A major locus controls a biologically active pheromone component in *Heliconius melpomene*

Kelsey J.R.P. Byers[1,2,9], Kathy Darragh[1,2,9], Jamie Musgrove[2], Diana Abondano Almeida[2,3], Sylvia Fernanda Garza[2,4], Ian A. Warren[1], Pasi Rastas[5], Marek Kučka[6], Yingguang Frank Chan[6], Richard M. Merrill[7], Stefan Schulz[8], W. Owen McMillan[2], Chris D. Jiggins[1,2,10]

[1] *Department of Zoology, University of Cambridge, Cambridge, United Kingdom*

[2] *Smithsonian Tropical Research Institute, Panama, Panama*

[3] *Present address: Institute for Ecology, Evolution and Diversity, Goethe Universität, Frankfurt, Germany*

[4] *Present address: Department of Collective Behaviour, Max Planck Institute of Animal Behaviour, Konstanz, Germany & Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany*

[5] *Institute of Biotechnology, University of Helsinki, Helsinki, Finland*

[6] *Friedrich Miescher Laboratory of the Max Planck Society, Tuebingen, Germany*

[7] *Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, Munich, Germany*

[8] *Institute of Organic Chemistry, Department of Life Sciences, Technische Universität Braunschweig, Braunschweig, Germany*

[9] *These authors contributed equally to this work*

[10] *To whom correspondence should be addressed: c.jiggins@zoo.cam.ac.uk*

Running title: Genetics of bioactive pheromones in *Heliconius*

Abstract

Pheromones are important for courtship and mate choice in many insects, but we know relatively little of their role in butterflies. The butterfly *Heliconius melpomene* uses a complex blend of wing androconial compounds during courtship. Electroantennography in *H. melpomene* and its close relative *H. cydno* showed that responses to androconial extracts were not species-specific. Females of both species responded more strongly to the *H. cydno* extract, suggesting conservation of peripheral nervous system elements across the two species. Individual blend components provoked little to no response, with the exception of octadecanal, a major component of the *H. melpomene* blend. Supplementing octadecanal on the wings of octadecanal-rich *H. melpomene* males led to an increase in the time until mating, demonstrating the bioactivity of octadecanal in *Heliconius*. Using quantitative trait locus (QTL) mapping, we identified a single locus on chromosome 20 responsible for 41% of the parental species' difference in octadecanal production. This QTL does not overlap with any of the major wing color or mate choice loci, nor does it overlap with known regions of elevated or reduced F_{ST}. A set of 16 candidate fatty acid biosynthesis genes lies underneath the QTL.

Introduction

Chemical communication is the oldest form of sensory communication, and is found across the tree of life from bacteria (1) to humans (2,3). In Lepidoptera, work on pheromones has largely focused on female pheromones of nocturnal moths, where relatively simple alterations to pheromones can produce drastic differences in mate attraction. For example, different populations of the corn borer moth *Ostrinia nubialis* exhibit a simple *cis* to *trans* switch in the pheromone 11-tetradecenyl acetate (4) that leads to partial reproductive isolation between them (5). Sex pheromones have been less well studied in day-flying butterflies, where visual signaling is often assumed to play a more dominant role in mate choice (6). However, male butterflies also emit close-range pheromone bouquets which may act in concert with other wing pattern and behavioral cues (7), and are important in mate choice (8).

Despite the potential importance of pheromones in butterflies, pheromones have so far been identified in only a handful of butterfly species (9,10,11,12,13,14,15,16). Within butterflies, the genetic basis of pheromone production is known only in *Bicyclus anynana* (17). As in other diurnal butterflies, the chemical bouquets of *Heliconius* (8,18), are often complex, both in identity and quantity of compounds; however, just a few individual compounds may be biologically active pheromones. Identifying which

51 components of these complex chemical bouquets are responsible for pheromonal communication is a
52 considerable challenge. This is particularly true as even minor compounds can have major effects
53 (19,20). Determining pheromone bioactivity requires screening compounds via physiological activity
54 followed by behavioral verification, and in many cases pheromone bouquet composition is known but
55 the bioactive components remain unidentified.

56 To better characterize male butterfly pheromone production and the role it plays in mating and species
57 recognition, we first comprehensively analyzed pheromone bouquets of the closely related *Heliconius*
58 *melpomene* and *H. cydno* butterflies to determine the most important bioactive compounds. We further
59 mapped a genetic locus likely responsible for the differential biosynthesis of the main fatty-acid
60 compound between the two species. The two species are closely related yet strongly reproductively
61 isolated (21). Over the past decade there has been considerable research into genomic architecture of
62 differences in wing pattern and mate preference (21,22). Surprisingly, both wing color and mating
63 preferences between these species have a relatively simple genetic basis with a large proportion of the
64 difference between parental forms being controlled by a small handful of loci of large effect (21,22). This
65 has important implications for speciation, as theory predicts that large effect loci should facilitate
66 speciation in the face of gene flow (23). Similarly, clustering of genetic loci that contribute to speciation
67 can also make speciation easier (24,25), and there is evidence for tight linkage of a gene controlling wing
68 pattern and a QTL underlying divergent preference behaviors (22). Here, we identify an
69 electrophysiologically active compound, octadecanal, explore its behavioral role, and investigate the
70 genetic basis of its production in *H. melpomene* using QTL mapping.

71 **Results**

72 <u>Androconial chemical bouquets of *Heliconius melpomene* and *H. cydno*</u>

73 In order to identify candidate pheromone compounds, we first investigated the distribution of chemical
74 compounds on the wings of *H. cydno* for comparison with published data on *H. melpomene* (Figure 1).
75 We found that the bouquet of *H. cydno* was simpler than that of *H. melpomene*, with 7 main compounds
76 versus 21 in *H. melpomene*. *Heliconius cydno* also had a less abundant overall bouquet, with an
77 individual total of 1787 ± 776 ng vs. 3174 ± 1040 ng in *H. melpomene* (Table S1). Most of the main
78 compounds (defined as those occurring in at least 90% of individuals) in *H. cydno* were linear alkanes (4
79 of 7 compounds), while *H. melpomene* has a more diverse array of different compound classes. Of the
80 five major compounds (> 100ng per individual) in *H. melpomene*, only syringaldehyde was found in

81 similar amounts in *H. cydno*; the major compounds octadecanal, 1-octadecanol, (*Z*)-11-icosenal, and (*Z*)-
82 11-icosenol were absent or found in very low amounts in *H. cydno*. Comparison with previously
83 published data for other *Heliconius* species, all of which lack octadecanal in the large amounts seen in *H.*
84 *melpomene*, demonstrates that this high level of octadecanal is an evolutionarily derived state in *H.*
85 *melpomene* (18,26). When focusing on the hindwing androconia of *H. cydno*, only two compounds
86 [syringaldehyde (24.7% of the hindwing androconial bouquet) and (*Z*)-11-icosenol (1.7%)] were specific
87 to this region. For details, please see SI text, Figures S1-S3, and Tables S1-S2.

88 Electroantennographic responses to con- and heterospecific pheromone stimulus sets

89 We next investigated the electroantennographic (EAG) response of both species to natural con- and
90 heterospecific pheromone bouquets extracted from adult male butterflies. Females of both *H.*
91 *melpomene* and *H. cydno* responded more strongly (i.e. showed a larger voltage displacement from the
92 antennal baseline) to both natural wing extracts (Mnat and Cnat, respectively) as compared with the
93 control solvent (dichloromethane plus 1 ng/µL 2-tetradecyl acetate, hereafter "DCM+IS") (Figure 2,
94 Figure S6; see Table S3 for statistical details). Males of both species also responded to both wing
95 extracts. Females of both species showed stronger responses to *H. cydno* wing extracts than to *H.*
96 *melpomene* wing extracts. Responses differed in males, with *H. melpomene* males responding equally to
97 both species' wing extracts, while males of *H. cydno* responded more strongly to wing extracts of *H.*
98 *melpomene*.

99 We then explored antennal responses to synthetic compound blends. These were based on the most
100 abundant compounds from each species (see Methods, Figure S4). Responses to the synthetic *H.*
101 *melpomene* wing blend were mixed in *H. melpomene* females. In the first stimulus set (comparison of
102 hetero- and conspecific stimuli) the synthetic blend (Msyn) yielded stronger responses than the natural
103 *H. melpomene* wing extracts (Figure 2, Table S3) but in the second stimulus set (individual pheromone
104 components) the synthetic blend yielded weaker responses than the natural *H. melpomene* wing
105 extracts (Figure 3, Table S3). A similar pattern was seen in male *H. melpomene*. Both sexes of *H. cydno*
106 produced only a modest response to the synthetic *H. cydno* wing blend (Csyn) when compared with the
107 hexane solvent. In all cases this response was lower than to natural *H. cydno* wing extract.

108 Electroantennographic responses to individual pheromone components from both species

109 Finally, we explored the responses to individual compounds to identify specific biologically active
110 pheromone components. Only octadecanal differed significantly from the controls in all four species-sex

111 combinations (Figure 3, Table 1, Table S3). Both sexes of *H. cydno* also showed altered responses to the
112 shared compound (*Z*)-11-icosenol, with an increased response in *H. cydno* males and a decreased
113 response in *H. cydno* females. Both *H. cydno* females and *H. melpomene* males had decreased responses
114 to (*Z*)-11-icosenal. *Heliconius cydno* females also showed a decreased response to (*Z*)-13-docosenal.
115 Male *H. cydno* had increased responses to 1-octadecanol and henicosane. No other compound-species-
116 sex combinations were significantly different from hexane. The shared compound syringaldehyde was
117 the only pheromone component to show no change in EAG amplitude relative to the solvents in any
118 species-sex combination.

