommatidia) across development in two sympatric *Heliconius* species. Within a region containing 200 genes, we identify five genes that are strongly associated with divergent visual preferences. Three of these have previously been implicated in key components of neural signalling (specifically an *ionotropic glutamate receptor* and two *regucalcins*), and overall our candidates suggest shifts in behaviour involve changes in visual integration or processing. This would allow preference evolution without altering perception of the wider environment.

The evolution and maintenance of new animal species often relies on the emergence of divergent mating preferences[1,2]. Changes in sensory perception or other neural systems must underlie differences in innate behaviours between species, and will ultimately have a genetic

31 basis. Although the significance of behavioural barriers for speciation has been recognized
32 since the Modern Synthesis[3], we know little of the genes underlying changes in mating
33 preferences, or variation in behaviours across natural populations more broadly[4,5]. Identifying
34 these genes will provide an important route towards understanding how behavioural
35 differences are generated, both during development and across evolutionary time.

36 Previous studies of isolating preference behaviours have largely been limited to the
37 identification of causal genomic regions, which almost invariably contain many genes[6,7,8].
38 Only a handful of studies have identified likely candidate genes that contribute to species
39 behavioural preferences. These are largely limited to chemosensory-guided mating
40 preferences[9,10,11] but see[12], and have identified changes at chemoreceptor genes. To our
41 knowledge, only two studies – in incipient fish species – have identified candidates for visual
42 preference evolution, albeit indirectly, both suggesting a role for sensory perception mediated
43 by changes in the peripheral visual system (e.g. opsins)[13,14]. Whether or not visual preference
44 evolution generally involves shifts at the sensory periphery, or in downstream processing,
45 remains unknown.

46 The closely related species *Heliconius melpomene* and *H. cydno* differ in warning
47 patterns, which are both under disruptive selection due to mimicry[15] and are important mating
48 cues[16]. As a result, these divergent patterns couple ecological and behavioural components of
49 reproductive isolation, which (as predicted by so-called "magic trait" models[17]) is expected to
50 facilitate speciation in the face of gene flow. In central Panama, *H. melpomene* shares the
51 black, red and yellow pattern of its local *Heliconius erato* co-mimic. In contrast, *H. cydno*
52 mimics the black and white patterns of *H. sapho*. *H. melpomene* and *H. cydno* remain separate
53 largely due to strong assortative mating[18]. Visual preferences for divergent patterns are
54 particularly apparent in males, which strongly prefer to court conspecific females[16,19,20].
55 Differences in warning pattern between *H. melpomene* and *H. cydno* are largely due to
56 expression differences in just three genes, specifically *optix*[21], *WntA*[22] and *cortex*[23].

57 Quantitative trait locus (QTL) mapping of *H. melpomene* and *H. cydno* has revealed
58 three genomic regions of major effect that influence the relative time males spend courting
59 red *H. melpomene* or white *H. cydno* females[20]. Notably, the best supported QTL was in the
60 same genomic region as *optix*, the gene responsible for presence and absence of red colour
61 pattern elements in *Heliconius*[21]. Genetic linkage will facilitate speciation by impeding the
62 breakdown of genetic associations between ecological and mating traits[23]. Nevertheless, this
63 QTL, and its associated candidate region, contain hundreds of genes, and the exact genes
64 responsible for differences in preference behaviour are not known.
65 Here, we first confirm that behavioural QTLs identified previously are associated with
66 variation in male courtship initiation. We then identify genes within the major QTL, which
67 were differentially expressed in the neural tissue (central brain, optic lobes and ommatidia) of
68 *H. melpomene* and *H. cydno*, or have protein coding changes predicted to alter protein
69 function. Out of 200 genes within the QTL region, we identify just five candidates likely to
70 underlie assortative mating behaviours.
71
72 **Results**
73 **Chromosome 18 is associated with differences in courtship initiation.**
74 Our previous results reveal that QTLs on chromosomes 1, 17 and 18 influence the relative
75 time hybrid males spend courting red *H. melpomene* or white *H. cydno* females[19]. However,
76 the time males spend courting a particular female might depend not only on male attraction,
77 but on the female's response (and in turn his response to her behaviour). To confirm that these
78 previously reported QTLs influence male approach behaviours (as opposed to other traits that
79 may influence courtship, for example male morphology[25]), we reanalysed our previous data,
80 by explicitly considering whether males initiated courtship towards *H. melpomene*, *H. cydno*
81 or both types of female during choice trials.

82 Consistent with our previous analyses[20], we found that males of both species, and in
83 particular *H. melpomene* males, show a strong preference for females of their own phenotype.
84 F1 and backcross-to-*melpomene* prefer to court *melpomene* females, whereas courtship
85 initiation behaviours segregate in the backcrosses to *cydno* (Figure 1). The QTL on
86 chromosome 1 was retained in our model of initiation behaviours (Supplementary figure 1; *n*
87 = 139, ΔELPD = -13.6 (SE±5.7), *i.e.* a change of 2.34 SE units). In contrast, the QTL on
88 chromosome 17 was not retained (*n* = 139, ΔELPD = -2.1 (S.E.±3.0)). Notably, backcrosses-
89 to-*cydno* males heterozygous (*i.e.* with a *H. melpomene* allele derived from the F1 father) at
90 the best supported QTL from our previous analysis[20] (on chromosome 18) initiated courtship
91 towards *H. melpomene* females more frequently than males homozygous for the *H. cydno*
92 allele (Figure 1, bottom left; *n* = 139, ΔELPD: -10.9 (S.E.±5.1), *i.e.* a change of 2.14 SE
93 units). Together with previous evidence that male hybrids bearing *H. melpomene* alleles at
94 *optix* prefer to court the artificial models of *H. melpomene* females over those of *H. cydno*[26],
95 these results suggest that the QTL on chromosome 18 harbours genes for visual attraction
96 behaviours towards females with the red pattern (and that the *H. melpomene* alleles are
97 dominant). As a result, we focused our subsequent analyses on this QTL on chromosome 18
98 (and also because tight linkage of *optix* allowed us to track the alleles at the preference locus
99 in hybrid crosses). A recent study[27] also reports no candidate chemosensory genes for
100 reproductive isolation at this QTL region on chromosome 18 (or at any other QTL).
101
102 **27 *genes* within the major QTL are differentially expressed in the brains and eyes of *H.***
103 ***cydno* and *H. melpomene*.**
104 We hypothesized that changes in gene regulation that determine differences in visual mate
105 preference behaviours might occur during pupal development (for instance, during visual
106 circuit assembly) or at the adult stage, and must involve changes in the peripheral and/or
107 central nervous system[28]. Therefore, we generated RNA-seq libraries for combined eye and

108 brain tissue, across two pupal stages (around the time of ommochrome pigment deposit and
109 half-way through pupal development) and one adult stage, for *H. melpomene* and *H. cydno*
110 and compared their gene expression levels. Across the QTL region on chromosome 18 (which
111 spans 2.75 Mb and contains 200 genes), we identified 27 genes that show differential
112 expression between *H. melpomene* and *H. cydno*, in at least one of the three developmental
113 stages. These were mostly located within the QTL peak (*i.e.* the genomic region with
114 strongest statistical association with male preference) or in close proximity to *optix* (Figure 2).
115 The same genes were frequently differentially expressed across development (Supplementary
116 table 1), with 11 genes being differentially expressed in more than one stage.
117 The genomic region between the start of chromosome 18 and *optix* (comprising the
118 QTL peak) is highly divergent between *H. melpomene* and *H. cydno*[29], and divergent gene
119 sequences within this region could also introduce mapping biases of RNA-seq reads. To
120 account for this, we repeated the analysis having mapped to both the *H. melpomene* reference
121 genome[30] and to a *H. cydno* genome[31]. Generally, we found similar patterns of differential
122 expression when mapping to the *H. cydno* genome (Supplementary figure 2, Supplementary
123 table 2A). Nevertheless, in subsequent analyses we excluded two genes, HMEL034187g1 and
124 HMEL034229g1, which showed reversal of the fold change or did not show differential
125 expression when mapping to the *H. cydno* genome respectively. Results for the QTL on
126 chromosome 1 are reported in the supplementary materials (Supplementary table 1).
127
128 **A *regucalcin* and an *ionotropic glutamate receptor* are upregulated in both *H. melpomene***
129 **and F1 hybrid males, across independent sequencing experiments.**
130 Our previous behavioural experiments suggest that the alleles for the *H. melpomene*
131 behaviour are dominant over the *H. cydno* alleles[20,26] (Figure 1). Given this pattern of
132 dominance, we predicted that genes underlying variation in male preference to be up- or
133 down- regulated in the brains of both *H. melpomene* and first generation (F1) hybrid males,

134 with respect to *H. cydno*. Of the putative genes differentially expressed between *H. cydno* and
135 *H. melpomene* reported above, only four, within the QTL candidate region, were differentially
136 expressed between the F1 hybrids and *H. cydno* (Figure 2). These included two *regucalcins*
137 (also called *senescence marker proteins-30*: HMEL013552g1, HMEL034199g1), an
138 *ionotropic glutamate receptor* (HMEL009992g4), which is a putative ortholog of *Grik2*, and
139 one gene with no annotated function (HMEL009992g1). We obtained the same results
140 regardless of whether we considered both males and females together, or males alone. Further
141 inspection of spliced mRNA-reads indicated that the two annotated *regucalcins* were in fact a
142 single gene (from now on referred to as *regucalcin2*). This was also the case for the
143 *ionotropic glutamate receptor* and the gene with no annotated function (from now on referred
144 to as *Grik2*).
145 To ensure that the apparent fragmentation of a few gene models in the *H. melpomene*
146 (Hmel2.5) annotation[32] did not introduce inaccuracies in estimates of differential gene
147 expression, we produced a new transcript-based annotation of the *melpomene* genome using
148 Cufflinks's RABT[33]. Repeating all comparative transcriptomic analyses using this new
149 annotation (where exons previously considered to be distinct genes are now assigned correctly
150 to single genes), we confirmed that both *regucalcin2* and *Grik2* were differentially expressed
151 in both species and hybrids comparisons. Finally, to determine whether differential expression
152 of these genes was repeatable across sequencing experiments, we generated RNA-seq
153 libraries for additional *H. melpomene, H. cydno and* F1 hybrids adult brains, sampled five
154 years after our initial tissue collection and sequenced independently. Differential expression
155 of *Grik2* and *regucalcin2* is consistently detected both between *H. melpomene* and *H. cydno*
156 and between F1 hybrids and *H. cydno*, both when these datasets are analysed separately, and
157 when combined.
158

159 **Differential expression of *Grik2* and *regucalcin2* in adults is likely due to cis-regulatory**
160 **effects.**
161 Causal changes in gene regulation underlying phenotypic variation associated with the QTL
162 must result from *cis*- rather than *trans*-regulation. In other words, if changes in expression of
163 *Grik2* and *regucalcin2* account for the observed shifts in behaviour associated with the QTL
164 these must be due to changes within the *cis*-regulatory regions of the genes themselves (as
165 opposed to of other *trans*-acting genes elsewhere in the genome); and these causal mutations
166 should be within the QTL region on chromosome 18. To determine whether differences in
167 gene expression levels between parental species were due to *cis*- or *trans*-regulatory changes,
168 we conducted allele specific expression (ASE) analyses in adult F1 hybrids (from both
169 sequencing batches). In F1 hybrids, both parental alleles are exposed to the same *trans*-
170 environment, and consequently *trans*-acting factors will act on alleles derived from each
171 species equally (unless there is a change in the *cis*-regulatory regions of the respective
172 alleles). Therefore, differences in allele specific expression indicate changes in *cis*-regulatory
173 regions[34]. For both candidate genes (*Grik2* and *regucalcin2*) the *H. melpomene* allele was
174 significantly more highly expressed relative to the *H. cydno* allele ($p<0.001$, Wald test),
175 suggesting *cis*-regulatory effects (Figure 3).
176
177 ***Grik2* is differentially expressed in hybrids that essentially differ only for allelic**
178 **composition at the behavioural QTL region.**
179 In order to study the specific effects that *H. melpomene* derived alleles at the QTL on
180 chromosome 18 had on gene-expression, we introgressed this region into a *H. cydno*
181 background through multiple backcrosses (crossing design in Supplementary figure 3). We
182 wanted to investigate whether differences at this QTL regulated expression of any specific
183 genetic pathway during development, and more generally what changes in genome-wide
184 transcription were observed in hybrids differing (mostly) just at this QTL region. We

185 compared 6 *cyd/melp* vs. 10 *cyd*/*cyd* (at the QTL region on chromosome 18) BC3 hybrids
186 sampled at 156 hours after pupal formation (APF) (Supplementary figure 5A), and 8 *cyd/melp*
187 vs. 9 *cyd*/*cyd* for those at 60h APF.
188 Across the entire genome, only 23 genes at 156h APF and 29 genes at 60h APF were
189 differentially expressed between third-generation backcross hybrids with *cyd*/*cyd* and
190 *cyd/melp* genotypes at the QTL. Within the QTL candidate region, *Grik2* was the only gene
191 detected as differentially expressed between species and hybrids at these pupal stages (at 60h
192 APF, Supplementary figure 4). Of the remaining 10 genes that were differentially expressed
193 between both species and backcross hybrids at either stage, seven were located within the
194 introgressed region (0 - 6.3 Mb, Supplementary table S3), and it seems most likely that these
195 are regulated by introgressed *H. melpomene cis*-acting elements. We discuss the possibility of
196 a *cis*-regulatory element within the QTL candidate region acting on a gene outside the QTL in
197 the Supplementary information.
198 It is possible that causative loci (e.g. expressed RNA or protein factors) within the QTL could
199 act on other genes in *trans* (both on chromosome 18 and elsewhere in the genome), and
200 identifying these could provide insight into the mode of action of causative genetic elements.
201 Three genes differentially expressed both in species and backcross hybrid comparisons, and
202 located outside of introgressed regions, including HMEL014795g1 and HMEL015842g1 (at
203 10.2 Mb and 13.5 Mb on chromosome 18, respectively) and HMEL030024g1 (on
204 chromosome 1) might be considered good candidates for *trans*-regulation. However, in our
205 backcrosses, the region on chromosome 18 introgressed from *H. melpomene* into a *H.* cydno
206 background extends ~3.6 Mb beyond the QTL candidate region, making it difficult to
207 determine whether these genes are regulated by loci associated with variation in behaviour.
208 Furthermore, no genetic pathway was enriched for gene expression differences between these
209 hybrids at either pupal stage (PANTHER enrichment test[35]), suggesting that overall this QTL
210 harbours a few, modular changes in gene regulation in the developing brain/eyes of *H. cydno*

211 and *H. melpomene,* or at least not impacting on other genes in *trans* with a clear mode of
212 action.

