



**HAL**  
open science

## Comparative transcriptome analysis at the onset of speciation in a mimetic butterfly-The Ithomiini *Melinaea marsaeus*

Florence Piron-Prunier, Emma Persyn, Fabrice Legeai, Melanie Mcclure, Camille Meslin, Stéphanie Robin, Susete Alves Carvalho, Ammara Mohammad, Corinne Blugeon, Emmanuelle Jacquin-joly, et al.

### ► To cite this version:

Florence Piron-Prunier, Emma Persyn, Fabrice Legeai, Melanie Mcclure, Camille Meslin, et al.. Comparative transcriptome analysis at the onset of speciation in a mimetic butterfly-The Ithomiini *Melinaea marsaeus*. *Journal of Evolutionary Biology*, 2021, 34 (11), pp.1704-1721. 10.1111/jeb.13940 . hal-03381525

**HAL Id: hal-03381525**

**<https://hal.science/hal-03381525>**

Submitted on 17 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Comparative transcriptome analysis at the onset of speciation in a  
2 mimetic butterfly, the Ithomiini *Melinaea marsaeus*

3

4 Short title: Transcriptomics in mimetic *Melinaea marsaeus*

5

6 Florence Piron-Prunier (1)§, Emma Persyn (2)§, Fabrice Legeai (3,4)§, Melanie McClure  
7 (1,5), Camille Meslin (2), Stéphanie Robin (3,4), Susete Alves-Carvalho (4), Ammara  
8 Mohammad (6), Corinne Blugeon (6), Emmanuelle Jacquin-Joly (2), Nicolas Montagné (2),  
9 Marianne Elias\*† (1) and Jérémy Gauthier\*† (4,7)

10 § co-first

11 \* co-last

12 † corresponding authors: [marianne.elias@mnhn.fr](mailto:marianne.elias@mnhn.fr) and [jeremy.gauthier@ville-ge.ch](mailto:jeremy.gauthier@ville-ge.ch)

13

14

15 1. Institut de Systématique, Evolution, Biodiversité, MNHN, CNRS, Sorbonne Université,  
16 EPHE, Université des Antilles, Paris, France

17 2. Institute of Ecology and Environmental Sciences of Paris, Sorbonne Université, INRAE,  
18 CNRS, IRD, UPEC, Université de Paris, Paris, France

19 3. BIPAA, IGEPP, INRAE, Institut Agro, Univ Rennes, 35000, Rennes, France

20 4. Univ Rennes, Inria, CNRS, IRISA, 35000, Rennes, France

21 5. Laboratoire Écologie, Évolution, Interactions des Systèmes Amazoniens (LEEISA),  
22 Université de Guyane, CNRS, IFREMER, 97300 Cayenne, France

23 6. Genomics core facility, Institut de Biologie de l'ENS (IBENS), Département de biologie,  
24 École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France

25 7. Geneva Natural History Museum, 1 Route de Malagnou, 1208 Geneva, Switzerland

26

27 **Authors contributions**

28 ME designed the study. MMC performed the sampling, breeding and dissections. FPP  
29 performed RNA extraction. AM and CB performed library construction and sequencing. JG,  
30 FPP and FL performed most of the analyses with contributions from EP, CM, SR, SAC, EJJ,  
31 NM and ME. All authors took part in discussions concerning the analyses and result  
32 interpretations. JG, FPP and ME wrote the paper, with contributions from all authors.

33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Acknowledgements**

We thank the Peruvian authorities for research permits (236-2012-AG-DGFFS-DGEFFS, 201-2013-MINAGRI-DGFFS/DGEFFS, 002-2015-SERFOR-DGGSPFFS and 373-2017-SERFOR-DGGSPFFS), the Gobierno Regional San Martín PEHCBM (permit: 124-2016-GRSM/PEHCBM-DMA/EII-ANP/JARR) and the Museo de Historia Natural and Prof. Gerardo Lamas for their support with research permits. We also thank Mario Tuanama and Ronald Moripezo for their precious help in the field. This work was funded by SPECREP, CLEARWING and PRISM ANR projects (ANR-14-CE02-0011, ANR-16-CE02-0012 and ANR-16-CE02-0003), by an HFSP research grant (RGP0014/2016), by an *Action Thématique du MNHN* grant (ATM RNADAPT 2018), by le Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT) as a Postdoctoral Fellowship and an “Investissements d’Avenir” grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01), and was supported by the France Génomique national infrastructure, funded as part of the "Investissements d'Avenir" program managed by the Agence Nationale de la Recherche (contract ANR-10-INBS-0009).

66 **Abstract**

67 Ecological speciation entails divergent selection on specific traits, and ultimately on the  
68 developmental pathways responsible for these traits. Selection can act on gene sequences, but  
69 also on regulatory regions responsible for gene expression. Mimetic butterflies are a relevant  
70 system for speciation studies because wing color pattern (WCP) often diverges between closely  
71 related taxa, and is thought to drive speciation through assortative mating and increased  
72 predation on hybrids. Here we generate the first transcriptomic resources for a mimetic butterfly  
73 of the tribe Ithomiini, *Melinaea marsaeus*, to examine patterns of differential expression  
74 between two subspecies and between tissues that express traits that likely drive reproductive  
75 isolation; WCP and chemosensory genes. We sequenced whole transcriptomes of three life  
76 stages to cover a large catalogue of transcripts and we investigated differential expression  
77 between subspecies in pupal wing discs and antennae. Eighteen known WCP genes were  
78 expressed in wing discs and 115 chemosensory genes were expressed in antennae, with a  
79 remarkable diversity of chemosensory protein genes. Many transcripts were differentially  
80 expressed between subspecies, including two WCP genes and one odorant receptor. Our results  
81 suggest that in *M. marsaeus* the same genes as in other mimetic butterflies are involved in traits  
82 causing reproductive isolation, and point at possible candidates for the differences in those traits  
83 between subspecies. Differential expression analyses of other developmental stages and body  
84 organs and functional studies are needed to confirm and expand these results. Our work  
85 provides key resources for comparative genomics in mimetic butterflies, and more generally in  
86 Lepidoptera.

87 **Significance statement:**

88 Ecological speciation entails divergent selection on specific traits, but the underlying  
89 developmental pathways remain poorly known. We examined patterns of differential  
90 expression in two recently diverged subspecies of the mimetic butterfly *M. marsaeus*  
91 (Ithomiini), which differ in traits likely driving speciation, wing color pattern and pheromone  
92 blend. Many transcripts were differentially expressed between subspecies, including two wing  
93 color pattern genes and one odorant receptor, likely candidate genes responsible for the  
94 variation of traits involved in speciation.

95

96 **Keywords:** transcriptomics, wing color pattern, chemosensory genes, reproductive isolation,  
97 mimicry, Lepidoptera

98 **Introduction**

99 When coupled with reproductive isolation, ecological diversification is one of the main  
100 processes that can explain the observed diversity of species in nature. Recently, studies have  
101 investigated the traits responsible for reproductive isolation in closely related taxa that span the  
102 speciation continuum, such as population or species pairs, which are under divergent ecological  
103 selection (Nosil 2012). However, few studies have investigated the molecular mechanisms  
104 responsible for differentiation from one species into distinct lineages early on in the process. In  
105 diverging lineages exhibiting little genetic difference overall, trait divergence may stem from  
106 subtle differences, such as genetic variations in gene sequences, but also differences in  
107 regulatory regions, thereby inducing differential expression of those genes (Eyres et al. 2016;  
108 van Schooten et al. 2020).

109 Müllerian mimetic butterflies, whereby multiple co-occurring chemically-defended species  
110 harbor convergent warning color patterns (Müller 1879), are excellent study systems to unravel  
111 differential patterns of gene expression during the early stages of speciation, because species  
112 often diverge for wing color patterns, which is thought to be one of the main drivers of  
113 speciation because it can drive reproductive isolation (Jiggins et al. 2006; Kozak et al. 2015).  
114 Indeed, offspring of crosses between individuals of different color patterns typically have  
115 intermediate, non-mimetic color patterns, and suffer increased predation because they are not  
116 recognized as unpalatable (Merrill et al. 2012; Arias et al. 2016). Moreover, wing color patterns  
117 are also involved in mate choice in mimetic butterflies, resulting in assortative mating for color  
118 patterns (Jiggins et al. 2001; Chamberlain et al. 2009; Merrill et al. 2011; McClure et al. 2019).  
119 In the well-studied mimetic butterfly genus *Heliconius*, color pattern variation is largely  
120 controlled by a small set of homologous loci across the genus, dubbed the ‘mimicry toolkit’  
121 (Gilbert 2003; Watt & Boggs 2003; M. Joron et al. 2006), some of which have been functionally  
122 characterized (e. g., transcription factors *optix* (Reed et al. 2011) and *aristaless* (Westerman et  
123 al. 2018), signaling ligand *WntA* (Martin et al. 2012; Mazo-Vargas et al. 2017) and cycle-cell  
124 regulator *cortex* (Nadeau et al. 2016; Saenko et al. 2019). Thus, the establishment of the color  
125 patterns takes place through a specific kinetic of these genes during metamorphosis and wing  
126 formation (Hines et al. 2012, Connahs et al. 2016; Livraghi et al. 2021).

127 Other traits that may contribute to reproductive isolation in mimetic butterflies include sex  
128 pheromones (González-Rojas et al. 2020; McClure et al. 2019; Schulz et al. 2004; Darragh et  
129 al. 2020), which can be of particular importance for mate recognition in co-mimetic species

130 (Mérot et al. 2015). This makes the study of chemosensory genes (i.e. genes involved in  
131 chemical communication) especially relevant in the study of speciation in mimetic butterflies.  
132 In insects, including butterflies, the detection of chemical signals is ensured by neurons housed  
133 in chemosensory sensilla located on different organs, but most notably the antennae. Three  
134 types of membrane receptors named Odorant Receptors (ORs), Gustatory Receptors (GRs) and  
135 Ionotropic Receptors (IRs), encoded by diverse multigenic families, bind chemicals and allow  
136 for signal transduction in olfactory and gustatory neurons (Robertson 2019). Secreted proteins,  
137 such as Odorant-Binding Proteins (OBPs) and Chemosensory Proteins (CSPs), are also thought  
138 to play a role in the detection of chemicals by solubilizing and transporting them within the  
139 sensillar lymph (Pelosi et al. 2006). There is extensive literature depicting the role of specific  
140 lineages of the OR gene family in the detection of volatile moth sex pheromones, i.e. long-  
141 chain aliphatics emitted by females that attract males from a distance (Montagné et al. 2021).  
142 However, almost nothing is currently known of the molecular bases of pheromone detection in  
143 butterflies, whereby males, rather than females, produce aphrodisiac compounds of various  
144 chemical structure and detected by females at close range (Nieberding et al. 2008; Sarto i  
145 Monteys et al. 2016).

146 In the mimicry literature, two butterfly clades belonging to the family Nymphalidae stand out  
147 as important study systems: the genus *Heliconius* and the tribe Ithomiini. Both clades are  
148 neotropical and consist of important adaptive radiations (Kozak et al. 2015; Chazot et al. 2019).  
149 Notably, the tribe Ithomiini, which comprises 393 species, is the largest clade of mimetic  
150 butterflies known to date. Ithomiini numerically dominate butterfly communities in neotropical  
151 forests, and are believed to be instrumental in the formation of mimicry rings in those habitats  
152 (G. W. Beccaloni 1997). Yet, studies of these two groups have mostly targeted different  
153 questions, in large part as a result of how amenable they are to captive rearing. Most studies of  
154 *Heliconius* are done at the species level, including the study of population structure (e.g.,  
155 Nadeau et al. 2014), mate choice (e. g., Jiggins et al. 2001) and the genetic basis of wing pattern  
156 variation (e. g., Mathieu Joron et al. 2006), the latter relying on the production of very large  
157 broods. By contrast, Ithomiini studies have mostly been multi-specific in nature, and focused  
158 on community ecology (George W. Beccaloni 1997; Devries et al. 1999; Elias et al. 2008; Hill  
159 2010; Willmott et al. 2017) and macroevolutionary patterns of diversification (Chazot et al.  
160 2016, 2018, 2019; De-Silva et al. 2016; Lisa De-Silva et al. 2017). However, recent studies  
161 have characterized trait and genetic structure at the population level, demonstrating genetic,  
162 wing color pattern (Gauthier et al. 2020; McClure & Elias 2016; McClure et al. 2019) and

163 pheromone (Stamm et al. 2019; Mann et al. 2020) differentiation between parapatric  
164 subspecies. Ithomiini are difficult to breed in captivity, and the first experimental test of traits  
165 involved in mate choice was only recently completed in a handful of Ithomiini species (McClure  
166 et al. 2019). These experiments have shown that, similarly to *Heliconius*, both color pattern and  
167 sex pheromones likely play a key role in speciation. Specifically, closely related ithomiine taxa  
168 (i.e., subspecies or closely related species) that differ for wing color patterns and sex  
169 pheromones exhibit assortative mating for those traits (McClure et al. 2019). This raises the  
170 question of the molecular bases of the variation of these traits. As these species are thought to  
171 drive mimicry in many other mimetic butterflies, including some *Heliconius* species (Joron et  
172 al. 1999), this also raises the question as to whether the loci underlying wing color pattern  
173 variation in these species are homologous to those found in *Heliconius* species (i.e., the mimicry  
174 toolkit).