119 In females of both species, response to octadecanal was stronger than response to the conspecific
120 synthetic mixture. By contrast, male *H. melpomene* responded equally to the conspecific synthetic
121 mixture and octadecanal, while male *H. cydno* responded equally to the conspecific synthetic mixture
122 and both its components syringaldehyde and henicosane, with no evidence for a synergistic mixture
123 effect. Individual components (with the exception of octadecanal) were not equivalent to the synthetic
124 blends for females of both *H. melpomene* and *H. cydno*, suggesting a synergistic mixture effect. Antennal
125 responses to a given stimulus can change over time, and this may reflect biological processes of
126 neuronal adaptation. Female *H. melpomene* adapted equally quickly to octadecanal and the natural and
127 synthetic *H. melpomene* pheromones, while adaptation to other stimuli was equal to the control, further
128 supporting octadecanal's salience as the main pheromone in *H. melpomen*e (see SI text, Figures S7-S8,
129 and Table S4).

130 <u>Behavioral response to octadecanal supplementation in *H. melpomene*</u>

131 We next confirmed a behavioral response to the most physiologically active substance, octadecanal, one
132 of the dominant compounds in *H. melpomene*. A total of 29 behavioral experiments were conducted,
133 with mating observed in 18 (62%); one trial was excluded due to wing damage, leaving 17 successful
134 matings. With our small sample size, we found no evidence that females showed a preference for either
135 treatment, mating with the hexane male 11 times (65%) and with the octadecanal male the remaining 6
136 times (35%) ($p = 0.332$). However, mating latency (time from experiment onset to mating) was
137 significantly longer for the octadecanal matings (average 88.5 minutes) than for the hexane matings
138 (average 43.7 minutes; $t = 2.7848$, $df = 8.2491$, $p = 0.023$). There was no evidence that this mating
139 latency was due to evaporation of the octadecanal treatment, as there was no detectable drop in
140 octadecanal quantity in the hindwing androconia over time [comparison of 30 minute and 2 hour
141 treatments (Figure S9)]. Furthermore, little octadecanal rubbed off onto the forewing overlap region.

142 Interestingly, although about 5.5x as much octadecanal as normal should have been present on the
143 wings of treated males, only about 2-2.5x was seen after 30 minutes (the time at which the female
144 would be introduced to the two males), suggesting some of the added octadecanal was lost before the
145 start of the behavioral experiments, perhaps due to oxidization or pheromone hydrolysis, as has been
146 shown in some moths (27).

147 Genetic basis of octadecanal production in *H. melpomene*

148 Analysis of octadecanal production by F_1 males showed that the *H.* cydno octadecanal phenotype (little
149 to no octadecanal) is dominant over the octadecanal-rich *H. melpomene* phenotype, and octadecanal
150 production segregates in backcrosses to *H. melpomene* (Figure S10). Using the variance within the
151 backcross individuals, we calculated a Castle-Wright estimator of 0.81 loci, suggesting a potentially
152 monogenic basis for octadecanal production in *H. melpomene*. Quantitative trait locus (QTL) mapping
153 revealed a single significant peak on chromosome 20 (Figure 4a, Figure S11). The chromosome 20 peak
154 remained significant when kinship was taken into account and explained 41.31% of the difference
155 between the two parent species. Bayesian confidence intervals for the peak on chromosome 20 were
156 identical with and without kinship, spanning a range of 46.9-56.37cM with the peak at 47.66cM (with or
157 without kinship), corresponding to a physical range of 3.4Mb along chromosome 20. To ensure that our
158 findings were replicable across individual families, we also constructed effect plots at the kinship peak
159 for each family separately, and all showed the same directionality (Figure 4b). The peak on chromosome
160 20 does not overlap with any of the major wing color loci (21,28), nor does it overlap with mate choice
161 QTL (22). The confidence interval region contains 160 genes, all of which represent potential candidates
162 for octadecanal production.

163 We next evaluated the evidence in support of the 160 genes in this interval to identify top candidates
164 for octadecanal production. In total, 14 were putative fatty-acyl CoA reductases (FARs), which catalyze
165 the conversion of fatty acids to alcohols via a bound aldehyde intermediate (Table S5). Octadecanal is
166 most likely produced via this pathway (Figure 4c), either as a direct product of a FAR-catalyzed
167 conversion of 18-carbon stearic acid (by releasing the bound intermediate directly) or as a product of a
168 further dehydrogenation of the alcohol intermediate (octadecanol) to the aldehyde product. Two
169 candidate alcohol dehydrogenases, which might catalyze this reaction, were also contained within the
170 region, yielding a total of 16 candidates. To ascertain whether octadecanol might serve as the precursor
171 to octadecanal in *H. melpomene*, we also searched for QTL underlying octadecanol production (Castle-
172 Wright estimator of 0.71 loci), and found a very similar pattern to octadecanal, with a single QTL peak

173 on chromosome 20 (Figure 4d) explaining 25.36% of the difference between the two parent species.
174 This peak broadly overlapped the octadecanal peak, with a much broader confidence interval from
175 10.91-56.37cm (12.9Mb) and a peak at 51.82cM regardless of whether kinship was taken into account
176 (Figure S11). The 14 FARs in the region are highly clustered, with a set of eight found within a 133kb
177 region. Sequence comparison between the *H. melpomene* and *H. cydno* alleles showed that nearly all of
178 these genes harbor nonsynonymous SNPs between the two species (Table S5, Supplementary Data 1).
179 One gene showed no coding SNPs between the two species; nine had between three and 24
180 nonsynonymous SNPs; and four had more substantial changes, including missing exons and frameshifts.
181 The final two genes (one alcohol dehydrogenase and one FAR) could not be found in the *H. cydno*
182 genome, and may instead represent annotation or assembly errors in *H. melpomene* or, alternately,
183 deletions in *H. cydno*. Nearly all of the intact genes displayed purifying selection (ω between 0.001 and
184 0.2459), with only the remaining alcohol dehydrogenase (ω = 1.2478) under positive selection (Table
185 S5). Taken together, these results suggest that either of the alcohol dehydrogenase candidates may
186 underlie the production of bioactive octadecanal from octadecanol in *H. melpomene*, although
187 functional experiments are required to confirm this hypothesis. Alternately, as QTL for octadecanal and
188 its likely precursor octadecanol overlap, a single FAR may be responsible for producing both volatiles.

189 **Discussion**

190 Previous work has shown that male *Heliconius* butterflies use pheromones during courtship, and the
191 presence of these pheromones is necessary for successful mating to take place (8). Here we
192 demonstrate that the two closely related species with strong mating differences, *H. melpomene* and *H.*
193 *cydno*, show strong differences in chemical bouquets with a number of major compounds including
194 octadecanal, 1-octadecanol, (*Z*)-11-icosenal, and (*Z*)-11-icosenol found in high concentrations in *H.*
195 *melpomene* but largely absent in *H. cydno*, consistent with prior results (18,26). Somewhat surprisingly,
196 despite these strongly divergent pheromone signals, there was little difference in antennal response to
197 both overall wing chemical bouquets and various chemical components between the species. Indeed,
198 only octadecanal elicited a significant response in *H. melpomene*. Somewhat surprisingly, octadecanal
199 was also physiologically active in *H. cydno*, where it is largely absent from the male wing bouquet.

200 These data stand in marked contrast to similar studies in other insect species and suggest that the
201 peripheral nervous system architecture of *H. melpomene* and *H. cydno* has not diverged alongside their
202 wing pheromone components. By contrast, the moth *Ostrinia nubialis*, whose *E*- and *Z*-strains diverged
203 approximately 75,000-150,000 years ago (29), have opposite topologies in the antennal lobe and

204 antennal sensillae (30,31), and similar divergence in peripheral nervous system architecture has been
205 seen in *Rhagoletis pomonella* (32,33) and *Drosophila mojavensis* (34,35) despite much shorter
206 divergence times than between *H. melpomene* and *H. cydno*. Our results in *Heliconius* parallel what is
207 seen in *Colias* butterflies, where *C. eurytheme* and *C. philodice* show very similar female
208 electrophysiological responses to the con- and heterospecific pheromone compounds, despite a
209 behavioral effect of treating males with heterospecific pheromones (11). However, *Colias* use multiple
210 signals when choosing between mates (36), so rapid divergence in peripheral nervous system elements
211 may not be necessary for full reproductive isolation. A similar pattern most likely occurs in *Heliconius*,
212 with EAG responses very similar between species, and probable differences in the antennal lobe and
213 higher brain regions (e.g. the mushroom body or lateral protocerebrum) driving behavioral differences
214 (see e.g. (37)). Female *Heliconius* butterflies likely integrate multiple signals (including pheromones,
215 male courtship flights, and visual cues) in these higher brain regions when making the decision to mate.

216 We have identified octadecanal as the major pheromone component in *H. melpomene* and showed that
217 responses to it are conserved across both species despite its general absence in *H. cydno*. In the
218 Lepidoptera, the activity of octadecanal as a pheromone has been tested in eight species across a
219 variety of families (16,38,39,40,41,42,43,44,45). Only in *Cerconota anonella* (44), and now in *Heliconius*
220 *melpomene*, has some electrophysiological and behavioral activity been seen. The closely related
221 hexadecanal is a major pheromone component in the butterfly *Bicyclus anynana* (15), and differs from
222 octadecanal only in its origin from palmitic rather than stearic acid and carbon number, so octadecanal's
223 role as a pheromone in *Heliconius* is perhaps not entirely unexpected from a biochemical perspective. In
224 contrast, we have failed to discover the physiologically active pheromone(s) in *H. cydno*. This is not
225 entirely unexpected, as we focused on the most abundant compounds, but minor components have
226 been shown to be important in other systems (19,20).

