Neural processing genes are linked to divergence in visual mate preference in sympatric *Heliconius* butterflies

Matteo Rossi[a,b], Timothy J. Thurman[b,c], Alexander E. Hausmann[a], Stephen H. Montgomery[d], Riccardo Papa[b,e,f], Chris D. Jiggins[g] , W. Owen McMillan[b] & Richard M. Merrill[a,b]

[a] Division of Evolutionary Biology, Ludwig Maximilian University of Munich, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany

[b] Smithsonian Tropical Research Institute, Gamboa, 0843-03092, Panama

[c] Division of Biological Sciences, University of Montana, 32 Campus Dr., Missoula, MT 59812, US

[d] School of Biological Sciences, University of Bristol, 24 Tyndall Ave, Bristol BS8 1TQ, UK

[e] Department of Biology, University of Puerto Rico, Rio Piedras, San Juan 00925, Puerto Rico

[f] Molecular Sciences and Research Center, University of Puerto Rico, San Juan 00926, Puerto Rico

[g] Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Corresponding author: merrill@bio.lmu.de

Short title: Genetics of visual mate preference in *Heliconius*

Abstract

Divergence in mating behaviors plays a major role during speciation, but we know little about the genetic mechanisms underlying the evolution of such traits. Reproductive isolation between the Neotropical butterflies *Heliconius melpomene* and *H. cydno* depends on shifts in mimetic warning patterns, which also act as mating cues. Preferences for conspecific warning patterns have a strong genetic component, but unlike the warning pattern cues, the exact genes responsible remain unknown. Here, we reduce the number of candidate genes associated with a large effect QTL for changes in visual mate preference from 200 to just 6. First, we confirm that genomic regions previously associated with male courtship modulate preference behaviors, rather than other traits that may influence courtship time. We then investigate gene expression in the brains of *H. melpomene*, *H. cydno*, and their hybrids, across development. We identify two genes, *regucalcin2* and an ionotropic glutamate receptor, that are differentially expressed both in species and hybrid comparisons, in a manner consistent with known patterns of dominance. We also find substitutions in the protein-coding regions of 4 genes predicted to affect protein function. These six candidates are located in regions resistant to interspecific gene flow, as expected for genes that contribute to reproductive isolation. Our candidates have been implicated in key components of neural signaling, suggesting shifts in visual mating preference do not rely on primary photoreceptor genes, but instead involve changes in visual integration or processing. Such genetic changes would allow mate preference evolution without altering perception of the wider environment.

Keywords: speciation | behavioral isolation | mate preferences | behavioral genetics

Significance statement

Many species remain separate not because they fail to produce viable hybrids but because they “choose” not to mate. Although these changes in behavior can be hardwired in the genome, we know little about the specific genes involved. *Heliconius* butterflies display a striking diversity of warning color patterns, which they use as cues to recognize conspecifics. We analyze a genomic region associated with changes in these visual behaviors in two *Heliconius* species. We couple quantitative trait locus (QTL) studies with gene expression and population genomic analyses to identify candidate genes associated with these visual preferences. This is an important step towards understanding how behavioral differences, crucial to speciation, are generated both during development and across evolutionary time.

Introduction

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 behaviors between species, and will ultimately have a genetic basis. However, although the significance of behavioral barriers for speciation has been recognized since the Modern Synthesis[3], we still know little of the genetic mechanisms underlying changes in mating preferences, or variation in behaviors across natural populations more broadly[4,5]. Identifying the genes underlying species-specific mating preferences provides an important route towards understanding how behavioral differences are generated, both during development and across evolutionary time.

Previous genetic studies of isolating preference behaviors have largely been limited to the identification of causal genomic regions, which almost invariably contain many genes[6,7,8,9]. Only a handful of studies have identified likely candidate genes that contribute to behavioral preferences causing reproductive isolation between species. These are largely limited to studies of chemosensory-guided mating preferences[10,11,12,13], which have identified changes at chemoreceptor genes as the underlying evolutionary mechanisms. To our knowledge, only two studies – in incipient fish species – have identified candidates for visual mate preference evolution, albeit indirectly, both suggesting a role for sensory perception mediated by changes in the peripheral visual system[14,15]. However, whether or not visual mate preferences generally involve shifts at the sensory periphery, or shifts in downstream processing, is not known.

The Neotropical *Heliconius* butterflies are well known for their diversity of warning color patterns. These patterns are often associated with Müllerian mimicry, where distantly related species converge on the same warning signals[16,17]. However, within a community, closely related species often look very different. Not only are warning patterns under disruptive selection due to increased predation of intermediate hybrid phenotypes[18], but these divergent patterns are also used as mating cues. In particular, males almost invariably prefer to court and attempt to mate live females, and artificial models, that share their own warning pattern[19]. This leads to varying degrees of premating isolation between taxa with divergent warning patterns[20,21].

The closely related species *Heliconius melpomene* and *H. cydno* differ in warning patterns, and although male hybrids are fertile, the two species remain separate largely due to strong assortative mating[22]. In central Panama, *Heliconius melpomene rosina* shares the black, red and yellow pattern of its local *Heliconius erato* co-mimic. In contrast, *Heliconius*

106 *cydno chioneus* mimics the black and white patterns of *H. sapho*. Differences in warning
107 pattern between *melpomene* and *cydno* are largely due to expression differences in just three
108 genes, specifically *optix*[23], *WntA*[24] and *cortex*[25]. Visual preferences for these divergent
109 patterns are particularly apparent in males, which strongly prefer to court conspecific
110 females[26,27,28].

111 Quantitative trait locus (QTL) mapping of *H. melpomene* and *H. cydno* has revealed
112 three genomic regions of major effect that influence the relative time males spend courting
113 red *melpomene* or white *cydno* females[28]. Notably, the best supported QTL was in the same
114 genomic region (within 4.8cM) as *optix*, the gene responsible for presence or absence of the
115 red color pattern elements in *Heliconius*[23]. Genetic linkage between ecological and mating
116 traits will facilitate speciation by acting against breakdown of these genetic associations by
117 recombination[29]. Nevertheless, this QTL, and its associated candidate region, contain
118 hundreds of genes, and the exact genes responsible for differences in preference behavior
119 remain unknown.

120 To identify specific genes that may underlie behavioral differences, we first confirm
121 that the QTLs identified previously are associated with variation in male courtship initiation
122 behaviors. We next identified genes within the major behavioral QTL, which were i)
123 differentially expressed in the neural tissue (central brain, optic lobes and ommatidia) of *H.*
124 *melpomene* and *H. cydno*, or ii) have protein coding changes predicted to alter protein
125 function. Out of 200 genes within the QTL region, we identify just 6 candidates likely to
126 underlie assortative mating behaviors.

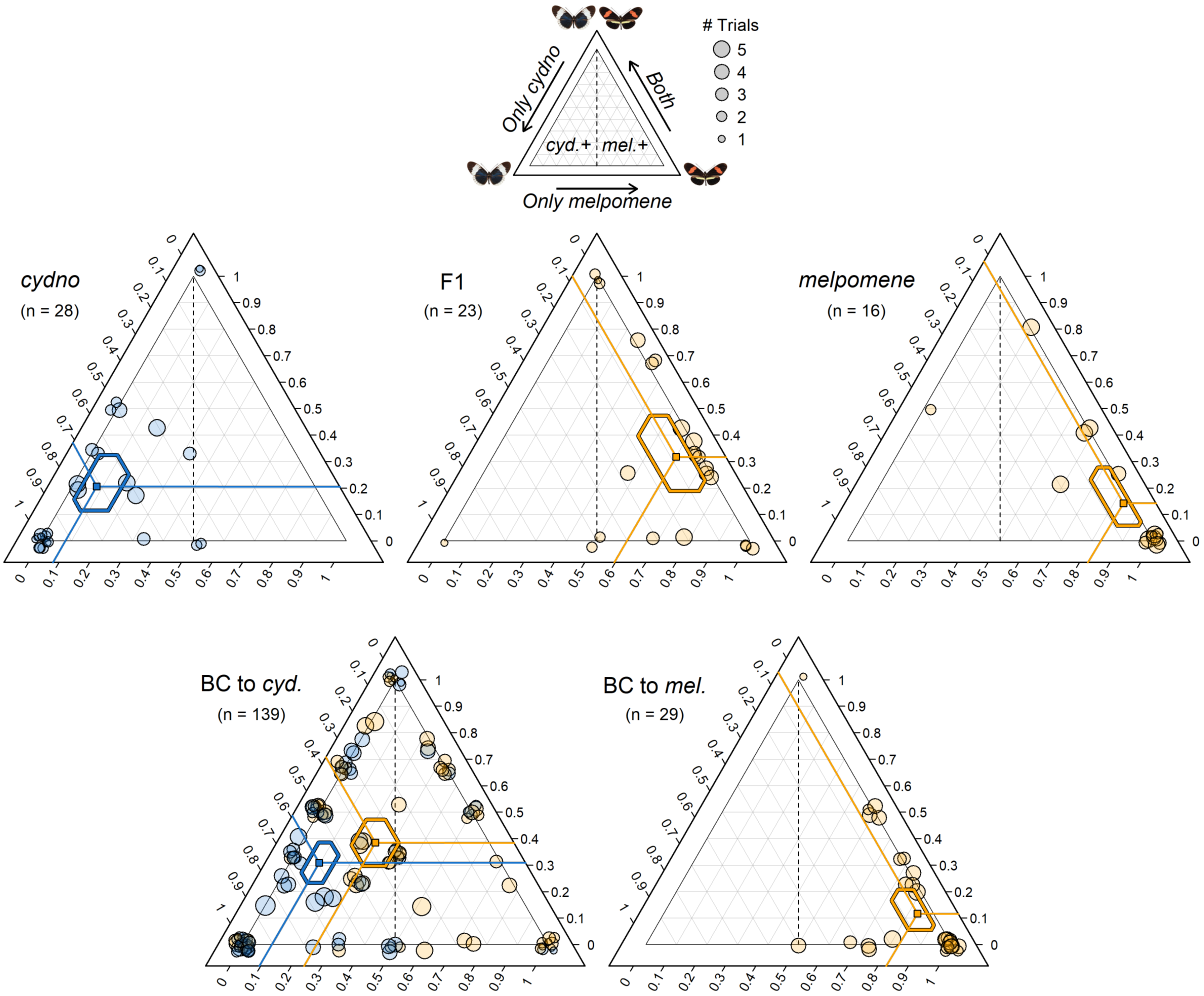
127

128 **Results**

129 **Chromosome 18 is associated with differences in courtship initiation.**

130 Our previous results reveal that QTLs on chromosomes 1, 17 and 18 influence the relative
131 time hybrid males spend courting red *melpomene* or white *cydno* females[28]. However, the
132 time males spend courting a particular female might depend not only on male attraction, but
133 also on the female's response (and in turn his response to her behavior). To confirm that these
134 previously reported QTLs influence male approach *behaviors* (as opposed to other traits that
135 may influence courtship, for example male morphology[30]), we reanalyzed our previous data,
136 this time explicitly considering whether males initiated courtship towards *melpomene*, *cydno*
137 or both types of female during choice trials. Consistent with our previous analyses[28], we
138 found that F1 and backcross-to-*melpomene* males initiated courtship more frequently with
139 *melpomene* females, whereas courtship initiation behaviors segregate in the backcrosses to