175 The ithomiine genus *Melinaea* is particularly well suited to address these questions, because it  
176 has undergone a rapid radiation (Dasmahapatra et al. 2010; Chazot et al. 2019), concomitantly  
177 with wing pattern diversification (McClure & Elias 2016; McClure et al. 2019). Moreover,  
178 *Melinaea* species engage in mimetic interactions with multiple *Heliconius* species, notably with  
179 *H. numata*, whose different morphs are nearly indistinguishable from different *Melinaea*  
180 species (Joron et al. 1999; Llaurens et al. 2014). One of these species, *Melinaea marsaeus*,  
181 consists of at least seven subspecies (S. Brown 1977; McClure & Elias 2016), two of which,  
182 *phasiana* and *rileyi*, form a contact zone in the transitional forests found between the Andes  
183 and the Amazon in Peru, near the city of Tarapoto (Fig. 1). The two subspecies harbor distinct  
184 wing color patterns and significantly different male pheromonal bouquets (McClure et al.  
185 2019). It is not known whether butterflies are able to discriminate the two subspecies based on  
186 male pheromones (McClure et al. 2019) but mate choice experiments between these two  
187 subspecies have demonstrated strong assortative mating (McClure et al. 2019), resulting in a  
188 low number of putative hybrids in the wild (McClure & Elias 2016; McClure et al. 2019).

189 In this paper, we address the question regarding the molecular bases for variation in color  
190 pattern and chemosensory traits in *M. marsaeus* by focusing on gene expression in the tissues  
191 displaying these traits. To this end, we sequenced RNA from multiple tissues and  
192 developmental stages to generate a reference transcriptome for *M. marsaeus* - the first to date  
193 for an ithomiine species - as a tool to investigate gene expression in the two subspecies *phasiana*  
194 and *rileyi*. We focused on two stages of pupal wing discs, where color patterns form in  
195 butterflies (Hines et al. 2012, Connahs et al. 2016; Livraghi et al. 2021), and in adult female

196 antennae, where chemical signals are detected, to screen for differentially expressed genes  
197 between subspecies and throughout development. We also undertook a candidate gene approach  
198 and looked more specifically at the expression of genes known to be involved in color pattern  
199 variation and in chemosensory activity in other Lepidoptera. Our data also enable us to compare  
200 the expansion of chemosensory genes in *M. marsaeus* with those of other Lepidoptera.

201

## 202 **Material and Methods**

### 203 *Sample collection*

204 Tissue samples were obtained in 2012-2013 from individuals reared in captivity under ambient  
205 conditions in Tarapoto, Peru (San Martin). Stocks were built from wild *M. marsaeus rileyi* and  
206 *M. marsaeus phasiana* females captured in Shucushyacu (Peru, Loreto) (W 5° 57' 48'' ; S 75°  
207 53' 24'') and Shapaja (Peru, San Martin) (W 76° 15' 39'' ; S 6° 34' 48''), respectively . Females  
208 were given *Juanulloa parasitica* for egg laying, and progeny were reared as per McClure and  
209 Elias 2017. Adults were fed sugar water and pollen and larvae were reared on *J. parasitica*.

210 A first set of samples from *M. marsaeus rileyi* was used to encompass the main developmental  
211 stages and tissues. It consisted of one 5<sup>th</sup> instar larva (gut was removed), one pupae and one  
212 adult female divided into three tissue samples – abdomen, thorax and head. To assess  
213 differentially expressed genes between subspecies and tissue types, female antennae, pupal  
214 wing discs dissected at 24h after pupation and at 48h after pupation were obtained for both  
215 subspecies. Each developmental stage/tissue type (pupal wing discs of 24h and 48h, antennae)  
216 had three to five biological replicates each (Fig. 1; Table S1). Organisms were anesthetized by  
217 chilling before dissection, and tissue preserved in *RNAlater* at 4°C according to the  
218 manufacturer's instructions (Qiagen, Hilden, Germany), then stored at -80°C until RNA  
219 extraction.

### 220 *Total RNA extraction*

221 Tissue samples were homogenised in 600 µl of RLT buffer with TissueLyser (Qiagen, Hilden,  
222 Germany). Total RNA was then extracted according to the manufacturer's protocol (RNeasy  
223 Mini kit, Qiagen, Hilden, Germany) and eluted in 30 µl of RNase-free water. To avoid genomic  
224 contamination, RNase-free DNase treatment (Qiagen, Hilden, Germany) was performed during  
225 RNA extraction. The quality of the isolated RNA was checked on 0.8% agarose gel for the  
226 presence of 28S and 18S bands. The quality and quantity of RNA was further analyzed using



227 Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) and RNA integrity was confirmed using  
228 an Agilent Bioanalyser 2100 (Agilent Technologies, CA, USA).

#### 229 *RNAseq library preparation and sequencing*

230 Library preparation was performed at IBENS (Institut de Biologie de l'Ecole Normale  
231 Supérieure, Paris, France) genomics facility, using the Illumina TruSeq Stranded RNA sample  
232 preparation kit according to the manufacturer's specifications (Illumina, San Diego, CA, USA).  
233 Sequencing was carried out on a NextSeq 500 platform; the first set of five libraries was  
234 sequenced in paired-end, 150-bp reads while the 28 libraries for differential gene expression  
235 analyses (wing discs and antennae) were sequenced in single-end, 75-bp reads (Table S1).

#### 236 *Reads pre-processing*

237 GC content and over-representation of sequences were checked with the FastQC software  
238 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), revealing no evidence of  
239 contamination. To obtain high-quality reads, 3' ends with quality values < 30 were trimmed (-  
240 q 30) and adapters were removed (-a AGATCGGAAGAGC -A AGATCGGAAGAGC) with  
241 Cutadapt version 1.11 (Martin 2011). Moreover, reads shorter than 25 bp were discarded (-m  
242 25). A total of 1278 million raw reads (single-end and paired-end) were then used for  
243 subsequent steps. The Ribopicker tool version 0.4.3 was used for automated identification and  
244 removal of ribosomal RNA sequences (Schmieder, Lim et al. 2012). For paired-end reads, non-  
245 rRNA reads were synchronized to associate R1 and R2 pairs and unpaired reads were discarded.  
246 After filtering, a total of 1194 million reads (corresponding to 93% of the raw reads) were  
247 retained (Table S2).

248

#### 249 *De novo reference transcriptome assembly*

250 In order to generate a reference transcriptome for *M. marsaeus* (hereafter, transcriptome), high-  
251 quality reads from all *M. marsaeus* libraries (paired reads and single reads) were assembled *de*  
252 *novo* using the trinity v2.4.0 transcriptome assembler with default parameters (Haas et al. 2013).  
253 Completeness of the assembled transcriptome was assessed using the BUSCO v4.0.6 software  
254 (Benchmarking Universal Single-Copy Orthologs) (Seppey et al. 2019), which tests the  
255 assembly for the presence of 1 367 single-copy genes highly conserved in insects  
256 (insect\_odb10).

#### 257 *Functional annotation and classification*

258 Open reading frames (ORFs) above 50 bp were predicted from the transcriptome using  
259 TransDecoder (<https://github.com/TransDecoder/>) and only those encoding proteins exhibiting  
260 a blastp hit (e-value < 1e-5) with a protein from Lepbase (Challi et al. 2016) were conserved.  
261 Protein motifs and domains were scanned with interproscan v5.29-68 with the options -  
262 iprlookup -goterms --pathways (Jones et al. 2014). BLASTP (version 2.9.0, with options -  
263 evaluate 1e-8 -max\_target\_seqs 10 -soft\_masking false -word\_size 3 -matrix BLOSUM62 -  
264 gapopen 11 -gapextend 1 -seg no) of the ORFs against NR (version 2020-5-29) and interproscan  
265 results were imported to the BLAST2GO suite for Gene ontology (GO) annotation of transcripts  
266 (Conesa et al. 2005). Finally, orthogroups were created with Orthofinder v2.4.0 (Emms & Kelly  
267 2015) based on the diamond (Buchfink et al. 2015) comparisons of the transdecoder predicted  
268 proteins and 28 proteomes from Lepbase (October 2020 version).

269

#### 270 *Identification of wing color pattern (WCP) genes*

271 To identify genes potentially involved in the development of wing pigmentation, we selected a  
272 list of 20 *Danaus plexippus* (the closest relative of *M. marsaeus* for which a reference genome  
273 is available) candidate genes associated with wing color patterns that have been previously  
274 characterized in other insects using multiple approaches including transcriptomics (Reed et al.  
275 2011; Saenko et al. 2019), linkage and QTL mapping (Martin et al. 2012; Westerman et al.  
276 2018), in situ hybridization (Martin et al. 2012) and CRISPR knockout (Westerman et al. 2018;  
277 Zhang & Reed 2016). (Table S3a). This includes *optix*, a transcription factor that acts as a switch  
278 for the ommochrome pathway and is responsible for red, orange or brown patches (Reed et al.  
279 2011); *WntA*, a ligand that determines the size and shape of color pattern elements (Martin et  
280 al. 2012); *cortex*, a cell-cycle regulator that switches yellow and white color on and off (Nadeau  
281 et al. 2016), and that can also induce switches between full color patterns in the polymorphic  
282 species *H. numata*, a co-mimic of *M. marsaeus* (Saenko et al. 2019); and *aristaless* (Westerman  
283 et al. 2018), a transcription factor that controls the switch between yellow and white colors.  
284 Color pattern variation in other Lepidoptera is also due to many of these same genes, but also  
285 include *doublesex* (wing pattern switch in females of *Papilio polytes*, (Kunte et al. 2014)),  
286 *distal-less* (eye-spot and melanization in *Bicyclus anynana*, (Beldade et al. 2002; Reed & Serfas  
287 2004; Monteiro et al. 2013; Dhungel et al. 2016; Zhang & Reed 2016) and *apterousA* (involved  
288 in dorso-ventral pattern differentiation in *B. anynana*, (Prakash & Monteiro 2018)). The  
289 developmental genes *domeless* and *wingless* are also candidate genes for color patterning  
290 (Kronforst et al. 2006; Jiggins et al. 2017). Finally, many other genes are also directly involved

291 in the pathway for melanin synthesis (*yellow*, *yellow\_d*, *yellow\_h2*, *tan*, *pale*, *black*,  
292 *Ddc\_dopa\_decarboxylase*, *ebony* and *dopamine\_N\_acetyltransferase* (Hori et al. 1984; Koch  
293 et al. 1998; Ferguson et al. 2010; Hines et al. 2012; Daniels et al. 2014; Zhang et al. 2017;  
294 Kuwalekar et al. 2020), and in the ommochrome synthesis pathway (*cinnabar* and *kynurenine*  
295 *formamidase*, (Hines et al. 2012; Reed et al. 2008; Daniels et al. 2014). We then extracted the  
296 *M. marsaeus* proteins and corresponding transcripts from the Orthofinder Orthogroups.

297

### 298 *Annotation of candidate chemosensory genes*

299 For each chemosensory gene family investigated (OR, GR, IR, OBP, CSP), a dataset was  
300 created with amino acid sequences annotated from the genomes of the following lepidopteran  
301 species: *Danaus plexippus*, *Heliconius melpomene*, *Helicoverpa armigera* and *Bombyx mori*.  
302 These sequences were used as queries to search the *M. marsaeus* reference transcriptome using  
303 tBLASTn v2.5 (with an e-value threshold of 0.001) as implemented in the Galaxy web interface  
304 (Cock et al. 2015). To eliminate false positive results, amino acid sequences translated from the  
305 transcripts that were identified were used as queries to search the NCBI nr database using  
306 BLASTp (Johnson et al. 2008). To rebuild the OR, GR, IR and OBP phylogenies, candidate *M.*  
307 *marsaeus* amino acid sequences were aligned with sequences of the four species mentioned  
308 above. For the CSP phylogeny, *M. marsaeus* amino acid sequences were aligned with sequences  
309 from *D. plexippus*, *H. melpomene*, *B. mori*, *Spodoptera frugiperda*, in addition to the  
310 Nymphalidae *Bicyclus anynana*, *Vanessa tameamea* and *Maniola hyperantus*, available on the  
311 NCBI GenBank database. OR and GR alignments were performed with Muscle (Edgar 2004)  
312 as implemented in Seaview v4.7 (Gouy et al. 2010). IR, OBP and CSP alignments were  
313 performed with MAFFT v7 (Kato et al. 2019). Best-fit models of amino acid substitutions  
314 were determined with SMS (Lefort et al. 2017) and maximum-likelihood phylogenies were  
315 calculated using PhyML v3.0 (Guindon et al. 2010). Node support was assessed using SH-like  
316 approximate likelihood-ratio tests (Anisimova & Gascuel 2006).

317

### 318 *Differential gene expression analysis (DGE)*

319 The clean reads corresponding to the 28 pupal wing discs and adult female antennae samples  
320 were mapped to the *de novo* assembled transcriptome using Bowtie 2 (2.2.7) (Langmead &  
321 Salzberg 2012) with default parameters. Raw counts (numbers of fragments mapped to a  
322 transcript) were used as input in EdgeR (Robinson et al. 2010) implemented in AskOR pipeline  
323 (<https://github.com/askomics/askoR>), and only transcripts with at least 0.5 CPM (counts per

324 million) on 3 of the replicates were kept for further analyses. Sample variability and correlations  
325 were assessed using Multi-dimensional Scaling (MDS) and hierarchical clustering. For relevant  
326 contrasts (i.e. comparisons of the two subspecies for each tissue, comparison of the two  
327 subspecies for all tissues, comparison of the wing discs at 24h and 48h for each subspecies and  
328 for all subspecies) GLM differential expression analyses with quasi-likelihood (QL) method  
329 (with Benjamini-Hochberg correction for false discovery rate) were applied on the trimmed  
330 mean of M-values (TMM)-normalized counts corrected by the dispersion estimation. All  
331 possible contrasts between subspecies and tissues were performed with EdgeR and lists of  
332 differentially expressed transcripts were obtained for each comparison at a minimum false  
333 discovery rate (FDR) of 0.05. Finally, a negative binomial Generalized Linear Model (GLM)  
334 has been used to test interaction effect between subspecies and wing disc conditions (24h and  
335 48h). Enrichment in transcript differentially expressed in specific condition, tissue or  
336 subspecies, has been tested by Chi-squared test. Gene Ontology enrichment analyses of  
337 differentially expressed genes (DEGs) against the transcriptome were performed using the  
338 Fisher exact test using topGO (Alexa and Rahnenfuhrer 2019).