213 To verify that differential expression of candidate genes within the QTL region is
214 driven by *H. melpomene* alleles on chromosome 18 and not by other *H. melpomene* alleles at
215 *trans*-acting genes on other chromosomes, we compared gene expression levels between
216 hybrids carrying *cyd/melp* vs. *cyd/cyd* regions on chromosomes chr1, chr4 and chr15, chr20
217 (Supplementary figure 5A). In these comparisons, there was no signal of differential
218 expression on chromosome 18**.** This supports the *cis*-regulatory activity of the *melpomene*
219 allele of candidate genes on chromosome 18. To test this further, we conducted another allele
220 specific expression (ASE) study in the BC3 hybrids, which suggested *trans*-only regulatory
221 effects for *Grik2* at these pupal stages, and *cis*-regulatory effects for *regucalcin2* at 60h APF
222 ($p<0.038$, Wald test) (Supplementary figure 6). Since causal gene/s might exert an effect on
223 behaviour due to their action during development or in adult form, and this action might in
224 turn be differently (*cis*- vs *trans*-) regulated, we still considered both genes as strong
225 candidates.

226

227 **4 genes with protein-coding substitutions within the QTL candidate region have**
228 **predicted effects on protein function**.

229 Because shifts in behavioural phenotypes could be due to changes in protein-coding regions,
230 we additionally considered protein-coding substitutions between *H. melpomene* and *H. cydno.*
231 Overall, we found 152 protein-coding substitutions, spanning 54 of the 200 genes across the
232 entire QTL candidate region. We then studied whether these variants were predicted to have
233 non-neutral effects on protein function with PROVEAN[36]. The PROVEAN algorithm predicts
234 the functional effect of protein sequence variations based on how they affect alignments to
235 different homologous protein sequences**.** We found 4 genes with such predicted effects
236 (PROVEAN score < -2.5): Specifically, a *WD40*-repeat domain containing protein

237 (HMEL013551g3), a *cysteine protease* (HMEL009684g2), a *MORN* motif containing protein
238 (HMEL006660g1), and another *regucalcin* (HMEL013551g4) adjacent to, but distinct from,
239 that found to be differentially expressed above (from now on referred to as *regucalcin1*).
240
241 **Candidate genes occur in regions with reduced gene flow.**
242 Of our six candidate genes for preference behaviours that contribute to reproductive isolation
243 between *H. cydno* and *H. melpomene* (*regucalcin2*, *Grik2* and the four genes with protein
244 coding modifications), five are found within the QTL peak (Figure 4). Genetic changes
245 causing reproductive isolation between populations are expected to reduce localized gene
246 flow in their genomes. Therefore, we compared the position of our candidate genes to
247 estimated levels of admixture proportions (f_d)[37] between *H. melpomene* and *H. cydno* across
248 the QTL candidate region[38]. We found that candidate genes were located in genomic regions
249 with low f_d values (Figure 4), suggesting localized resistance to gene flow between *H.*
250 *melpomene* and *H. cydno* at these genes and their putative *cis*-regulatory regions.
251
252 **Discussion**
253 Behavioural isolation is frequently implicated in the formation of new species[1,39], and often
254 involves the correlated evolution of both mating cues and mating preference. Here we have
255 analysed a genomic region in a pair of closely related sympatric butterflies, *H. cydno* and *H.*
256 *melpomene*, that contains genes for divergence in both an ecologically relevant mating cue
257 and the corresponding preference. Physical linkage between ecological and mating traits will
258 facilitate speciation by allowing different barriers to act in concert to restrict gene flow[40,41].
259 Although the genes underlying changes in the warning pattern cue in *Heliconius* are well
260 characterized[21,22,23,42] (e.g. *optix*), those underlying the corresponding shift in behaviour have
261 not previously been identified[20,43,44]. We have pinpointed a small number of genes that fall
262 within the QTL peak, which show either expression (*regucalcin2* and *Grik2*) or protein

263 coding differences (HMEL013551g3, HMEL009684g2, and *regucalcin1*) and fall within a
264 region of reduced admixture, that are strong candidates for modulating mating behaviour.
265 Two broad neural mechanisms could underlie the evolution of divergent visual
266 preferences, involving changes in either i) detection at the sensory periphery or ii) the
267 processing and/or integration of visual information. Although *H. melpomene* and *H. cydno*
268 have the same retinal mosaics/class of photoreceptors[45], spectral sensitivity in the *Heliconius*
269 eyes could be altered by filtering pigments[46], or other physiological processes taking place at
270 the photoreceptors/sensory periphery, eventually shifting sensitivity towards different
271 wavelengths (and possibly colour patterns). It has previously been hypothesized that the gene
272 regulatory networks for ommochrome deposition in the *Heliconius* eyes might have been co-
273 opted in the wings[47], where *optix* plays a central role, and therefore that *optix* might play a
274 role in eye pigmentation in *Heliconius.* However, the protein product of *optix* has not been
275 detected in pupal or adult retinas of various *Heliconius* species tested[48], and therefore has no
276 obvious link to ommochrome deposition in the eyes. More generally, the underlying
277 evolutionary mechanism is unlikely to involve detection at photoreceptors, as this would
278 probably have a broad effect on downstream processing[2] and alter the visual perception of the
279 animal's wider environment.
280 The second mechanism, involving changes in the processing, and/or integration, of
281 visual information, could act through an alteration of neuronal activity or connectivity. For
282 instance, different levels of gene expression in conserved neural circuits between *H.*
283 *melpomene* and *H. cydno* may affect overall synaptic weighting and determine whether a
284 signal (e.g. colour and motion) elicits a motor pattern (response towards a female) or not.
285 Consistent with this scenario, the composition of ionotropic receptors at post-synapses is a
286 key modulator of synaptic transmission[49], implicating *Grik2*. Differential expression of
287 ionotropic receptors is also associated with variation in female preference behaviours in
288 fish[50,51.52], raising the possibility that ion channels might provide a likely route to modulate

289 mate preferences across taxa more broadly. *Regucalcins* are involved in calcium signalling[53],
290 which regulates synaptic excitability and plasticity[54], and has an important role in axon
291 guidance[55] (albeit alongside additional roles across a broad range of biological processes),
292 making the two *regucalcins* we identify strong candidates for behaviour.

293 Changes in the regulation of genes with pleiotropic effects are likely to be less
294 detrimental compared to changes in their protein-coding sequences[56] (although emerging
295 evidence has begun to suggest that enhancer/repressor elements may be more pleiotropic than
296 previously thought[57,58]). Furthermore, there is considerable evolutionary potential in the co-
297 option of transcription factors/networks[56] that regulate neural patterning or neuron-type
298 activity, possibly resulting in novel adaptive expression patterns. In line with this, *regucalcin2*
299 and *Grik2,* which are differentially expressed in the eyes and brain in both our species and
300 hybrid comparisons, are likely to be involved in multi-functional processes, such as calcium
301 signalling and ion transport, and likely have pleiotropic alleles. We also found evidence of
302 *cis*-regulatory effects for both genes, which would be required of the causal genetic change
303 within the QTL, if it were to be in gene regulation.

304 Neither *regucalcin2* or *Grik2* show male-biased gene expression, which might be
305 expected of candidate genes for a behaviour that is evident only in males. It is possible that
306 the lack of sex-biased expression indicates that visual cues are similarly important in female
307 mating preference in this species pair. Indeed, Chouteau *et al*[59] report that female preferences
308 contribute to *disassortative* mating between colour pattern morphs of *H. numata,* and it is
309 possible that in *H. cydno* and *H. melpomene* females share the same genetic basis for colour
310 pattern-based discrimination as males. Alternatively, if visual preference behaviour is
311 restricted to males, changes in gene expression may be integrated differently in the female
312 and male nervous systems. The role of female preference in *Heliconius* mate choice remains
313 poorly understood. Although emerging data suggests that female choice does contribute to

314 reproductive isolation between *melpomene-cydno* clade taxa[60,61], it remains to be tested if
315 there is a strong visual component to this preference similar to that observed in males.
316 Despite expectations that non-coding, regulatory loci may provide a flexible route to
317 divergent mating preferences, we also found substitutions in coding regions at the QTL,
318 which are predicted to have an effect on protein functioning and therefore remain strong
319 candidates. These genes include *regucalcin1*, which is distinct from, but located next to,
320 *regucalcin2* (which is differentially expressed). Notably, the eye transcript of *regucalcin1* was
321 recently characterized as fast-evolving across *Heliconius* species[62]. Other candidates include a
322 *cysteine protease*, which functions in protein degradation, and might be linked to behaviour
323 for example through degradation of neurotransmitters, a *MORN* motif containing protein
324 (function unknown), and a *WD40* containing protein. *WD-repeat* containing proteins have
325 been implicated in a wide array of functions ranging from signal transduction to apoptosis
326 (https://www.ebi.ac.uk/interpro).
327 Although preference for red colouration and the *optix* gene are tightly linked, we find
328 no evidence that *optix* is differentially expressed in the eyes or brains of our two species. It is
329 also not located within the QTL peak (and it contains no non-synonymous changes in protein
330 coding regions[21]). It seems unlikely therefore that changes in cue and preference are
331 pleiotropic effects of the same allele. More generally, although we have pinpointed the
332 strongest candidates yet identified for assortative mating behaviours in *Heliconius*, it is
333 possible that actual causal changes in gene regulation are restricted to developmental stages
334 other those sampled, or restricted to a few neuronal populations not detected with
335 transcriptomic data from eyes and whole brain tissue. Nonetheless, by sampling at two pupal
336 stages (around the time of *optix* expression/ommochrome pigment deposit in the wing/eye and
337 halfway through pupal development) and at the adult stage, we should have captured
338 important transitions for the behavioural programming of the two species.

339 Work in the past decade has shown that complex innate behavioural differences
340 between species can be encoded in relatively few genetic modules[63,64], but very few
341 studies[65,66,67] have identified specific genes underlying behavioural evolution. In particular,
342 traditional laboratory organisms continue to provide important insights into the evolution and
343 genetics of behaviour,[28,66,68], however, comparative approaches are required to determine if
344 developmental principles can be broadly applied, and also to incorporate a wider range of
345 phenotypic variation and sensory modalities. The challenge now is to increase the resolution
346 of studies in non-traditional systems, in order to link individual genetic elements to
347 behaviours, *and* the sensory and/or neurological structures through which they are mediated.
348 In this light, we have identified a small handful of strong candidate genes associated with the
349 evolution of visual mate preference behaviours in *Heliconius*. These genes are in tight
350 physical linkage with the locus for the corresponding shifts in an ecologically relevant mating
351 cue, providing an important opportunity to investigate the build-up of genetic barriers crucial
352 to speciation. The candidate genes identified seem more likely to alter visual processing or
353 integration, rather than detection at photoreceptors, consistent with permitting changes in
354 mate preference without altering perception of the animal's wider environment.
355
356 **Methods**
357 **Courtship initiation analyses.** Butterfly rearing, crossing design and genotyping are
358 described in detail elsewhere[20]. In brief, we assayed male preference behaviours for *H.*
359 *melpomene*, *H. cydno,* their first generation (F1) hybrids and backcross to hybrids to both
360 parental species in standardized choice trials. Males were introduced into outdoor
361 experimental cages (1x1x2m) with a virgin female of each species and courtship behaviours
362 recorded. Whenever possible, trials were repeated for each male (median = 5 trials). To
363 determine whether previously identified QTLs for courtship time contribute to variation in
364 courtship *initiation* behaviours, we performed a *post-hoc* analysis using categorical models in