227 Intriguingly, despite the strong EAG response, there was a marked negative behavioral response to
228 increased octadecanal. A plausible explanation for the negative behavioral response to octadecanal
229 supplementation is that disruption of the normal mixture ratios of *H. melpomene* may inhibit the female
230 response, as seen in the butterfly *Pieris napi*, where synergistic processing of two volatile components in
231 the male bouquet is necessary for acceptance behavior (46). Octadecanal may also experience a dose-
232 response curve with an aversive response to higher concentrations and an attractive response at lower
233 ones. Potential mixture or dosage effects suggest that female *H. melpomene* may use octadecanal
234 quantity or relative abundance to assess male quality or choose between courting males. The increased

235 mating latency with octadecanal-treated males may reflect females undergoing a period of adjustment,
236 either in the peripheral or central nervous system, to the higher dose of octadecanal; this would be
237 consistent with our results showing long-term adaptation to octadecanal. We remain uncertain of what
238 effect, if any, octadecanal would have on the behavior of *H. cydno*, where it is largely absent from the
239 male pheromone bouquet. It may be used to avoid courtship with *H. melpomene*, supporting other
240 divergent signals in maintaining reproductive isolation.

241 Given the strong physiological and behavioral response to octadecanal, and its possible role in
242 reproductive isolation between *H. melpomene* and *H. cydno*, we sought to understand its genetic
243 underpinnings. Fatty-acid derived compounds comprise the largest category of Lepidoptera sex
244 pheromones (47), and are produced from fatty acyl-CoA precursors via the action of several enzymes.
245 Since these pheromones are secondary metabolites derived from primary metabolic pathways, their
246 production is likely to be relatively labile in evolutionary terms, allowing simple genetic changes to drive
247 the wide diversity of lepidopteran sex pheromones. Even though we have a broad knowledge of
248 pheromone diversity in Lepidoptera, our understanding of the genetics of pheromone biosynthesis is
249 relatively weak. To date, only a handful of genes have been identified, mostly desaturases and fatty acyl-
250 CoA reductases (48).

251 Using a QTL mapping approach, we have shown that the production of octadecanal has a relatively
252 simple genetic basis, with a region on chromosome 20 corresponding to production of both octadecanal
253 and its likely precursor octadecanol in *Heliconius*. This locus therefore likely represents a region under
254 divergent selection between *H. melpomene* and *H. cydno* that is unlinked to previously identified species
255 differences in color and mate choice (21,22). Patterns of F_{ST} between the species are highly
256 heterogeneous and were not especially informative in further delimiting the locus (data from (49)). Due
257 to our small mapping population, the confidence intervals for these QTL therefore remain large: the
258 octadecanal QTL spans 3.4Mb and contains 160 genes. Of these, we identified 16 likely candidate genes
259 based on known biosynthetic pathways in moths and the butterfly *Bicyclus anynana* (17): 14 fatty acyl-
260 CoA reductases and two alcohol dehydrogenases. Fatty acyl-CoA reductases have previously been
261 identified in *H. melpomene* by (17), who noted lineage-specific duplications within *H. melpomene* on
262 two scaffolds corresponding to *H. melpomene* chromosomes 19 and 20. All but one of the candidate
263 FARs found on chromosome 20 were identified by Lienard et al, but all fall outside their clade of
264 pheromone gland FARs. The identified *Bicyclus* FAR that produces hexadecanol does not also produce
265 the major pheromone hexadecanal, implying the presence of an additional as yet undescribed alcohol

266 dehydrogenase in *Bicyclus*, and the biochemical similarity between hexadecanal and octadecanal
267 suggests *Heliconius* may also use an alcohol dehydrogenase to produce octadecanal. By contrast, the
268 overlapping octadecanol and octadecanal QTL on chromosome 20 in *Heliconius* suggest the presence of
269 a bifunctional FAR that produces both the alcohol and aldehyde together. Further studies, including
270 functional assays and location of wing pheromone biosynthesis, will be required to tease apart our
271 potential candidates.

272 Our studies of the electrophysiological and behavioral responses of *Heliconius* butterflies and the
273 genetic basis of pheromone production add to the growing body of literature suggesting that
274 pheromonal communication in Lepidoptera is not limited to nocturnal moths but can be found in day-
275 flying butterflies that also use striking visual signals. *Heliconius* butterflies can detect con- and
276 heterospecific wing compound bouquets, and a major component, octadecanal, is physiologically and
277 behaviorally active in *H. melpomene* and its genetic basis appears relatively simple, consistent with
278 other pheromone shifts found in insects (50). Along with their striking wing color patterns, male
279 *Heliconius* use chemistry to affect female mate choice, combining courtship behaviors, and chemistry in
280 a dance to elicit female mating responses (7,51). Despite our human bias towards visual signals, we are
281 now beginning to understand how such visually striking butterflies communicate using chemistry.

282

283 **Methods**

284 Data and analysis code are deposited in Dryad under accession XXX. Sequencing reads leading to linkage
285 map construction are deposited in the European Nucleotide Archive (ENA) under accession YYY.

286 <u>Butterflies</u>

287 Stocks from central Panama of *Heliconius melpomene rosina* and *H. cydno chioneus* (hereafter *H.*
288 *melpomene* and *H. cydno*) were used for all experiments (8). Butterflies were reared in insectaries at the
289 Smithsonian Tropical Research Institute, Gamboa, Panama under ambient temperature and light
290 conditions. Eggs were collected from these breeding stocks and the resulting larvae fed on *Passiflora*
291 *platyloba var. williamsi*, *P. biflora*, *P. menispermifolia*, and *P. vitifolia* until pupation. Data from (52)
292 indicate that larval diet does not affect the major compounds found in *H. melpomene*, suggesting that
293 this dietary variation is unlikely to affect results. Newly-eclosed adult butterflies were separated by sex
294 to ensure virgin status and supplied with flowers from *Psiguria warscewiczii*, *Psiguria triphylla*, *Gurania*

295 *eriantha*, *Psychotria poeppigiana*, *Stachytarpheta mutabilis*, and *Lantana* sp. (most likely *L. camara*) as
296 pollen sources, as well as a ~20% sucrose solution as a nectar source. All experiments used virgin
297 butterflies. For assessment of *H. cydno* wing bouquets, male butterflies were between 10-12 days post
298 eclosion and had fed exclusively on *Passiflora platyloba var. williamsi*. For electrophysiology, female
299 butterflies were between 1 and 20 days post eclosion, and males between 10-20 days post eclosion to
300 ensure male sexual maturity (8). For behavior, female butterflies were used the day after eclosion and
301 males were between 10 and 59 days post eclosion. Natural wing extracts of both species were extracted
302 from males 10-12 days post eclosion as described in (8) using dichloromethane plus 1 ng/µL 2-tetradecyl
303 acetate (hereafter "DCM+IS") and concentrated approximately 10x prior to use under room air. All
304 samples were stored at -20°C before use.

305 Identification and quantification of androconial compounds

306 To identify species specific compounds among our two species, the chemical composition of the *H.*
307 *cydno* androconial bouquet was investigated in samples from 26 adult male *H. cydno* and compared with
308 31 adult male *H. melpomene*, the latter including samples previously analyzed in (8,52), all collected as
309 above. Samples were assessed using gas chromatography-mass spectrometry (GC-MS) with an Agilent
310 7890B gas chromatograph coupled with an Agilent 5977 mass spectrometer with electron ionization
311 (Agilent Technologies, California, USA). The GC utilized an Agilent HP-5MS capillary column (30m length,
312 0.25mm inner diameter), helium carrier gas at 1.2 mL/minute, and an Agilent ALS 7693 autosampler.
313 Injection was splitless with an inlet temperature of 250°C. The temperature ramp was isothermal at 50°C
314 for 5 minutes, then increased at 5°C/minute to 320°C and was then isothermal for 5 minutes. Samples
315 were identified using a custom MS library and quantified by comparison with the internal standard.

316 In line with (8), wings from eight male *H. cydno* were dissected into four regions: hindwing androconia,
317 forewing overlap region, hindwing rest-of-wing, and forewing rest-of-wing, and all extracted identically
318 after dissection. Wing region area was quantified by photographing the wings before dissection and
319 measuring the total pixel area of each wing region in the GNU Image Manipulation Program v.2.8.20
320 (GIMP Development Team), with the pixel-mm conversion via measurement of a ruler in the
321 photograph. Quantified compounds in each wing region for each individual were scaled by the area of
322 the relevant wing region in that individual.

323 Chemicals

324 Syringaldehyde (3,5-dimethoxy-4-hydroxybenzaldehyde), 1-octadecanol, and henicosane were obtained
325 commercially (Sigma-Aldrich). The aldehydes octadecanal, (*Z*)-11-icosenal, and (*Z*)-13-docosenal were
326 obtained from the respective alcohols 1-octadecanol, (*Z*)-11-icosen-1-ol, and (*Z*)-13-docosen-1-ol by
327 oxidation with iodoxybenzoic acid (IBX) in ethyl acetate according to (53). The required alcohols (*Z*)-11-
328 icosen-1-ol and (*Z*)-13-docosen-1-ol were in turn obtained by lithium aluminum hydride reduction from
329 commercially available (*Z*)-11-icosenoic acid and methyl (*Z*)-13-docosenoate (Larodan) according to (54).
330 The seven target compounds (see Figures S4 and S5 for structures and reaction scheme) were chosen
331 due to their quantitative dominance in the chemical profiles of *H. melpomene* and *H. cydno*. The solvent
332 for all synthesized compounds was hexane, with the exception of the polar syringaldehyde, which was
333 diluted in a 1:10 mixture of dichloromethane and hexane.

334 Synthetic blends of *H. melpomene* and *H. cydno* male wing bouquets were prepared from these
335 synthesized compounds. Synthetic *H. melpomene* contained 23.22 ng/µL syringaldehyde, 23.33 ng/µL
336 octadecanal, 6.92 ng/µL 1-octadecanol, 4.67 ng/µL (*Z*)-11-icosenal, 20.33 ng/µL (*Z*)-11-icosenol, and 4.83
337 ng/µL (*Z*)-13-docosenal. Synthetic *H. cydno* contained 47 ng/µL syringaldehyde and 93.33 ng/µL
338 henicosane. Floral direct extractions from *Lantana* sp. (most likely *L. camara*) growing wild in Gamboa,
339 Panama were used as a positive control. Single umbels were removed from plants at dawn and placed in
340 a scintillation vial to which 400uL of DCM+IS was added. After 1 hour, the DCM+IS was removed to a
341 glass vial and kept at -20°C before use.