339

## 340 **Results**

### 341 *Sequencing and transcriptome statistics*

342 A total of 1,248 million reads were obtained after sequencing all thirty-two libraries on  
343 the Illumina NextSeq 500 platform. All libraries were of good quality and satisfactory for GC  
344 distribution, quality of sequences and redundancy. Trimming and rRNA removal eliminated  
345 6.7% of the reads before assembly (Table S2). The *de novo* transcriptome assembly obtained  
346 with Trinity consisted of 179,833 transcripts, of which 82,469 ORFs > 50 bp were identified  
347 (details are given in Table 1). The average and median transcript length was reduced to 620 bp  
348 and 342 bp respectively (Table 1), which suggests fragmentation of the transcripts into smaller  
349 fragments and explains the large number of transcripts generated. This fragmentation does not  
350 seem to impact the completeness of the transcriptome as BUSCO's assessment of transcriptome  
351 completeness found more than 90% complete genes (single copy + duplicates) in the *de novo*  
352 transcriptome of *M. marsaeus*, using either the transcripts or the ORFs (Table 1). Identified  
353 protein sequences were searched against the NCBI non-redundant (nr) protein database using  
354 BLASTP, resulting in the annotation of 57,313 sequences. The mean and median length of the  
355 protein sequences not getting any hits (69 and 62 nucleotides in length respectively) were much

356 smaller than those sequences that did get a hit (239 and 134 nucleotides respectively),  
357 suggesting that most of these contigs do not overlap with the whole CDS part of the transcript.  
358 Most of the best hits were found against *Danaus plexippus* (53.13%), followed by *Vanessa*  
359 *tameamea* (9.05%) and *Bicyclus anynana* (5.10%), which is consistent with the fact that *M.*  
360 *marsaeus* is more closely related to *D. plexippus* than to any of the other Lepidoptera species  
361 available in Lepbase.

362 Overall, 36,683 sequences (44.48%) were assigned to a putative function and one or  
363 more GO terms, which were allocated to major categories (Biological Processes, Cellular  
364 Components and Molecular Function) and subcategories (details in Figure S1). Enzyme codes  
365 could be assigned to 8.71% of the sequences (Figure S1).

366

### 367 *Candidate gene annotation*

368 We identified 68 transcripts corresponding to 19 wing color pattern genes in the *M.*  
369 *marsaeus* transcriptome, based on homology with *D. plexippus* genes. This manual annotation  
370 enabled grouping of the transcripts that corresponded to the same gene. Most of them (11 genes)  
371 were represented by only one transcript, while other WCP genes were represented by several  
372 transcripts, with a maximum of 13 transcripts for *Dopamine-N-acetyltransferase*. Finally, we  
373 did not find any *M. marsaeus* orthologous gene for kynurenine formamidase (Table S3).

374 We annotated 51 candidate ORs (125 transcripts), including the coreceptor Orco, 22  
375 candidate IRs (78 transcripts) and 21 candidate GRs (46 transcripts) (Table S3). The large  
376 diversity of ORs present in the reference transcriptome of *M. marsaeus* is mirrored by the fact  
377 that the ORs (hereafter, MmarORs) were identified within almost every paralogous lineage of  
378 the Lepidoptera OR phylogeny, with the notable exception of the so-called pheromone receptor  
379 clade (Figure S2). That said, MmarOR35 and MmarOR38 clustered within clades that have  
380 recently been shown to also contain sex pheromone receptors (Bastin-Héline et al. 2019; Li et  
381 al. 2017). We also identified five members of an OR lineage specific to Papilionoidea  
382 (“butterfly-specific expansion” in Figure S2). A similar diversity was found for MmarIRs, as  
383 we identified all four coreceptors (IR8a, IR25a, IR76b, IR93a) and all but one of the highly  
384 conserved antennal IRs. On the other hand, we identified only four divergent IRs, known to be  
385 expressed in gustatory tissues in *Drosophila* (Sánchez-Alcañiz et al. 2018). In regard to GRs,  
386 we identified transcripts encoding for candidate CO<sub>2</sub> and sugar receptors as well as homologs  
387 of the *Drosophila* fructose receptor GR43a, but annotated only a few MmarGRs belonging to

388 other lepidopteran paralogous lineages (Figure S2), whose expression is generally higher in  
389 gustatory tissues such as legs or proboscis (Briscoe et al. 2013; Guo et al. 2017; van Schooten  
390 et al. 2020). In addition to chemoreceptors, we also annotated 32 candidate OBPs (50  
391 transcripts) and 40 candidate CSPs (103 transcripts) (Table S3). The *M. marsaeus* OBP  
392 repertoire was rather similar to those annotated from the genomes of *D. plexippus* and *H.*  
393 *melpomene* (Zhan et al. 2011; Heliconius Genome Consortium 2012), with only two moderate  
394 gene expansions (MmarOBP17-21 and MmarOBP26-29, see Figure S2). We identified four  
395 members of the PBP/GOBP subfamily involved in sex pheromone detection in moths (Vogt et  
396 al. 2015). Contrary to the other gene families, the CSP repertoire of *M. marsaeus* was more  
397 divergent (Figure S2). Most notably we identified a large CSP gene expansion within a single  
398 lineage (MmarCSP24-40), likely the result of recent and repeated gene duplication. This  
399 expansion would explain the unprecedented number of CSPs identified here.

400

#### 401 *Differential Gene Expression*

402         Sequenced reads from the wing-discs and antennae libraries were mapped on the  
403 reference transcriptome in order to measure the expression levels of each transcript in each  
404 sample and perform differential expression analyses. After quality trimming and ribosomal  
405 RNA removal, 94.2% to 96.3% of the reads were mapped to the reference transcriptome and  
406 69.3 to 76.5% were assigned to a unique transcript, according to those libraries. Most of the  
407 residual reads were removed because they could be equivalently mapped to multiple transcripts.  
408 The Trimmed Mean of M Values (TMM)-normalized counts per million (CPM) was used to  
409 assess the similarity between replicates, using a Multidimensional Scaling plot (MDS) and a  
410 heatmap of the correlation matrix. A single sample (*M. marsaeus rileyi*, wing-disc 24h replicate  
411 1) did not cluster correctly and was removed from any further analysis. All other samples  
412 clustered correctly, with the first MDS axis (explaining 50.19% of the sum of eigen-values)  
413 discriminating antennae from wing-discs, and the second (6.83%) and third axes (6.52%)  
414 separated 24h from the 48h wing-discs as well as subspecies (second axis for 24h wings discs,  
415 and third axis for antennae and 48h wing discs) (Figure S3). Hierarchical clustering confirms  
416 that variation between subspecies is smaller than variation between tissues (Figure S3).

417         The detailed analysis of wing color pattern genes revealed that of the 68 WCP transcripts  
418 identified in the transcriptome, 49 were expressed in the wing discs (CPM>0.5 in 3 samples),  
419 corresponding to 18 genes (Table S3). Only one WCP gene identified in the transcriptome, i.e.

420 *cortex*, was not expressed. Among the 94 candidate chemoreceptor genes, 65 were found to be  
421 consistently expressed in the adult antennae, including 39 ORs (63 transcripts), 20 IRs (50  
422 transcripts) and 6 GRs (10 transcripts) (Table S3). Finally, of the 32 candidate OBPs (50  
423 transcripts) and 40 candidate CSPs (103 transcripts) annotated, 22 OBPs (37 transcripts) and  
424 28 CSPs (64 transcripts) were expressed in the adult antennae (Table S3).

425

#### 426 *Patterns of gene Expression during wing disc development*

427 The comparison between wing discs and adult antennae revealed 38,076 transcripts (59.8% of  
428 the expressed transcripts) and 39,013 transcripts (61.3% of the expressed transcripts)  
429 differentially expressed compared to wing discs at 24h and 48h respectively, highlighting the  
430 strong difference in molecular pathways between the two types of tissue (Fig. 2). Differences  
431 in expression were notably found in wing color pattern genes (Fig. 3) and in chemosensory  
432 genes (Fig. 4 and 5). For WCP genes, a total of 35 (71%, 23 up and 12 down regulated) and 41  
433 (84%, 33 up and 8 down regulated) transcripts were differentially expressed between wing discs  
434 at 24h and adult antennae, and wing discs at 48h and adult antennae, respectively. These  
435 proportions were similar for chemosensory genes: 200 transcripts (88%) were differentially  
436 expressed between wing discs at 24h and adult antennae, and 218 transcripts (96%) between  
437 wing discs at 48h and adult antennae. As expected, the vast majority of these genes were up  
438 regulated in the adult antennae. However, one OR (MmarOR49), one IR (MmarIR68a), four  
439 OBPs and 11 CSPs were most expressed in the two wing disc developmental stages.

440 The analysis of the genes differentially expressed in wing discs at the two developmental stages  
441 sampled, 24h and 48h after pupation, both taxa combined, revealed fewer differences than the  
442 comparison with adult antennae. In this comparison, 4,851 transcripts (7.6% of the expressed  
443 transcripts) were differentially expressed, 88 of which were specific to this comparison (Fig.  
444 2). Approximately one third of them (i.e., 1,823 out of 4,851 transcripts) were more expressed  
445 at 48h than 24h, and the 3,028 remaining transcripts were most expressed at 24h. There was a  
446 large difference in the number of differentially expressed transcripts between the two wing disc  
447 stages in each subspecies, with a total of 3,494 differentially expressed transcripts for *M.*  
448 *marsaeus phasiana* versus 898 transcripts for *M. marsaeus rileyi* (Fig. 2). Among these, a large  
449 proportion was shared between the two subspecies (693), corresponding to 77% of the  
450 differentially expressed transcripts in *M. marsaeus rileyi* and 20% in *M. marsaeus phasiana*.  
451 When the two subspecies were combined, we found more up regulated (3,028) than down

452 regulated transcripts (1,823) in the comparison between wing discs at 24h and 48h. In the  
453 context of wing disc development (comparison between 24h and 48h), the automated functional  
454 annotation using Gene Ontology identified an enrichment in key cellular contents, such as  
455 “Chitin-based extracellular matrix”, and key biological processes, such as “Taurine metabolic  
456 process” (Figure S4). For WCP genes, 13 transcripts were significantly differentially expressed  
457 between 24h and 48h (significant enrichment Chi-squared test p-value = 2.069e-5) , only one  
458 was up regulated (corresponding to the *optix* gene) and 12 were down regulated (corresponding  
459 to *black*, *Dopamine-N-acetyltransferase* and *dopa\_decarboxylase* genes) (Fig. 3).

460

#### 461 *Transcriptomic differences between subspecies*

462 Comparisons of differentially expressed genes between the two subspecies and for the  
463 three tissues (wing disc at 24h, at 48h and adult antennae) showed different patterns (Fig. 2).  
464 The comparison between adult antennae revealed a large number of differentially expressed  
465 genes, with a total of 1,028 transcripts. This number was higher than for the wing discs at both  
466 time points, with 64 transcripts at 24h and 495 at 48h. The number of shared genes across the  
467 three tissues was very low. Notably, there were only 29 differentially expressed transcripts  
468 shared between the two wing disc developmental stages. By combining the three tissues, i.e.  
469 increasing sample size, the number of differentially expressed transcripts was much higher,  
470 with 4,545 transcripts in total, but 3,475 (76%) were specific to this comparison alone.

471 The ten most differentially expressed transcripts between the two subspecies in each of  
472 the four comparisons, wing discs at 24 and 48h, adult antennae, and the combination of the  
473 three libraries, were extracted (Table 2). Of these 40 transcripts, 27 transcripts were unique.  
474 Some of them were specific to certain tissues, such as DN21106\_c0\_g1\_i1 and  
475 DN49364\_c0\_g1\_i2, which were differentially expressed in adult antennae and  
476 DN61874\_c0\_g1\_i1 and DN73911\_c7\_g1\_i2, which were specific to wing discs. By contrast,  
477 some were shared between different libraries, found in the comparison of the merged libraries,  
478 such as DN74456\_c3\_g1\_i2, or identified in the four comparisons such as DN65831\_c0\_g1\_i1.  
479 Two transcripts were identified as differentially expressed by the interaction between  
480 subspecies and wing disc conditions (24h and 48h), DN61874\_c0\_g1\_i1 and  
481 DN69040\_c0\_g1\_i4. Of these 28 transcripts, 11 proteins have been predicted and 7 had a blast  
482 hit on the nr database. Their putative annotation highlights possible functions in traits other than  
483 color pattern and odorant detection (Table 2).



484 Statistical analyses of the wing color pattern genes for the two subspecies did not find  
485 any transcripts differentially expressed in any of the tissue type (i.e. 24h- and 48h wing discs,  
486 and antennae). However, two transcripts corresponding to the genes *pale* and *dopa-N-ac* were  
487 significantly differentially expressed between the two species when comparing the merged  
488 libraries (Fig. 3). The former was down regulated in *M. marsaeus phasiana* while the latter was  
489 up regulated. This difference in results was likely due to the higher statistical power provided  
490 by combining all three tissue types and therefore increasing the total number of samples.

491 Results were similar for chemosensory genes, with only two transcripts differentially  
492 expressed when analyzed in separate tissues, i.e. *MmarOBP22* and *MmarCSP33*.  
493 Unexpectedly, this significant difference between the two subspecies occurred in the wing  
494 discs, at 48h for *MmarOBP22* and 24h for *MmarCSP33*. When the libraries from different  
495 tissues were combined, five transcripts corresponding to five genes were significantly  
496 differentially expressed: *MmarOBP8*, *MmarCSP5*, *MmarCSP15*, *MmarCSP31* and  
497 *MmarCSP32*. Regarding antennae, the comparisons of the two subspecies revealed transcripts  
498 with large fold changes, associated with a nearly significant test for differential expression  
499 (FDR < 0.1 but > 0.05, Table S3). These transcripts were associated with the *MmarOR31*,  
500 *MmarGR18*, *MmarIR68a*, *MmarOBP19*, *MmarOBP20*, *MmarCSP36* and *MmarCSP37* genes.  
501 Among these, *MmarGR18* and *MmarOBP19* were most expressed in *M. marsaeus rileyi*  
502 whereas the other transcripts were most expressed in *M. marsaeus phasiana*.