365 a Bayesian framework with a multinomial error structure, using the R package *brms*. All
366 models were run under default priors (non- or very weakly informative). In contrast to our
367 previous analysis[20], in which we considered the number of minutes (*i.e.* time) for which
368 courtship was directed towards *H. cydno* or *H. melpomene* females, here the response variable
369 was number of trials in which male courtship was *initiated* towards *H. cydno* females only, *H.*
370 *melpomene females* only*,* or both female types (hereafter referred to as "*initiation"*). Across
371 males the median number of trials with a response was 3. Using backcross-to-*cydno* males
372 only, we fitted *initiation* as a response variable to genotype (*cyd*/*cyd* or *cyd*/*melp*) at each
373 QTL, which were included as separate fixed effects. Individual ID was fitted as random
374 factor. To test the effect of each QTL on male *initiation*, we compared the saturated model
375 incorporating all three QTL with reduced models excluding each QTL in turn, using
376 approximate leave-one-out (LOO) cross-validation[69] as implemented in *brms*, and based on
377 expected log pointwise predictive density (ELPD). Normal distribution of ELPD can be a
378 straightforward approximation given our large samples sizes (n=139)[69]. Therefore, we
379 considered an absolute value of ELPD greater than 1.96 units of its standard error as
380 indicative of the reduced model being less-informative than the saturated model (95%
381 confidence). Males that did not initiate courtship to any female across trials were excluded
382 from analyses, resulting in a dataset of 139 males, from a total of 146 backcross males for
383 which we had genotype data. Finally, we extracted predictors and credibility intervals for
384 backcross males with differing genotypes from the minimum adequate model. Credibility
385 intervals for *H. melpomene*, *H. cydno,* F1 hybrid and backcross to *melpomene* males
386 displayed in Figure 1 were generated following the same procedures. Raw data and analysis
387 code are available in the following github repository:
388 https://github.com/SpeciationBehaviour/neural_genes_heliconius.git

389 **Butterfly collection, rearing and crossing design for expression analyses.** Wild *H.*
390 *melpomene rosina* and *H. cydno chioneus* individuals were caught along Pipeline Road near
391 Gamboa, Panama, in the Soberania National Park, and used to establish stocks at the
392 Smithsonian Tropical Research Institute insectaries in Gamboa. Butterflies were reared in
393 common garden conditions, in 2x2x2m cages, and provided with fresh *Psiguria* flowers and
394 10% sugar solution. Larvae were reared on fresh *Passiflora* shoots/leaves until pupation. *H.*
395 cydno, *H. melpomene* and hybrid individuals used for RNA-seq (see below) were reared
396 concurrently and under the same conditions. F1 hybrids were obtained by crossing a wild-
397 caught *H. m. rosina* male to an insectary-bred virgin *H. c. chioneus* female.
398 Third-generation backcross hybrids (BC3) were generated by outcrossing a hybrid
399 male with a red forewing band (crossing design shown in Supplementary figure 3) to virgin
400 *H. cydno* females, over three generations. The peak of the behavioural QTL reported
401 previously[20] on chromosome 18 (at 0cM) is in very tight linkage with the *optix* colour pattern
402 locus (at 1.2cM), which controls for the presence and absence of the red forewing band seen
403 in *H. melpomene rosina*. Presence of the red forewing band is dominant over its absence so
404 that segregation of the red band can be used to infer genotype at the *optix* locus. Specifically,
405 hybrid individuals with a red forewing band are heterozygotes for *H. melpomene/H. cydno*
406 alleles at the *optix* locus, whereas individuals lacking the red band are homozygous for the *H.*
407 *cydno* allele. Due to the tight linkage we expected little recombination between *optix* and
408 QTL peak even after three generations of introgression, allowing us to infer genotype at the
409 preference-*optix* locus (which we confirmed with genetic data, see below).

410 **Tissue dissection, RNA extraction and mRNA sequencing.** Eye (ommatidia and retinal
411 membrane) and brain tissue (central brain and optic lobes) were dissected out of the head
412 capsule (as a single combined tissue) in cold (4 ºC) 0.01M PBS solution, at two pupal stages:
413 60 hours after pupal formation (60h APF) and 156h APF; and in adults aged 9 - 13 days in

414 2013. We sampled adults at around 10 days of age because by this stage males are mature and
415 frequently court females[70]. Adult males and females sampled were sexually naive. We
416 decided to sample at 60h APF because this is the developmental stage at which *optix* is
417 expressed in the wing, so we hypothesized that it might had also been when *optix* is expressed
418 in the brain. We sampled at 156h APF as a putative stage halfway through pupal
419 development, and at this stage most of the major neural connections have just been
420 established in the *Heliconius* brain (Stephen Montgomery, unpublished data).
421 Tissues were stored in RNAlater at 4 ºC for 24 hours, and subsequently at -20 ºC, until
422 RNA extraction. Total RNA was extracted using TRIzol Reagent (Thermo Fisher, Waltham,
423 MA, USA) and a RNeasy Mini kit (Qiagen, Valencia, CA, USA). Samples were treated with
424 DNase I (Ambion, Darmstadt, Germany). Integrity of total RNA was checked either on an
425 agarose gel or using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA,
426 USA). RNA concentration was measured on a Nanodrop spectrophotomer. Illumina TruSeq
427 RNA-seq libraries were prepared and sequenced at Edinburgh Genomics (Edinburgh, UK)
428 with 100 bp paired-end reads (in 2014). To avoid lane effects the distribution of the species
429 samples was randomized on the sequencing platform. More detailed information about
430 individuals and sequencing yields can be found in the Supplementary dataset. In 2019, we
431 independently sampled a further 5 *H. melpomene, 5 H. cydno,* and 6 F1 hybrids males, of 10
432 days of age. These F1 hybrids were generated by crossing a wild caught *H. c. chioneus* male
433 to an insectary-bred virgin *H. m. rosina* female. Briefly, for these, RNA was extracted
434 similarly with Trizol Reagent and a PureLink RNA Mini Kit, with PureLink DNase digestion
435 on column (Thermo Fisher, Waltham, MA, USA). Illumina 150bp paired-end RNA-seq
436 libraries were prepared and sequenced at Novogene (Hong Kong, China).
437
438 **RNA-seq read mapping and differential gene expression analyses.** After a quality control
439 of RNA-seq reads with FastQC, we trimmed adaptor and low-quality bases using TrimGalore

440 v.0.4.4 (https://www.bioinformatics.babraham.ac.uk/projects/). RNA-seq reads were mapped
441 to the *H. melpomene* 2.5 genome assembly[30]/annotation[32] using STAR v.2.4.2a[71] in 2-pass
442 mode. We only kept reads that mapped in 'proper pairs' using Samtools[72]. The number of
443 reads mapping to each gene were estimated with HTseq v. 0.9.1[73] with model "union", thus
444 excluding ambiguously mapped reads. Differential gene expression analyses between
445 species/hybrids were conducted in DESeq2[74], including sequencing batch as random factor
446 when comparing species samples from both datasets. We considered only those genes
447 showing a 2-fold change in expression level, and at adjusted (false discovery rate 5%) p-
448 values < 0.05, to be differentially expressed, to exclude expression differences caused by
449 known differences in brain morphology[75], that in this species pair are clustered in the visual
450 system[76], although there is no evidence that visual mate preference is linked to these divergent
451 brain morphologies.
452
453 **Sexing pupae**. In all DESeq2 analyses, sex was included as a random factor. To sex pupae,
454 we first marked duplicate RNA mapped reads with Picard
455 (https://broadinstitute.github.io/picard/), and used GATK 3.8[77] to split uniquely mapped reads
456 into exon segments and trim sequences overhanging the intronic regions. We then used
457 Haplotype Caller on each individual, using calling and filtering parameters according to the
458 GATK Best Practices for variant calling on RNA-seq data. The sex of pupal samples was
459 inferred from the proportion of heterozygous (biallelic) SNPs using the R package SNPstats.
460 Males (ZZ) were expected to have >> 0% heterozygous sites, whereas females (ZW) to have
461 0%. Z-linked heterozygosity of the pupal samples (Supplementary table 4) were in line with
462 expectations (either ~ 0 for females or an order of magnitude higher for males), and matched
463 heterozygosity of either adult males or females, for which the sex was determined from
464 external morphology.
465

466 **Inference of gene function and transcript-based annotation.** Biological functions of
467 annotated genes were inferred with InterProScan v5[78], using the corresponding Hmel2.5
468 predicted protein sequences. InterProScan uses different databases like InterPro, Pfam,
469 PANTHER, and others, to infer functional protein domains and motifs (based on homology).
470 To study whether specific biological functions were enriched among genes showing
471 differential expression among hybrid types, we conducted the PANTHER enrichment test[35]
472 (with Bonferroni correction for multiple testing) using *Drosophila melanogaster* as the
473 reference gene function database.
474 Upon detailed inspection of the mapping coverage of spliced RNA-seq reads to the
475 Hmel2.5 gene annotation, we noticed that some gene models were fragmented, namely, a few
476 exons that appeared to be spliced together were incorrectly considered distinct genes. To
477 check that this did not introduced inaccuracies in our differential gene expression analyses,
478 we re-annotated the *melpomene* genome using the Cufflinks reference annotation-based
479 transcript (RABT) assembly tool[33] We used the transcriptomic data from both *H. melpomene*
480 and *H. cydno* to reannotate the *melpomene* genome, separately for every developmental stage,
481 and reconducted the differential gene expression analyses in DESeq2 as described above.
482 Repeating all comparative transcriptomic analyses using these new annotations (where exons
483 were correctly considered as part of single genes), we confirmed that both *regucalcin2* and
484 *Grik2* were differentially expressed in both species and hybrids comparisons.
485
486 **Inference of BC3 hybrids genome composition.** In order to perform comparative
487 transcriptomic analyses between third-generation backcross hybrids (BC3) segregating at the
488 QTL on chromosome 18 (crossing design in Supplementary figure 3), we first determined
489 which genomic regions in these hybrids were heterozygous (*cyd/melp*) or homozygous
490 (*cyd*/*cyd*). For this, we inferred variants from RNA-seq reads for each BC3 hybrid
491 (individually as above), and from the combined *H. melpomene* and *H. cydno* samples. For the

492 species, we used HaplotypeCaller[77] on RNA-seq samples from all developmental stages of
493 either species, to produce individual genomic records (gVCF), and then jointly genotyped *H.*
494 *melpomene* and *H. cydno* gVCFs (separately for the two species) using genotypeGVCFs with
495 default parameters. Genotype calls were filtered for quality by depth (QD) > 2, strand bias
496 (FS) < 30 and allele depth (DP) > 4. For further analyses we kept *biallelic* genotypes only.
497 We then used the *intersect* function of *bcftools*[72] to infer variants exclusive to *the H. cydno*
498 and to the *melpomene* samples.

499 We calculated the fraction of variants that each BC3 hybrid individual shared with the
500 *H. melpomene* and with the *H. cydno* samples, in non-overlapping 100kb windows. We
501 compared these to the fraction of variants that a F1 hybrid and a *H. cydno* individual (not
502 included in the combined genotyping of the *cydno* samples), shared with the same species
503 samples, and found that they matched either one of them, indicating heterozygous (*cyd/melp*)
504 or homozygous (*cyd*/*cyd*) regions (Supplementary figure 5B). In this analysis, we considered
505 only those 100kb windows where BC3 hybrids/F1 hybrid/*H. cydno* individuals shared more
506 than 30 variants with the *melpomene*/*cydno* samples.

507 To corroborate our findings, we repeated the same type of analysis, this time inferring
508 species-specific variants for *H. melpomene* and *H. cydno* using 10 *H. melpomene rosina* and
509 10 *H. cydno chioneus* genome resequencing samples. Variant calling files (vcf) were retrieved
510 from Martin et al[38]. We considered only *biallelic* genotype calls that had 10 < DP < 100 and
511 genotype quality (GQ) > 30. With this analysis we found the same heterozygous and
512 homozygous regions in BC3 hybrids**.**

513 The size and number of the introgressed regions were in line with expectations about
514 3rd generation backcross hybrids following our crossing design: segregating at the level of
515 chromosome 18 and at four other chromosomes. For the BC3 hybrids sampled at 156 hours
516 after pupal formation (APF) we had 6 *cyd/melp* and 10 *cyd*/*cyd* at the QTL region on
517 chromosome 18 (Supplementary figure 5A), for those at 60h APF, 8 *cyd/melp* and 9 *cyd*/*cyd*

518 hybrids at the same region. The average percentage of the genome that is heterozygous
519 (*cyd*/*melp)* as opposed to homozygous (*cyd*/*cyd)* outside of chromosome 18 was ~6% (close
520 to the expectation that a 3rd generation backcross genome should be 1/16 heterozygous
521 (*cyd*/*melp*)).
522
523 **Allele-specific expression (ASE) in hybrids.** In order to conduct ASE analyses we first
524 identified species specific variants, fixed in either *H. melpomene* and *H. cydno*. For this, we
525 took the quality filtered variants inferred from the species genome resequencing data, and
526 assigned those genotype calls in *H. cydno* and *H. melpomene* for which allele frequency (AF)
527 was > 0.9 as homozygous (we did not consider indels in this analysis). We then used *bcftools*
528 *intersect*[73] to get only those variants for which *H. cydno* and *H. melpomene* had opposite
529 alleles.
530 At the same time, we called variants from RNA-seq reads of F1 hybrid individuals
531 (using F1 hybrids from both datasets), again according to the GATK Best Practices (with the
532 exception of parameters -window 35 -cluster 3, to increase SNPs density), and selected only
533 heterozygous SNPs in F1s that matched the species-specific variants. Finally, we used
534 GATK's ASEReadCounter[77], with option "-drf DuplicateRead" (without deduplicating RNA
535 reads), to count RNA reads in the F1 hybrids (and later on in BC3 hybrids) that mapped to
536 either the *H. cydno* or the *H. melpomene* allele. We summed all reads mapping to either the *H.*
537 *cydno* or *H. melpomene* allele/variant within the same gene (both for gene models of the
538 Hmel2.5 gene annotation and for the Cufflinks annotation we assembled previously). To test
539 for allele specific expression (diffASE) we fitted the model "~0 + individual + allele" in
540 DESeq2[74], setting library size factors to 1 (thus not normalizing between samples, as the test
541 for diffASE is conducted within individuals). We only considered those alleles showing at
542 least a 2-fold change in expression and $p < 0.05$, as differentially expressed.