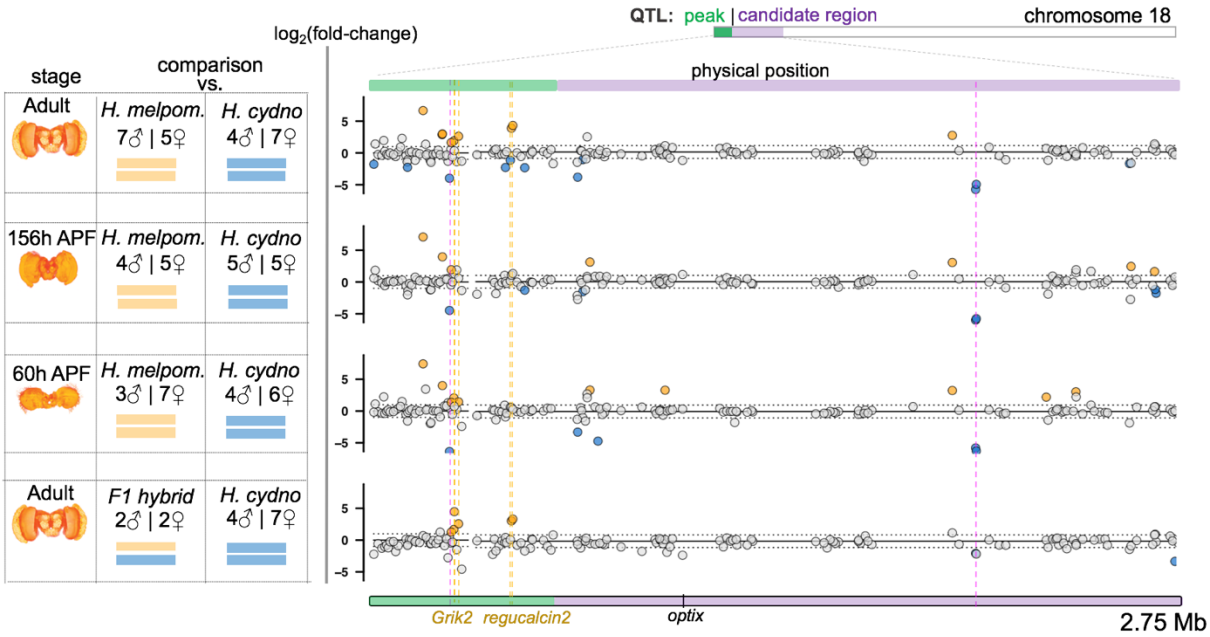
Figure 1. Genotype at the preference QTL on chromosome 18 influences courtship initiation. Ternary plots showing the number of 15-minute choice trials in which courtship was initiated towards *melpomene, cydno* or both females for different male types. Left ternary axis shows proportion of trials where courtship was initiated towards *H. cydno* female only, bottom axis towards *H. melpomene* female only, and right axis towards both female species. Orange points represent individuals that have inherited at least one *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 *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. 95% credibility intervals (CrIs) for all three proportions are shown as hexagons around predictors with lines projecting to corresponding values on the three axes.



cydno (**Figure 1**). Notably, backcrosses-to-*cydno* males heterozygous at the QTL on chromosome 18 (*i.e.* with a *melpomene* allele derived from the F1 father) initiated courtship towards *melpomene* females more frequently than males homozygous for the *cydno* allele (**Figure 1**, bottom left; n = 139, ΔELPD: -10.9 (S.E.±5.1), *i.e.* a change of 2.14 SE units). Together with previous evidence that male hybrids bearing *melpomene* alleles at *optix* prefer to court the artificial models of *melpomene* females over those of *cydno*[27], these results suggest that the QTL on chromosome 18 harbors genes for visual attraction behaviors towards females with the red pattern. Consequently, we focused our subsequent analyses on this QTL

161 on chromosome 18 (and also because tight linkage of *optix* allowed us to track the alleles at
162 preference-color locus in hybrid crosses). The QTL on chromosome 1 was also retained in our
163 model of initiation behaviors (**Supporting figure 1;** n = 139, ΔELPD = -13.6 (SE±5.7), *i.e.* a
164 change of 2.34 SE units), in contrast to the QTL on chromosome 17 which was not retained (n
165 = 139, ΔELPD = -2.1 (S.E.±3.0)). Results for the QTL on chromosome 1 are reported in the
166 supporting materials (**Supporting table 1**).
167
168 **27 *genes* within the major QTL are differentially expressed in the brains and eyes of *H.***
169 ***cydno* and *H. melpomene*.**
170 We hypothesized that changes in gene regulation that determine differences in visual mate
171 preference behaviors might occur during pupal development (for instance, during visual
172 circuit assembly) or in the imago, and must involve changes in the peripheral and/or central
173 nervous system[31]. Therefore, we generated RNA-seq libraries for combined eye and brain
174 tissue, across two pupal stages (around the time of ommochrome pigment deposit and half-
175 way through pupal development) and one adult stage, for *H. melpomene* and *H. cydno* and
176 compared their gene expression levels. We found considerable differential expression at the
177 QTL on chromosome 18 (the QTL spans 2.75 Mb, and contains 200 genes). We identified 27
178 genes within the QTL region that show differential expression between *melpomene* and
179 *cydno*, in at least one of the three developmental stages, and with a minimum fold change
180 threshold of 2. The same genes were frequently differentially expressed across development
181 (**Supporting table 1**), with 11 genes being differentially expressed in more than one stage.
182 Differentially expressed genes were mostly located within the QTL peak (*i.e.* the genomic
183 region with strongest statistical association with male preference) or in close proximity to
184 *optix* (**Figure 2**).
185 The genomic region between the start of chromosome 18 and *optix* (comprising the
186 QTL peak) is highly divergent between *melpomene* and *cydno*[32], and divergent coding
187 sequences within this region could also introduce mapping biases of RNA-seq reads. To
188 account for this, we repeated the analysis having mapped to both the *H. melpomene* reference
189 genome[33] and to a *H. cydno* genome[34]. Generally, we found similar patterns of differential
190 expression when mapping to the *H. cydno* genome (**Supporting figure 2, Supporting table**
191 **2**). Nevertheless, in subsequent analyses we excluded two genes, HMEL034187g1 and
192 HMEL034229g1, which showed reversal of the fold change or did not show differential
193 expression when mapping to the *H. cydno* genome respectively.
194

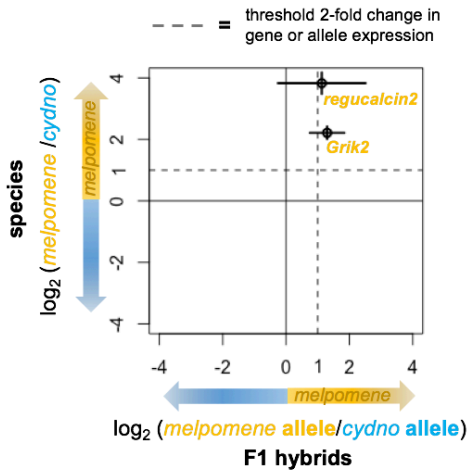
Figure 2. Differential expression at the preference QTL region on chromosome 18. Left: Summary of the comparative transcriptomic analyses with stage, number of samples and chromosome 18 composition. Right: the corresponding results, zooming in on the QTL region on chromosome 18. The *x*-axis represents 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 *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 *melpomene* and *cydno* AND between *cyd/melp* vs *cyd/cyd* individuals, at the same stage. 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).



A *regucalcin* and an *ionotropic glutamate receptor* are upregulated in both *H. melpomene* and F1 hybrid males.

Our previous behavioral experiments suggest that the alleles for the *melpomene* behavior are dominant over the *cydno* alleles[27,28] (**Figure 1**). Given this pattern of dominance, we predicted that genes underlying variation in male preference to be up- or down- regulated in the brains of both *melpomene* and first generation (F1) hybrid males, with respect to *cydno*. Of the putative genes differentially expressed between *cydno* and *melpomene* reported above, only four, within the QTL candidate region, were differentially expressed between the F1 hybrids and *cydno* (**Figure 2**). These included two *regucalcins* (also called *senescence marker proteins-30*: HMEL013552g1, HMEL034199g1), an *ionotropic glutamate receptor* (HMEL009992g4), which is a putative ortholog of *Grik2*, and one gene with no annotated

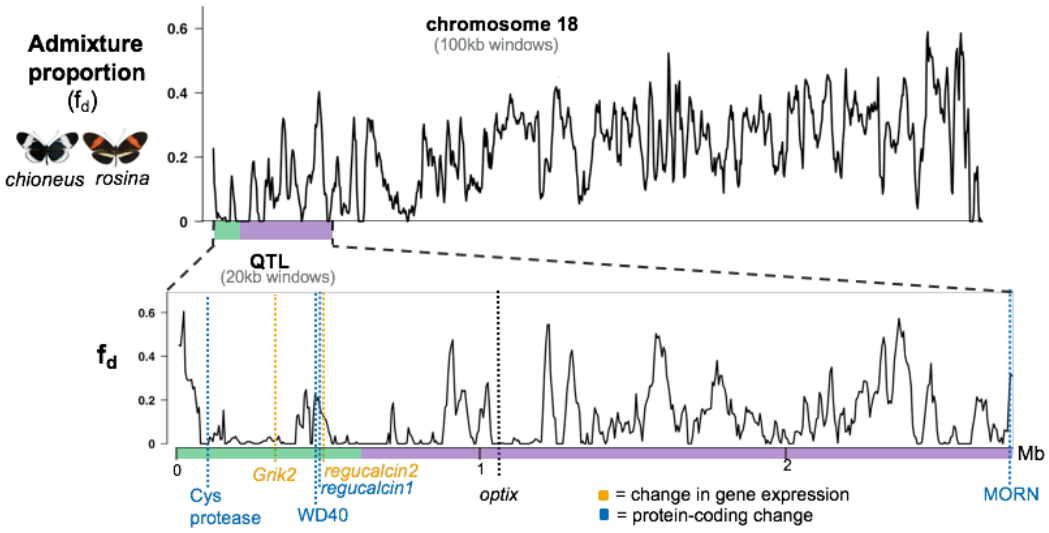
222 function (HMEL009992g1). We obtained the same results regardless of whether we
223 considered both males and females together, or males alone.
224 Further inspection of spliced mRNA-reads indicated that the two annotated
225 *regucalcins* were in fact a single gene (from now on referred to as *regucalcin2*). This was also
226 the case for the *ionotropic glutamate receptor* and the gene with no annotated function (from
227 now on referred to as *Grik2*). To ensure that apparent fragmentation of a few gene models in
228 the *H. melpomene* (Hmel2.5) annotation[35], in which exons belonging to the same isoform are
229 incorrectly considered distinct genes, did not introduce inaccuracies in estimates of
230 differential gene expression, we produced a new transcript-based annotation of the
231 *melpomene* genome using Cufflinks's RABT[36]. Repeating all comparative transcriptomic
232 analyses using this new annotation (where exons were correctly considered as part of single
233 genes), we confirmed that both *regucalcin2* and *Grik2* were differentially expressed in both
234 species and hybrids comparisons.
235
236 **Figure 3. *Grik2* and *Regucalcin2* show evidence of allele specific expression.** Points indicate the value, and bars
237 the standard error, of the (base 2) logarithmic fold change in expression between parental species (horizontal) and
238 the alleles in F1 hybrids (vertical), for candidate genes (as defined in the transcript-guided annotation). Dashed
239 lines indicate the threshold for a 2-fold change in expression for the genes in the species (horizontal), and for the
240 alleles in the hybrids (vertical).
241