503

## 504 Discussion

505 We generated the first transcriptional resources for an ithomiine species, *M. marsaeus*,  
506 which enabled us to look at gene expression in tissues of interest between diverging lineages  
507 during the early stages of speciation. We found differentially expressed genes, including genes  
508 associated with for variation in traits likely involved in reproductive isolation.

### 509 Reference transcriptome of *M. marsaeus*

510 Combining all libraries from the two subspecies to build the reference transcriptome  
511 resulted in a large amount of duplicates, but enabled us to obtain a very complete reference  
512 transcriptome, including transcripts potentially specific to each subspecies (Table 1). This  
513 transcriptome remains relatively fragmented since it consists of 179,833 transcripts with a N50  
514 of 922 bp and the annotated genes were often represented in several transcripts. This

515 fragmentation is apparent when comparing it with other butterfly transcriptomes. For instance,  
516 the first transcriptome of *H. melpomene*, generated from only wing disc tissue, consisted of  
517 82,000 contigs (Ferguson et al. 2010); the reference transcriptome of *Vanessa cardui* generated  
518 from various tissues consisted of 74,995 transcripts with a N50 length of 2062 bp (Zhang et al.  
519 2017) and eight eye transcriptomes for *Dryas iulia* and several *Heliconius* species had a number  
520 of transcripts ranging between 62,962 and 116,342 (Zhang et al. 2019). Despite this  
521 transcriptome fragmentation, a large proportion of the predicted proteins obtained a hit with  
522 reference databases (69% and 63% of proteins on nr and Lepbase databases, respectively). The  
523 first transcriptome of an ithomiine species generated by this study is therefore an important  
524 resource for future studies, both in the search for genes of interest and for population  
525 resequencing approaches, often used to investigate the evolutionary mechanisms involved in  
526 the divergence of subspecies.

527

#### 528 *Annotation of candidate WCP and chemosensory genes*

529 For this study, we identified specific candidate genes known to be important for color  
530 pattern formation in butterflies, some of which are responsible for color pattern variation in the  
531 mimetic *Heliconius* butterflies. Of these 20 WCP genes, only one, *kynurenine formamidase*,  
532 was not found in the reference transcriptome, suggesting that this gene is not expressed in our  
533 samples, and may even be absent in this species. As such, the WCP genes expressed in *M.*  
534 *marsaeus* are those classically observed in Lepidoptera.

535 The first genome assemblies of *D. plexippus* and *H. melpomene* have demonstrated that,  
536 among the chemosensory gene families, CSPs are especially diversified in butterflies compared  
537 to other Lepidoptera, with a “butterfly-specific expansion” (Zhan et al. 2011; Heliconius  
538 Genome Consortium 2012). The phylogenetic analysis carried out here revealed that among  
539 Nymphalidae, this diversification is especially apparent in the sub-family Danainae (Figure S2).  
540 Here, we found 17 MmarCSPs belonging to this lineage, a number much higher than in *D.*  
541 *plexippus* or any other lepidopteran species investigated thus far. This may indicate that CSPs  
542 have had an important role in the adaptation of the chemosensory system of Ithomiini, but it is  
543 important to note that the function of CSPs largely exceeds chemoreception. Indeed, CSPs are  
544 often widely expressed throughout the insect body and it has been proposed that they are  
545 involved in diverse functions including pheromone release, development, or carotenoid  
546 pigment transportation (Pelosi et al. 2018). Interestingly, RNAseq results show that several

547 MmarCSPs are more expressed in wing discs than in antennae, notably within the expanded  
548 lineage mentioned above. While their exact function in this tissue is unclear, it could be related  
549 to the fact that hindwings of male Ithomiini have androconia that produce pheromones (Schulz  
550 et al. 2004). This warrants further investigation.

551         Apart from CSPs, chemosensory gene repertoires identified in *M. marsaeus* appeared  
552 globally similar to what was shown in the analysis of the *D. plexippus* and *H. melpomene*  
553 genomes (Zhan et al. 2011; Heliconius Genome Consortium 2012). Notably, the annotation of  
554 candidate chemoreceptor genes expressed in antennae revealed a large diversity of ORs and  
555 antennal IRs, with no gene expansion that might be specific to Ithomiini. We did not identify  
556 any member of the so-called pheromone receptor clade in the *M. marsaeus* reference  
557 transcriptome, but we found two members of other OR lineages containing moth pheromone  
558 receptors (Bastin-Héline et al. 2019; Li et al. 2017). However, the chemical nature of  
559 compounds found in male Ithomiini androconia differ drastically from that of moth sex  
560 pheromone components (Schulz et al. 2004; Stamm et al. 2019; Mann et al. 2020), suggesting  
561 that the male ORs that bind these chemicals in Ithomiini probably do not belong to the same  
562 pheromone receptor lineages described in moths.

563

#### 564 *Differential expression during wing disc development and across tissues*

565         The analysis of the differentially expressed transcripts between the wing discs during  
566 metamorphosis and the adult antennae revealed how different these tissues are in terms of  
567 molecular pathways. Although some WCP genes, such as *optix* and *wntA*, are specific to wing  
568 discs, a large proportion of those genes are nevertheless highly expressed in adult antennae.  
569 Many WCP genes are involved in general functions, such as cycle-cell regulation and gene  
570 transcription, which may not be specific to the development of wing color patterns. Similarly,  
571 genes of the melanin pathway also influence cuticle sclerotization (Matsuoka & Monteiro  
572 2018), and it is not surprising to find them expressed in tissues other than wings, not to mention  
573 that many body parts, including antennae, are darkly pigmented. The detailed comparison  
574 between the wing discs at 24h and 48h also showed significant differences, indicating that this  
575 tissue is undergoing major changes, although we cannot rule out that some of this was due to  
576 slightly fluctuating rearing conditions in the field.

577         In wing disc libraries, all WCP genes were expressed except *cortex*. The gene *cortex* is  
578 expressed at the prepupal stage in *Biston betularia* (van't Hof et al. 2016) or at the caterpillar

579 stage in some *Heliconius* species (Livraghi et al. 2021; Nadeau et al. 2016), including the co-  
580 mimic *H. numata* (Saenko et al. 2019), two morphs of which have a remarkable resemblance  
581 to the two subspecies of *M. marsaeus* investigated here. Conversely, while *optix* does not seem  
582 to be expressed in wing discs of *H. numata*, it has a high level of expression in *M. marsaeus*.  
583 These results therefore suggest that the means of producing very similar color patterns in *M.*  
584 *marsaeus* and *H. numata* may involve different pathways. Overall, the kinetic gene expression  
585 we observed is comparable to that observed in *H. erato*, a species for which expression of the  
586 WCP genes was monitored at three time points during metamorphosis, 24h, 72h and 120h  
587 (Hines et al. 2012). For instance, we found in *M. marsaeus* an increase in the amount of  
588 expression of *optix* and a decrease in the expression of *pale* and *dopamine N-acetyltransferase*  
589 (*Dat1*) between 24h and 48h, and the same trends were observed in *H. erato* between 24h and  
590 72h. Other genes involved in melanisation, such as *yellow\_d*, *tan* or *black*, are usually expressed  
591 at a later stage (e. g., at day 5 in (Hines et al. 2012)), which may explain why we failed to find  
592 them in 24h and 48h wing discs. Overall, the kinetic expression of WCP genes in wing discs  
593 appears to be comparable between *M. marsaeus*, *H. erato* (Hines et al. 2012) and *Vanessa*  
594 *cardui* (Connahs et al. 2016), and are therefore potentially conserved within the Nymphalidae.

595

#### 596 *Differential expression at the early stages of speciation*

597 The major goal of this study was to examine differential expression at the onset of  
598 speciation in targeted tissues: wing discs and female antennae. These tissues are responsible for  
599 traits, color patterns and chemosensory detection, which are likely involved in reproductive  
600 isolation in *M. marsaeus* (McClure et al. 2019).

601 Overall, differential expression between subspecies is much less than that between  
602 tissues (Figure S3). We found differentially expressed genes between subspecies in all tissues,  
603 and more so in antennae than in wing discs, and at 48h than at 24h wing disc development.  
604 Moreover, patterns of gene expression in wing discs over time was markedly different between  
605 the two subspecies, with many more genes differentially expressed at 48h in *M. marsaeus*  
606 *phasiana* than in *M. marsaeus rileyi*. These results suggest that although the two subspecies  
607 have likely diverged recently, as testified by the low level of genetic differentiation between  
608 them (McClure et al. 2019), different developmental processes are at play in wing discs and  
609 antennae, and may contribute to the differences observed in the traits of interest. However,  
610 detailed analysis of the different tissues identified a limited number of differentially expressed  
611 transcripts between the two subspecies and hardly any candidate genes involved in wing pattern

612 and chemosensory variation. The analysis carried out by combining the different tissues  
613 allowed the identification of a larger number of transcripts. These latter results should however  
614 be considered with caution and could be linked to genetic drift, which could affect all tissues in  
615 a similar way, unlike gene expression related to specific differences in traits, which is expected  
616 to be tissue specific (Bierkhman et al. 2008). However, genes affected by drift are likely to  
617 show small differences in expression. Here, all the transcripts detected by the analysis have a  
618 substantial difference in expression ( $|\logFC| > 1$ ). Furthermore, genes affected by drift are  
619 expected to show similar expression trajectory across developmental stages. Our analysis that  
620 accounted for the interaction between subspecies and wing disc stages identified two candidate  
621 transcripts that had different expression patterns across stages between the two subspecies.  
622 More broadly, difficulties in identifying differentially expressed transcripts in each tissue could  
623 be related to small but important variations in developmental stage among the different  
624 replicates. Indeed, although the rearing and dissection conditions were controlled to the  
625 maximum, this difficult-to-breed species can only be reared close to the field and small  
626 fluctuations in environmental conditions (temperature, moisture) could have impacted the pace  
627 of development, thereby inducing variation in gene expression levels among biological  
628 replicates.

629         Examining the expression of candidate genes for these traits sheds further light on the  
630 pathways that lead to different phenotypes, and, ultimately, to reproductive isolation. Regarding  
631 color pattern, *M. marsaeus rileyi* and *phasiana* differ by the presence of yellow only in *M.*  
632 *marsaeus rileyi* (at the tip of the forewing) and slightly more melanized wings in this subspecies  
633 (Fig. 1, (McClure et al. 2019)). While several WCP genes showed different expression patterns  
634 between the two subspecies in wing discs (Fig. 3, Table S3), none of these differential  
635 expressions were statistically significant. However, when all tissues were pooled, which  
636 increased statistical power, two of these genes were differentially expressed between the  
637 subspecies: *pale* was downregulated in *M. marsaeus phasiana* compared to *M. marsaeus rileyi*,  
638 while *dopamine-N-acetyltransferase* was upregulated in *M. marsaeus phasiana*. Both genes are  
639 involved in the pathway of melanin synthesis in butterflies, and *pale* is also involved in cuticle  
640 formation (Zhang et al. 2017; Hines et al. 2012). The expression of these genes in pupal wing  
641 discs of the mimetic butterfly species *H. erato* for different color pattern elements and  
642 developmental stages (Hines et al. 2012) showed no association of *pale* with any particular  
643 color, but showed an increase in expression of *dopamine-N-acetyltransferase* in yellow-  
644 containing hindwings during melanin synthesis (i.e., at the very end of the pupal stage). In *H.*

645 *erato*, *dopamine-N-acetyltransferase* is also highly expressed in the early pupal stages (24h),  
646 but with no significant differences among color pattern elements. Unlike *H. erato*, in *M.*  
647 *marsaeus* the increased expression of *dopamine-N-acetyltransferase* is found in the subspecies  
648 that contains no yellow. However, we only have data for early pupal stages, and our data can  
649 therefore not be fully compared to those of Hines et al. (2012). It is possible that our analysis  
650 has failed to detect genes differentially expressed during other developmental stages, or because  
651 of differential expression taking place only at a very small scale, i.e. in specific wing areas  
652 corresponding to certain color pattern elements. Future investigations of WCP genes in *M.*  
653 *marsaeus* should extend to all relevant developmental stages, from the last larval instar up to  
654 the melanization stage in pupae, and should attempt to examine gene expression of specific  
655 wing areas (particularly the area containing the yellow spot in *M. marsaeus rileyi*).

656         Chemosensory traits have long been suspected to be important for the establishment or  
657 reinforcement of reproductive barriers in insects, which can occur for instance through adaptive  
658 divergence in host preference or in pheromone communication (Smadja & Butlin 2009).  
659 Although the genetic basis of chemosensory speciation remains largely unknown in insects, a  
660 combination of transcriptomics and population genomics carried out in a pair of recently  
661 diverged *Heliconius* species, *H. cydno* and *H. melpomene*, identified a few chemosensory genes  
662 differentially expressed between the two species and showing a low level of genetic admixture.  
663 One GR gene and one OBP gene were particularly likely to be involved in host plant and  
664 pheromone shifts, respectively (Eyres et al. 2016; van Schooten et al. 2020). Because *M.*  
665 *marsaeus rileyi* and *phasiana* use the same larval host plant (McClure & Elias 2016) but have  
666 been shown to diverge somewhat in their male pheromonal blend (McClure et al. 2019), we  
667 hypothesised that divergent expression patterns between the two subspecies would be more  
668 likely to have occurred in genes involved in pheromone detection in females. That said, in *M.*  
669 *marsaeus*, the difference in pheromonal blend is subtle, with substantial overlap between  
670 subspecies, and it is not known whether butterflies are able to discriminate the two subspecies  
671 based on male pheromones (McClure et al. 2019). Perhaps unsurprisingly, our differential  
672 expression analysis did not find any significant divergence in female antennae between the two  
673 subspecies. Seven chemosensory genes (one OR, two OBP and two CSP, which are likely  
674 involved in olfaction, and one GR, one IR) showed a trend for differential expression between  
675 the subspecies in this tissue. Notably, the OR (*MmarOR31*) appeared to be more than 10 times  
676 over-expressed in *M. marsaeus phasiana* (Table S3). This receptor belongs to a butterfly-  
677 specific OR lineage of unknown function, but the high duplication rate within this lineage

678 (Figure S2) suggests a link between these receptors and adaptation in these butterflies, possibly  
679 even to changes in male pheromone blends. Interestingly, the OBP and CSP genes also show a  
680 trend for differential expression in antennae of the two subspecies ( $0.05 < \text{FDR} < 0.1$ ), i.e.  
681 *MmarOBP19-20* and *MmarCSP36-37*, also belong to lineage-specific duplications (Figure S2).  
682 Many past studies have shown the importance of chemosensory gene duplication (followed by  
683 functional divergence) for the adaptation of insects to different host plants or pheromone blends  
684 (Briscoe et al. 2013; McKenzie et al. 2016; Anholt 2020; Montagné et al. 2021). Further  
685 functional studies are needed to clarify whether the difference in *M. marsaeus* male pheromone  
686 blends is perceived by the females, but if so, the genes listed here appear to be prime candidates  
687 involved in reproductive isolation.