543 In order to check that there were no biases in alleles assignment to one of the two
544 species, we analyzed the ratios of the species alleles, for every gene, and checked that they
545 were not systematically biased to either one of the two species. The $\log_2$ fold-changes of the
546 species alleles were centered around 0, suggesting no obvious bias in alleles assignment[79]
547 (Supplementary figure 7).
548
549 **Protein-coding substitutions and predicted effects on protein-function.** We inferred fixed
550 variants in protein-coding regions from the combined *H. melpomene* and *H. cydno* RNA
551 samples in order to include variants from genes for which we detected expression in the
552 brain/eyes across the 3 stages. We took the quality filtered variants called from the joint
553 genotyping of RNA-seq data of *H. cydno* and *H. melpomene* (from all stages, these did not
554 include adult samples from the more recent sequencing batch), and selected those genotype
555 calls for which allele frequency (AF) > 0.8, and where the allelic variant was present in at
556 least 7 individuals of the ~30 samples (for each species). We retained those
557 substitutions/indels validated with the genome resequencing data. For this, of the genotype
558 calls found in RNA reads from brain/eyes of different stages, we kept only those that were
559 also called in at least 8 of the 10 genome resequencing samples of each species. We
560 considered this overlapping set of variants as being fixed in *H. melpomene rosina* or *H. cydno*
561 *chioneus*. Following a similar approach to Bendesky *et al.*[67], we then restricted this set of
562 substitutions between *H. cydno* and *H. melpomene* to protein-coding regions, and selected
563 those non-synonymous substitutions that were considered to have moderate or high effect on
564 protein function, with SNPeff[80]. Finally, we used the PROVEAN algorithm[35], to further study
565 the functional effects of these substitutions on protein function. The PROVEAN algorithm
566 predicts the functional effect of protein sequence variations based on how they affect
567 alignments to homologous protein sequences (for this we used the PROVEAN protein

568 database online). We selected those amino acid changes with the suggested PROVEAN score
569 < -2.5, indicating non-neutral effect on protein function.
570
571 **Admixture analyses.** We retrieved estimated admixture proportions between *H. melpomene*
572 *rosina* and *H. cydno chioneus*, for 100kb and 20kb windows, from Martin et al.[38]
573
574 **Data availability.** RNA-seq data will be deposited on a public database
575 (https://www.ebi.ac.uk/ena) on acceptance (accession ID pending).
576
577 **Code availability.** Analysis scripts and behavioural data are available at:
578 https://github.com/SpeciationBehaviour/neural_genes_heliconius.git
579
580 **Author contributions:** R.M.M. and M.R. conceived the study and designed the experiments,
581 with input from W.O.M. and C.D.J.; M.R. analysed expression and sequence data; A.E.H.
582 analysed the behavioural data; M.R., T.J.T., S.H.M. and R.M.M. reared butterflies, dissected
583 neural tissue and extracted RNA; R.M.M., C.D.J. and W.O.M. secured funding, contributed
584 resources and provided supervision; R.P. additionally secured funding and contributed
585 resources; M.R. and R.M.M. wrote the manuscript with contributions from all authors.
586
587 **Acknowledgments**
588 We thank Chi-Yun Kuo, Liz Evans, Adriana Tapia, Moises Abanto and Morgan Oberweiser
589 for help and assistance in the insectaries. We are grateful to Ana Pinharanda for sharing the *H.*
590 *cydno* genome assembly and annotation and to Simon Martin for sharing admixture
591 proportions and comments on the manuscript. We thank the Smithsonian Tropical Research
592 Institute for providing research infrastructure, the Ministerio del Ambiente for permission to
593 collect butterflies in Panama, Edinburgh Genomics and Novogene for sequencing support and

Jochen Wolf who kindly provided computational resources. MR, AEH and RMM are supported by an Emmy Noether fellowship and research grant awarded to RMM by the Deutsche Forschungsgemeischaft (DFG) (Grant Number: GZ: ME 4845/1-1). RMM was also supported by a Junior Research Fellowship from King's College Cambridge, an Ernst Mayr fellowship from the Smithsonian Tropical Research Institute, a Varley-Gradwell fellowship from the Oxford University Museum of Natural History, and the Balfour-Browne Fund, University of Cambridge. SHM was supported by a NERC IRF (NE/N014936/1). This project was also supported by the Puerto Rico Science, Technology and Research Trust (ARG 2020-00138), the Institutional Development Award (IdeA) INBRE NIH/NIGMS (award number P20 GM103475) and a National Science Foundation EPSCoR RII Track-2 FEC (award number OIA-1736026) to RP, and a European Research Council (ERC) awarded to CDJ (Grant number: 339873).

References

1. Coyne, J. A., Orr, H. A. *Speciation* (Sunderland, MA: Sinauer 2004).
2. Rosenthal, G. G. *Mate choice* (Princeton University Press, 2017).
3. Mayr, E. *Animal species and evolution* (Harvard University Press, 1963).
4. Cande, J., Prud'homme, B. & Gompel, N. Smells like evolution: the role of chemoreceptor evolution in behavioural change. *Curr. Opin. Neurobiol.* **23**, 152–158 (2013).
5. Arguello, J. R. & Benton, R. Open questions: tackling Darwin's "instincts": the genetic basis of behavioural evolution. *BMC Biol.* **15**, 8–10 (2017).
6. Bay, R. A. *et al.* Genetic Coupling of female mate choice with polygenic ecological divergence facilitates Stickleback Speciation. *Curr. Biol.* **27**, 3344-3349 (2017).
7. Shahandeh, M. P., Pischedda, A., Rodriguez, J. M. & Turner, T. L. The genetics of male pheromone preference difference between *Drosophila melanogaster* and *Drosophila simulans*. *G3 Genes, Genomes, Genet.* **10**, 401–415 (2020).
8. Gould, F. *et al.* Sexual isolation of male moths explained by a single pheromone response QTL containing four receptor genes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8660–8665 (2010)
9. Leary, G. P. *et al.* Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14081–14086 (2012).
10. Fan, P. *et al.* Genetic and neural mechanisms that inhibit *Drosophila* from mating with other species. *Cell* **154**, 89–102 (2013).
11. Brand, P. *et al.* The evolution of sexual signaling is linked to odorant receptor tuning in perfume-collecting orchid bees. *Nat. Commun.* **11** (2020).
12. Xu, M. & Shaw, K. L. Genetic coupling of signal and preference facilitates sexual isolation during rapid speciation. *Proceedings of the Royal Society B*, ***286***, 20191607 (2019).

631 13. Seehausen, O. *et al.* Speciation through sensory drive in cichlid fish. *Nature* **455**, 620–626
632 (2008).
633 14. Hench, K., Vargas, M., Höppner, M. P., McMillan, W. O. & Puebla, O. Inter-chromosomal
634 coupling between vision and pigmentation genes during genomic divergence. *Nat. Ecol.*
635 *Evol.* **3**, 657–667 (2019).
636 15. Merrill, R. M. *et al.* Disruptive ecological selection on a mating cue. *Proc. R. Soc. B Biol.*
637 *Sci.* **279**, 4907–4913 (2012).
638 16. Jiggins, C. D., Naisbit, R. E., Coe, R. L. & Mallet, J. Reproductive isolation caused by
639 colour pattern mimicry. *Nature* **411**, 302–305 (2001).
640 17. Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M. & Nosil, P. Magic traits in
641 speciation: 'magic' but not rare?. *Trends Ecol. Evol.* ***26***, 389-397 (2011).
642 18. Jiggins, C. D. Ecological Speciation in Mimetic Butterflies. *Bioscience* **58**, 541–548 (2008).
643 19. Jiggins, C. D., Estrada, C. & Rodrigues, A. Mimicry and the evolution of premating
644 isolation in *Heliconius melpomene Linnaeus*. *J. Evol. Biol.* **17**, 680–691 (2004).
645 20. Merrill, R. M. *et al.* Genetic dissection of assortative mating behaviour. *PLoS Biol 17*,
646 e2005902 (2018).
647 21. Reed, R. D. *et al.* Optix drives the repeated convergent evolution of butterfly wing pattern
648 mimicry. *Science* ***333***, 1137-1141 (2011).
649 22. Martin, A. *et al.* Diversification of complex butterfly wing patterns by repeated regulatory
650 evolution of a *Wnt* ligand. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12632–12637 (2012).
651 23. Nadeau, N. J. *et al.* The gene cortex controls mimicry and crypsis in butterflies and moths.
652 *Nature* **534**, 106–110 (2016).
653 24. Felsenstein, J. Skepticism Towards Santa Rosalia, or why are there so few kinds of animals?
654 *Evolution* **35**, 124–138 (1981).
655 25. Massey, J. H., Chung, D., Siwanowicz, I., Stern, D. L. & Wittkopp, P. J. The *yellow* gene
656 influences *Drosophila* male mating success through sex comb melanization. *Elife* **8**, 1–20
657 (2019).
658 26. Merrill, R. M., Van Schooten, B., Scott, J. A. & Jiggins, C. D. Pervasive genetic
659 associations between traits causing reproductive isolation in *Heliconius* butterflies. *Proc.*
660 *R. Soc. B Biol. Sci.* **278**, 511–518 (2011).
661 27. Van Schooten, B. *et al.* Divergence of chemosensing during the early stages of speciation.
662 *Proc. Natl. Acad. Sci. U. S. A.* doi: 10.1073/pnas.1921318117 (2020).
663 28. Seeholzer, L. F., Seppo, M., Stern, D. L. & Ruta, V. Evolution of a central neural circuit
664 underlies *Drosophila* mate preferences. *Nature* **559**, 564–569 (2018).
665 29. Martin, S. H. *et al.* Genome-wide evidence for speciation with gene flow in *Heliconius*
666 butterflies. *Genome Res.* **23**, 1817–1828 (2013).
667 30. Davey, J. *et al.* Major improvements to the *Heliconius melpomene* genome assembly used
668 to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3* **6**,
669 695–708 (2015).
670 31. Darragh, K. *et al.* A novel terpene synthase produces an anti-aphrodisiac pheromone in the
671 butterfly *Heliconius melpomene*. *bioRxiv*, 779678 (2019).
672 32. Pinharanda, A. *et al.* Sexually dimorphic gene expression and transcriptome evolution
673 provide mixed evidence for a fast-Z effect in *Heliconius*. *J. Evol. Biol.* **32**, 194–204 (2019).
674 33. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in
675 annotated genomes using RNA-seq. *Bioinformatics* **27**, 2325–2329 (2011).
676 34. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in *cis* and *trans* gene
677 regulation. *Nature* **430**, 85–88 (2004).
678 35. Thomas, P. D. *et al.* Applications for protein sequence-function evolution data:
679 mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res.* **34**,
680 645–650 (2006).

36. Choi, Y., Sims, G. E., Murphy, S, Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7** (2012).
37. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
38. Martin, S. H., Davey, J. W., Salazar, C. & Jiggins, C. D. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* **17**, 1–28 (2019).
39. Nosil, P. *Ecological Speciation* (Oxford University Press, 2012).
40. Kopp, M. *et al.* Mechanisms of assortative mating in speciation with gene flow: connecting theory and empirical research. *Am. Nat.* **191**, 1–20 (2018).
41. Butlin, R. K. & Smadja, C. M. Coupling, reinforcement, and speciation. *Am. Nat.* **191**, 155–172 (2018).
42. Westerman, E. L. *et al. Aristaless* controls butterfly wing color variation used in mimicry and mate choice. *Current Biology*, ***28***, 3469-3474 (2018).
43. Kronfrost, M. R *et al.* Linkage of butterfly mate preference and wing color preference cue at the genomic location of *wingless*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6575-6580 (2006).
44. Chamberlain, N. L., Hill, R. I., Kapan, D. D., Gilbert, L. E., & Kronforst, M. R. Polymorphic butterfly reveals the missing link in ecological speciation. *Science*, ***326***, 847-850 (2009).
45. McCulloch, K. J. *et al.* Sexual dimorphism and retinal mosaic diversification following the evolution of a violet receptor in butterflies. *Mol. Biol. Evol.* **34**, 2271–2284 (2017).
46. Zaccardi, G., Kelber, A., Sison-Mangus, M. P. & Briscoe, A. D. Colour discrimination in the red range with only one long-wavelength sensitive opsin. *J. Exp. Biol.* **209**, 1944–1955 (2006).
47. Monteiro, A. Gene regulatory networks reused to build novel traits. *BioEssays* **34**, 181–186 (2012).
48. Martin, A. *et al.* Multiple recent co-options of *optix* associated with novel traits in adaptive butterfly wing radiations. *Evodevo* **5** (2014).
49. Kandel, E. R, Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A. & Hudspeth. A. J. *Principles of Neural Science*, 2012th Ed. (New York: McGraw Hill, 2000).
50. Ramsey, M. E., Vu, W., & Cummings, M. E. (2014). Testing synaptic plasticity in dynamic mate choice decisions: N-methyl d-aspartate receptor blockade disrupts female preference. *Proc. R. Soc. B Biol. Sci*. ***281***, 20140047 (2014).
51. Bloch, N. I. *et al.* Early neurogenomic response associated with variation in guppy female mate preference. *Nat. Ecol. Evol.* **2**, 1772–1781 (2018).
52. Delclos, P. J., Forero, S. A. & Rosenthal, G. G. Divergent neurogenomic responses shape social learning of both personality and mate preference. *J. Evol. Biol.* ***223*** (2020)
53. Yamaguchi, M. Role of regucalcin in brain calcium signaling. *Integr. Biol.* **4**, 825–837 (2012).
54. Berridge, M. J. Neuronal calcium signaling. *Neuron*, ***2*1**, 13-26 (1998).
55. Bashaw, G. J. & Klein, R. Signaling from axon guidance receptors. *Cold Spring Harb. Perspect. Biol.* **2**, 1–17 (2010).
56. Prud'homme, B., Gompel, N. & Carroll, S. B. Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* ***104*,** 8605-8612 (2007).
57. Preger-Ben Noon, E. *et al.* Comprehensive analysis of a *cis*-regulatory region reveals pleiotropy in enhancer function. *Cell Rep.* **22**, 3021–3031 (2018).
58. Lewis, J. *et al.* Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24174–24183 (2019).
59. Chouteau, M., Llaurens, V., Piron-Prunier, F. & Joron, M. Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures. *Proc. Natl. Acad. Sci. U. S. A.* ***114***, 8325-8329 (2017).