242
243
244 **Differential expression of *Grik2* in adults is likely due to cis-regulatory effects.**
245 To determine whether differences in gene expression levels between parental species were
246 due to *cis*- or *trans*-regulatory changes, we conducted allele specific expression (ASE)
247 analyses in adult F1 hybrids. In F1 hybrids, both parental alleles are exposed to the same
248 *trans*-environment, and consequently *trans*-acting factors will act on alleles derived from each
249 species equally (unless there is a change in the cis-regulatory regions of the respective

250 alleles). Therefore, differences in allele specific expression should indicate changes in *cis*-
251 regulatory regions[37]. Both candidate genes (*Grik2* and *regucalcin2*) had a very low number of
252 SNPs that could differentiate the *melpomene* and *cydno* allele (using both gene models of the
253 Hmel2.5 annotation and RABT annotation) and few reads mapped to these SNPs.
254 Nevertheless, for *Grik2,* the *melpomene* allele was significantly more highly expressed
255 relative to the *cydno* allele (p=0.017, Wald test), suggesting cis-regulatory effects (**Figure 3**).
256 For *regucalcin2,* although there was a tendency towards up-regulation of the *melpomene*
257 allele, consistent with *cis*-regulation, we did not have sufficient power-to rule out *trans*-only
258 regulatory effects (**Figure 3;** p=0.108, Wald test).
259
260 ***Grik2* is differentially expressed in hybrids that essentially differ only for allelic**
261 **composition at the behavioral QTL region.**
262 In order to study the specific effects that *melpomene* derived alleles at the QTL on
263 chromosome 18 had on gene-expression, we introgressed this region into a *cydno* background
264 through multiple backcrosses (crossing design in **Supporting figure 3**). We wanted to
265 investigate whether differences at this QTL regulated expression of any specific genetic
266 pathway during development, and more generally what changes in genome-wide transcription
267 were observed in hybrids differing (mostly) just at this QTL region.
268 Notably, in these third-generation backcross hybrid comparisons, across the entire
269 genome only 23 and 29 genes were differentially expressed (at 156h after pupal formation
270 (APF) and at 60hAPF, respectively). Of these, 20 and 19 genes (at 156hAPF and at 60hAPF,
271 respectively) were located on chromosome 18, indicating that gene expression differences in
272 these comparisons were mostly restricted to the preference-color region on chromosome 18,
273 segregating for *cyd/melp* or *cyd/cyd* alleles. No genetic pathway was enriched for gene
274 expression differences between these hybrids at either pupal stage (PANTHER enrichment
275 test[38]), suggesting that overall this QTL harbors a few, modular changes in gene regulation in
276 the developing brain/eyes of *cydno* and *melpomene***.** *Grik2* was the only gene detected as
277 differentially expressed between species and hybrids at these pupal stages (**Supporting figure**
278 **4**).
279 To verify that differential expression of candidate genes at the QTL region is driven
280 by *melpomene* alleles on chromosome 18 and not by other *melpomene* alleles at *trans*-acting
281 genes on other chromosomes, we compared gene expression levels between hybrids carrying
282 *cyd/melp* vs. *cyd/cyd* regions on chromosomes chr1, chr4 and chr15, chr20 (**Supporting**

283 **figure 5A**). In these comparisons, there was no signal of differential expression on
284 chromosome 18**.** This supports the *cis*-regulatory activity of the *melpomene* allele of
285 candidate genes on chromosome 18. To test this further, we conducted another allele specific
286 expression study in the BC3 hybrids, which suggested *trans*-regulatory effects for *Grik2* at
287 these pupal stages, but were less conclusive with regard to *regucalcin2* (**Supporting figure**
288 **6**). Since causal gene/s might exert an effect on behavior due to their action during
289 development or in adult form, and this action might in turn be differently (cis- vs trans-)
290 regulated, we still considered both genes as strong candidates.
291
292 **Figure 4**. **Gene flow at the QTL region for behavior.** Admixture proportion (f_d) values estimated in overlapping
293 100kb (top) and 20kb (bottom) windows for chromosome 18 (top) and the QTL region (bottom) between *H.*
294 *melpomene rosin*a and *H. cydno chioneus*, with candidate genes positions highlighted by a vertical dashed line,
295 and *optix* location displayed for reference. The *x*-axis represents physical position, the *y*-axis indicates the f_d value.
296 f_d values close to zero indicate that the proportion of shared derived alleles, and consequently gene flow, between
297 *melpomene* and *cydno* is small (or zero), implying localized selection against foreign alleles that introgress between
298 the two species.
299



300
301
302 **4 genes with protein-coding substitutions within the QTL candidate region have**
303 **predicted effects on protein function**.
304 Because shifts in behavioral phenotypes could be due to changes in protein-coding regions,
305 we additionally considered protein-coding substitutions between *melpomene* and *cydno.*
306 Overall, we found 152 protein-coding substitutions, spanning 54 of the 200 genes across the
307 entire QTL candidate region. We then studied whether these variants were predicted to have
308 non-neutral effects on protein function with PROVEAN[39]. The PROVEAN algorithm predicts
309 the functional effect of protein sequence variations based on how they affect alignments to
310 different homologous protein sequences**.** We found 4 genes with such predicted effects

311 (PROVEAN score < -2.5): Specifically, a *WD40*-repeat domain containing protein
312 (HMEL013551g3), a *cysteine protease* (HMEL009684g2), a *MORN* motif containing protein
313 (HMEL006660g1), and another *regucalcin* (HMEL013551g4) adjacent to, but distinct from,
314 that found to be differentially expressed above (from now on referred to as *regucalcin1*).
315
316 **Candidate genes occur in regions with reduced gene flow.**
317 Of our six candidate genes for preference behaviors that contribute to reproductive isolation
318 between *H. cydno* and *H. melpomene* (*regucalcin2*, *Grik2* and the four genes with protein
319 coding modifications), five are found within the QTL peak (**Figure 4**). Genetic changes
320 causing reproductive isolation between populations are expected to reduce localized gene
321 flow in their genomes. Therefore, we compared the position of our candidate genes to
322 estimated levels of admixture proportions (f_d)[40] between *H. melpomene* and *H. cydno* across
323 the QTL candidate region[41]. We found that all candidate genes were located in genomic
324 regions with low f_d values (**Figure 4**), suggesting localized resistance to gene flow between
325 *melpomene* and *cydno* at these genes and their putative *cis*-regulatory regions.
326
327 **Discussion**
328 Behavioral isolation is frequently implicated in the formation of new species, and involves the
329 correlated evolution of both mating cues and mating preference. Here we have analyzed a
330 genomic region in a pair of closely related sympatric butterflies, *H. cydno* and *H. melpomene*,
331 that contains genes for divergence in both an ecologically relevant mating cue and the
332 corresponding preference. Physical linkage between ecological and mating traits will facilitate
333 speciation by allowing different barriers to act in concert to restrict gene flow[42,43]. Although
334 the genes underlying changes in the warning pattern cue in *Heliconius* are well characterized
335 (e.g. *optix*), those underlying the corresponding shift in behavior have not been identified. We
336 have pinpointed a small number of genes that fall within the QTL peak, which show either
337 expression (*regucalcin2* and *Grik2*) or protein coding differences (HMEL013551g3,
338 HMEL009684g2, HMEL006660g1, and *regucalcin1*) and fall within a region of reduced
339 admixture, that are strong candidates for modulating mating behavior.
340 Two broad neural mechanisms could underlie the evolution of divergent visual
341 preferences, involving changes in either i) detection at the sensory periphery or ii) the
342 processing and/or integration of visual information. Although *H. melpomene* and *H. cydno*
343 have the same retinal mosaics/class of photoreceptors[44], spectral sensitivity in *Heliconius* eyes
344 could be altered by filtering pigments[45], or other physiological processes taking place at the

345 photoreceptors/sensory periphery, eventually shifting sensitivity towards different
346 wavelengths (and possibly color patterns). It has previously been hypothesized that the gene
347 regulatory networks for ommochrome deposition in the *Heliconius* eyes might have been co-
348 opted in the wings[46], where *optix* plays a central role, and therefore that *optix* might play a
349 role in eye pigmentation in *Heliconius.* However, the protein product of *optix* has not been
350 detected in pupal or adult retinas of various *Heliconius* species tested[47], and therefore has no
351 obvious link to ommochrome deposition in the eyes. More generally, the underlying
352 evolutionary mechanism is unlikely to involve detection at photoreceptors, as this would
353 probably have a broad effect on downstream processing[2] and alter the visual perception of the
354 animal's wider environment.

355 The second mechanism, involving changes in the processing, and/or integration, of
356 visual information, could act through an alteration of neuronal activity or connectivity. For
357 instance, different levels of gene expression in conserved neural circuits between *melpomene*
358 and *cydno* may affect overall synaptic weighting and determine whether a signal (e.g. color
359 and motion) elicits a motor pattern (response towards a female) or not. Consistent with this
360 scenario, the composition of ionotropic receptors at post-synapses is a key modulator of
361 synaptic transmission[48], implicating *Grik2*. Interestingly, differential expression of ionotropic
362 glutamate receptors is also associated with variation in social and aggressive behaviors in
363 vertebrates[49,50]. *Regucalcins* are involved in calcium signaling[51], which regulates synaptic
364 excitability and plasticity[52], and has an important role in axon guidance[53] (albeit alongside
365 additional roles across a broad range of biological processes), making the two *regucalcins* we
366 identify strong candidates for behavior.