688

## 689 *Conclusion*

690 We generated the first transcriptomic resource for an ithomiine butterfly, *M. marsaeus*,  
691 a co-mimic of certain *Heliconius* species, to assess whether gene expression in tissues of interest  
692 differed between two recently diverged subspecies that diverged in wing color pattern, and, to  
693 a lesser degree, male pheromone blend. We found that all but one known WCP gene were  
694 expressed in this species, of which all but one were expressed in wing discs. Two of the  
695 expressed WCP genes were differentially expressed between the two subspecies, suggesting  
696 that they may be involved in color pattern differentiation and, ultimately, mate choice and  
697 reproductive isolation. We also recovered a large number of chemosensory genes. One of them  
698 was slightly upregulated in one of the subspecies, and may play a role in pheromone detection  
699 and mate discrimination. Our results complement recent experimental findings that different  
700 color patterns and perhaps male pheromones drive reproductive isolation in *M. marsaeus*. Our  
701 study is also the first step towards future investigations aiming at deciphering the genetic bases  
702 that underlie wing color pattern and chemosensory variation in this species, and are a significant  
703 contribution for comparative genomics in Lepidoptera, and mimetic butterflies.

## 704 **Data Availability Statement:**

705 Scripts and methods used to perform RNA-Seq analyses are available on Github:  
706 [https://github.com/flegeai/Melinaea\\_marsaeus\\_askoR](https://github.com/flegeai/Melinaea_marsaeus_askoR). Raw reads are available on the SRA  
707 repository BioProject ID PRJNA725991.

708

709

710 **Table and figure captions:**

711 **Table 1.** Transcriptome statistics on the sequences and the annotation and their respective  
712 completeness (BUSCO results).

713 **Table 2.** Blast hit results from the seven proteins predicted on the 27 unique transcripts from  
714 the Top10 differentially expressed transcripts (top highest FDR) in the four comparisons  
715 between the two subspecies: the comparison of the adult antennae from the two subspecies  
716 (Mmp\_aa/Mmr\_aa), the wing discs at 24h (Mmp\_1/Mmr\_1), the wing discs at 48h  
717 (Mmp\_2/Mmr\_2) and when combining the three previous libraries (Mmp/Mmr). Transcripts  
718 without predicted protein or protein without blastp hits are not shown.

719 **Fig 1. A** Map of the study area around Tarapoto in Peru, including sampling localities and  
720 expected distribution of the two subspecies. **B** and **C** Representation of the tissue samples used  
721 for the RNAseq libraries: **B** one 5th instar larva (gut was removed), one pupa and one adult  
722 body (separated into three parts: head, thorax and abdomen) from *M. marsaeus rileyi* and **C**  
723 female antennae and wing discs from pupae at two different developmental stages of both *M.*  
724 *marsaeus rileyi* and *M. marsaeus phasiana*, and used for differential expression analyses.

725 **Fig 2.** Venn diagrams with the number of transcripts differentially expressed. The “Tissue  
726 comparison” is the comparison between nymphal wing discs at 24h (WD24h) and 48h (WD48h)  
727 and adult antennae (AA) for all samples of *M. marsaeus phasiana* (p) and *M. marsaeus rileyi*  
728 (r) combined. A small additional Venn diagram details the transcripts differentially expressed  
729 in the wing discs at 24h and 48h in the two subspecies separately. The “Subspecies comparison”  
730 is the comparison between the two subspecies for each tissue and for all tissue types combined.

731 **Fig 3.** Heatmap of the expression levels ( $\log_2(\text{CPM}+1)$ ) of 18 wing color pattern genes.  
732 Legend: WD = wing discs, AA = adult antennae, r = *M. marsaeus rileyi*, p = *M. marsaeus*  
733 *phasiana*. The first column of 3 points correspond to the “Tissue comparison” as follows:  
734 WD24h vs WD48h, WD24h vs AA, WD48h vs AA. The following columns of four points  
735 correspond to the “Subspecies comparison”, with separate comparisons of the two subspecies  
736 for the three tissue types, lastly followed by a comparison of all pooled tissue types. Green =  
737 up-regulated and red = down-regulated, according to the direction of the comparison indicated  
738 at the top of the column.



739 **Fig 4.** Heatmap for the level of expression ( $\log_2(\text{CPM}+1)$ ) of 65 chemosensory receptor genes.  
740 Legend: WD = wing discs, AA = adult antennae, r = *M. marsaeus rileyi*, p = *M. marsaeus*  
741 *phasiana*. The first series of points correspond to “Tissue comparison” with the following  
742 comparisons: WD24h vs WD48h, WD24h vs AA, WD48h vs AA. The four subsequent points  
743 correspond to the “Subspecies comparison”, with comparisons between the two subspecies for  
744 the three tissues, first individually and then pooled. Green = up-regulated and red = down-  
745 regulated, according to the direction of the comparison indicated at the top of the column.

746 **Fig 5.** Heatmap for the level of expression ( $\log_2(\text{CPM}+1)$ ) of 49 OBP and CSP genes. Legend:  
747 WD = wing discs, AA = adult antennae, r = *M. marsaeus rileyi*, p = *M. marsaeus phasiana*,  
748 first points correspond to the “Tissue comparison” with the comparisons: WD24h vs WD48h,  
749 WD24h vs AA, WD48h vs AA, the four points correspond to the “Subspecies comparison”  
750 with the comparisons between the two subspecies for the three tissues separated and combined.  
751 Green = up-regulated and red = down-regulated, according to the direction of the comparison  
752 indicated at the top of the column.

753

#### 754 **Supplementary material:**

755 **Table S1.** Number of samples and RNA sequencing methods used for each condition ordered  
756 by subspecies, stage and tissue.

757 **Table S2.** Sequencing statistics for each sample including raw reads number, cleaning (quality  
758 and rRNA) and mapping statistics.

759 **Table S3.** Candidate gene list with references and differential expression statistics: Fold  
760 Change and FDR for each comparison. Colors are similar to those from Figure 3,4 and 5: green  
761 = up-regulated and red = down-regulated, according to the direction of the comparison indicated  
762 at the top of the column.

763 **Figure S1. A** Pie chart showing the proportions of sequences which were successfully  
764 annotated in comparison to those that did not get a blast hit, mapping or GO annotation step.  
765 Blasted without Hits: number of sequences without Blast hits; With Blast Hits: number of  
766 sequences with Blast hits; With GO Mapping: Number of sequences that mapped to the  
767 Blast2GO database; B2G Annotated: Number of sequences that did retrieve one or more GO  
768 annotations from the Blast2GO database. **B** Number of transcripts associated to the main  
769 enzyme code level identified by interproscan and classified by Enzyme Commission classes.

770 **CD** Number of transcripts associated to the top 20 Gene Ontology (GO) terms at level 2 (C)  
771 and level 3 (D) ordered by Biological Processes (BP), Molecular Function (MF) and Cellular  
772 Component (CC).

773 **Figure S2.** Phylogenies of **A** Odorant Receptors, **B** Gustatory Receptors, **C** Ionotropic  
774 Receptors, **D** Odorant-Binding Proteins and **E** Chemosensory Proteins including *M. marsaeus*  
775 (in red), *D. plexippus* (in orange), *H. melpomene* (in green), *S. frugiperda* (in light blue), *M.*  
776 *hyperantus* (in pink), *B. anynana* (in purple), *V. tameamea* (in brown) and *B. mori* (in blue) and  
777 performed with PhyML v3.0. Node support is represented by grey dots on nodes with a  
778 approximate likelihood-ratio (aLRT)  $\geq 0.95$ .

779 **Figure S3.** **A** Density plots for the level of expression of filtered genes with at least 0.5 CPM  
780 (counts per million) in at least 3 samples. Samples names corresponds to Supplementary Table  
781 2. **B** Boxplot of the expression level distribution of the filtered genes using the TMM method.  
782 Samples names corresponds to Supplementary Table 2. **CD** Sample distribution along 1,2 and  
783 3 axes of the Multidimensional scaling (MDS) based on differential expression (DE). For *M.*  
784 *marsaeus phasiania* wing discs 24h samples are in brown, 48h in green and adult antennae in  
785 red, for *M. marsaeus rileyi* wing discs 24h samples are in blue, 48h in pink and adult antennae  
786 in turquoise. **E** Hierarchical clustering of samples and heatmap of sample correlation matrix.  
787 Samples names corresponds to Supplementary Table 2.

788 **Figure S4.** Enriched GO terms: the 10 GO for Biological Processes, Cellular Components and  
789 Molecular Function with the largest gene ratios. The size of the dots represents the number of  
790 genes in the significant DE gene list associated with the GO term for the three “Tissue  
791 comparisons”.

792

## 793 **References**

794 Alexa A, Rahnenfuhrer J. 2019. topGO: Enrichment Analysis for Gene Ontology. R package  
795 version 2.37.0.

796 Anholt RRH. 2020. Chemosensation and Evolution of Drosophila Host Plant Selection.  
797 iScience. 23:100799.

798 Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: A fast,  
799 accurate, and powerful alternative. Syst. Biol. 55:539–552.

800 Arias M et al. 2016. Crossing fitness valleys: empirical estimation of a fitness landscape  
801 associated with polymorphic mimicry. Proc. Biol. Sci. 283.

802 Bastin-Héline L et al. 2019. A novel lineage of candidate pheromone receptors for sex  
803 communication in moths. *Elife*. 8.

804 Beccaloni GW. 1997. Ecology, natural history and behaviour of Ithomiine butterflies and their  
805 mimics in Ecuador (Lepidoptera: Nymphalidae: Ithomiinae). *Trop. Lepid. Res.* 8:103–124.

806 Beccaloni GW. 1997. Vertical stratification of ithomiine butterfly (Nymphalidae: Ithomiinae)  
807 mimicry complexes: the relationship between adult flight height and larval host–plant height.  
808 *Biol. J. Linn. Soc. Lond.* 62:313–341.

809 Beldade P, Brakefield PM, Long AD. 2002. Contribution of Distal-less to quantitative  
810 variation in butterfly eyespots. *Nature*. 415:315–318.

811 Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y. 2008. Gene regulation in primates  
812 evolves under tissue-specific selection pressures. *PLoS Genet.* 4(11):e100027

813 Briscoe AD et al. 2013. Female Behaviour Drives Expression and Evolution of Gustatory  
814 Receptors in Butterflies. *PLoS Genet.* 9:e1003620.

815 Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND.  
816 *Nat. Methods.* 12:59–60.

817 Challi RJ, Kumar S, Dasmahapatra KK, Jiggins CD, Blaxter M. 2016. Lepbase: the  
818 Lepidopteran genome database. Cold Spring Harbor Laboratory. 056994.

819 Chamberlain NL, Hill RI, Kapan DD, Gilbert LE, Kronforst MR. 2009. Polymorphic butterfly  
820 reveals the missing link in ecological speciation. *Science*. 326:847–850.

821 Chazot N et al. 2018. Contrasting patterns of Andean diversification among three diverse  
822 clades of Neotropical clearwing butterflies. *Ecol. Evol.* 8:3965–3982.

823 Chazot N et al. 2016. Into the Andes: multiple independent colonizations drive montane  
824 diversity in the Neotropical clearwing butterflies Godyridina. *Mol. Ecol.* 25:5765–5784.

825 Chazot N et al. 2019. Renewed diversification following Miocene landscape turnover in a  
826 Neotropical butterfly radiation. *Glob. Ecol. Biogeogr.* 1118-1132

827 Cock PJA, Chilton JM, Grüning B, Johnson JE, Soranzo N. 2015. NCBI BLAST+ integrated  
828 into Galaxy. *Gigascience.* 4:39.

829 Conesa A et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in  
830 functional genomics research. *Bioinformatics.* 21:3674–3676.

831 Connahs H, Rhen T, Simmons RB. 2016. Transcriptome analysis of the painted lady butterfly,  
832 *Vanessa cardui* during wing color pattern development. *BMC Genomics.* 17:270.

833 Daniels EV, Murad R, Mortazavi A, Reed RD. 2014. Extensive transcriptional response  
834 associated with seasonal plasticity of butterfly wing patterns. *Mol. Ecol.* 23:6123–6134.

835 Darragh K et al. 2020. Species specificity and intraspecific variation in the chemical profiles  
836 of *Heliconius* butterflies across a large geographic range. *Ecol. Evol.* 10:3895–3918.

837 Dasmahapatra KK, Lamas G, Simpson F, Mallet J. 2010. The anatomy of a ‘suture zone’ in  
838 Amazonian butterflies: a coalescent-based test for vicariant geographic divergence and  
839 speciation. *Mol. Ecol.* 19:4283–4301.

840 De-Silva DL, Elias M, Willmott K, Mallet J, Day JJ. 2016. Diversification of clearwing  
841 butterflies with the rise of the Andes. *J. Biogeogr.* 43:44–58.