60. Southcott, L. & Kronforst, M. R. Female mate choice is a reproductive isolating barrier in *Heliconius* butterflies. *Ethology*, ***124***, 862-869 (2018).
61. González-Rojas, M. F. *et al*. Chemical signals act as the main reproductive barrier between sister and mimetic *Heliconius* butterflies. *Proc. R. Soc. B Biol.* ***287***, 20200587 (2020)
62. Zhang, W. *et al.* Comparative transcriptomics provides insights into reticulate and adaptive evolution of a butterfly radiation. *Genome Biol. Evol.* **11**, 2963–2975 (2019).
63. Weber, J. N., Peterson, B. K. & Hoekstra, H. E. Discrete genetic modules are responsible for complex burrow evolution in *Peromyscus* mice. *Nature* **493**, 402–405 (2013).
64. Cande, J., Andolfatto, P., Prud'homme, B., Stern, D. L. & Gompel, N. Evolution of multiple additive loci caused divergence between *Drosophila yakuba* and *D. santomea* in wing rowing during male courtship. *PLoS One* **7**, 1–10 (2012).
65. McBride, C. S. *et al.* Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **515**, 222–227 (2014).
66. Ding, Y., Berrocal, A., Morita, T., Longden, K. D. & Stern, D. L. Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature* **536**, 329–332 (2016).
67. Bendesky, A. *et al.* The genetic basis of parental care evolution in monogamous mice. *Nature* **544**, 434–439 (2017).
68. Auer, T. O. *et al.* Olfactory receptor and circuit evolution promote host specialization. *Nature*, 1-7 (2020).
69. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
70. Jiggins, C. D. *The ecology and evolution of Heliconius butterflies* (Oxford University Press, 2016).
71. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
72. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
73. Anders, S., Pyl, P. T. & Huber, W. HTSeq- a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
74. Love, M. I. Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
75. Montgomery, S. H. & Mank, J. E. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Mol. Ecol.* **25**, 5114–5128 (2016).
76. Montgomery, S. H., Rossi, M, McMillan, W. O & Merrill, R. Neural divergence and hybrid disruption between ecologically isolated *Heliconius* butterflies. *bioRxiv*, 182337 (2020)
77. McKenna, A. *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
78. Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199 (2017).
79. York, R. A. *et al.* Behaviour-dependent *cis* regulation reveals genes and pathways associated with bower building in cichlid fishes. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1081–1090 (2018).
80. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff *Fly.* **6**, 80–92 (2012).

Figures

Figure 1. Genotype at the preference QTL on chromosome 18 influences courtship initiation. Ternary plots showing the proportion of 15-minute choice trials in which single male individuals initiated courtship towards *H. melpomene, H. cydno* or both females. Left ternary axis shows the proportion of trials where courtship was initiated towards *H. cydno* female only, bottom axis towards *H. melpomene* female only, and right axis towards both female species. Trials without male response were removed from the dataset. Each point represents a single individual and the location of the point in the triangle is a way of representing these three proportions at once. Lines project the three predicted proportions to corresponding values on the three axes and 95% credibility intervals (CrIs) for these proportions are shown as hexagons. Orange points represent individuals that have inherited at least one *H. melpomene* derived allele at the preference QTL on chromosome 18 (i.e. either *melp*/*melp* or *cyd*/*melp)*; and blue points represent individuals that are homozygous for *H. cydno* alleles at the preference QTL on chromosome 18 (i.e. *cyd*/*cyd*). Point size is scaled to the number of trials in which the male showed a response and a 'jitter' function has been applied (leading to some dots being jittered to outside the triangle).

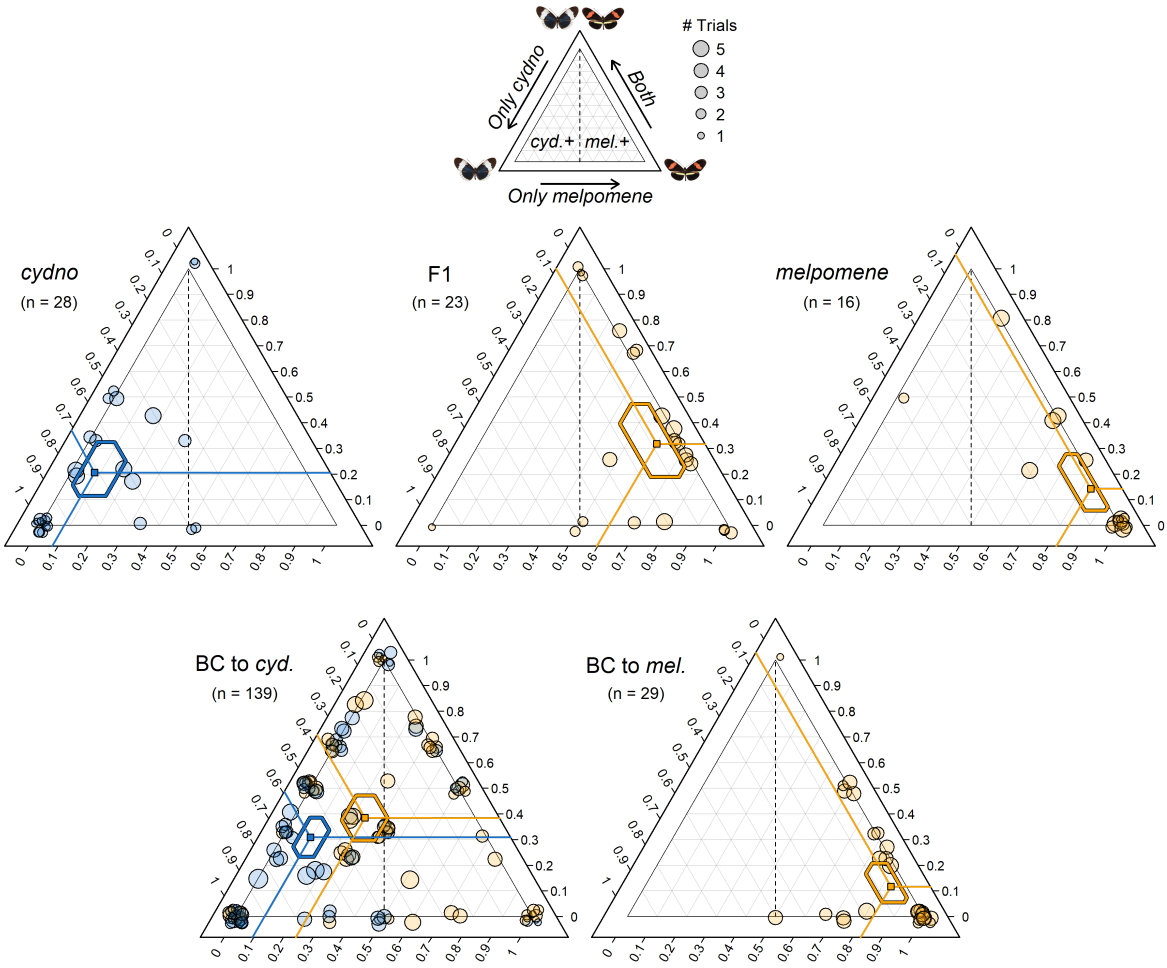


Figure 2. Differential expression at the preference QTL region on chromosome 18. Left: Summary of the comparative transcriptomic analyses with stage, number of samples and chromosome 18 composition. Right: the corresponding results, zooming in on the QTL region on chromosome 18. The *x*-axis represents physical position. The QTL peak, and the rest of the QTL 1.5 LOD candidate region (from[28]) are shown in green and purple, respectively. Points correspond to individual genes, with the *y*-axis indicating the log_2(fold-change) for each comparison. The two horizontal dashed lines (at *y*-values of 1 and -1) indicate a 2-fold change in expression. Genes showing a significant 2-fold+ change in expression level between groups are highlighted in orange and blue, where orange indicates higher levels in *H. melpomene* or in the hybrids *cyd*/*melp* (blue if in *H. cydno* – hybrids *cyd*/*cyd*). Vertical dashed lines highlight those genes that are differentially expressed between *H. melpomene* and *H. cydno* AND between *cyd/melp* vs *cyd/cyd* individuals, at the same stage (*Grik2* and *regucalcin2* at the adult stage, *Grik2* at 60h APF). Two genes highlighted by dashed fuchsia vertical lines were excluded because they did not show differential expression, or showed reversal of the fold change when mapping RNA-seq reads to the *H. cydno* genome. Note that *Heliconius* brain (reconstruction) images, added for reference, do not include the eyes (ommatidia and retinal membrane).

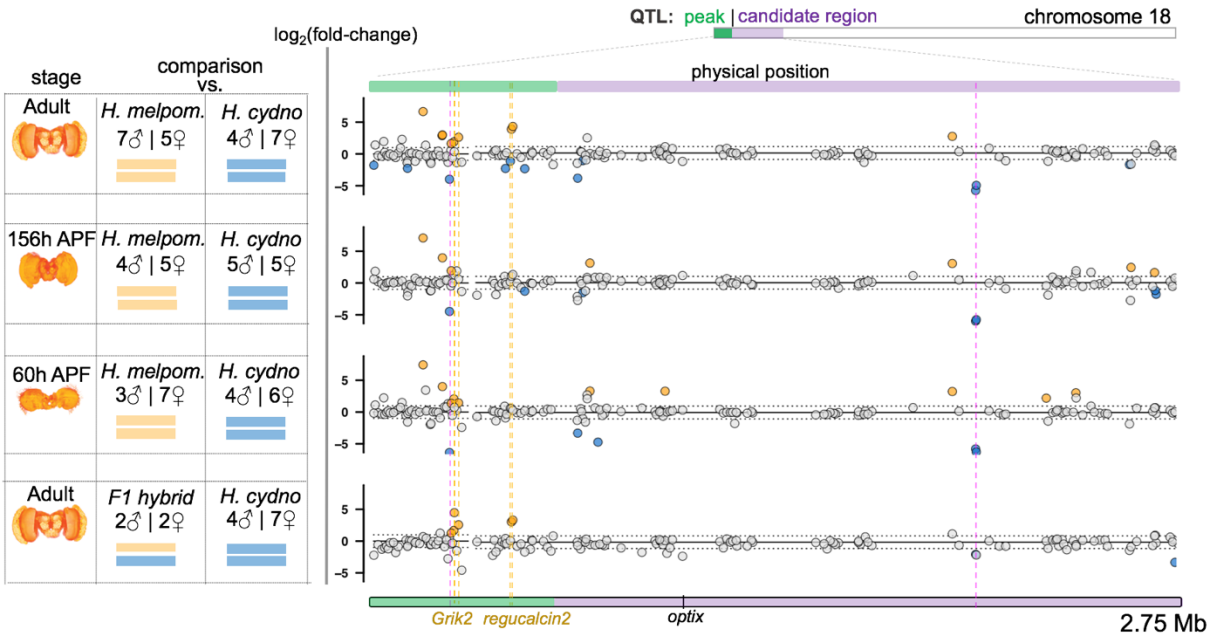
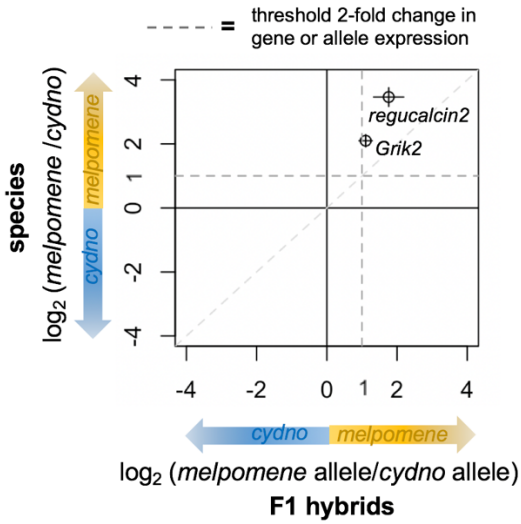
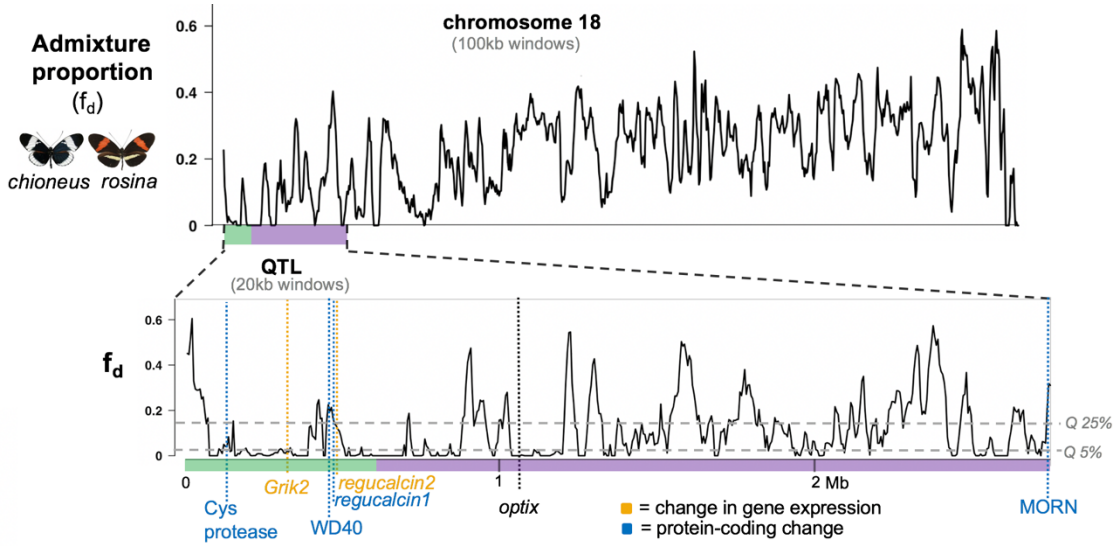