367 Changes in the regulation of genes with pleiotropic effects are likely to be less
368 detrimental compared to changes in their protein-coding sequences[54] (although emerging
369 evidence has begun to suggest that enhancer/repressor elements may be more pleiotropic than
370 previously thought[55,56]). Furthermore, there is considerable evolutionary potential in the co-
371 option of transcription factors/networks[54] that regulate neural patterning or neuron-type
372 activity, possibly resulting in novel adaptive expression patterns. In line with this,
373 *Regucalcin2* and *Grik2,* which are differentially expressed in the eyes and brain in both our
374 species and hybrid comparisons, are likely to be involved in multi-functional processes, such
375 as calcium signaling and ion transport, and likely have pleiotropic alleles. We also found
376 evidence of cis-regulatory effects for both genes (albeit not significant for *regucalcin2*),
377 which would be required of the causal genetic change within the QTL, if it were to be in gene
378 regulation.

379 Despite expectations that non-coding, regulatory loci may provide a flexible route to
380 divergent mating preferences, we also found substitutions in coding regions at the QTL,
381 which are predicted to have an effect on protein functioning and therefore remain strong
382 candidates. These genes include *regucalcin1*, which is distinct from, but located next to,
383 *regucalcin2* (which is differentially expressed). Notably, the eye transcript of *regucalcin1* was
384 recently characterized as fast-evolving across *Heliconius* species[57]. Other candidates include a
385 *cysteine protease*, which functions in protein degradation, and might be linked to behavior for
386 example through degradation of neurotransmitters, a *MORN* motif containing protein
387 (function unknown), and a *WD40* containing protein. *WD-repeat* containing proteins have
388 been implicated in a wide array of functions ranging from signal transduction to apoptosis
389 (https://www.ebi.ac.uk/interpro).
390 Although preference for red coloration and the *optix* gene are tightly linked, we find no
391 evidence that *optix* is differentially expressed in the eyes or brains of our two species. It is
392 also not located within the QTL peak (and it contains no non-synonymous changes in protein
393 coding regions[23]). It seems unlikely therefore that changes in cue and preference are
394 pleiotropic effects of the same allele. More generally, although we have pinpointed the
395 strongest candidates yet identified for assortative mating behaviors in *Heliconius*, it is
396 possible that actual causal changes in gene regulation are restricted to developmental stages
397 other those sampled, or restricted to a few neuronal populations not detected with
398 transcriptomic data from eyes and whole brain tissue. Nonetheless, by sampling at two pupal
399 stages (around the time of *optix* expression/ommochrome pigment deposit in the wing/eye and
400 halfway through pupal development) and at the adult stage, we should have captured
401 important transitions for the behavioral programming of the two species.
402 Functional studies of candidate genes would of course be desirable in order to clarify
403 the link between individual genes and behavioral divergence. For example, can a distinct
404 genetic change at a single locus trigger a shift in behavior, or are multiple interacting loci
405 necessary to affect it? Although CRISPR-Cas9 technology has been successfully used to
406 produce knock-out individuals[54,58,59] for color patterning genes in *Heliconius*, "behavioral
407 genes" will present a greater challenge. Gene knock-outs could broadly impair an individual's
408 behavioral phenotype, permitting only tentative inference about the link between the action of
409 a gene and a specific behavior/perception. While posing a significant challenge, in the future,
410 this or other emerging technologies that allow switching of specific alleles will allow direct
411 functional experiments on behavioral phenotypes for the candidates we have identified.

412 Work in the past decade has shown that complex innate behavioral differences between
413 species can be encoded in relatively few genetic modules[60,61], but very few studies[62,63,64] have
414 identified specific genes underlying behavioral evolution. Although traditional laboratory
415 organisms continue to provide important insights into the evolution and genetics of behavior,
416 comparative approaches are required to determine if developmental principles can be broadly
417 applied, and also to incorporate a wider range of phenotypic variation. The challenge now is
418 to increase the resolution of studies in non-traditional systems, in order to link individual
419 genetic elements to behaviors, *and* the sensory and/or neurological structures through which
420 they are mediated. In this light, we have identified a small handful of strong candidate genes
421 associated with the evolution of visual mate preference behaviors in *Heliconius*. These genes
422 are in tight physical linkage with the locus for the corresponding shifts in an ecologically
423 relevant mating cue, providing an important opportunity to investigate the build-up of genetic
424 barriers crucial to speciation. The candidate genes identified seem more likely to alter visual
425 processing or integration, rather than detection at photoreceptors, consistent with permitting
426 changes in mate preference without altering perception of the animal's wider environment.
427

428 **Materials and Methods**

429 **Courtship initiation analyses.** Butterfly rearing, crossing design and genotyping are
430 described in detail elsewhere[28]. In brief, we assayed male preference behaviors for *H.*
431 melpomene, *H. cydno,* their first generation (F1) hybrids and backcross to hybrids to both
432 parental species in standardized choice trials. Males were introduced into outdoor
433 experimental cages (1x1x2m) with a virgin female of each species and courtship behaviors
434 recorded. Whenever possible, trials were repeated for each male (median = 5 trials). To
435 determine whether previously identified QTLs for courtship time contribute to variation in
436 courtship *initiation* behaviors, we performed a *post-hoc* analysis using categorical models in a
437 Bayesian framework with a multinomial error structure, using the R package *brms*[65]. In
438 contrast to our previous analysis[28], in which we considered the number of minutes (*i.e.* time)
439 for which courtship was directed towards *H. cydno* or *H. melpomene* females, here the
440 response variable was number of trials in which male courtship was *initiated* towards *H.*
441 *cydno* females only, *H. melpomene females* only*,* or both female types (hereafter referred to as
442 "*initiation"*). Across males the median number of trials with a response was 3. Using
443 backcross-to-*cydno* males only, we fitted *initiation* as a response variable to genotype
444 (*cyd*/*cyd* or *cyd*/*melp*) at each QTL, which were included as separate fixed effects. Individual
445 ID was fitted as random factor. To test the effect of each QTL on male *initiation*, we

446 compared the saturated model incorporating all three QTL with reduced models excluding
447 each QTL in turn, using approximate leave-one-out (LOO) cross-validation as implemented in
448 *brms*[66], and based on expected log pointwise predictive density (ELPD). Normal distribution
449 of ELPD can be a straightforward approximation given our large samples sizes (n=139)[66].
450 Therefore, we considered an absolute value of ELPD greater than 1.96 units of its standard
451 error as indicative of the reduced model being less-informative than the saturated model (95%
452 confidence). Males that did not initiate courtship to any female across trials were excluded
453 from analyses, resulting in a dataset of 139 males, from a total of 146 backcross males for
454 which we had genotype data. Finally, we extracted predictors and credibility intervals for
455 backcross males with differing genotypes from the minimum adequate model. Credibility
456 intervals for *H. melpomene*, *H. cydno,* F1 hybrid and backcross to *melpomene* males
457 displayed in Figure 1 were generated following the same procedures. Raw data and analysis
458 code are available in the following github repository:
459 https://github.com/SpeciationBehaviour/neural_genes_heliconius.git

460 **Butterfly collection, rearing and crossing design for expression analyses.** Wild *H.*
461 *melpomene rosina* and *H. cydno chioneus* individuals were caught along Pipeline Road near
462 Gamboa, Panama, in the Soberania National Park, and used to establish stocks at the
463 Smithsonian Tropical Research Institute insectaries in Gamboa. Butterflies were reared in
464 common garden conditions, in 2x2x2m cages, and provided with fresh *Psiguria* flowers and
465 10% sugar solution. Larvae were reared on fresh *Passiflora* shoots/leaves until pupation. *H.*
466 cydno, *H. melpomene* and hybrid individuals used for RNA-seq (see below) were reared
467 concurrently and under the same conditions. F1 hybrids were obtained by crossing a wild-
468 caught *H. m. rosina* male to an insectary-bred virgin *H. c. chioneus* female.
469 The introgression line was generated by outcrossing a hybrid male with a red forewing
470 band (crossing design shown in **Supporting figure 3**) to virgin *H. cydno* females, over three
471 generations. The peak of the behavioral QTL reported previously[28] on chromosome 18 (at
472 0cM) is in very tight linkage with the *optix* color pattern locus (at 1.2cM), which controls for
473 the presence and absence of the red forewing band seen in *H. melpomene rosina*. Presence of
474 the red forewing band is dominant over its absence so that segregation of the red band can be
475 used to infer genotype at the *optix* locus. Specifically, hybrid individuals with a red forewing
476 band are heterozygotes for *H. melpomene/H. cydno* alleles at the *optix* locus, whereas
477 individuals lacking the red band are homozygous for the *cydno* allele. Due to the tight linkage
478 we expected little recombination between *optix* and QTL peak even after three generations of

479 introgression, allowing us to infer genotype at the preference-*optix* locus (which we
480 confirmed with genetic data, see below).

481 **Tissue dissection, RNA extraction and mRNA sequencing.** Eye (ommatidia and retinal
482 membrane) and brain tissue (central brain and optic lobes) were dissected out of the head
483 capsule in cold (4 ºC) 0.01M PBS solution, at two pupal stages: 60 hours after pupal
484 formation (60h APF) and 156h APF; and in adults aged 9 - 13 days. We sampled adults at
485 around 10 days of age because by this stage males are mature and frequently court females[17].
486 Adult males and females sampled were sexually naive. We decided to sample at 60h APF
487 because this is the developmental stage at which *optix* is expressed in the wing, so we
488 hypothesized that it might had also been when *optix* is expressed in the brain. We sampled at
489 156h APF as a putative stage halfway through pupal development, and at this stage most of
490 the major neural connections have just been established in the *Heliconius* brain (Stephen
491 Montgomery, unpublished data).
492 Tissues were stored in RNAlater at 4 ºC for 24 hours, and subsequently at -20 ºC, until RNA
493 extraction. Total RNA was extracted using TRIzol Reagent (Thermo Fisher, Waltham, MA,
494 USA) and a RNeasy Mini kit (Qiagen, Valencia, CA, USA). Samples were treated with
495 DNase I (Ambion, Darmstadt, Germany). Integrity of total RNA was checked either on an
496 agarose gel or using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA,
497 USA). RNA concentration was measured on a Nanodrop spectrophotomer. Illumina TruSeq
498 RNA-seq libraries were prepared and sequenced at Edinburgh Genomics (Edinburgh, UK)
499 with 100 bp paired-end reads. To avoid lane effects the distribution of the species samples
500 was randomized on the sequencing platform. More detailed information about individuals and
501 sequencing yields can be found in the Supporting file 1.
502
503 **RNA-seq read mapping and differential gene expression analyses.** After a quality control
504 of RNA-seq reads with FastQC, we trimmed adaptor and low-quality bases using TrimGalore
505 v.0.4.4 (https://www.bioinformatics.babraham.ac.uk/projects/). RNA-seq reads were mapped
506 to the *H. melpomene* 2.5 genome[33]/annotation[35] using STAR v.2.4.2a[67] in 2-pass mode. We
507 only kept reads that mapped in 'proper pairs' using Samtools[68]. The number of reads mapping
508 to each gene were estimated with HTseq v. 0.9.1[69] with model "union", thus excluding
509 ambiguously mapped reads. Differential gene expression analyses between species/hybrids
510 were conducted in DESeq2[70]. We considered only those genes showing a 2-fold change in
511 expression level, and at adjusted (false discovery rate 5%) p-values < 0.05, to be differentially