842 Devries PJ, Lande R, Murray D. 1999. Associations of co-mimetic ithomiine butterflies on  
843 small spatial and temporal scales in a Neotropical rainforest. *Biol. J. Linn. Soc.* 67:73.

844 Dhungel B et al. 2016. Distal-less induces elemental color patterns in Junonia butterfly wings.  
845 *Zoological Letters.* 2:4.

846 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high  
847 throughput. *Nucleic Acids Res.* 32:1792–1797.

848 Elias M, Gompert Z, Jiggins C, Willmott K. 2008. Mutualistic interactions drive ecological  
849 niche convergence in a diverse butterfly community. *PLoS Biol.* 6:2642–2649.

850 Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome  
851 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.

852 Eyres I et al. 2016. Differential gene expression according to race and host plant in the pea  
853 aphid. *Mol. Ecol.* 25:4197–4215.

854 Ferguson L et al. 2010. Characterization of a hotspot for mimicry: assembly of a butterfly  
855 wing transcriptome to genomic sequence at the HmYb/Sb locus. *Mol. Ecol.* 1:240–254.

856 Gauthier J et al. 2020. Contrasting genomic and phenotypic outcomes of hybridization  
857 between pairs of mimetic butterfly taxa across a suture zone. *Mol. Ecol.* 29:1328–1343.

858 Gilbert LE. 2003. Adaptive novelty through introgression in *Heliconius* wing patterns:  
859 Evidence for shared genetic ‘tool box’ from synthetic hybrid zones and a theory of  
860 diversification. In *Butterflies*. University of Chicago Press.

861 González-Rojas MF et al. 2020. Chemical signals act as the main reproductive barrier  
862 between sister and mimetic *Heliconius* butterflies. *Proc. Biol. Sci.* 287:20200587.

863 Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user  
864 interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27:221–  
865 224.

866 Guindon S et al. 2010. New algorithms and methods to estimate maximum-likelihood  
867 phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–321.

868 Guo H et al. 2017. Expression map of a complete set of gustatory receptor genes in  
869 chemosensory organs of *Bombyx mori*. *Insect Biochem. Mol. Biol.* 82:74–82.

870 Haas BJ et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the  
871 Trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–1512.

872 *Heliconius* Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of

873 mimicry adaptations among species. *Nature*. 487:94–98.

874 Hill RI. 2010. Habitat segregation among mimetic ithomiine butterflies (Nymphalidae). *Evol.*  
875 *Ecol.* 24:273–285.

876 Hines HM et al. 2012. Transcriptome analysis reveals novel patterning and pigmentation  
877 genes underlying *Heliconius* butterfly wing pattern variation. *BMC Genomics*. 13:288.

878 van't Hof AE et al. 2016. The industrial melanism mutation in British peppered moths is a  
879 transposable element. *Nature*. 534:102–105.

880 Hori M, Hiruma K, Riddiford LM. 1984. Cuticular melanization in the tobacco hornworm  
881 larva. *Insect Biochem.* 14:267–274.

882 Jiggins CD, Mallarino R, Willmott KR, Bermingham E. 2006. The phylogenetic pattern of  
883 speciation and wing pattern change in neotropical ithomiid butterflies (Lepidoptera:  
884 Nymphalidae). *Evolution*. 60:1454–1466.

885 Jiggins CD, Naisbit RE, Coe RL, Mallet J. 2001. Reproductive isolation caused by colour  
886 pattern mimicry. *Nature*. 411:302–305.

887 Jiggins CD, Wallbank RWR, Hanly JJ. 2017. Waiting in the wings: what can we learn about  
888 gene co-option from the diversification of butterfly wing patterns? *Philos. Trans. R. Soc.*  
889 *Lond. B Biol. Sci.* 372.

890 Johnson M et al. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36:W5–9.

891 Jones P et al. 2014. InterProScan 5: genome-scale protein function classification.  
892 *Bioinformatics*. 30:1236–1240.

893 Joron M et al. 2006. A conserved supergene locus controls colour pattern diversity in  
894 *Heliconius* butterflies. *PLoS Biol.* 4:e303.

895 Joron M, Jiggins CD, Papanicolaou A, McMillan WO. 2006. *Heliconius* wing patterns: an  
896 evo-devo model for understanding phenotypic diversity. *Heredity*. 97:157–167.

897 Joron M, Wynne IR, Lamas G, Mallet J. 1999. Variable Selection and the Coexistence of  
898 Multiple mimetic forms of the Butterfly *Heliconius numata*. *Evol. Ecol.* 13:721–754.

899 Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence  
900 alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20:1160–1166.

901 Koch PB et al. 1998. Regulation of dopa decarboxylase expression during colour pattern  
902 formation in wild-type and melanic tiger swallowtail butterflies. *Development*. 125:2303–  
903 2313.

904 Kozak KM et al. 2015. Multilocus species trees show the recent adaptive radiation of the  
905 mimetic *Heliconius* butterflies. *Syst. Biol.* 64:505–524.

906 Kronforst MR, Kapan DD, Gilbert LE. 2006. Parallel genetic architecture of parallel adaptive  
907 radiations in mimetic *Heliconius* butterflies. *Genetics*. 174:535–539.

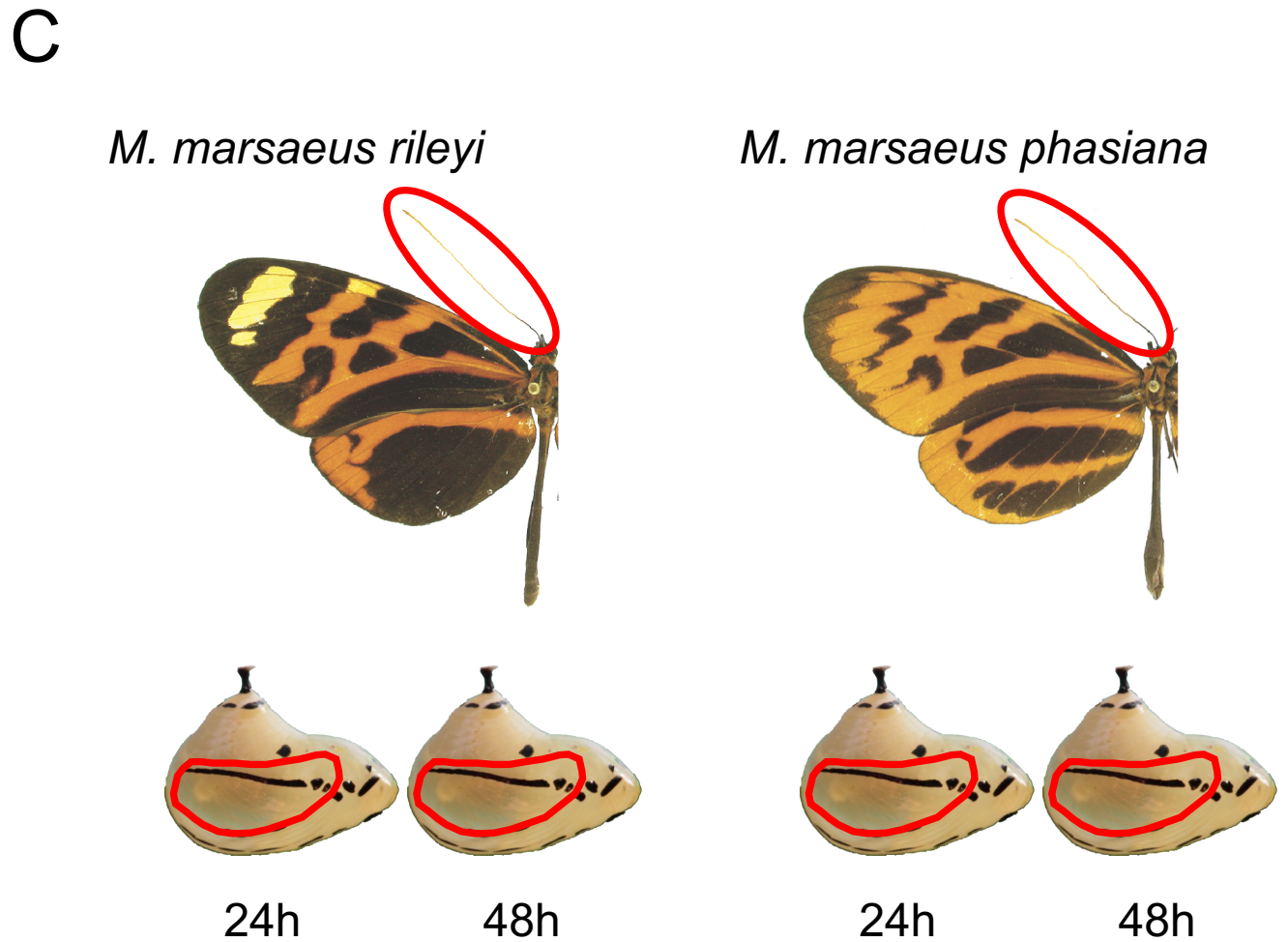
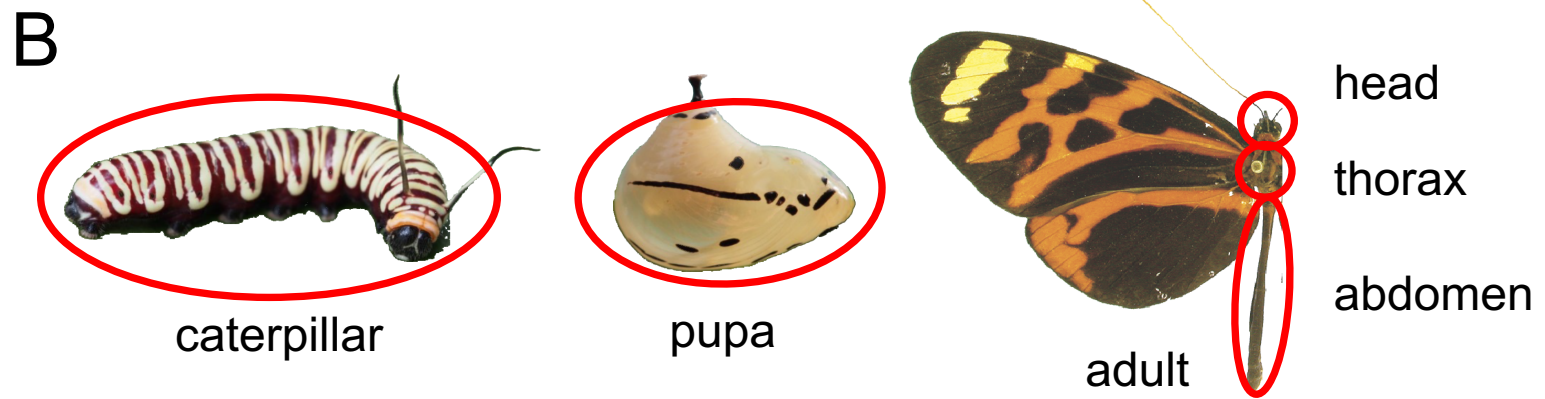
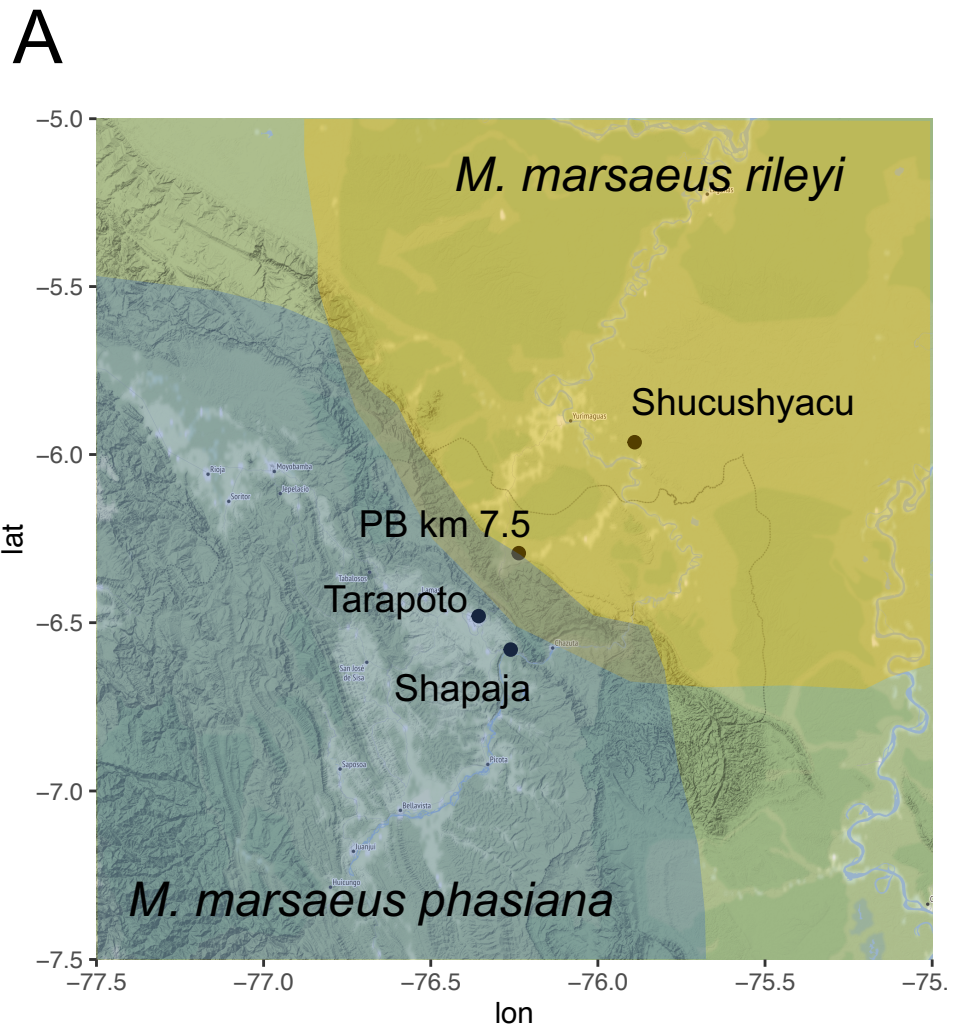
- 908 Kunte K et al. 2014. doublesex is a mimicry supergene. *Nature*. 507:229–232.
- 909 Kuwalekar M, Deshmukh R, Padvi A, Kunte K. 2020. Molecular Evolution and  
910 Developmental Expression of Melanin Pathway Genes in Lepidoptera. *Frontiers in Ecology*  
911 *and Evolution*. 8:226.
- 912 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*.  
913 9:357–359.
- 914 Lefort V, Longueville J-E, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. *Mol.*  
915 *Biol. Evol.* 34:2422–2424.
- 916 Lisa De-Silva D et al. 2017. North Andean origin and diversification of the largest ithomiine  
917 butterfly genus. *Sci. Rep.* 7:45966.
- 918 Livraghi L et al. 2021. Cortex cis-regulatory switches establish scale colour identity and  
919 pattern diversity in *Heliconius*. *eLife* 2021;10:e68549.
- 920 Li Z-Q et al. 2017. Chemosensory Gene Families in *Ectropis grisescens* and Candidates for  
921 Detection of Type-II Sex Pheromones. *Front. Physiol.* 8:953.
- 922 Llaurens V, Joron M, Théry M. 2014. Cryptic differences in colour among Müllerian mimics:  
923 how can the visual capacities of predators and prey shape the evolution of wing colours? *J.*  
924 *Evol. Biol.* 27:531–540.
- 925 Mann F et al. 2020. 3-Acetoxy-fatty acid isoprenyl esters from androconia of the ithomiine  
926 butterfly *Ithomia salapia*. *Beilstein J. Org. Chem.* 16:2776–2787.
- 927 Martin A et al. 2012. Diversification of complex butterfly wing patterns by repeated  
928 regulatory evolution of a Wnt ligand. *Proc. Natl. Acad. Sci. U. S. A.* 109:12632–12637.
- 929 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing  
930 reads. *EMBnet.journal.* 17:10–12.
- 931 Matsuoka Y, Monteiro A. 2018. Melanin Pathway Genes Regulate Color and Morphology of  
932 Butterfly Wing Scales. *Cell Rep.* 24:56–65.
- 933 Mazo-Vargas A et al. 2017. Macroevolutionary shifts of WntA function potentiate butterfly  
934 wing-pattern diversity. *Proc. Natl. Acad. Sci. U. S. A.* 114:10701–10706.
- 935 McClure M et al. 2019. Does divergent selection predict the evolution of mate preference and  
936 reproductive isolation in the tropical butterfly genus *Melinaea* (Nymphalidae: Ithomiini)? *J.*  
937 *Anim. Ecol.* 88:940–952.
- 938 McClure M, Elias M. 2016. Ecology, life history, and genetic differentiation in  
939 Neotropical *Melinaea* (Nymphalidae: Ithomiini) butterflies from north-eastern Peru. *Zool. J.*  
940 *Linn. Soc.*
- 941 McKenzie SK, Fetter-Pruneda I, Ruta V, Kronauer DJC. 2016. Transcriptomics and  
942 neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in  
943 chemical communication. *Proc. Natl. Acad. Sci. U. S. A.* 113 (49):14091–14096.