342 <u>Electroantennography</u>

343 Electrophysiological preparations were assembled as follows: antennae were cut from the head of a
344 virgin butterfly using fine scissors and the final 6.5 segments (approximately 1.5mm) cut off with a
345 scalpel. Both antennae were then placed in parallel across an antenna fork (Syntech, Buchenbach,
346 Germany) and held in place using electrode gel (Spectra 360 Electrode Gel, Parker Laboratories Inc.,
347 Fairfield, NJ, USA). The antenna fork was mounted on a Syntech EAG CombiProbe with 10x internal gain,
348 and signal from this was routed through a Syntech IDAC4. EAG waveforms recorded using Syntech
349 GcEad/2014 software. Stimulus pulses were delivered using a Syntech CS-55 Stimulus Controller with
350 foot pedal trigger. Both continuous unhumified room air and stimulus pulses were delivered to the
351 preparation at 1.5 liters/min. The stimulus pulses were delivered in triplets of 0.5 seconds each,
352 separated by 5 seconds, with triplets initiated every 30 seconds. Stimulus delivery used odor cartridges
353 assembled from Pasteur pipettes with a strip of filter paper plugged with cotton when not in use; each

354 stimulus cartridge was 'charged' with 10uL of stimulus solution for each experiment. Each antennal
355 preparation was used only once.

356 Two sets of stimuli were delivered: a species comparison set and a synthetic compound set. Both sets
357 used air (nothing added to filter paper), hexane, and DCM+IS as negative controls and *Lantana* extract as
358 a positive control. The species comparison set included male wing extracts from *H. melpomene* and *H.*
359 *cydno* ("Mnat" and "Cnat" respectively) and synthetic blends representing the two species ("Msyn" and
360 Csyn"). The synthetic compound set included air, hexane, DCM+IS, the conspecific male wing extract,
361 the conspecific synthetic blend, and the seven synthetic compounds. Presentation order was
362 randomized before each experiment. Species comparison experiments consisted of sixteen pulses each
363 of the seven stimuli, interspersed with five pulses of *Lantana* extract at the start, between every three
364 stimuli, and at the end. Synthetic compounds were similar, with eleven pulses of each of the twelve
365 stimuli, interspersed with four pulses of *Lantana* at the start, between every four stimuli, and at the end.
366 For analysis, the first triplet of each stimulus set was removed, leaving fifteen and ten pulses
367 respectively. At least ten female and five male butterflies of each species were used with each
368 experiment.

369 Onset time and amplitude of EAG responses were marked using GcEad/2014. To control for antennal
370 signal degradation over time, a time-dependent correction factor was calculated using linear
371 interpolation between each *Lantana* set and this applied to the absolute value of the EAG response
372 amplitude. These corrected amplitudes were then scaled to the amplitude of the initial *Lantana* set to
373 partially control for differences between preparations.

374 <u>Analysis of antennal adaptation</u>

375 Short-term adaptation (STA), as defined in (55), is adaptation occurring only over a very brief window
376 from the initial stimulus that then resolves quickly with no further stimulation (e.g. the 10 second
377 interval given for salamanders). By contrast, long-term or long-lasting adaptation (LTA) are defined in
378 the same source as persisting over an extended period of time, up to several minutes in vertebrates and
379 insects (56). Within our electrophysiological data set, there is the potential to measure both types of
380 adaptation. STA, if present, should be evident within an individual triplet, as the stimuli within a triplet
381 are separated by 5 seconds (and thus STA is likely to persist within a triplet), whereas LTA should be
382 evident across an individual stimulus set, as these lasted 5.5-8 minutes with maximum intervals of 30
383 seconds between stimuli, insufficient for LTA to be abolished if we assume similar mechanisms as

384 vertebrates. We assessed antennal responses for LTA by pooling all triplets within a stimulus-species-sex
385 combination. We pooled these data within stimulus set types (species comparison and synthetic
386 compound), treating butterfly preparation identity as a random effect. For each stimulus, the change in
387 antennal response was assessed over the time since initial presentation of the stimulus. STA was
388 assessed by looking for significant changes in the residuals from this analysis between members of a
389 triplet.

390 Behavior

391 To test the potential role of octadecanal in *H. melpomene* female mate choice, behavioral experiments
392 were conducted in insectaries at STRI, Gamboa, Panama, between April and July 2018. One day old
393 virgin females were presented with both an octadecanal treated and a control *H. melpomene* male for
394 two hours. Males were at least ten days old and were selected based on similarity of size, wing-wear
395 and age, with treatment being allocated randomly by coin flip. Either 25 µl octadecanal (140 ng/µl
396 octadecanal in hexane, thus adding 3500ng to the existing average 773.4ng for approximately a 5.5x
397 dose) or 25 µl pure hexane (both evaporated down to a smaller volume under room air) was applied to
398 the hindwing androconial region of each male, and males were then allowed 30 minutes to settle before
399 beginning the two hour experiment period. Experiments began at or close to 9am, with observations
400 being made every 15 minutes or until mating occurred.

401 To test the persistence of the octadecanal treatment on the wings of live butterflies, a separate set of *H.*
402 *melpomene* males was treated as above with either hexane or octadecanal. Separate males were
403 sampled at 30 minutes post treatment and two hours post treatment by extraction of the forewing
404 overlap region (8) and the hindwing androconia in DCM+IS as above, with two males per treatment-time
405 combination. Octadecanal was then measured using GC-MS as above and quantified by peak area
406 comparison with the 2-tetradecyl acetate internal standard.

407 Quantitative trait locus mapping for octadecanal production

408 To map the genetic basis for octadecanal production in *H. melpomene*, we took advantage of the fact
409 that *H. cydno* produces little to no octadecanal. Bidirectional F_1 crosses between the two species
410 revealed that the *H. cydno* phenotype (low to no octadecanal) is dominant over the high octadecanal
411 production found in *H. melpomene*, so we constructed backcross families by crossing F_1 males to female
412 *H. melpomene* from our existing stocks. A total of ten families (nine with a female *H. melpomene*

413 grandparent and one with a female *H. cydno* grandparent) were constructed, with each offspring
414 representing a single recombination event from the F_1 father. Butterflies were reared and wing extracts
415 collected and analyzed from male offspring as described above, except that all larvae were reared on
416 *Passiflora platyloba var. williamsi*. Bodies of male offspring were collected into dimethyl sulfoxide
417 (DMSO) for later library preparation. The Castle-Wright estimators for octadecanal and octadecanol
418 production were calculated using the phenotypic variance of the backcross individuals as the estimated
419 segregation variance (57).

420 Qiagen DNeasy kits (Qiagen) were used for DNA extraction. Individuals were genotyped either by RAD-
421 sequencing as previously described (22,58), or low-coverage whole genome sequencing using nextera-
422 based libraries. For the nextera-based libraries a secondary purification was performed using magnetic
423 SpeedBeads™ (Sigma) dissolved in 0.44mM PEG8000, 2.5M NaCl, 1mM Tris-Cl pH=8, and 0.1mM EDTA
424 pH=8.0. High-throughput sequencing libraries were generated using a method based on the Nextera
425 DNA Library Prep (Illumina, Inc.) with purified Tn5 transposase (59). Sample barcoding was performed
426 using PCR extension with an i7-index primer (N701–N783) and the N501 i5-index primer. Libraries were
427 purified and size selected using the same beads as above. Pooled libraries were sequenced by HiSeq
428 3000 (Illumina) by BGI (China).

429 Linkage mapping was conducted using standard Lep-MAP3(LM3) pipeline (60) . First, individual fastq
430 files were mapped to the melpomene reference genome using BWA MEM (61) and then sorted bams
431 were created using SAMtools (62). The input genotype likelihoods were constructed by SAMtools
432 mpileup and pileupParser2+pileup2posterior from LM3. The pedigree of individuals was validated and
433 corrected using IBD (identity-by-descent) module and the sex of individuals was validated and corrected
434 according to the coverage on the Z chromosome and autosomes using SAMtools depth. Then,
435 ParentCall2 (parameter ZLimit=2) and Filtering2 (dataTolerance=0.001) modules were called on the
436 input data and a random subset of 25% of markers (to speed up analysis) was used for the subsequent
437 steps.

438 Initial linkage groups (chrX.map) and marker orders (orderX.txt) were constructed based on melpomene
439 genome for each of 21 chromosomes. SeparateChromosomes2 was run on each of these groups with
440 the default parameters (lodLimit=10) except for map=chrX.map (for X=1..21). Finally, OrderMarkers2
441 was run on each chromosome in the constructed order, with parameter scale=0.05, recombination2=0,
442 evaluateOrder=orderX.txt and map=result_from_SeparateChromsomes2.chrX.txt. Another evaluation
443 was done with data in the grandparental phase (additional parameter grandparentPhase=1). The phased

444 data of these orders were matched using phasematch script (LM3) and obtaining all markers from the
445 first evaluation in the grandparental phase. This obtained result was used as the final map.