512 expressed, to exclude expression differences caused by known differences in brain
513 morphology[71] (Montgomery et al., in prep).
514
515 **Sexing pupae**. In all DESeq2 analyses, sex was included as a random factor. To sex pupae,
516 we first marked duplicate RNA mapped reads with Picard
517 (https://broadinstitute.github.io/picard/), and used GATK 3.8[72] to split uniquely mapped reads
518 into exon segments and trim sequences overhanging the intronic regions. We then used
519 Haplotype Caller on each individual, using calling and filtering parameters according to the
520 GATK Best Practices for variant calling on RNA-seq data. The sex of pupal samples was
521 inferred from the proportion of heterozygous (biallelic) SNPs using the R package SNPstats.
522 Males (ZZ) were expected to have >> 0% heterozygous sites, whereas females (ZW) to have
523 0%. Z-linked heterozygosity of the pupal samples (**Supporting table 3**) were in line with
524 expectations (either ~ 0 for females or an order of magnitude higher for males), and matched
525 heterozygosity of either adult males or females, for which the sex was determined from
526 external morphology.
527
528 **Inference of gene function and transcript-based annotation.** Biological functions of
529 annotated genes were inferred with InterProScan v5[73], using the corresponding Hmel2.5
530 predicted protein sequences. InterProScan uses different databases like InterPro, Pfam,
531 PANTHER, and others, to infer functional protein domains and motifs (based on homology).
532 To study whether specific biological functions were enriched among genes showing
533 differential expression among hybrid types, we conducted the PANTHER enrichment test[38]
534 (with Bonferroni correction for multiple testing) using *Drosophila melanogaster* as the
535 reference gene function database.
536 Upon detailed inspection of the mapping coverage of spliced RNA-seq reads to the
537 Hmel2.5 gene annotation, we noticed that some gene models were fragmented, namely, a few
538 exons that appeared to be spliced together were incorrectly considered distinct genes. To
539 check that this did not introduced inaccuracies in our differential gene expression analyses,
540 we re-annotated the *melpomene* genome using the Cufflinks reference annotation-based
541 transcript (RABT) assembly tool[36]. We used the transcriptomic data from both *melpomene*
542 and *cydno* to reannotate the *melpomene* genome, separately for every developmental stage,
543 and reconducted the differential gene expression analyses in DESeq2 as described above.
544

545 **Inference of BC3 hybrids genome composition.** In order to perform comparative
546 transcriptomic analyses between third-generation backcross hybrids (BC3) segregating at the
547 QTL on chromosome 18 (crossing design in **Supporting figure 3**), we first determined which
548 genomic regions in these hybrids were heterozygous (*cyd/melp*) or homozygous (*cyd*/*cyd*).
549 For this, we inferred variants from RNA-seq reads for each BC3 hybrid (individually as
550 above), and from the combined *melpomene* and *cydno* samples. For the species, we used
551 HaplotypeCaller[72] on RNA-seq samples from all developmental stages of either species, to
552 produce individual genomic records (gVCF), and then jointly genotyped *melpomene* and
553 *cydno* gVCFs (separately for the two species) using genotypeGVCFs with default parameters.
554 Genotype calls were filtered for quality by depth (QD) > 2, strand bias (FS) < 30 and allele
555 depth (DP) > 4. For further analyses we kept *biallelic* genotypes only. We then used the
556 *intersect* function of *bcftools*[68] to infer variants exclusive to the *cydno* and to the *melpomene*
557 samples.
558 We calculated the fraction of variants that each BC3 hybrid individual shared with the
559 *melpomene* and with the *cydno* samples, in non-overlapping 100kb windows. We compared
560 these to the fraction of variants that a F1 hybrid and a *H. cydno* individual (not included in the
561 combined genotyping of the *cydno* samples), shared with the same species samples, and found
562 that they matched either one of them, indicating heterozygous (*cyd/melp*) or homozygous
563 (*cyd*/*cyd*) regions (**Supporting figure 5B**). In this analysis, we considered only those 100kb
564 windows where BC3 hybrids/F1 hybrid/*H. cydno* individuals shared more than 30 variants
565 with the *melpomene*/*cydno* samples.
566 To corroborate our findings, we repeated the same type of analysis, this time calling
567 species-specific variants for *melpomene* and *cydno* using 10 *H. melpomene rosina* and 10 *H.*
568 *cydno chioneus* genome resequencing samples[41]. For this, genomic reads were mapped to the
569 *melpomene* genome with BWA v0.7.2[74] using default parameters, genotypes were called with
570 GATK's HaplotypeCaller and GenotypeGVCFs with default parameters, and we considered
571 only *biallelic* genotype calls that had 10 < DP < 100 and genotype quality (GQ) > 30. With
572 this analysis we found the same heterozygous and homozygous regions in BC3 hybrids**.**
573 The size and number of the introgressed regions were in line with expectations about
574 3rd generation backcross hybrids following our crossing design: segregating at the level of
575 chromosome 18 and at four other chromosomes. For the BC3 hybrids sampled at 156 hours
576 after pupal formation (APF) we had 6 *cyd/melp* and 10 *cyd*/*cyd* at the QTL region on
577 chromosome 18 (**Supporting figure 5A**), for those at 60h APF, 8 *cyd/melp* and 9 *cyd*/*cyd*
578 hybrids at the same region.

579

580 **Allele-specific expression (ASE) in hybrids.** In order to conduct ASE analyses we first
581 identified species specific variants, fixed in either *melpomene* and *cydno*. For this, we took
582 the quality filtered variants inferred from the species genome resequencing data, and assigned
583 those genotype calls in *cydno* and *melpomene* for which allele frequency (AF) was > 0.9 as
584 homozygous (we did not consider indels in this analysis). We then used *bcftools intersect*[68] to
585 get only those variants for which *cydno* and *melpomene* had opposite alleles.

586 At the same time, we called variants from RNA-seq reads of F1 hybrid individuals,
587 again according to the GATK Best Practices (with the exception of parameters -window 35 -
588 cluster 3, to increase SNPs density), and selected only heterozygous SNPs in F1s that matched
589 the species-specific variants. Finally, we used GATK's ASEReadCounter, with default
590 parameters, to count RNA reads in the F1 hybrids (and later on in BC3 hybrids) that mapped
591 to either the *cydno* or the *melpomene* allele. We summed all reads mapping to either the *cydno*
592 or *melpomene* allele/variant within the same gene (both for gene models of the Hmel2.5 gene
593 annotation and for the Cufflinks annotation we assembled previously). To test for allele
594 specific expression (diffASE) we fitted the model "~0 + individual + allele" in DESeq2,
595 setting library size factors to 1 (thus not normalizing between samples, as the test for diffASE
596 is conducted within individuals). We only considered those alleles showing at least a 2-fold
597 change in expression and $p < 0.05$, as differentially expressed.

598 In order to check that there were no biases in alleles assignment to one of the two
599 species, we analyzed the ratios of the species alleles, for every gene, and checked that they
600 were not systematically biased to either one of the two species. The $\log_2$ fold-changes of the
601 species alleles were centered around 0, suggesting no obvious bias in alleles assignment[75]
602 (**Supporting figure 7**).

603

604 **Protein-coding substitutions and predicted effects on protein-function.** We inferred fixed
605 variants in protein-coding regions from the combined *melpomene* and *cydno* RNA samples in
606 order to include variants from genes for which we detected expression in the brain/eyes across
607 the 3 stages. We took the quality filtered variants called from the joint genotyping of RNA-
608 seq data of *cydno* and *melpomene* (from all stages), and selected those genotype calls for
609 which allele frequency (AF) > 0.8, and where the allelic variant was present in at least 7
610 individuals of the ~30 samples (for each species). We retained those substitutions/indels
611 validated with the genome resequencing data. For this, of the genotype calls found in RNA
612 reads from brain/eyes of different stages, we kept only those that were also called in at least 8

613 of the 10 genome resequencing samples of each species. We considered this overlapping set
614 of variants as being fixed in *H. melpomene rosina* or *H. cydno chioneus*. Following a similar
615 approach to Bendesky *et al.*[64], we then restricted this set of substitutions between *cydno* and
616 *melpomene* to protein-coding regions, and selected those non-synonymous substitutions that
617 were considered to have moderate or high effect on protein function, with SNPeff[76]. Finally,
618 we used the PROVEAN algorithm[39], to further study the functional effects of these
619 substitutions on protein function. The PROVEAN algorithm predicts the functional effect of
620 protein sequence variations based on how they affect alignments to homologous protein
621 sequences (for this we used the PROVEAN protein database online). We selected those amino
622 acid changes with the suggested PROVEAN score < -2.5, indicating non-neutral effect on
623 protein function.
624
625 **Admixture analyses.** We retrieved estimated admixture proportions between *H. melpomene*
626 *rosina* and *H. cydno chioneus*, for 100kb and 20kb windows, from Martin et al.[41]
627
628 **Data accessibility.** RNA-seq data will be deposited on a public database
629 (https://www.ebi.ac.uk/ena) on acceptance. Analysis scripts and behavioral data are available
630 at: https://github.com/SpeciationBehaviour/neural_genes_heliconius.git
631
632 **Acknowledgments**
633 We thank Liz Evans and Adriana Tapia for assistance in the insectaries. We are very grateful
634 to Ana Pinharanda for sharing the *H. cydno* genome assembly and annotation and to Simon
635 Martin for sharing admixture proportions and comments on the manuscript. We thank Ricardo
636 Pereira, Nicolas Gompel and Jochen Wolf for input on the project. Jochen Wolf also kindly
637 provided computational resources. We thank the Smithsonian Tropical Research Institute for
638 providing research infrastructure, the Ministerio del Ambiente for permission to collect
639 butterflies in Panama, and Edinburgh Genomics for sequencing support. MR, AEH and RMM
640 are supported by an Emmy Noether fellowship and research grant awarded to RMM by the
641 Deutsche Forschungsgemeischaft (DFG) (Grant Number: GZ: ME 4845/1-1). RMM was also
642 supported by a Junior Research Fellowship from King's College Cambridge, an Ernst Mayr
643 fellowship from the Smithsonian Tropical Research Institute, a Varley-Gradwell fellowship
644 from the Oxford University Museum of Natural History, and the Balfour-Browne Fund,
645 University of Cambridge. SHM was supported by a NERC IRF (NE/N014936/1). This project
646 was also supported by an Institutional Development Award (IdeA) INBRE from the National

Institute of General Medical Sciences (NIGMS) awarded to RP (Grant number: P20GM103475), and a European Research Council (ERC) awarded to CDJ (Grant number: 339873).