Figure 3. *Grik2* and *regucalcin2* show evidence of allele specific expression. Points indicate the value, and bars the standard error, of the (base 2) logarithmic fold change in expression between parental species (vertical) and the alleles in F1 hybrids (horizontal), for candidate genes (as defined in the transcript-guided annotation). Dashed lines indicate the threshold for a 2-fold change in expression for the genes in the species (horizontal), and for the alleles in the hybrids (vertical). Both genes seem to be regulated by a combination of *cis- and trans*-acting factors, rather than *cis*-acting factors alone (which would be indicated by y=x).



859 **Figure 4**. **Gene flow at the QTL region for behaviour.** Admixture proportion (f_d) values
860 estimated in overlapping 100kb (top) and 20kb (bottom) windows for chromosome 18 (top) and
861 the QTL region (bottom) between *H. melpomene rosin*a and *H. cydno chioneus*, with candidate
862 genes positions highlighted by a vertical dashed line, and *optix* location displayed for reference.
863 The *x*-axis represents physical position, the *y*-axis indicates the f_d value. f_d values close to zero
864 indicate that the proportion of shared derived alleles, and consequently gene flow, between *H.*
865 *melpomene* and *H. cydno* is small (or zero), implying localized selection against foreign alleles
866 that introgress between the two species. 5% and 25% quantile of the f_d distribution are indicated
867 by horizontal grey dotted lines.
868
869



870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888

Supplementary information: Visual mate preference evolution during butterfly speciation is linked to neural processing genes

Matteo Rossi[1,2], Alexander E. Hausmann[1], Timothy J. Thurman[2,3], Stephen H. Montgomery[4], Riccardo Papa[2,5,6], Chris D. Jiggins[7], W. Owen McMillan[2] & Richard M. Merrill[1,2*]

1. Division of Evolutionary Biology, LMU, Munich, Germany; 2. Smithsonian Tropical Research Institute, Panama City, Panama; 3. Division of Biological Sciences, University of Montana, USA; 4. School of Biological Sciences, University of Bristol, Bristol, UK; 5. Department of Biology, University of Puerto Rico, San Juan, Puerto Rico; 6. Molecular Sciences and Research Center, University of Puerto Rico, San Juan, Puerto Rico; 7. Department of Zoology, University of Cambridge, Cambridge UK. *corresponding authors**:**
rossi@bio.lmu.de, merrill@bio.lmu.de

927 **Supplementary methods and results**
928
929
930 **Mapping RNA-seq reads to the *Heliconius cydno* genome.** To determine whether the *H.*
931 *melpomene* reference genome introduced mapping biases of RNA-seq reads, possibly
932 affecting differential expression estimates, we also mapped to a *H. cydno*
933 assembly/annotation[1]. Generally, we found similar patterns of differential expression when
934 mapping to the two genomes. Since **i)** we observed an equal decrease (~ 40 %) of genes
935 showing 2-fold changes in *melpomene* and *cydno* when mapping to *H. cydno,* at every stage
936 ($P > 0.05$ at every stage, Fisher's Exact test, Table S2A), and **ii)** this decrease was widespread
937 throughout the genome, we concluded that the *H. melpomene* reference genome did not bias
938 differential gene expression analyses. We report the number of reads mapping to both
939 genomes for each adult sample (Table S2B).
940
941 **Allele-specific expression in the introgression line.** BC3 hybrids had different combinations
942 of chromosomes segregating for the *melpomene* alleles in a *H. cydno* background. Therefore,
943 in principle, we could not infer *cis*- or *trans*- gene regulatory effects genome-wide from the
944 profiles of allele specific expression (ASE) in these hybrids as for F1 hybrids, due to the
945 diverse trans-acting environments. However, analyses (comparing gene expression levels
946 between hybrids carrying *cyd/melp* vs. *cyd/cyd* regions on chromosomes other than 18) imply
947 that differential expression of the candidate genes seems to be driven by the *H. melpomene*
948 copy difference within the introgressed region on chromosome 18. Therefore, ASE analyses
949 of candidate genes in BC3 hybrids carrying *cyd/melp* alleles on chromosome 18 should
950 indicate whether the differences are due to cis- or trans-regulatory effects from within the
951 introgressed region (Figure S6).
952

953 In BC3 hybrids sampled at 156h APF and 60hAPF, the *H. melpomene* and *H. cydno* alleles of
954 the *ionotropic glutamate receptor* (*Grik2*) are expressed at very similar levels (P > 0.05 at
955 both stages, Wald test), suggesting trans-only regulatory effects at these stages for *Grik2*. We
956 detected diffASE expression of *regucalcin2* at 60h APF (p <0.05, Wald test), (and at 156h
957 APF there was a tendency towards up-regulation of the *H. melpomene* allele), but at these
958 stages *regucalcin2* was not detected as differentially expressed between pure species. Thus,
959 although there is evidence for *cis*-regulatory effects for species differences in *regucalcin2*
960 expression during development, the possible effect of *regucalcin2* on behaviour at these
961 stages is less clear.
962
963 The region introgressed into *H. cydno* extended ~3.6 Mb beyond the QTL candidate region
964 and seven genes located within this region were differentially expressed (at either stage) in
965 both species and backcross hybrid comparisons (Table S3). We conducted ASE analyses on
966 these seven genes and found evidence for *cis*-regulation for only one (HMEL010030g1,
967 p<0.001, Wald test), which had no annotated function. Unfortunately, we were unable to
968 detect allele-informative reads for the other six genes including a serine protease inhibitor
969 (HMEL014931g1), a CUB domain containing protein (HMEL002560g1), a methyltransferase
970 (HMEL010030g2), a gene with a reverse transcriptase domain (HMEL034294g1), an
971 ionotropic glutamate receptor (HMEL034304g1), and a major facilitator superfamily
972 transporter (HMEL015745g1). (A further two genes on chromosome 18 were located outside
973 of the introgressed region, for which we could not conduct ASE analyses because they were
974 in *cyd*/*cyd* regions).
975 Our expectation was that *cis*-regulatory elements will normally act on genes in close
976 proximity (the average distance between regulatory elements such as enhancers/repressors
977 and the genes they regulate has been estimated to be less than 100kb[2]), making differentially
978 expressed genes within the QTL peak – and to a lesser extent the candidate region – the best

979 candidates. The QTL might conceivably harbour *cis*-regulatory element(s) acting on these
980 gene(s) at a longer distance on chromosome 18 (*i.e.* outside of the QTL region). Although the
981 closest is at >2.6Mb beyond the QTL peak, we currently cannot completely rule out those
982 genes, differentially expressed in both species and backcross hybrid comparisons on
983 chromosome 18 (Table S3), as – albeit far less well supported – candidates.
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004

Supplementary figures

Figure S1. Ternary plots showing the number of 15-minute choice trials in which courtship was initiated towards *melpomene, cydno* or both females for backcross-to-*cydno* males, with different genotypes at the two QTLs retained in our model (on chromosome 1 and chromosome 18). Left ternary axis shows proportion of trials where courtship was initiated towards *H. cydno* female only, bottom axis towards *H. melpomene* female only, and right axis towards both female species. Lines project the three predicted proportions to corresponding values on the three axes and 95% credibility intervals (CrIs) for these proportions are shown as hexagons. Point size is scaled to the number of trials in which the male showed a response and a 'jitter' function has been applied (leading to some dots being jittered to outside the triangle).

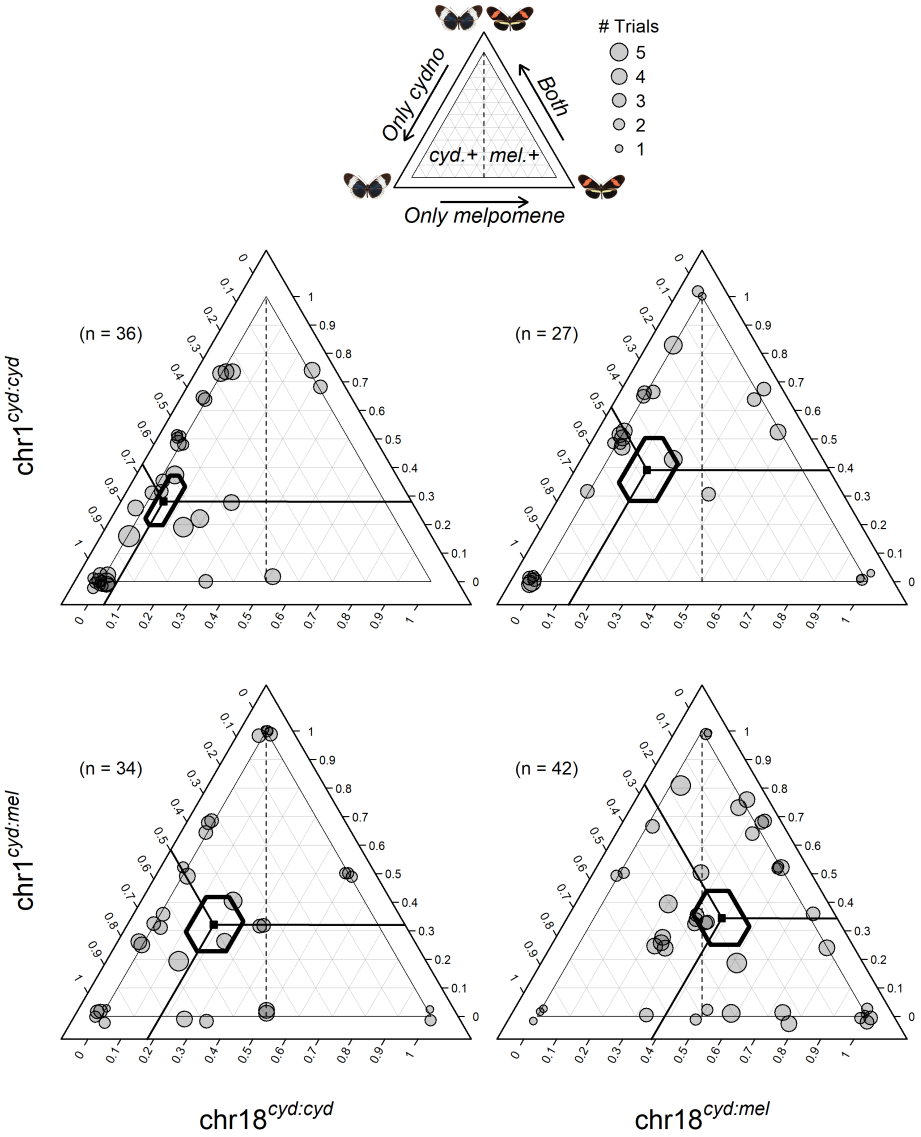


Figure S2. Results of comparative transcriptomic analyses between *H. melpomene* and *H. cydno* (in the imago) when mapping RNA-seq reads to the *H. melpomene* assembly/annotation (top), and to the *H. cydno* assembly/annotation (bottom), zooming in on the QTL region on chromosome 18. The *x*-axis represents physical position. Points correspond to individual genes, with the *y*-axis indicating the log_2(fold-change) for each comparison. The two horizontal dashed lines (at *y*-values of 1 and -1) indicate a 2-fold change in expression. Genes showing a significant 2-fold+ change in expression level between groups are highlighted in orange and blue, where orange indicates higher levels in *melpomene*, blue if in *cydno*. Genes detected as differentially expressed mapping to both *melpomene* and *cydno* genomes are labelled with gene names. dehydr.=2-oxoisovalerate dehydrogenase, OBP=odorant-binding protein.