446 Map construction resulted in the retention of 447,818 SNP markers across 89 individuals with
447 phenotype data. To facilitate computation, markers were thinned evenly by a factor of ten, resulting in
448 44,782 markers with no missing data. Octadecanal production was log-transformed to obtain normality,
449 then regressed against marker position using the R/qtl2 R library (63). Significance thresholds were
450 obtained by permutation testing in R/qtl2 with 1000 permutations, and QTL confidence intervals
451 obtained using the bayes_int command. To account for the family structure present in our QTL mapping
452 populations, we additionally included a kinship matrix calculated by R/qtl2 using the LOCO (leave one
453 chromosome out) method in the marker regression and recalculated significance thresholds and
454 confidence intervals with the kinship term included. Percent variance explained was calculated as the
455 difference in phenotype means of individuals of each genotype divided by the difference in the parental
456 phenotype. Since the genetic linkage map was based on whole genome data, we were able to obtain
457 physical positions of QTL confidence interval endpoints. The physical positions of the kinship-included
458 confidence interval were used to query Lepbase (64) for potential candidate genes from the *H.*
459 *melpomene* genome. To identify putative functions for each potential candidate, protein sequences
460 from Lepbase were searched against the nr (non-redundant) protein database using BLASTp (65). For
461 each candidate with a promising functional annotation, exons were pulled out of the *H. cydno* genome
462 (66) and aligned to the *H. melpomene* genes using BLASTn with each *H. melpomene* exon to search for
463 SNPs between the two species. Selection was tested at the sequence level using codeml (67) in pairwise
464 mode with the F3X4 codon frequency model.

465 Statistical analysis

466 Differences in individual compounds between *H. melpomene* and *H. cydno* were assessed using a
467 Welch's *t*-test. Wing region differences in *H. cydno* were assessed for each individual compound found
468 in at least four of eight samples of at least one wing region using a linear mixed model with wing region
469 as a fixed effect and butterfly identity as a random effect using the package nlme v.3.1.137 (68).
470 Statistical tests for these models were assessed with the Anova function in the package car v.3.0.0 (69),
471 and comparisons between wing regions were performed using the emmeans package v.1.3.1 (70).

472 For electroantennography, species comparison sets and synthetic compound sets were analyzed
473 separately. Corrected and scaled EAG responses (for each experiment within sexes and species) were

474 compared between stimuli with a linear mixed model for *H. melpomene* with stimulus as a fixed effect
475 and butterfly preparation identity as a random effect using nlme as above. EAG responses for *H. cydno*
476 were better represented as an inverse gamma distribution, so were assessed in the same fashion with a
477 generalized linear mixed model (glmmPQL) using the package MASS v.7.3.50 (71). Statistical tests for
478 these models were assessed with the Anova and emmeans (version 1.1.3) functions as above.

479 Long-term adaptation was assessed using robust linear mixed models with the package robustlmm
480 v.2.2.1 (72), with corrected but unscaled amplitude as the response variable, time since initial
481 presentation of the stimulus as a fixed variable, and butterfly preparation as a random variable.
482 Responses were pooled across samples within a species-sex combination and considered for each
483 stimulus separately. As robustlmm does not provide p-values, confidence intervals were used to assess
484 significance and difference between sample-species combinations. Short-term adaptation was assessed
485 using the residuals from the same regression. Differences in the residuals between triplets within a
486 triplet set were tested using a one-sample t-test with a hypothesized mean value of zero (i.e. no
487 difference between residuals), performed on the subtractive difference between the residuals of the
488 third (last) and first triplets.

489 Female mate choice was assessed using a binomial test, and treatment differences in time until mating
490 were assessed with a two-sided t-test assuming unequal variances. Octadecanal persistence was not
491 assessed statistically due to the small sample size. All statistical tests were performed in R v.3.5.0,
492 v.3.5.1, or v.3.5.2 (73).

493

494 **Acknowledgments**

495 The authors wish to thank Bill Wcislo and Callum Kingwell for advice on electroantennography. We also
496 thank the insectary team in Panama including Oscar Paneso and Chi-Yun Kuo, as well as the
497 administrative support of Oris Acevedo and Melissa Cano. We also thank S. Dilek for technical
498 assistance. Permits for research and collection of butterfly stocks were provided by the government of
499 Panama. KJRPB, KD, JM, IAW, and CDJ were funded by the European Research Council (FP7-IDEAS-ERC
500 339873); KD was additionally funded by a Natural Research Council Doctoral Training Partnership and a
501 Smithsonian Tropical Research Institute Short Term Fellowship; YFC and MK are supported by the
502 European Research Council Grant No. 639096 "HybridMiX" and by the Max Planck Society; WOM is

503 funded by the Smithsonian Tropical Research Institute; and SS by the Deutsche Forschungsgemeinschaft
504 (grant DFG Schu984/13-1).

505 **Author contributions**

506 KJRPB designed, performed, and analyzed electrophysiology experiments, performed QTL mapping and
507 downstream analysis, and wrote the manuscript. KD designed and performed QTL mapping crosses,
508 designed behavioral experiments, and prepared sequencing libraries. KD, SFG, DAA, KJRPB, and JM
509 reared butterflies for all experiments. JM performed and analyzed behavioral experiments. IAW
510 prepared sequencing libraries. YFC and MK provided Tn5 transposase for sequencing. SS provided
511 synthetic compounds used in electrophysiology and behavior experiments as well as mass spectrometry
512 libraries. PR analyzed sequencing data and constructed linkage maps. RMM helped design the QTL
513 experiment. CDJ and WOM contributed funding, resources, and helped design the overall study. All
514 authors contributed to the manuscript editing process.

515 **Tables**

516 Table 1: Electroantennographic responses of the different sex-species combinations to individual
517 compounds compared with hexane (P2-7) or hexane and dichloromethane (P1). Only significant
518 alterations are shown. No asterisk: $p \leq 0.05$; *: $p \leq 0.01$; **: $p \leq 0.001$; ***: $p \leq 0.0001$. P1:
519 syringaldehyde, P2: octadecanal, P3: 1-octadecanol, P4: henicosane, P5: (*Z*)-11-icosenal, P6: (*Z*)-11-
520 icosenol, P7: (*Z*)-13-docosenal. *Heliconius melpomene* wing bouquet contains P1-7; *H. cydno* wing
521 bouquet contains P1, P4, and P6 (Table S1). Artificial *H. melpomene* wing blend contains the dominant
522 compounds P1-P3 and P5-P7; artificial *H. cydno* wing blend contains dominant compounds P1 and P4.

Species	**Sex**	**n**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
H. melpomene	Female	11	-	↑***	-	-	-	-	-
H. melpomene	Male	5	-	↑***	-	-	↓**	-	-
H. cydno	Female	10	-	↑***	-	-	↓	↓***	↓
H. cydno	Male	5	-	↑***	↑*	↑***	-	↑*	-

523

524 SI Table 1: Comparison of *Heliconius cydno* and *H. melpomene* wing androconia bouquet. Amounts given
525 are the mean in nanograms across 31 (*H. melpomene*) and 26 (*H. cydno*) samples, ± SD. Compounds are
526 only included if found in at least 1/3 of samples for at least one species at levels of at least 1ng. #:

527 compound found in at least 90% of samples of that species. Bold: species differ significantly. dof:
528 Welch's two sample t-test degrees of freedom.

529 SI Table 2: Wing region specificity of *Heliconius cydno* androconial compounds. Amounts given are the
530 mean across eight samples ± SD. Numbers in parentheses after compound amounts indicate the number
531 of samples the compound was found in. Bold: wing regions differ significantly. NA: post-hoc test not
532 conducted as original linear model not significant.

533 SI Table 3: Details of statistical comparisons of stimuli from electrophysiological experiments. P1-P7: as
534 in the main text and Figure 4.

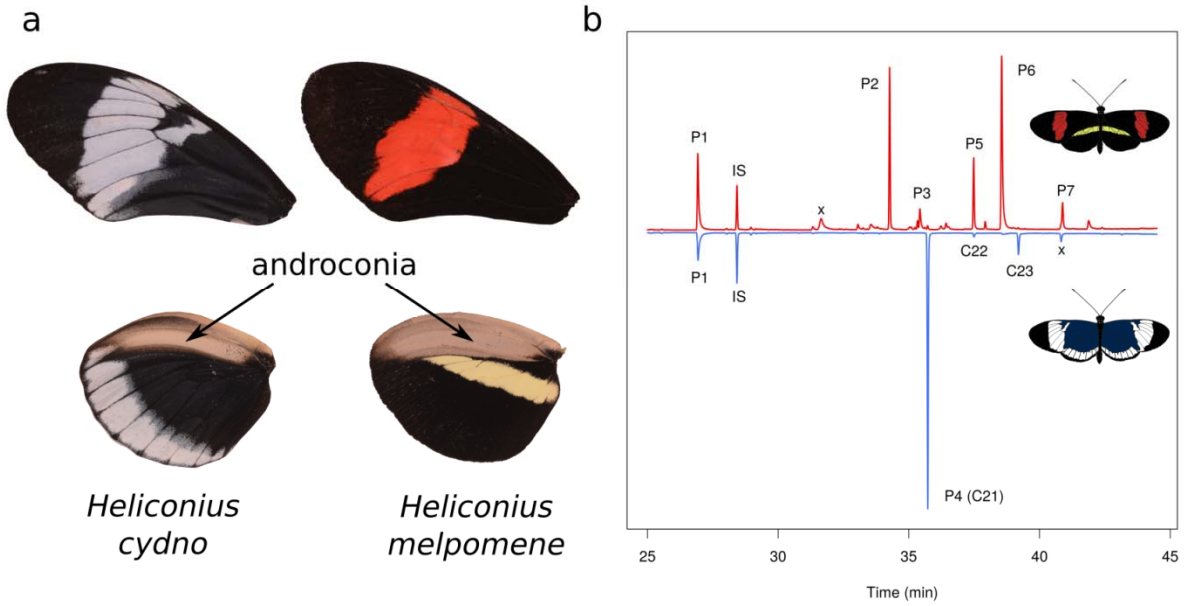
535 SI Table 4: Short-term adaptation to different stimuli in *H. melpomene* and *H. cydno* males and females.
536 t1, t2, and t3: first, second, and third members of a stimulus triplet. dof: degrees of freedom. P1-P7: as
537 in the main text and Figure 4.

538 SI Table 5: Potential candidate genes for octadecanal production underlying the QTL peak on
539 chromosome 20 in *Heliconius melpomene*.