References

1. J. A. Coyne, H. A. Orr, *Speciation* (Sunderland, MA: Sinauer 2004).
2. G. Rosenthal, *Mate choice* (Princeton University Press, 2017).
3. E. Mayr, *Animal species and evolution* (Harvard University Press, 1963).
4. J. Cande, B. Prud'homme, N. Gompel, Smells like evolution: The role of chemoreceptor evolution in behavioral change. *Curr. Opin. Neurobiol.* **23**, 152–158 (2013).
5. J. R. Arguello, R. Benton, Open questions: Tackling Darwin's "instincts": The genetic basis of behavioral evolution. *BMC Biol.* **15**, 8–10 (2017).
6. A. J. Moehring, *et al.*, Quantitative trait loci for sexual isolation between Drosophila simulans and D. mauritiana. *Genetics* ***167***, 1265-1274 (2004).
7. K. L. Shaw, S. C. Lesnick, Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9737-9742 (2009).
8. R. A. Bay, *et al.*, Genetic Coupling of Female Mate Choice with Polygenic Ecological Divergence Facilitates Stickleback Speciation. *Curr. Biol.* **27**, 3344-3349 (2017).
9. M. P. Shahandeh, A. Pischedda, J. M. Rodriguez, T. L. Turner, The genetics of male pheromone preference difference between Drosophila melanogaster and Drosophila simulans. *G3 Genes, Genomes, Genet.* **10**, 401–415 (2020).
10. F. Gould, *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)
11. G. P. Leary, *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).
12. P. Fan, *et al.*, Genetic and neural mechanisms that inhibit drosophila from mating with other species. *Cell* **154**, 89–102 (2013).
13. P. Brand, *et al.*, The evolution of sexual signaling is linked to odorant receptor tuning in perfume-collecting orchid bees. *Nat. Commun.* **11** (2020).
14. O. Seehausen, *et al.*, Speciation through sensory drive in cichlid fish. *Nature* **455**, 620–626 (2008).
15. K. Hench, M. Vargas, M. P. Höppner, W. O. McMillan, O. Puebla, Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nat. Ecol. Evol.* **3**, 657–667 (2019).
16. R. M. Merrill, *et al.*, The diversification of Heliconius butterflies: What have we learned in 150 years? *J. Evol. Biol.* **28**, 1417–1438 (2015).
17. C. D. Jiggins, *The ecology and evolution of Heliconius butterflies* (Oxford University Press, 2016).
18. R. M. Merrill, *et al.*, Disruptive ecological selection on a mating cue. *Proc. R. Soc. B Biol. Sci.* **279**, 4907–4913 (2012).
19. C. D. Jiggins, R. E. Naisbit, R. L. Coe, J. Mallet, Reproductive isolation caused by colour pattern mimicry. *Nature* **411**, 302–305 (2001).
20. R. M. Merrill, *et al.,* Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution*, ***65***, 1489-1500 (2011).
21. C. Mérot, C. Salazar, R. M. Merrill, C. D. Jiggins, M. Joron, What shapes the continuum of reproductive isolation? Lessons from Heliconius butterflies. *Proc. R. Soc. B Biol. Sci.* **284** (2017).

696 22. C. D. Jiggins, Ecological Speciation in Mimetic Butterflies. *Bioscience* **58**, 541–548 (2008).
697 23. R. D. Reed, *et al.,* Optix drives the repeated convergent evolution of butterfly wing pattern
698 mimicry. *Science 333*, 1137-1141 (2011).
699 24. A. Martin, *et al.*, Diversification of complex butterfly wing patterns by repeated regulatory
700 evolution of a Wnt ligand. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12632–12637 (2012).
701 25. N. J. Nadeau, *et al.*, The gene cortex controls mimicry and crypsis in butterflies and moths.
702 *Nature* **534**, 106–110 (2016).
703 26. C. D. Jiggins, C. Estrada, A. Rodrigues, Mimicry and the evolution of premating isolation
704 in Heliconius melpomene Linnaeus. *J. Evol. Biol.* **17**, 680–691 (2004).
705 27. R. M. Merrill, B. Van Schooten, J. A. Scott, C. D. Jiggins, Pervasive genetic associations
706 between traits causing reproductive isolation in Heliconius butterflies. *Proc. R. Soc. B Biol.*
707 *Sci.* **278**, 511–518 (2011).
708 28. R. M. Merrill, *et al.*, Genetic dissection of assortative mating behavior. *PLoS Biol 17*,
709 e2005902 (2018).
710 29. J. Felsenstein, Skepticism Towards Santa Rosalia, or Why Are There So Few Kinds of
711 Animals? *Evolution* **35**, 124–138 (1981).
712 30. J. H. Massey, D. Chung, I. Siwanowicz, D. L. Stern, P. J. Wittkopp, The yellow gene
713 influences drosophila male mating success through sex comb melanization. *Elife* **8**, 1–20
714 (2019).
715 31. L. F. Seeholzer, M. Seppo, D. L. Stern, V. Ruta, Evolution of a central neural circuit
716 underlies Drosophila mate preferences. *Nature* **559**, 564–569 (2018).
717 32. S. H. Martin, *et al.*, Genome-wide evidence for speciation with gene flow in Heliconius
718 butterflies. *Genome Res.* **23**, 1817–1828 (2013).
719 33. J. Davey, *et al.*, Major improvements to the Heliconius melpomene genome assembly used
720 to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3* **6**,
721 695–708 (2015).
722 34. K. Darragh, *et al.*, A novel terpene synthase produces an anti-aphrodisiac pheromone in the
723 butterfly Heliconius melpomene. *bioRxiv*, 779678 (2019).
724 35. A. Pinharanda, *et al.*, Sexually dimorphic gene expression and transcriptome evolution
725 provide mixed evidence for a fast-Z effect in Heliconius. *J. Evol. Biol.* **32**, 194–204 (2019).
726 36. A. Roberts, H. Pimentel, C. Trapnell, L. Pachter, Identification of novel transcripts in
727 annotated genomes using RNA-seq. *Bioinformatics* **27**, 2325–2329 (2011).
728 37. P. J. Wittkopp, B. K. Haerum, A. G. Clark, Evolutionary changes in cis and trans gene
729 regulation. *Nature* **430**, 85–88 (2004).
730 38. P. D. Thomas, *et al.*, Applications for protein sequence-function evolution data:
731 mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res.* **34**,
732 645–650 (2006).
733 39. Y. Choi, G. E. Sims, S. Murphy, J. R. Miller, A. P. Chan, Predicting the Functional Effect
734 of Amino Acid Substitutions and Indels. *PLoS One* **7** (2012).
735 40. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA-BABA statistics to
736 locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
737 41. S. H. Martin, J. W. Davey, C. Salazar, C. D. Jiggins, Recombination rate variation shapes
738 barriers to introgression across butterfly genomes. *PLoS Biol.* **17**, 1–28 (2019).
739 42. M. Kopp, *et al.*, Mechanisms of assortative mating in speciation with gene flow: Connecting
740 theory and empirical research. *Am. Nat.* **191**, 1–20 (2018).
741 43. R. K. Butlin, C. M. Smadja, Coupling, reinforcement, and speciation. *Am. Nat.* **191**, 155–
742 172 (2018).
743 44. K. J. McCulloch, *et al.*, Sexual Dimorphism and Retinal Mosaic Diversification following
744 the Evolution of a Violet Receptor in Butterflies. *Mol. Biol. Evol.* **34**, 2271–2284 (2017).
745 45. G. Zaccardi, A. Kelber, M. P. Sison-Mangus, A. D. Briscoe, Color discrimination in the red
746 range with only one long-wavelength sensitive opsin. *J. Exp. Biol.* **209**, 1944–1955 (2006).

46. A. Monteiro, Gene regulatory networks reused to build novel traits. *BioEssays* **34**, 181–186 (2012).
47. A. Martin, *et al.*, Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. *Evodevo* **5** (2014).
48. E. R Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, A. J. Hudspeth. *Principles of Neural Science*, 2012th Ed. (New York: McGraw Hill, 2000).
49. X. Wang, *et al.*, Genomic responses to selection for tame/aggressive behaviors in the silver fox (Vulpes vulpes). *Proc. Natl. Acad. Sci. U. S. A.* **115**, 10398–10403 (2018).
50. N. I. Bloch, *et al.*, Early neurogenomic response associated with variation in guppy female mate preference. *Nat. Ecol. Evol.* **2**, 1772–1781 (2018).
51. M. Yamaguchi, Role of regucalcin in brain calcium signaling. *Integr. Biol.* **4**, 825–837 (2012).
52. M. J. Berridge, Neuronal calcium signaling. *Neuron*, ***2*1**, 13-26 (1998).
53. G. J. Bashaw, R. Klein, Signaling from axon guidance receptors. *Cold Spring Harb. Perspect. Biol.* **2**, 1–17 (2010).
54. B. Prud'homme, N. Gompel, S. B. Carroll, Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* ***104*,** 8605-8612 (2007).
55. E. Preger-Ben Noon, *et al.*, Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer Function. *Cell Rep.* **22**, 3021–3031 (2018).
56. J. Lewis, *et al.*, Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24174–24183 (2019).
57. W. Zhang, *et al.*, Comparative Transcriptomics Provides Insights into Reticulate and Adaptive Evolution of a Butterfly Radiation. *Genome Biol. Evol.* **11**, 2963–2975 (2019).
58. A. Mazo-Vargas, *et al.*, Macroevolutionary shifts of WntA function potentiate butterfly wing-pattern diversity. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10701–10706 (2017).
59. C. Concha, *et al.*, Interplay between Developmental Flexibility and Determinism in the Evolution of Mimetic Heliconius Wing Patterns. *Curr. Biol.* **29**, 3996-4009 (2019).
60. J. N. Weber, B. K. Peterson, H. E. Hoekstra, Discrete genetic modules are responsible for complex burrow evolution in Peromyscus mice. *Nature* **493**, 402–405 (2013).\
61. J. Cande, P. Andolfatto, B. Prud'homme, D. L. Stern, N. Gompel, 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).
62. C. S. McBride, *et al.*, Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **515**, 222–227 (2014).
63. Y. Ding, A. Berrocal, T. Morita, K. D. Longden, D. L. Stern, Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature* **536**, 329–332 (2016).
64. A. Bendesky, *et al.*, The genetic basis of parental care evolution in monogamous mice. *Nature* **544**, 434–439 (2017).
65. P. C. Bürkner, Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, **10**, 395-411 (2018).
66. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
67. A. Dobin, *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
69. S. Anders, P. T. Pyl, W. Huber, HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
70. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

71. S. H. Montgomery, J. E. Mank, Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Mol. Ecol.* **25**, 5114–5128 (2016).
72. A. McKenna, *et al.*, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
73. R. D. Finn, *et al.*, InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199 (2017).
74. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
75. R. A. York, *et al.*, Behavior-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).
76. P. Cingolani, *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff *Fly.* 6, 80–92 (2012).

Supporting information: Neural processing genes are linked to divergence in visual mate preference in sympatric *Heliconius* butterflies.