- 944 Mérot C, Frérot B, Leppik E, Joron M. 2015. Beyond magic traits: Multimodal mating cues in  
945 *Heliconius* butterflies. *Evolution*. 69:2891–2904.
- 946 Merrill RM et al. 2012. Disruptive ecological selection on a mating cue. *Proc. Biol. Sci.*  
947 279:4907–4913.
- 948 Merrill RM et al. 2011. Mate preference across the speciation continuum in a clade of  
949 mimetic butterflies. *Evolution*. 65:1489–1500.
- 950 Montagné N, Wanner K, Jacquin-Joly E. 2021. Olfactory genomics within the Lepidoptera. In  
951 *Insect Pheromone Biochemistry and Molecular Biology (Second Edition)*. pp469-505.
- 952 Monteiro A et al. 2013. Distal-less regulates eyespot patterns and melanization in *Bicyclus*  
953 butterflies. *J. Exp. Zool. B Mol. Dev. Evol.* 320:321–331.
- 954 Müller F. 1879. Ituna and Thyridia; a remarkable case of mimicry in butterflies. *Royal Ent.*  
955 *Soc. London Trans.*
- 956 Nadeau NJ et al. 2014. Population genomics of parallel hybrid zones in the mimetic  
957 butterflies, *H. melpomene* and *H. erato*. *Genome Res.* 24:1316–1333.
- 958 Nadeau NJ et al. 2016. The gene cortex controls mimicry and crypsis in butterflies and moths.  
959 *Nature*. 534:106–110.
- 960 Nieberding CM et al. 2008. The male sex pheromone of the butterfly *Bicyclus anynana*:  
961 towards an evolutionary analysis. *PLoS One*. 3:e2751.
- 962 Nosil P. 2012. *Ecological Speciation*. Oxford University Press.
- 963 Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR. 2018. Beyond chemoreception: diverse tasks  
964 of soluble olfactory proteins in insects. *Biol. Rev. Camb. Philos. Soc.* 93:184–200.
- 965 Pelosi P, Zhou J-J, Ban LP, Calvello M. 2006. Soluble proteins in insect chemical  
966 communication. *Cell. Mol. Life Sci.* 63:1658–1676.
- 967 Prakash A, Monteiro A. 2018. *apterous A* specifies dorsal wing patterns and sexual traits in  
968 butterflies. *Proc. Biol. Sci.* 285.
- 969 Reed RD et al. 2011. *optix* drives the repeated convergent evolution of butterfly wing pattern  
970 mimicry. *Science*. 333:1137–1141.
- 971 Reed RD, McMillan WO, Nagy LM. 2008. Gene expression underlying adaptive variation in  
972 *Heliconius* wing patterns: non-modular regulation of overlapping cinnabar and vermilion  
973 prepatterns. *Proc. Biol. Sci.* 275:37–45.
- 974 Reed RD, Serfas MS. 2004. Butterfly wing pattern evolution is associated with changes in a  
975 Notch/Distal-less temporal pattern formation process. *Curr. Biol.* 14:1159–1166.
- 976 Robertson HM. 2019. Molecular Evolution of the Major Arthropod Chemoreceptor Gene  
977 Families. *Annu. Rev. Entomol.* 64:227–242.
- 978 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for

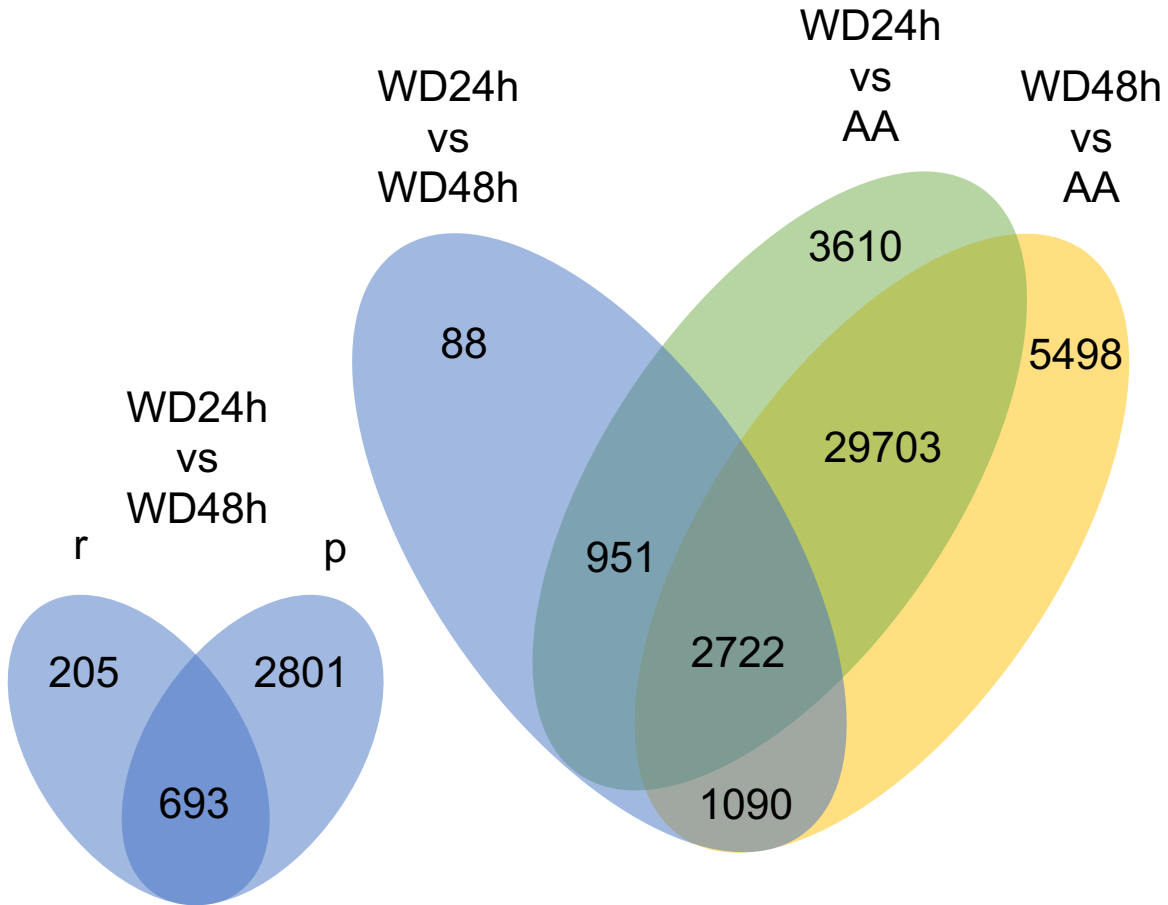
- 979 differential expression analysis of digital gene expression data. *Bioinformatics*. 26:139–140.
- 980 Saenko SV et al. 2019. Unravelling the genes forming the wing pattern supergene in the  
981 polymorphic butterfly *Heliconius numata*. *Evodevo*. 10:16.
- 982 Sánchez-Alcañiz JA et al. 2018. An expression atlas of variant ionotropic glutamate receptors  
983 identifies a molecular basis of carbonation sensing. *Nat. Commun*. 9:4252.
- 984 Sarto i Monteys V, Quero C, Santa-Cruz MC, Rosell G, Guerrero A. 2016. Sexual  
985 communication in day-flying Lepidoptera with special reference to castniids or ‘butterfly-  
986 moths’. *Bull. Entomol. Res.* 106:421–431.
- 987 S. Brown K. 1977. Geographical patterns of evolution in Neotropical Lepidoptera:  
988 differentiation of the species of *Melinaea* and *Mechanitis* (Nymphalidae, Ithomiinae). *Syst.*  
989 *Entomol.* 2:161–197.
- 990 van Schooten B et al. 2020. Divergence of chemosensing during the early stages of speciation.  
991 *Proc. Natl. Acad. Sci. U. S. A.* 117:16438–16447.
- 992 Schulz S et al. 2004. Semiochemicals derived from pyrrolizidine alkaloids in male ithomiine  
993 butterflies (Lepidoptera: Nymphalidae: Ithomiinae). *Biochem. Syst. Ecol.* 32:699–713.
- 994 Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and  
995 Annotation Completeness. In: *Gene Prediction: Methods and Protocols*. Springer New York:  
996 New York, NY pp. 227–245.
- 997 Smadja C, Butlin RK. 2009. On the scent of speciation: the chemosensory system and its role  
998 in premating isolation. *Heredity*. 102:77–97.
- 999 Stamm P, Mann F, McClure M, Elias M, Schulz S. 2019. Chemistry of the Androconial  
1000 Secretion of the Ithomiine Butterfly *Oleria onega*. *J. Chem. Ecol.* 45:768–778.
- 1001 Vogt RG, Große-Wilde E, Zhou J-J. 2015. The Lepidoptera Odorant Binding Protein gene  
1002 family: Gene gain and loss within the GOBP/PBP complex of moths and butterflies. *Insect*  
1003 *Biochem. Mol. Biol.* 62:142–153.
- 1004 Watt WB, Boggs CL. 2003. Butterflies as model systems in ecology and evolution: Present  
1005 and future. In book: *Butterflies: Ecology and Evolution Taking Flight* Publisher: University of  
1006 Chicago Press.
- 1007 Westerman EL et al. 2018. Aristaless Controls Butterfly Wing Color Variation Used in  
1008 Mimicry and Mate Choice. *Curr. Biol.* 28:3469–3474.e4.
- 1009 Willmott KR, Robinson Willmott JC, Elias M, Jiggins CD. 2017. Maintaining mimicry  
1010 diversity: optimal warning colour patterns differ among microhabitats in Amazonian  
1011 clearwing butterflies. *Proc. Biol. Sci.* 284.
- 1012 Zhang L et al. 2017. Genetic Basis of Melanin Pigmentation in Butterfly Wings. *Genetics*.  
1013 205:1537–1550.
- 1014 Zhang L, Reed RD. 2016. Genome editing in butterflies reveals that spalt promotes and  
1015 Distal-less represses eyespot colour patterns. *Nat. Commun.* 7:11769.