540

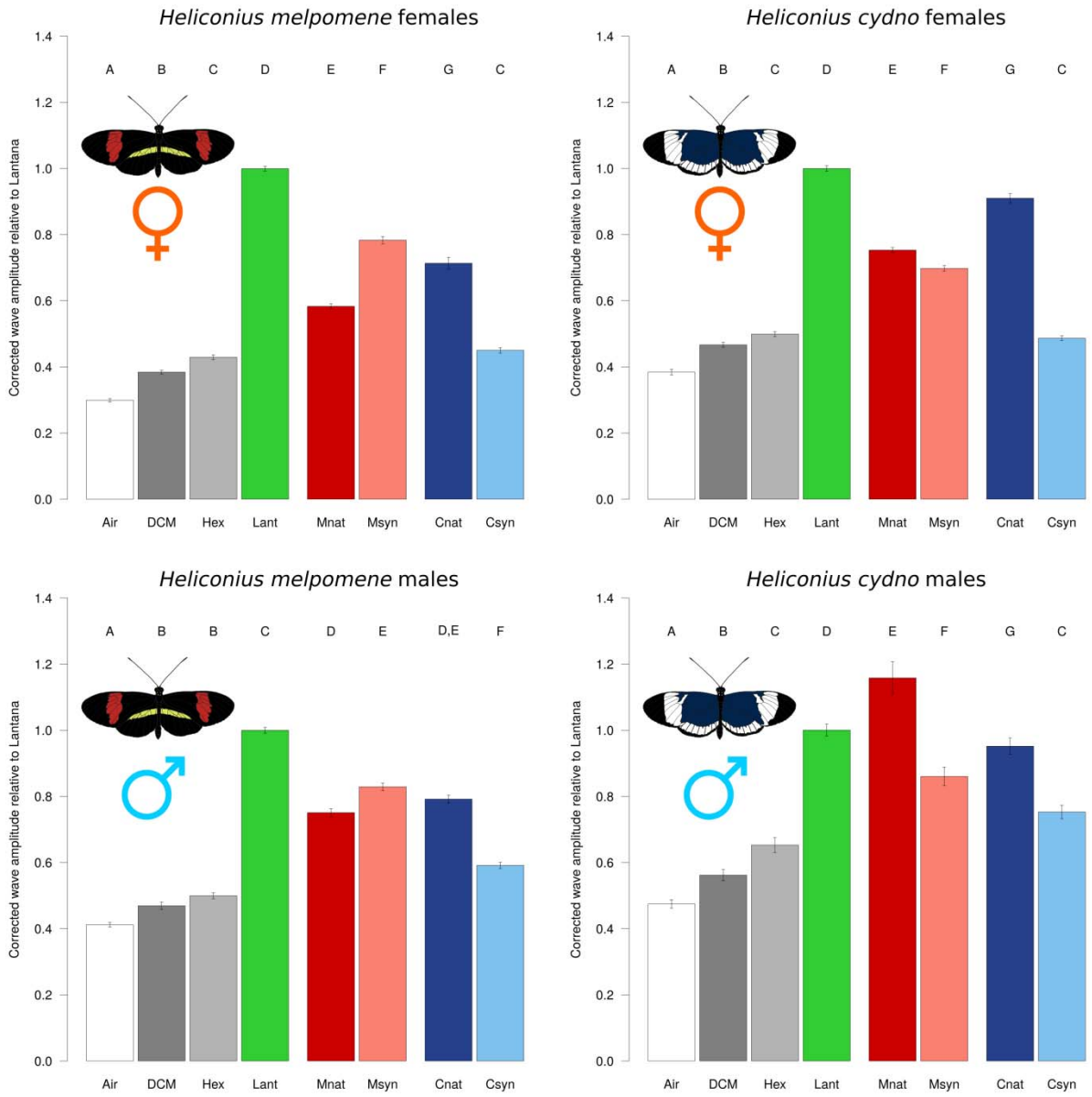
541 **Figures**

542 Figure 1: Chemical ecology of *Heliconius melpomene* and *H. cydno*. a: dorsal forewing and hindwing of
543 each species showing the silvery androconial region of the hindwing used during male courtship. b: Total
544 ion chromatogram of *H. melpomene* (top) and *H. cydno* (bottom) wing androconia. P1: syringaldehyde;
545 P2: octadecanal; P3: 1-octadecanol; P4: henicosane; P5: (*Z*)-11-icosenal; P6: (*Z*)-11-icosenol; P7: (*Z*)-13-
546 docosenal; IS: internal standard (2-tetradecyl acetate); x: contaminant; C21: henicosane; C22: docosane;
547 C23: tricosane.



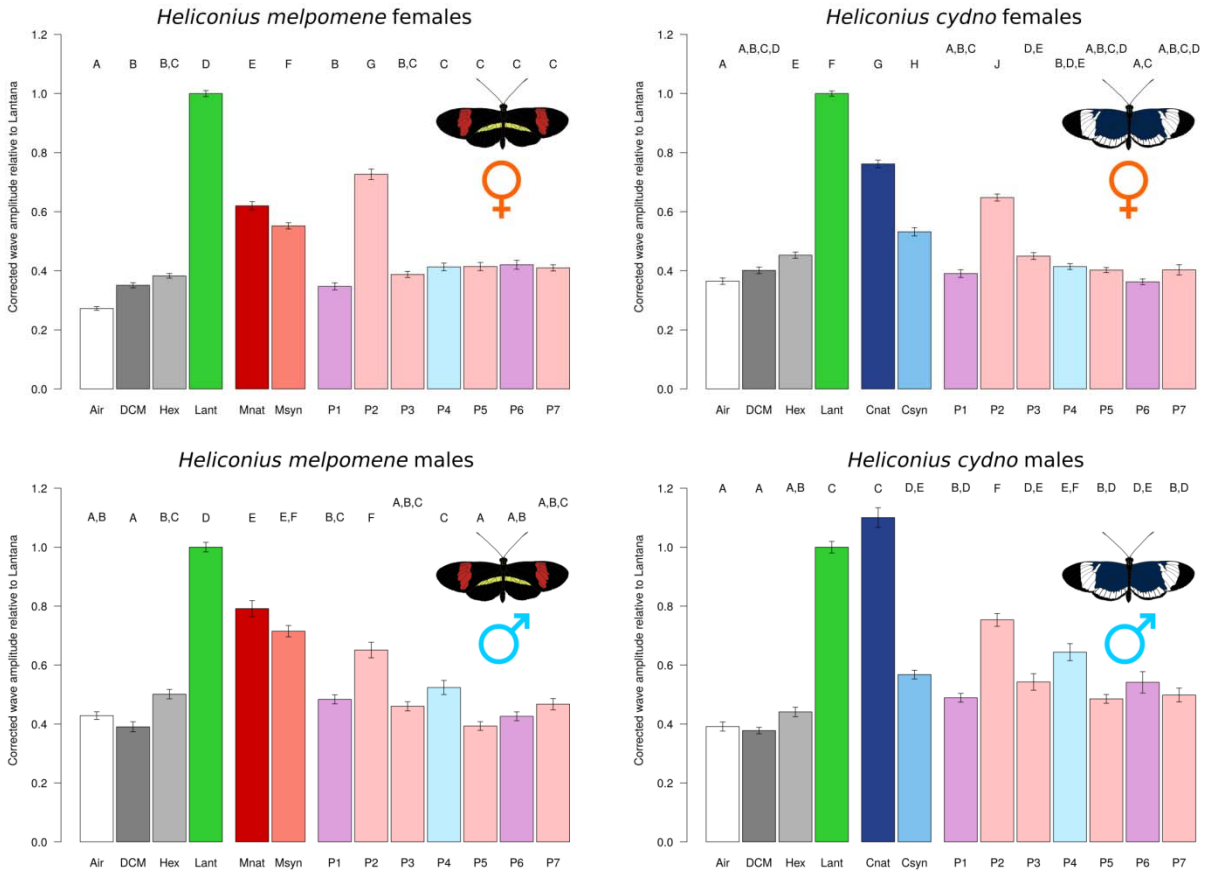
548

549 Figure 2: Electroantennographic responses of *Heliconius* butterflies to conspecific and heterospecific
550 wing extracts. Stimuli: air, dichloromethane plus 2-tetradecyl acetate (internal standard), hexane,
551 *Lantana* extract, natural male *H. melpomene* wing extract, synthetic *H. melpomene* blend, natural male
552 *H. cydno* wing extract, synthetic *H. cydno* blend. Bars: average of normalized corrected amplitude ± SEM.