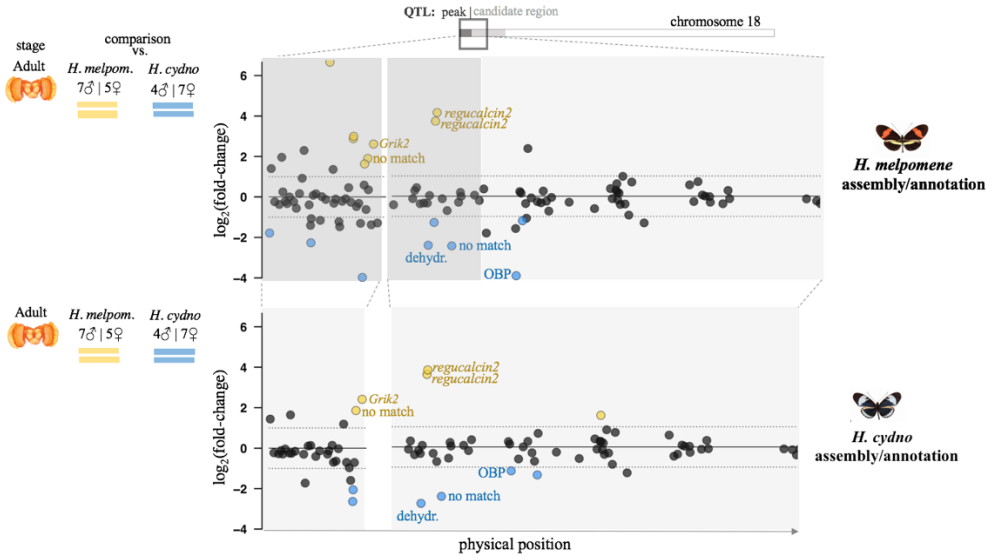


Figure S3. Crossing design for producing backcross hybrids segregating at the QTL on chromosome 18. This introgression line was created by outcrossing a male hybrid to *H. cydno* females over three generations, selecting a hybrid male that showed a red band on the wing at each generation. This meant that these males carried one copy of the *H. melpomene* allele at the *optix* locus. We expected that, following recombination (which occurs in males), by the fourth generation we would remain with two types of individuals: either *cyd*/*melp* or *cyd*/*cyd* at the level of the *optix* region (which approximately corresponds to the region associated with male preference behaviour).

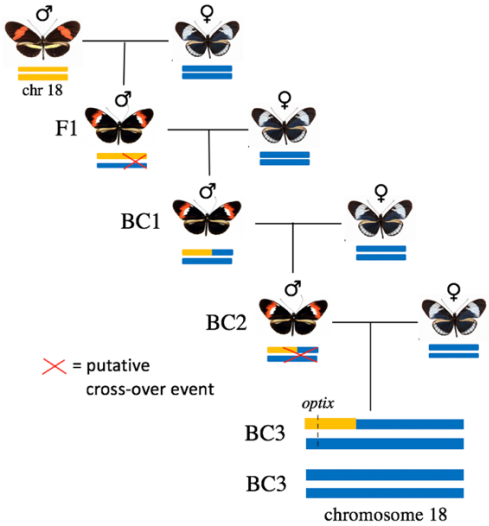


Figure S4. Differential gene expression at the QTL region at pupal stages. Left: summary of the comparative transcriptomic analyses with stage, number of samples and chromosome 18 composition. Right: the corresponding results, zooming in on the QTL region on chromosome 18. The *x*-axis represents physical position. The QTL peak, and the rest of the QTL 1.5 LOD candidate region are shown in green and purple, respectively. Points correspond to individual genes, with the *y*-axis indicating the log_2(fold-change) for each comparison. The two horizontal dashed lines (at *y*-values of 1 and -1) indicate a 2-fold change in expression. Genes showing a significant 2-fold+ change in expression level between groups are highlighted in orange and blue, where orange indicates higher levels in *melpomene* or in the hybrids *cyd*/*melp* (blue if in *cydno* – hybrids *cyd*/*cyd*). Vertical dashed lines highlight those genes that are differentially expressed between *H. melpomene* and *H. cydno* AND between *cyd/melp* vs *cyd/cyd* individuals, at the same stage. One gene highlighted by a dashed fuchsia vertical line was excluded because it showed reversal of the fold change when mapping RNA-seq reads to the *H. cydno* genome.

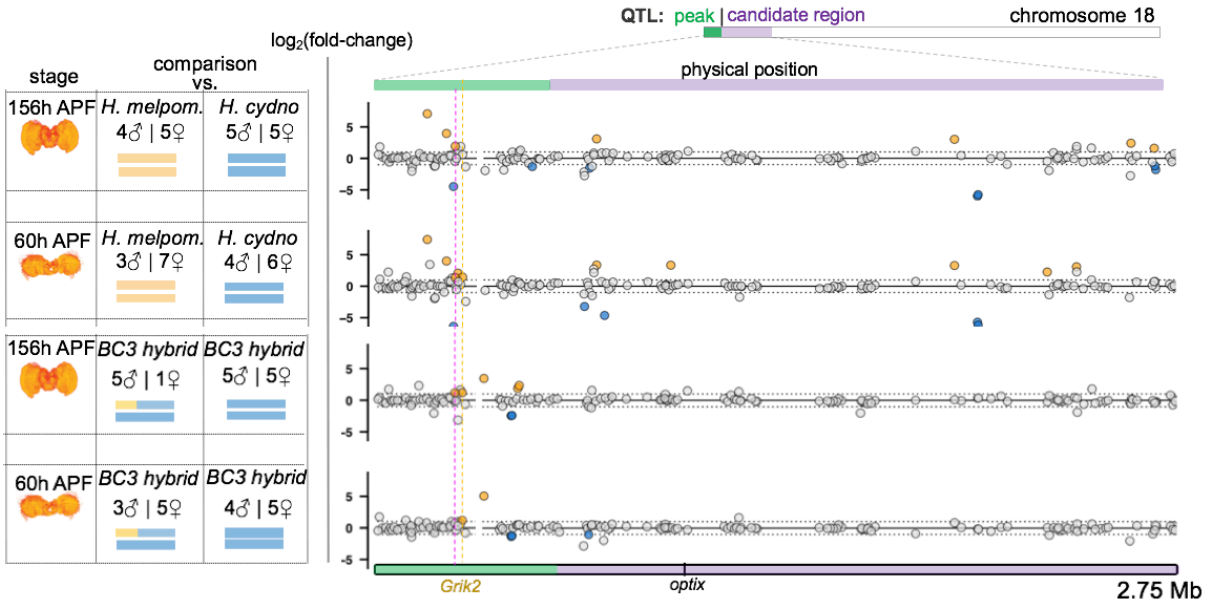
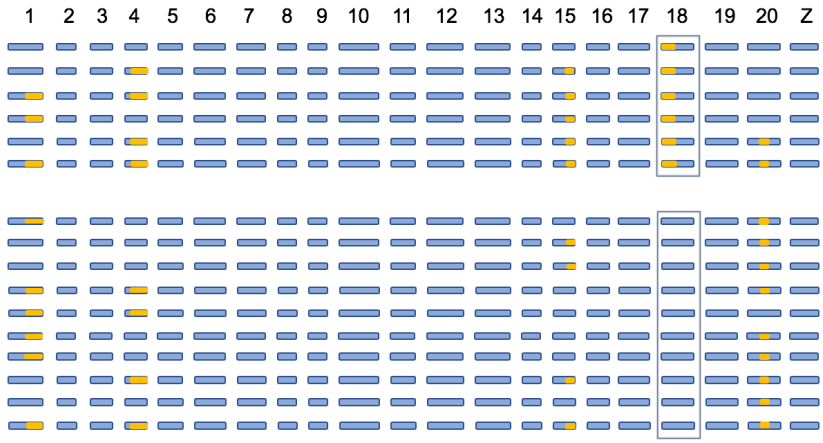
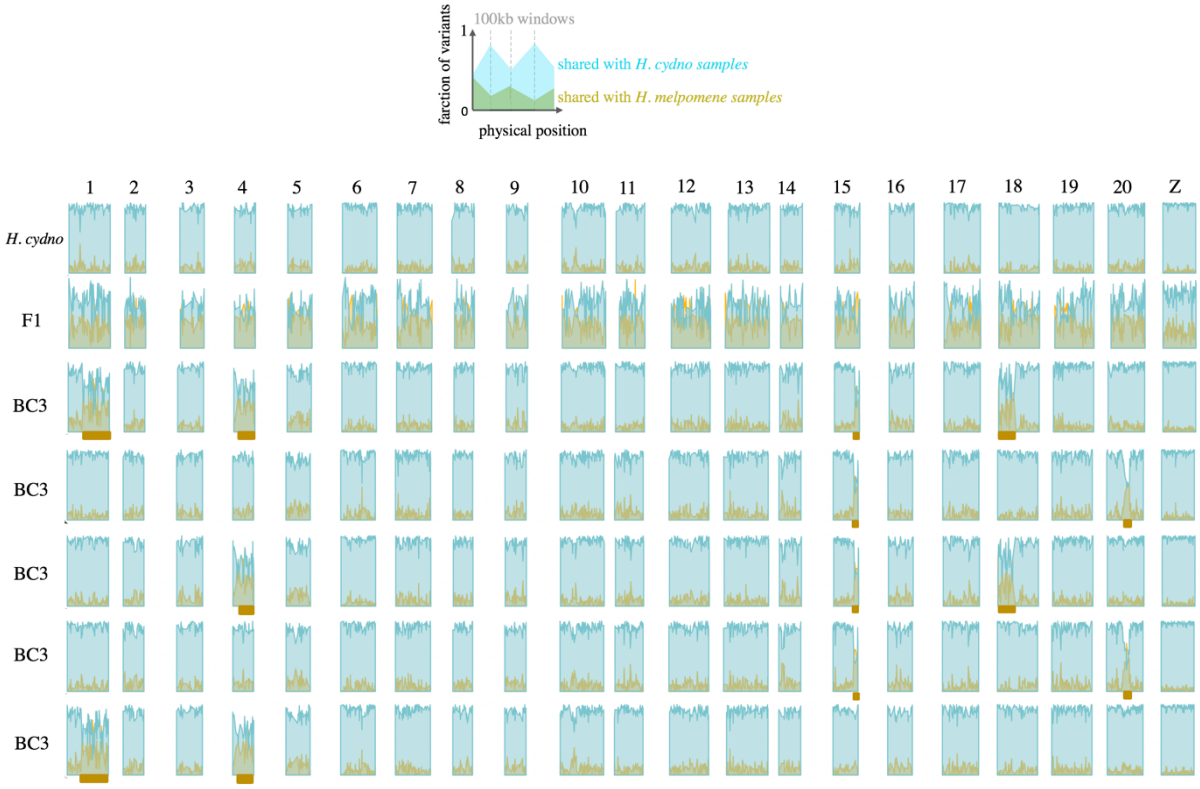


Figure S5. A) Schematic representation of hybrid pupae (sampled at 156h APF) genome composition. Columns represent chromosomes, rows represent individuals, orange indicates *cyd/melp* regions, blue indicates *cyd/cyd* regions. B) Genome composition of (a subset of) BC3 hybrids. We calculated the fraction of SNPs and indels that each BC3 hybrid, one *cydno* and one F1 hybrid samples shared with *H. melpomene* and *H. cydno* samples, in non-overlapping 100kb windows. x-axes represent physical position (for each chromosome), y-axes fractions of shared variants with *melpomene* (in gold) and with *cydno* (in light blue). Matching variant fractions between BC3 hybrids and the F1 hybrid, indicating heterozygous regions, are highlighted with a gold bar underneath. Note that the general trend of higher number of variants shared with *H. cydno* in heterozygous regions is due to the fact that we inferred variants by mapping to the *H. melpomene* genome (and used variant sites only for this analysis).