Matteo Rossi[a,b], Timothy J. Thurman[b,c], Alexander E. Hausmann[a], Stephen H. Montgomery[d], Riccardo Papa[b,e,f], Chris D. Jiggins[g] , W. Owen McMillan[b] & Richard M. Merrill[a,b]

[a] Division of Evolutionary Biology, Ludwig Maximilian University of Munich, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany

[b] Smithsonian Tropical Research Institute, Gamboa, 0843-03092, Panama

[c] Division of Biological Sciences, University of Montana, 32 Campus Dr., Missoula, MT 59812, US

[d] School of Biological Sciences, University of Bristol, 24 Tyndall Ave, Bristol BS8 1TQ, UK

[e] Department of Biology, University of Puerto Rico, Rio Piedras, San Juan 00925, Puerto Rico

[f] Molecular Sciences and Research Center, University of Puerto Rico, San Juan 00926, Puerto Rico

[g] Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

870 **Supporting methods and results**
871
872 **Mapping RNA-seq reads to the *Heliconius cydno* genome.** To determine whether the *H.*
873 *melpomene* reference genome introduced mapping biases of RNA-seq reads, possibly
874 affecting differential expression estimates, we also mapped to a *H. cydno*
875 assembly/annotation. Generally, we found similar patterns of differential expression when
876 mapping to the two genomes. Since **i)** we observed an equal decrease (~ 40 %) of genes
877 showing 2-fold changes in *melpomene* and *cydno* when mapping to *H. cydno,* at every stage
878 (p-value=0.317 at adult stage, p-value=0.800 at 156h APF, p-value=0.897 at 60h APFP,
879 Fisher's Exact test, **Table S2**), and **ii)** this decrease was widespread throughout the genome,
880 we concluded that the *melpomene* reference genome did not bias differential gene expression
881 analyses.
882
883 **Allele-specific expression in the introgression line.** BC3 hybrids had different combinations
884 of chromosomes segregating for the *melpomene* alleles in a *cydno* background. Therefore, in
885 principle, we could not infer *cis*- or *trans*- gene regulatory effects genome-wide from the
886 profiles of allele specific expression (ASE) in these hybrids as for F1 hybrids, due to the
887 diverse trans-acting environments. However, previous analyses (comparing gene expression
888 levels between hybrids carrying *cyd/melp* vs. *cyd/cyd* regions on chromosomes other than 18)
889 imply that differential expression of the candidate genes seems to be driven by the *melpomene*
890 copy difference within the introgressed region on chromosome 18. Therefore, ASE analyses
891 of candidate genes in BC3 hybrids carrying *cyd/melp* alleles on chromosome 18 should
892 indicate whether the differences are due to cis- or trans-regulatory effects from within the
893 introgressed region (**Figure S6**).
894
895 In BC3 hybrids sampled at 156h APF and 60hAPF, the *melpomene* and *cydno* alleles of the
896 *ionotropic glutamate receptor* (*Grik2*) are expressed at very similar levels (at 156hAPF p
897 value=0.841, at 60hAPF p value=0.579, Wald test), suggesting trans-only regulatory effects at
898 these stages for *Grik2*. For *regucalcin1* we again had very few allele-informative read counts
899 in hybrids at 156h APF. Although there was a tendency towards up regulation of the
900 *melpomene* allele, there was no statistical significance to support this (p=0.174, Wald test).
901 We detected diffASE expression of *regucalcin2* only at 60h APF (p <0.001, Wald test), but at
902 this stage *regucalcin2* was not detected as differentially expressed between pure species.
903 Thus, although there is tentative evidence for *cis*-regulatory effects for species differences in
904 *regucalcin* expression during development, it is not conclusive.

Supporting 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. Each point corresponds to a male (and a 'jitter' function applied). Point size is scaled to the number of trials in which the male showed a response. 95% credibility intervals (CrIs) for all three proportions are shown as hexagons around predictors extrapolated from the statistical model with the best fit, with lines projecting to corresponding values on the three axes.

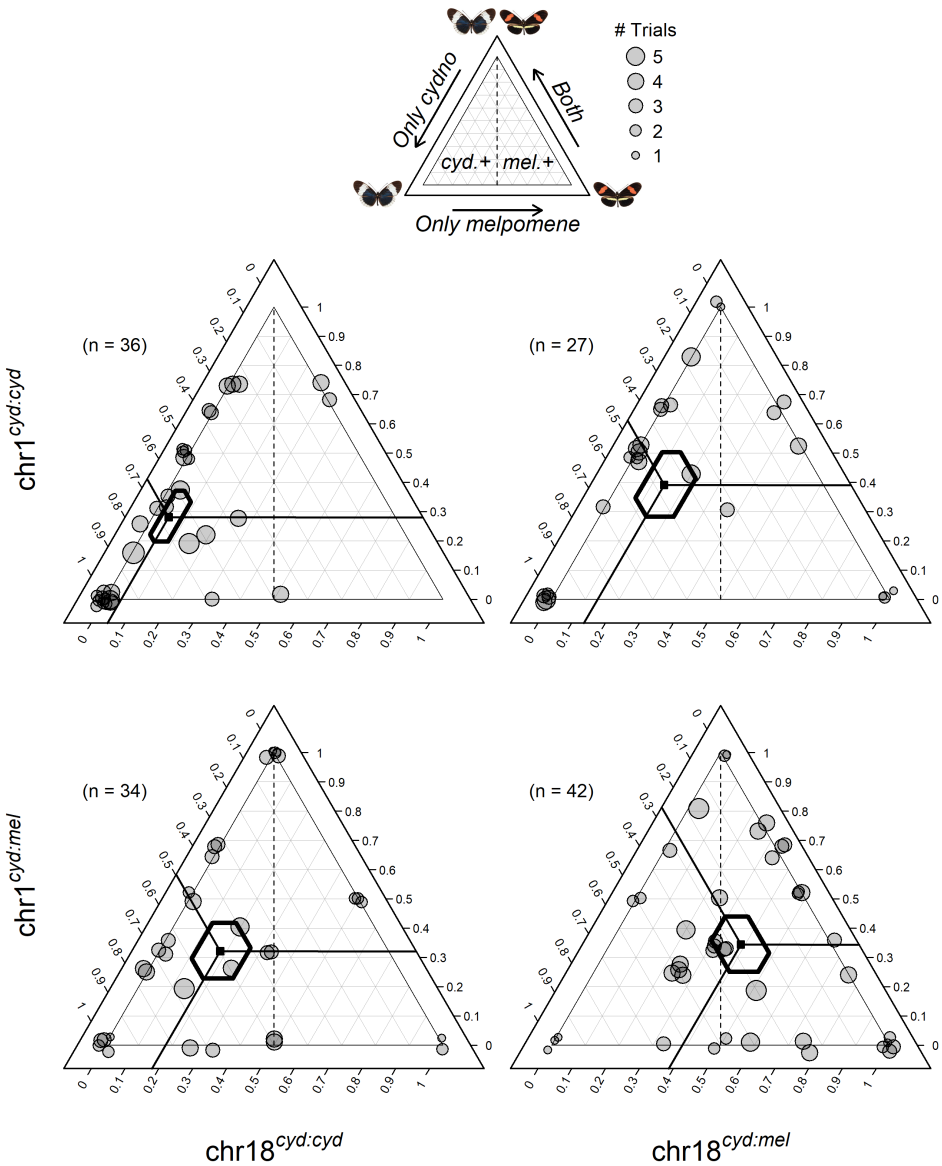


Figure S2. Results of comparative transcriptomic analyses between *melpomene* and *cydno* (in the imago) when mapping RNA-seq reads to the *H. melpomene* assembly/annotation (top), and to the *H. cydno* assembly/annotation (bottom), zooming in on the QTL region on chromosome 18. The *x*-axis represents 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.

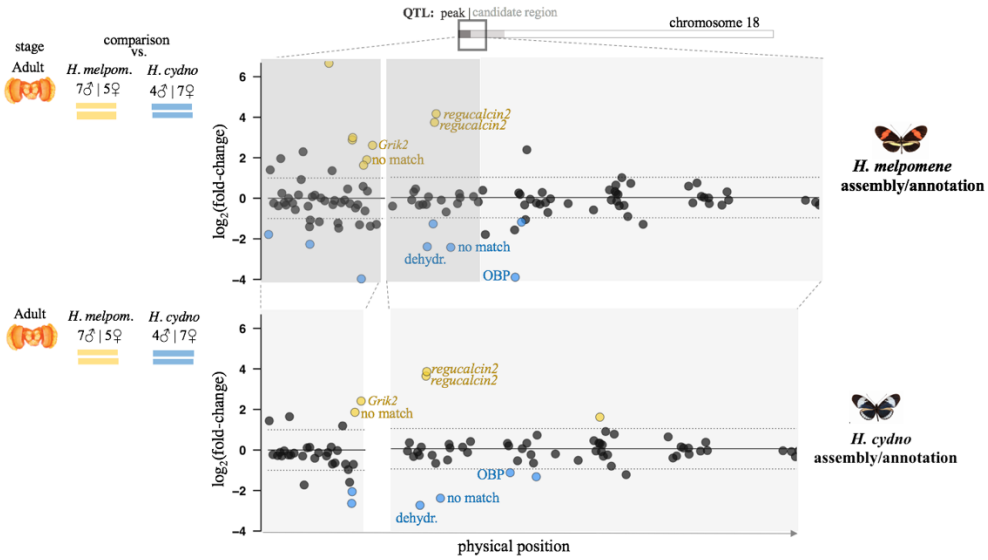


Figure S3. Crossing design for producing backcross hybrids segregating at the QTL on chromosome 18. This introgression line was created by outcrossing a male hybrid to *H. cydno* females over three generations, selecting a hybrid male that showed a red band on the wing at each generation. This meant 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 behavior).

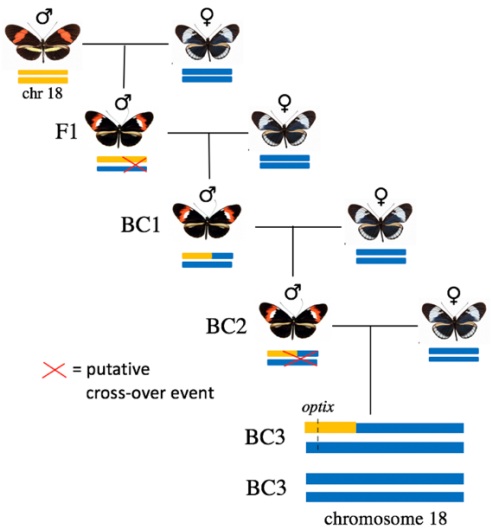


Figure S4. Differential gene expression at the QTL region at pupal stages. Left: summary of the comparative transcriptomic analyses with stage, number of samples and chromosome 18 composition. Right: the corresponding results, zooming in on the QTL region on chromosome 18. The *x*-axis represents physical position. The QTL peak, and the rest of the QTL 1.5 LOD candidate region (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 *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 *melpomene* and *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.