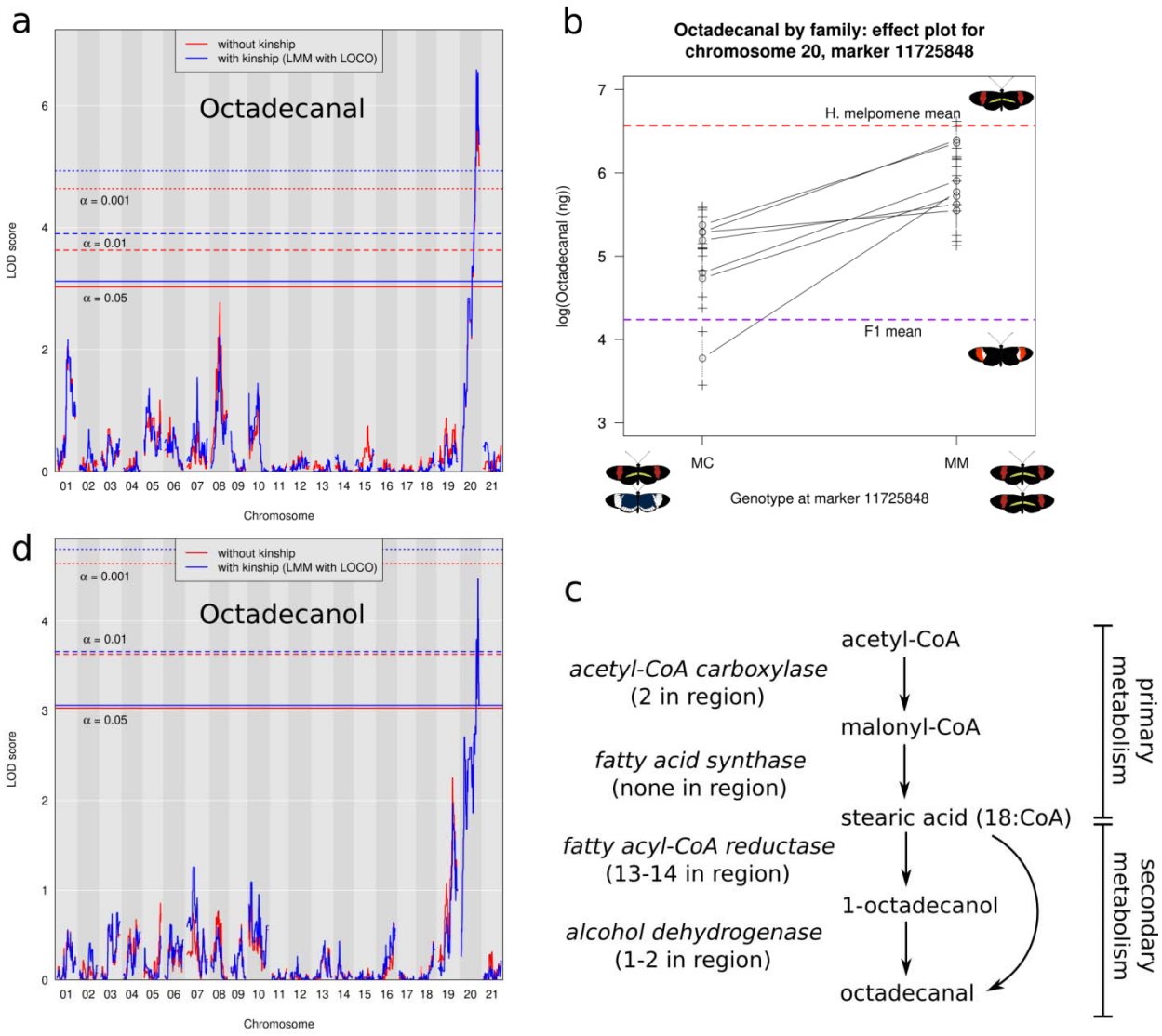
553

554 Figure 3: Electroantennographic responses of *Heliconius* butterflies to synthetic wing compounds.
555 Stimuli: air, dichloromethane plus 2-tetradecyl acetate (internal standard), hexane, *Lantana* extract,
556 natural male *H. melpomene* or *H. cydno* wing extract, synthetic *H. melpomene* or *H. cydno* blend. P1-P7
557 as in Figure 1. Bars: average of normalized corrected amplitude ± SEM. Light red/light blue: compound
558 found in all *H. melpomene* (red) or *H. cydno* (blue) and >1% of natural bouquet makeup; purple:
559 compound found in both species and >1% of natural bouquet makeup in at least one species.



560

561 Figure 4: QTL mapping of octadecanal and octadecanol production in *Heliconius melpomene*. a: QTL map
562 for production of octadecanal. b: Effect plots at the peak of the locus on chromosome 20 for the seven
563 individual backcross mapping families with at least 5 individuals. c: Potential biosynthetic pathway for
564 octadecanal production. d: QTL map for production of octadecanol, a potential precursor of
565 octadecanal.



566

567 SI Figure 1: Comparison of the major compounds in *Heliconius melpomene* and *H. cydno*. Compounds
568 shown are those contributing at least 1% of the total bouquet amount in either species. Numbers under
569 each bar indicate how many samples (out of 31 for *H. melpomene* and 26 for *H. cydno*) the compound
570 was found in. Significantly different compounds: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. n.s., not
571 significant.

572 SI Figure 2: Absolute and relative abundance of different compound classes in *H. melpomene* and *H.*
573 *cydno*. n.s., not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. The two unknown categories were not
574 tested as the compound types are not known.

575 SI Figure 3: The seven compounds found in at least 0.1ng/mm^2 of wing tissue in at least one wing region
576 in *Heliconius cydno*. a: Compound abundance per square millimeter of tissue. b: Compound abundance
577 without tissue area correction. Numbers under each bar indicate how many samples (out of eight) the
578 compound was found in; letters above bars indicate significant differences between regions. n.s., not
579 significant. A: hindwing androconia; O: forewing overlap region; H: hindwing excluding androconia; F:
580 forewing excluding overlap region.

581 SI Figure 4: Structures and names of major components of the androconia of *H. melpomene* and *H.*
582 *cydno* used in electrophysiological experiments.

583 SI Figure 5: Synthesis of target compounds used in electrophysiological experiments. IBX: iodosobenzoic
584 acid; LiAlH: lithium aluminum hydride.

585 SI Figure 6: *Heliconius melpomene* responds to electrophysiological stimuli. Top to bottom:
586 dichloromethane plus 2-tetradecyl acetate (internal standard) (negative control), *Lantana* extract
587 (positive control), natural *H. melpomene* male wing extract, depiction of stimulus pulse timing. Data
588 from a single virgin female.

589 SI Figure 7: Long-term adaptation to natural and synthetic stimuli in *Heliconius* butterflies. The 95%
590 confidence intervals of the robust LMM slope are shown; a negative slope means that responses to that
591 stimulus drop over time. P1-P7: see Figure 2.

592 SI Figure 8: Strength of long-term adaptation correlates with amplitude of EAG response in a sex-specific
593 fashion. In females, a stronger response to a given stimulus correlates with a stronger degree of LTA
594 both overall and for the synthetic compound set. In males the same is seen overall and for the natural
595 extract set in *H. cydno*, but not in *H. melpomene*.

596 SI Figure 9: Octadecanal persistence in treated males over time. Bars show individual males, with two
597 males per treatment-time point combination. Hindwing androconia and forewing overlap region males
598 are shown in the same positions.

599 SI Figure 10: Octadecanal in *H. melpomene*, *H. cydno*, two F_1 families (one in each crossing direction),
600 and the ten backcross to *H. melpomene* families used in QTL mapping.

601 SI Figure 11: Chromosome 20 QTL map for production of octadecanal and octadecanol in *H. melpomene*.
602 Shaded regions indicate the Bayesian confidence intervals with kinship structure taken into account and
603 black line indicates the peak of the QTL.

604

605 **Literature cited**

606 01. Taga, M.E. & Bassler, B.L. Chemical communication among bacteria. *Proc. Natl. Acad. Sci. USA* **100,**
607 14549-14554 (2003).
608
609 02. Jacob, S., McClintock, M.K., Zelano, B. & Ober, C. Paternally inherited HLA alleles are associated with
610 women's choice of male odor. *Nat. Genet.* **30,** 175–179 (2002).
611
612 03. Wysocki, C.J. & Preti, G. Facts, fallacies, fears, and frustrations with human pheromones. *Anatomical*
613 *Record Part A* **281A,** 1201-1211 (2004).
614
615 04. Kochansky, J., Cardé, R.T., Liebherr, J. & Roelofs, W.L. Sex pheromone of the European corn borer,
616 *Ostrinia nubialis* (Lepidoptera: Pyralidae), in New York. *J. Chem. Ecol.* **1,** 225-231 (1975).
617
618 05. Lassance, J.-M. et al. Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex
619 pheromones. *Nature* **466,** 486-489 (2010).
620
621 06. Vane-Wright, R.I. & Boppré, M. Visual and chemical signaling in butteflies: functional and
622 phylogenetic perspectives. *Phil. Trans. R. Soc. Lond. B* **340,** 197-205 (1993).
623
624 07. Mérot, C., Frérot, B., Leppik, E. & Joron, M. Beyond magic traits: Multimodal mating cues in
625 *Heliconius* butterflies. *Evolution* **69,** 2891-2904 (2015).
626
627 08. Darragh, K. et al. Male sex pheromone components in *Heliconius* butterflies released by the
628 androconia affect female choice. *PeerJ* **5,** e3953 (2017).
629
630 09. Meinwald, J., Meinwald, Y.C. & Mazzocchi, P.H. Sex pheromone of the queen butterfly: chemistry.
631 *Science* **164,** 1174-1175 (1969).
632
633 10. Pliske, T.E. & Eisner, T. Sex pheromone of the queen butterfly: biology. *Science* **164,** 1170-1172
634 (1969).

11. Grula, J.W., McChesney, J.D. & Taylor, O.R., Jr. Aphrodisiac pheromones of the sulfur butterflies *Colias eurytheme* and *C. philodice* (Lepidoptera, Pieridae). *J. Chem. Ecol.* **6,** 241-256 (1980).

12. Nishida, R. et al. Male sex pheromone of a giant danaine butterfly, *Idea leuconoe*. *J. Chem. Ecol.* **22,** 949-972 (1996).

13. Schulz, S. & Nishida, R. The pheromone system of the male danaine butterfly, *Idea leuconoe*. *Bioorg. Med. Chem.* **4,** 341-349 (1996).

14. Andersson, J., Borg-Karlson, A.-K., Vongvanich, N. & Wiklund, C. Male sex pheromone release and female mate choice in a butterfly. *J. Exp. Biol.* **210,** 964-970 (2007).

15. Nieberding, C.M. et al. The male sex pheromone of the butterfly *Bicyclus anynana*: towards an evolutionary analysis. *PLOS ONE* **3,** e2751 (2008).

16. Yildizhan, S. et al. Aphrodisiac pheromones from the wings of the Small Cabbage White and Large Cabbage White butterflies, *Pieris rapae* and *Pieris brassicae*. *ChemBioChem* **10,** 1666-1677 (2009).