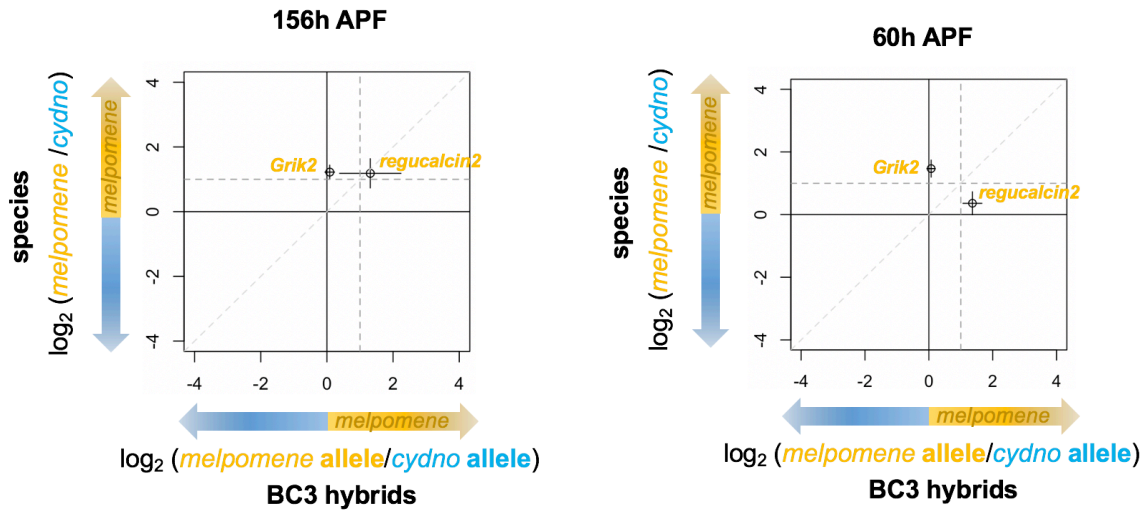
A)



B)

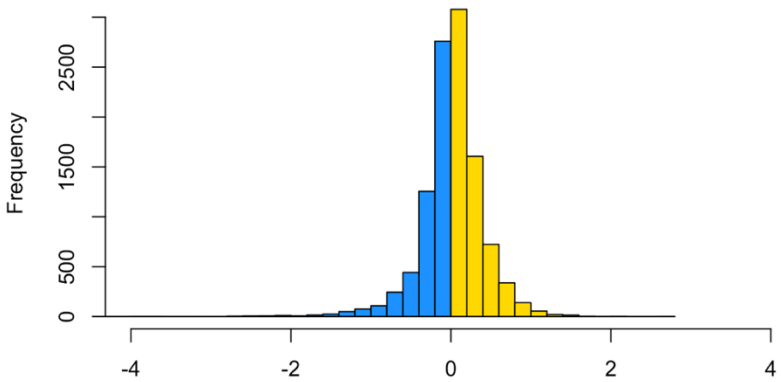


1135 **Figure S6.** Allele specific expression profiles of candidate genes at pupal stages. Points indicate
1136 the value, and bars the standard error, of the (base 2) logarithmic fold change in expression
1137 between parental species (vertical) and the alleles in F1 hybrids (horizontal), for candidate
1138 genes (as defined in the transcript-guided annotation). Dashed lines indicate the threshold for a
1139 2-fold change in expression for the genes in the species (horizontal), and for the alleles in the
1140 hybrids (vertical).
1141



1142
1143
1144

1145 **Figure S7.** Distribution of the (base 2) logarithmic fold change in allele expression. Coloured
1146 bars indicate the number of genes showing a bias in expression for the *H. cydno* allele (in blue)
1147 and for the *H. melpomene* allele (in yellow). Values departing from 0 on the x-axis, indicate an
1148 increase in the fold change for the *H. cydno* allele (negative values) or for the *H. melpomene*
1149 allele (positive values), respectively.
1150



1151
1152
1153
1154
1155
1156

Supplementary tables

Table S1. List of differentially expressed genes in species and hybrids comparisons.

A) QTL chromosome 1. Orange indicates genes up-regulated in *H. melpomene*, and blue those up-regulated in *H. cydno.*

#	Gene name (Hmel2.5)	Annotated function	Species comparison			F1 Hybrids
			60h APF	156h APF	Adult	Adult
1	HMEL002973g1	No match	✓	✓	✓	✓
2	HMEL003796g1	Regulation of enolase protein 1	✓	✗	✓	✓
3	HMEL011272g1	no match	✗	✓	✓	✗
4	HMEL030024g1	Ribonuclease H superfamily	✓	✓	✓	✗
5	HMEL030042g1	SWR1-complex protein 5	✗	✗	✓	✗
6	HMEL030052g1	reverse transcriptase	✓	✓	✓	✗
7	HMEL005260g1	unknown	✗	✓	✗	✗
8	HMEL030040g1	No match	✗	✓	✗	✗
9	HMEL030037g1	No match	✓	✓	✗	✗
10	HMEL010076g1	Amino acid transporter	✓	✗	✗	✗

1191 **B**) QTL chromosome 18. Those genes found to be differentially expressed when also mapping
1192 to the *H. cydno* genome are highlighted in bold. Genes annotated as distinct but sharing the
1193 same number in the table (#) were later found to be single genes (see second paragraph of the
1194 Results section).
1195
1196

#	Gene name (Hmel2.5)	Annotated function	Species comparison			Hybrids comparison		
			60h APF	156h APF	Adult	60h APF	156h APF	Adult
1	**HMEL009992g1**	**No match**	✓	✗	✓	✗	✗	✓
1	**HMEL009992g4**	***Ionotropic glutamate receptor***	✓	✗	✓	✓	✓	✓
2	**HMEL009996g1**	***Gag-related protein***	✓	✓	✓	✗	✗	✗
3	HMEL034168g1	unknown	✗	✗	✓	✗	✗	✗
4	HMEL034173g1	*SWR1-complex protein 5*	✗	✗	✓	✗	✗	✗
5	HMEL034176g1	*Aspartic peptidase*	✓	✓	✓	✗	✗	✗
6	HMEL034184g1	No match	✗	✗	✓	✗	✗	✗
7	**HMEL034185g1**	**No match**	✓	✓	✓	✗	✗	✗
8	HMEL034187g1	*Major facilitator superfamily (MFS) transporter*	✓	✓	✓	✗	✓	✓
9	**HMEL003176**	***Odorant binding protein***	✗	✓	✓	✓	✗	✗
10	HMEL013551g1	*2-oxoisovalerate dehydrogenase*	✗	✗	✗	✓	✓	✗
10	**HMEL013551g2**	***2-oxoisovalerate dehydrogenase***	✗	✗	✓	✓	✗	✗
11	HMEL013551g4	*SMP-30/regucalcin*	✗	✗	✓	✗	✗	✗
12	**HMEL013552g1**	***SMP-30/regucalcin***	✗	✗	✓	✗	✓	✓
12	**HMEL034199g1**	***SMP-30/regucalcin***	✗	✗	✓	✗	✓	✓
13	HMEL014202g1	*Catalase*	✗	✗	✓	✗	✗	✗
14	**HMEL014202g3**	***Catalase***	✗	✓	✗	✗	✗	✗
15	**HMEL034201g1**	***No match***	✗	✗	✓	✗	✗	✗
16	**HMEL034205g1**	***No match***	✓	✗	✓	✗	✗	✗
17	HMEL034227g1	*Ribonuclease H superfamily*	✓	✓	✓	✗	✗	✗
18	HMEL034229g1	*Endonuclease/exonuclease/phosphatase superfamily*	✓	✓	✓	✗	✗	✓
19	HMEL034230g1	*No match*	✓	✓	✓	✗	✗	✗
20	**HMEL003863g1**	***Vacuolar protein sorting-associated (VPS) protein***	✓	✓	✗	✗	✗	✗
21	HMEL003863g3	*No match*	✓	✗	✗	✗	✗	✗
22	**HMEL006662g1**	***Serpin family protein***	✗	✓	✗	✗	✗	✗
23	**HMEL006663**	***Odorant binding protein***	✗	✓	✗	✗	✗	✗
24	**HMEL022553**	***Odorant binding protein***	✗	✓	✗	✗	✗	✗
25	HMEL001038g1	*Monocarboxylate transporter*	✓	✗	✗	✗	✗	✗
26	HMEL014190g1	*unknown function*	✓	✗	✗	✗	✗	✗
27	HMEL034236g1	*No match*	✓	✗	✗	✗	✗	✗
28	HMEL034189g1	*PiggyBac transposable element-derived protein*	✗	✗	✗	✗	✗	✓
29	HMEL034246g1	*No match*	✗	✗	✗	✗	✗	✓
30	HMEL034195g1	*Gag-related protein*	✗	✗	✗	✓	✓	✗

Table S2.

A) Number of genes showing significant >2-fold change in expression, at different stages, mapping to the *H. melpomene* and to the *H. cydno* genomes. Note that the considerable reduction in the number of genes detected as differentially expressed when mapping to *H. cydno* is most likely a result of the lower quality/completeness of the *H. cydno* genome assembly.

Stage	Mapping to:	Up-regulated in *H. melpomene*	Up-regulated in *H. cydno*
Adult	*H. melpomene*	694	733
	H. cydno	390	451
156h APF	*H. melpomene*	837	667
	H. cydno	518	403
60h APF	*H. melpomene*	846	642
	H. cydno	490	376

B) Number of reads mapping to the *H. melpomene* and the *H. cydno* genome for every species sample at the adult stage.

H. melpomene			***H. cydno***		
ID	**mapped to *H. mel***	**mapped to *H. cyd***	**ID**	**mapped to *H. mel***	**mapped to *H. cyd***
45	14660303	14399385	**50**	17617312	10276127
47	7157775	7106443	**51**	14965350	11157953
53	14013806	13804267	**57**	17460091	10958479
70	11850181	11656033	**58**	20474961	14651415
71	11218116	11023267	**67**	27415308	15742957
78	14114255	13618292	**68**	22159665	12175286
80	14551439	14097670	**81**	34947798	19085351
83	13616160	13260160	**82**	15525206	10264627
100	13771837	13458527	**84**	13139892	9836625
104	12802894	12664628	**98**	18846622	12585701
128	13840154	13489851	**99**	20309368	13271219
218	16942305	16812556			

Table S3. List of differentially expressed genes in both species and backcross hybrid comparisons (at 60h APF or 156h APF). Genes annotated as distinct but sharing the same number in the table (#) were later found to be single genes. *gene excluded because it showed reversal of the fold change when mapping to the *H. cydno* genome.

#	Gene name (Hmel2.5)	Annotated function	60h APF	156h APF	chromo some	Start (bp) position	End (bp) position
1	HMEL009992g4	*Ionotropic glutamate receptor*	✓	✗	18	334856	345144
*	HMEL034187g1	*no match*	✗	✓	18	314734	315069
2	HMEL014931g1	Serine endopeptidase inhibitor	✓	✓	18	3260635	3264491
3	HMEL034294g1	Reverse transcriptase domain	✓	✓	18	4369133	4370180
4	HMEL002560g1	CUB domain	✓	✗	18	4679193	4684506
5	HMEL034304g1	*Ionotropic glutamate receptor*	✓	✗	18	4935938	4944768
6	HMEL010030g2	*Methyltransferase*	✓	✗	18	5503482	5505157
6	HMEL010030g3	no match	✓	✗	18	5506399	5508526
7	HMEL010030g1	no match	✓	✓	18	5496842	5498456
8	HMEL015745g1	*Major facilitator superfamily*	✗	✓	18	5976551	5982484
9	HMEL014795g1	*Major facilitator superfamily*	✗	✓	18	10194935	10200873
10	HMEL015842g1	*Alpha crystallin/heat shock protein*	✗	✓	18	13546671	13547216
11	HMEL030024g1	*Ribonuclease H superfamily*	✗	✓	1	1009751	1013056

Table S4. Heterozygosity on the *Z*-chromosome. Heterozygosity is calculated as proportion of variants (SNPs and indels) which are heterozygous, in each sample, rounded at the second decimal place (note that variant sites were inferred having mapped to the *H. melpomene* genome).

H. melpomene				***H. cydno***			
Males		Females		Males		Females	
ID	Het.	**ID**	Het.	**ID**	Het.	**ID**	Het.
Adults							
45	0.49	**53**	0.04	**57**	0.23	**50**	0.02
47	0.46	**78**	0.05	**82**	0.23	**51**	0.02
70	0.47	**80**	0.04	**98**	0.25	**58**	0.02
71	0.48	**128**	0.04	**99**	0.24	**67**	0.02
83	0.46	**218**	0.05			**68**	0.01
100	0.49					**81**	0.02
104	0.46					**84**	0.02
156h APF							
5	0.47	**6**	0.05	**4**	0.26	**13**	0.02
14	0.49	**18**	0.05	**8**	0.26	**21**	0.04
17	0.49	**24**	0.05	**142**	0.30	**30**	0.03
184	0.50	**150**	0.05	**151**	0.29	**137**	0.04
		220	0.06	**156**	0.29	**168**	0.04
60h APF							
92	0.48	**87**	0.06	**85**	0.26	**90**	0.02
97	0.49	**95**	0.05	**86**	0.27	**118**	0.02
115	0.47	**117**	0.05	**119**	0.27	**125**	0.02
		124	0.04	**146**	0.26	**144**	0.02
		149	0.06			**162**	0.02
		164	0.05			**200**	0.02
		208	0.06				

Hybrids							
Males		Females		Males		Females	
ID	Het.	**ID**	Het.	**ID**	Het.	**ID**	Het.
F1 hybrids (adults)							
42	0.64	**56**	0.05				
49	0.64	**69**	0.04				
Introgression line -156h APF				**Introgression line – 60h APF**			
105	0.26	**108**	0.02	**152**	0.27	**161**	0.02
116	0.27	**123**	0.02	**193**	0.27	**165**	0.02
126	0.27	**136**	0.02	**198**	0.25	**179**	0.02
131	0.27	**139**	0.02	**199**	0.27	**187**	0.02
133	0.26	**140**	0.02	**201**	0.26	**188**	0.02
154	0.26	**166**	0.02	**212**	0.28	**192**	0.02
155	0.26			**215**	0.27	**197**	0.02
183	0.27					**209**	0.03
185	0.27					**214**	0.02
189	0.27					**224**	0.02

1256 **Supplementary References**

1257

1258 1. Darragh, K. *et al.* A novel terpene synthase produces an anti-aphrodisiac pheromone in
1259 the butterfly *Heliconius melpomene*. *bioRxiv*, 779678 (2019).\
1260 2. Gaffney, D. J. Mapping and predicting gene–enhancer interactions. *Nat. Gen.* ***51***,
1261 1662-1663 (2019).
1262
1263

1264