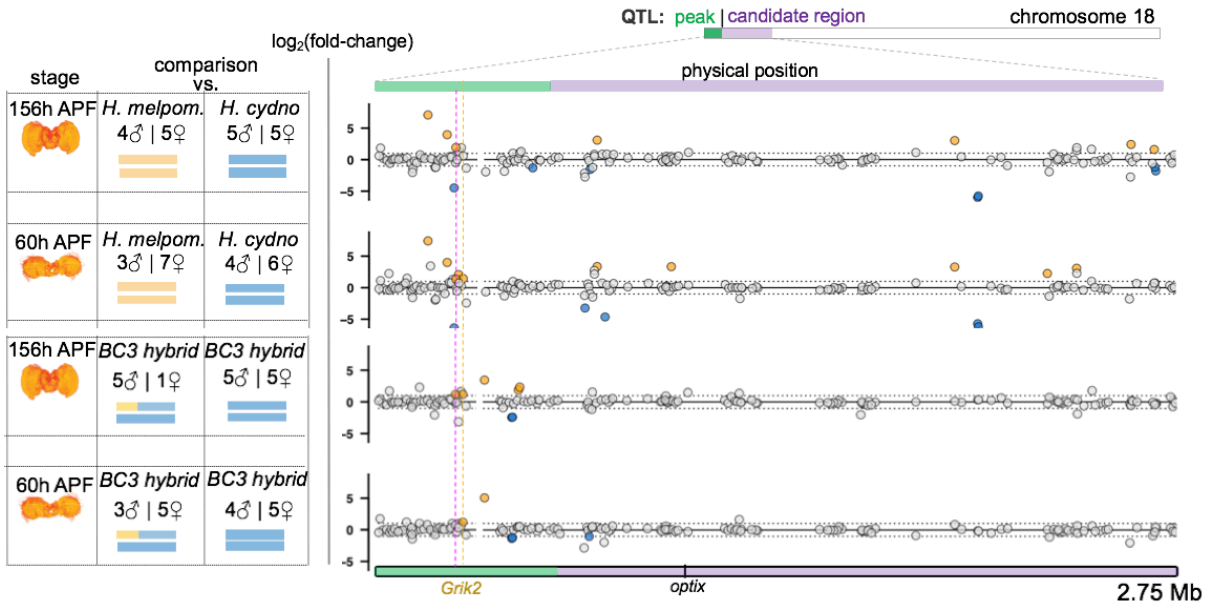
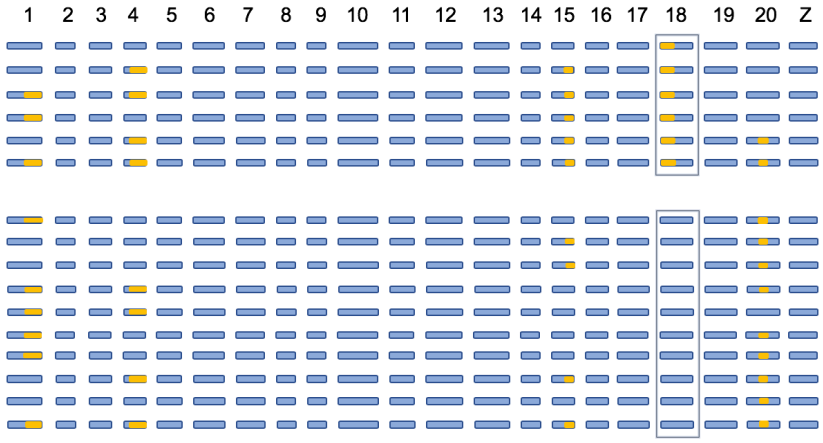


Figure S5. A) Schematic representation of hybrid pupae (sampled at 156h APF) genome composition. Columns represent chromosomes, rows represent individuals, orange indicates *cyd/melp* regions, blue indicates *cyd/cyd* regions. **B)** Genome composition of (a subset of) BC3 hybrids. We calculated the fraction of SNPs and indels that each BC3 hybrid, one *cydno* and one F1 hybrid samples shared with *melpomene* and *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)

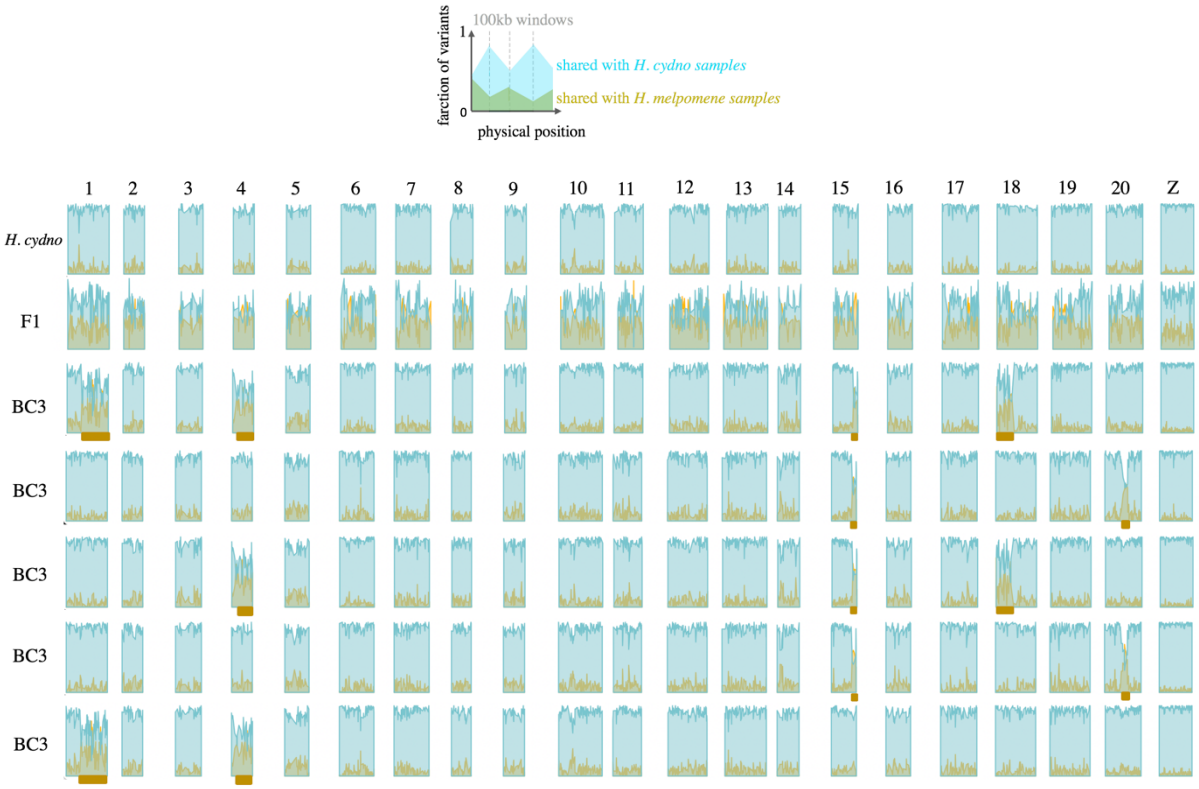


Figure S6. Allele specific expression profiles of candidate genes at pupal stages. Points indicate the value, and bars the standard error, of the (base 2) logarithmic fold change in expression between parental species (horizontal) and the alleles in F1 hybrids (vertical), for candidate genes (as defined 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).

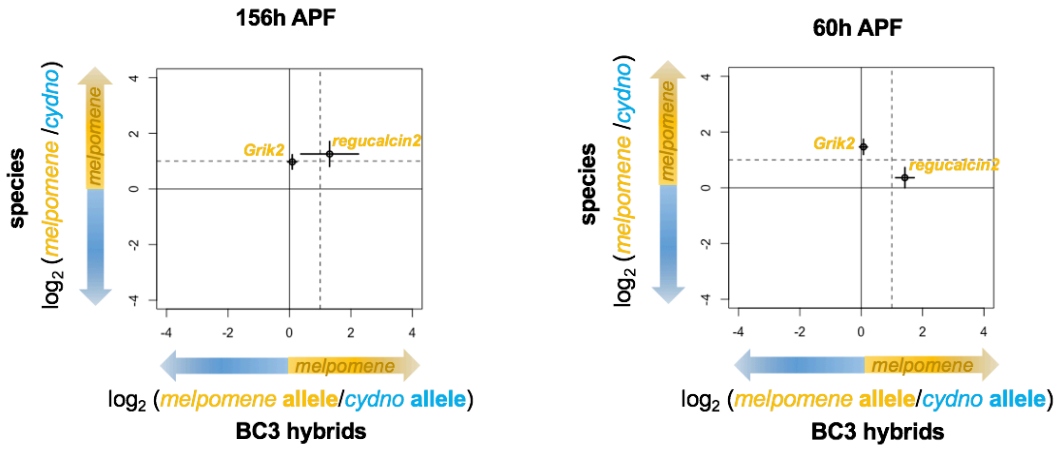
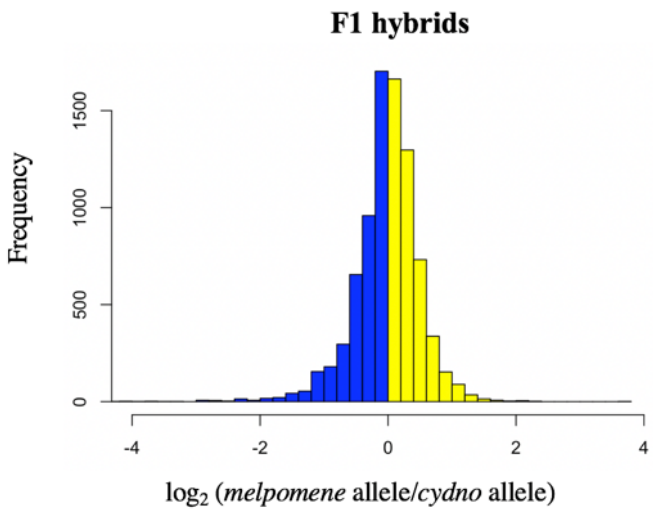


Figure S7. Distribution of the (base 2) logarithmic fold change in allele expression. Colored bars indicate the number of genes showing a bias in expression for the *cydno* allele (in blue) and for the *melpomene* allele (in yellow). Values departing from 0 on the x-axis, indicate an increase in the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



Supporting 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		
			60h APF	156h APF	Imago
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	✓	✗	✗

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

#	Gene name (Hmel2.5)	Annotated function	Species comparison			Hybrids comparison		
			60h APF	156h APF	Imago	60h APF	156h APF	Imago
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*	✗	✗	✗	✓	✓	✗

1091 **Table S2.** Number of genes showing significant >2-fold change in expression, at different
1092 stages, mapping to the *melpomene* and to the *cydno* genomes. Note that the considerable
1093 reduction in the number of genes detected as differentially expressed when mapping to *H. cydno*
1094 is most likely a result of the lower quality/completeness of the *H. cydno* genome assembly.
1095

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

1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124

Table S3. 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