17. Liénard, M.A., Wang, H.-L., Lassance, J.-M. & Löfstedt, C. Sex pheromone biosynthetic pathways are conserved between moths and the butterfly *Bicyclus anynana*. *Nat. Comms*. **5,** 3957 (2014).

18. Mann, F. et al. The scent chemistry of *Heliconius* wing androconia. *J. Chem. Ecol.* **43,** 843-857 (2017).

19. McCormick, A.C., Gershenzon, J. & Unsicker, S.B. Little peaks with big effects: establishing the role of minor plant volatiles in plant–insect interactions. *Plant Cell Env.* **37,** 1836-1844 (2014).

20. Chen, Q.-H. et al. Minor components play an important role in interspecific recognition of insects: a basis to pheromone based electronic monitoring tools for rice pests. *Insects* **9,** 192 (2018).

21. Jiggins, C.D. *The Ecology and Evolution of* Heliconius *Butterflies*. Oxford University Press (2017).

22. Merrill, R.M. et al. Genetic dissection of assortative mating behavior. *PLOS Biology* **17,** e2005902 (2019).

23. Via, S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Phil. Trans. R. Soc. Lond. B* **367,** 451-460 (2012).

24. Merrill, R.M., Van Schooten, B., Scott, J.A. & Jiggins, C.D. Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. *Proc. R. Soc. Lond. B* **278,** 511-518 (2010).

25. Smadja, C.M. & Butlin, R.K. A framework for comparing processes of speciation in the presence of gene flow. *Mol. Ecol.* **20,** 5123-5140 (2011).

26. Darragh, K. et al. Species specificity and intraspecific variation in the chemical profiles of *Heliconius* butterflies across a large geographic range. Preprint at https://www.biorxiv.org/content/10.1101/573469v2.abstract (2019).

683 27. Ferkovich, S.M., Oliver, J.E. & Dillard, C. Pheromone hydrolysis by cuticular and interior esterases of
684 the antennae, legs, and wings of the cabbage looper moth, *Trichoplusia ni* (Hübner). *J. Chem. Ecol.* **8,**
685 859-866 (1982).
686
687 28. Van Belleghem, S.M. et al. Complex modular architecture around a simple toolkit of wing pattern
688 genes. *Nature Ecol. Evol.* **1,** 0052 (2017).
689
690 29. Malausa, T. et al. Molecular differentiation at nuclear loci in French host races of the European corn
691 borer (*Ostrinia nubilalis*). *Genetics* **176,** 2343-2355 (2007).
692
693 30. Kárpáti, Z., Dekker, T. & Hansson, B.S. Reversed functional topology in the antennal lobe of the male
694 European corn borer. *J. Exp. Biol.* **211,** 2841-2848 (2007).
695
696 31. Koutroumpa, F.A. et al. Shifts in sensory neuron identity parallel differences in pheromone
697 preference in the European corn borer. *Frontiers Ecol. Evol.* **2,** 65 (2014).
698
699 32. Frey, J.E. & Bush, G.L. *Rhagoletis* sibling species and host races differ in host odor recognition.
700 *Entomol. Exp. Appl.* **57,** 123-131 (1990).
701
702 33. Tait, C., Batra, S., Ramaswamy, S.S., Feder, J.L. & Olsson, S.B. Sensory specificity and speciation: a
703 potential neuronal pathway for host fruit odour discrimination in *Rhagoletis pomonella*. *Proc. R. Soc.*
704 *Lond. B* **283,** 20162101 (2016).
705
706 34. Date, P. et al. Divergence in olfactory host plant preference in *D. mojavensis* in response to cactus
707 host use. *PLOS ONE* **8,** e70027 (2013).
708
709 35. Crowley-Gall, A. et al. Population differences in olfaction accompany host shift in *Drosophila*
710 *mojavensis*. *Proc. R. Soc. B* **283,** 20161562 (2016).
711
712 36. Papke, R.S., Kemp, D.J. & Rutowski, R.L. Multimodal signaling: structural ultraviolet reflectance
713 predicts male mating success better than pheromones in the butterfly *Colias eurytheme*. *Anim. Behav.*
714 **73,** 47-54 (2007).
715
716 37. Montgomery, S.H. & Merrill, R.M. Divergence in brain composition during the early stages of
717 ecological specialization in *Heliconius* butterflies. *J. Evol. Biol.* **30,** 571-582 (2017).
718
719 38. Tatsuki, S. et al. Sex pheromone of the rice stem borer, *Chilo suppressalis* (Walker) (Lepidoptera:
720 Pyralidae): the third component, *Z*-9-hexadecenal. *Appl. Entomol. Zool.* **18,** 443-446 (1983).
721
722 39. Cork, A. et al. Components of female sex pheromone of spotted bollworm, *Earias vittella* F.
723 (Lepidoptera: Noctuidae): identification and field evaluation in Pakistan. *J. Chem. Ecol.* **14,** 929-945
724 (1988).
725
726 40. Tumlinson, J.H. et al. Identification of a pheromone blend attractive to *Manduca sexta* (L.) males in a
727 wind tunnel. *Arch. Insect Biochem. Physiol.* **10,** 255-271 (1989).
728

41. Ho, H.Y. et al. Isolation, identification, and synthesis of sex pheromone components of female tea cluster caterpillar, *Andraca bipunctata* Walker (Lepidoptera: Bombycidae) in Taiwan. *J. Chem. Ecol.* **22,** 271-285 (1996).

42. McElfresh, J.S., Chen, X., Ross, D.W. & Millar, J.G. Sex pheromone blend of the pandora moth (Lepidoptera: Saturniidae), an outbreak pest in pine forests (Pinaceae). *Canadian Entomologist* **132,** 775-787 (2000).

43. El-Sayed, A.M., Mitchell, V.J., Manning, L.M. & Suckling, D.M. New sex pheromone blend for the Lightbrown Apple Moth, *Epiphyas postvittana*. *J. Chem. Ecol.* **37,** 640-646 (2011).

44. Pires, E.V. et al. Identification and field and laboratory tests of the sex pheromone of *Cerconota anonella* Sepp. (Lepidoptera: Oecophoridae). *J. Appl. Entomol.* **140,** 72-80 (2015).

46. Larsdotter-Mellstrom, H. et al. It's all in the mix: blend-specific behavioral response to a sexual pheromone in a butterfly. *Frontiers Physiol.* **7,** 68 (2016).

47. Ando, T. & Yamakawa, R. Analyses of lepidopteran sex pheromones by mass spectrometry. *Trends Anal. Chem.* **30,** 990-1002 (2011).

48. Yew, J.Y. & Chung, H. Insect pheromones: An overview of function, form, and discovery. *Prog. Lipid Res.* **59,** 88-105 (2015).

49. Martin, S.H. et al. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* **23,** 1817-1828 (2013).

50. Symonds, M.R.E. & Elgar, M.A. The evolution of pheromone diversity. *Trends Ecol. Evol.* **23,** 220-228 (2007).

51. Klein, A.L. & de Araújo, A.M. Courtship behavior of *Heliconius erato phyllis* (Lepidoptera, Nymphalidae) towards virgin and mated females: conflict between attraction and repulsion signals? *J. Ethology* **28,** 409-420 (2010).

52. Darragh, K. et al. Male pheromone composition depends on larval but not adult diet in *Heliconius* melpomene. *Ecol. Entomol.* **44,** 397-405 (2019).

53. More, J.D. & Finney, N.S. A simple and advantageous protocol for the oxidation of alcohols with o-iodoxybenzoic acid (IBX). *Org. Lett.* **4,** 3001–3003 (2002).

54. Cha, J.S. & Brown, H.C. Reaction of sodium aluminum hydride with selected organic compounds containing representative functional groups. Comparison of the reducing characteristics of lithium and sodium aluminum hydrides. *J. Org. Chem.* **58,** 4727–4731 (1993).

55. Zufall, F. & Leinders-Zufall, T. The cellular and molecular basis of odor adaptation. *Chem. Senses* **25,** 473-481 (2000).

56. Stengl, M. Pheromone transduction in moths. *Frontiers Cell. Neurosci.* **4,** 133 (2010).

57. Jones, C.D. Extension of the Castle–Wright effective factor estimator to sex linkage and haplodiploidy. *J. Hered.* **92,** 274-276 (2001).

58. Davey, J.W. et al. No evidence for maintenance of a sympatric Heliconius species barrier by chromosomal inversions. *Evol. Lett.* **1,** 138-154 (2017).

60. Rastas, P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **33,** 3726-3732 (2017).

61. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).

62. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475,** 493-496 (2011).

63. Broman, K.W. et al. R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multi-parent populations. *Genetics* **211,** 495-502 (2018).

64. Challis, R.J., Kumar, S., Dasmahapatra, K.K., Jiggins, C.D. & Blaxter, M. Lepbase: the Lepidopteran genome database. Preprint at https://www.biorxiv.org/content/10.1101/056994v1.abstract (2016).

65. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403-410 (1990).

66. Pessoa Pinharanda, A.L. The genomic basis of species barriers in *Heliconius* butterflies (Doctoral thesis). https://doi.org/10.17863/CAM.16857 (2017).

67. Yang, Z. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586-1591 (2007).

68. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. nlme: linear and nonlinear mixed effects models. R package version 3.1-137, URL: https://CRAN.R-project.org/package=nlme (2018).

69. Fox, J. & Weisberg, S. *An R Companion to Applied Regression, Second Edition*. Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion (2011).

70. Lenth, R. emmeans: estimated marginal means, aka least-squares means. R package version 1.3.1. URL https://CRAN.R-project.org/package=emmeans (2018).

71. Venables, W.N. & Ripley, B.D. *Modern Applied Statistics with S. Fourth Edition*. Springer (2002).

72. Koller, M. robustlmm: An R package for robust estimation of linear mixed-effects models. *J. Stat. Software* **75,** 1-24 (2016).

73. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ (2013).