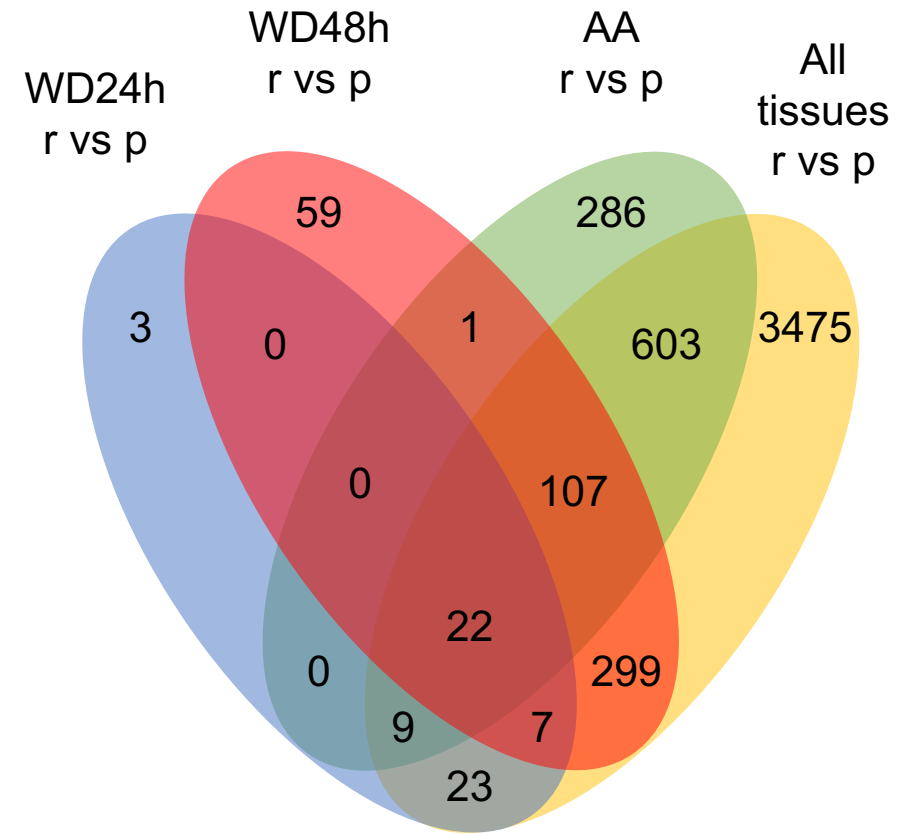
- 1016 Zhang W et al. 2019. Comparative Transcriptomics Provides Insights into Reticulate and  
1017 Adaptive Evolution of a Butterfly Radiation. *Genome Biol. Evol.* 11:2963–2975.
- 1018 Zhan S, Merlin C, Boore JL, Reppert SM. 2011. The monarch butterfly genome yields  
1019 insights into long-distance migration. *Cell.* 147:1171–1185.

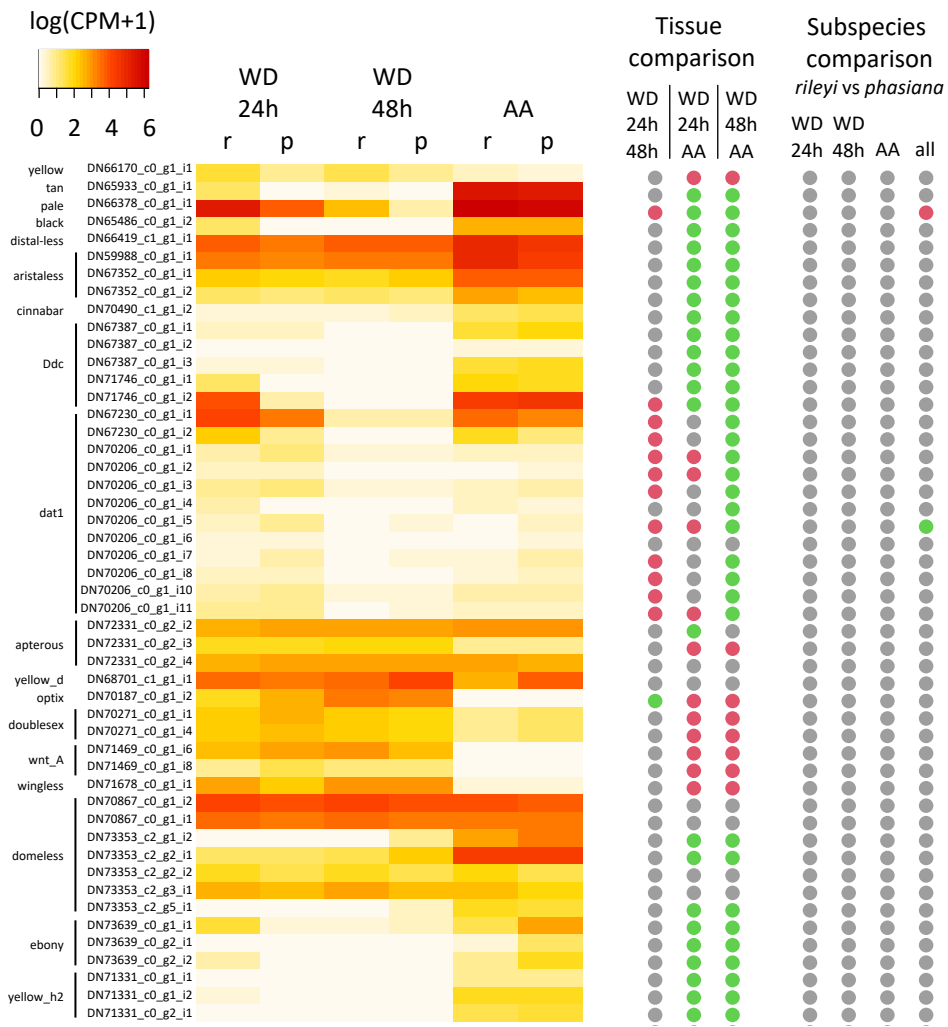


## Tissue comparison

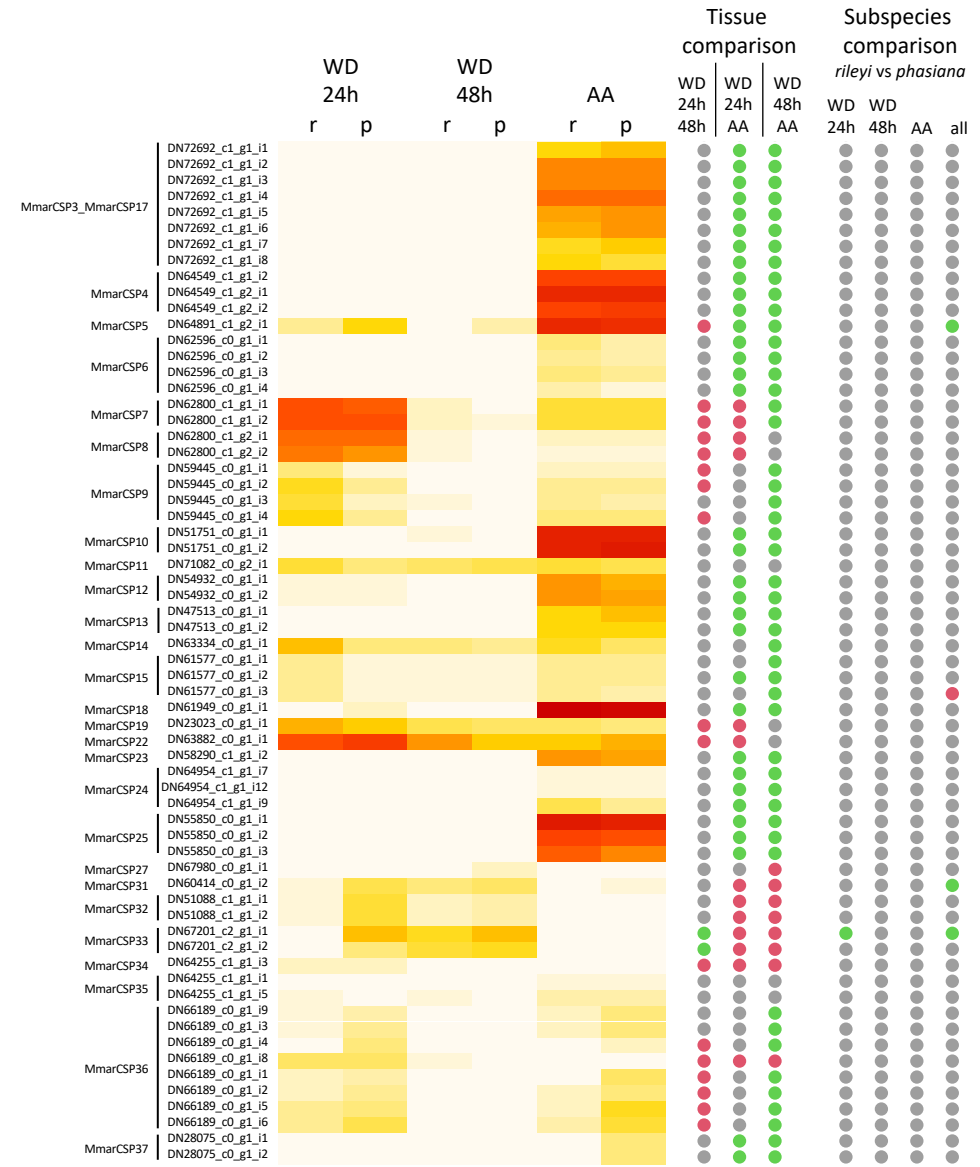
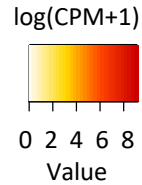
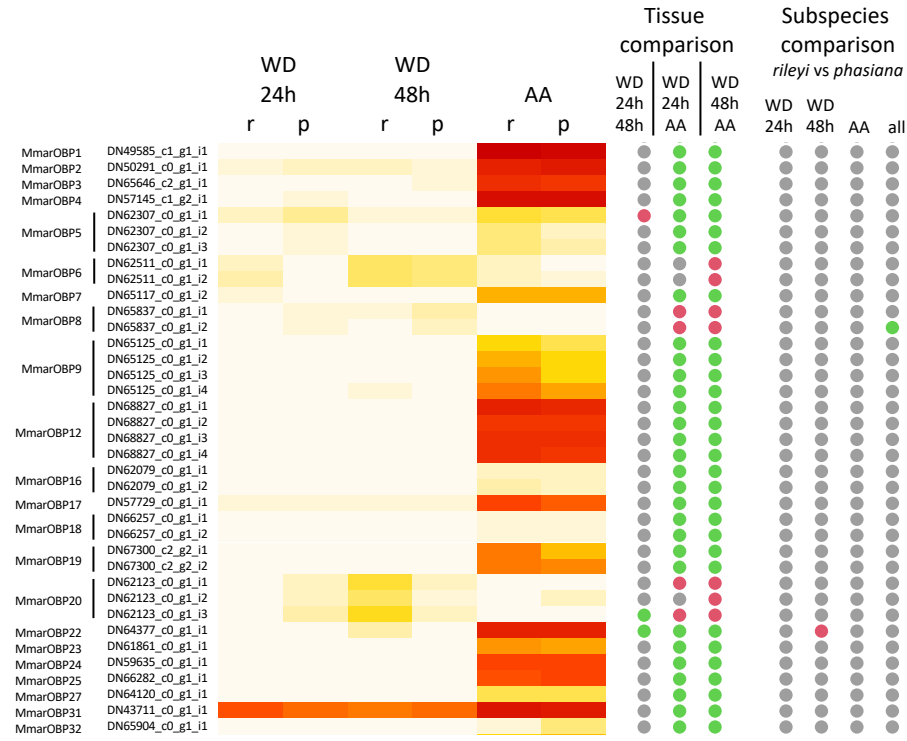


## Subspecies comparison









**Transcriptome assembly statistics**

Total transcripts	179 833
Total length assembled bases	111 561 423
Average contig length (bases)	620
Median contig length	342
Max length (bases)	33 457
Min length (bases)	201
GC (%)	38.21
Contig N50 (bases)	922

Complete and single-copy BUSCOs	939 (68.7%)
Complete and duplicated BUSCOs	271 (19.8%)
Fragmented BUSCOs	79 (5.8%)
Missing BUSCOs	78 (5.7%)

**Predicted ORF / Protein statistics**

# ORFs / Proteins	82 469
-------------------	--------

Complete and single-copy BUSCOs	969 (70.9%)
Complete and duplicated BUSCOs	271 (19.8%)
Fragmented BUSCOs	91 (6.7%)
Missing BUSCOs	36 (2.6%)

**Functional annotation statistics**

# Contig with a match to nr (NCBI)	57 313
# Contig with a match to lepbases	52 228
# ORFs annotated in GO	36 683
Sequences with enzyme code assigned	7 182

Table 1

Comparison Top10	Protein predicted	Length	Score	%id	Evalue	Accession	Definition
Mmp/Mmr; Mmp_aa/Mmv_aa	DN21106_c0_g1_i1.p1	466	-18.99	68.243	0.0	XP_023940701	nose resistant to fluoxetine protein 6-like [Bicyclus anynana]
Mmp_aa/Mmv_aa	DN49364_c0_g1_i2.p1	108	11.16	76.636	6.92e-47	XP_032517920	sortilin-related receptor isoform X2 [Danaus plexippus plexippus]
Mmp_1/Mmv_1	DN61874_c0_g1_i1.p1	457	70.67	75.817	0.0	XP_026497302	serpin B3 [Vanessa tameamea]
Mmp/Mmr; Mmp_aa/Mmv_aa; Mmp_1/Mmv_1; Mmp_2/Mmv_2	DN65831_c0_g1_i1.p1	86	10.88	80.000	4.54e-39	OWR54905	endoplasmic reticulum resident protein 29 [Danaus plexippus plexippus]
Mmp_1/Mmv_1	DN67201_c2_g1_i1.p1	123	7.52	61.947	2.44e-47	XP_032519987	chemosensory protein 7 precursor [Danaus_plexippus_plexippus]
Mmp_2/Mmv_2	DN73911_c7_g1_i2.p1	91	7.75	79.570	1.60e-46	KPJ02711	putative tyrosyl-tRNA synthetase, mitochondrial [Papilio xuthus]
Mmp/Mmr	DN74456_c3_g1_i2.p1	105	33.18	80.769	4.66e-52	XP_032524437	CCAAT/enhancer-binding protein zeta [Danaus plexippus plexippus]

Table 2