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.

Manuscript published as:
Piron-Prunier F, Persyn E, Legeai F, McClure M, Meslin C, Robin S, Alves-Carvalho S, Mohammad A, Blugeon C, Jacquin-Joly E, Montagné N, Elias M*, Gauthier J*. 2021. Comparative transcriptome analysis at the onset of speciation in a mimetic butterfly, the Ithomiini Melinaea marsaeus. J Evol Biol, doi: 10.1111/jeb.13940. Online ahead of print. * co-last

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

Short title: Transcriptomics in mimetic *Melinaea marsaeus*

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

§ co-first

* co-last

† corresponding authors: marianne.elias@mnhn.fr and jeremy.gauthier@ville-ge.ch

1. Institut de Systématique, Evolution, Biodiversité, MNHN, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France
2. Institute of Ecology and Environmental Sciences of Paris, Sorbonne Université, INRAE, CNRS, IRD, UPEC, Université de Paris, Paris, France
3. BIPAA, IGEPP, INRAE, Institut Agro, Univ Rennes, 35000, Rennes, France
4. Univ Rennes, Inria, CNRS, IRISA, 35000, Rennes, France
5. Laboratoire Écologie, Évolution, Interactions des Systèmes Amazoniens (LEEISA), Université de Guyane, CNRS, IFREMER, 97300 Cayenne, France
6. Genomics core facility, Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
7. Geneva Natural History Museum, 1 Route de Malagnou, 1208 Geneva, Switzerland

Authors contributions

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

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 Mori-Pezo 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).

Abstract

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

Significance statement:

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

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

Introduction

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

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

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

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

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

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

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

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

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

Material and Methods

Sample collection

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

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

Total RNA extraction

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

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

RNAseq library preparation and sequencing

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

Reads pre-processing

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

De novo reference transcriptome assembly

In order to generate a reference transcriptome for *M. marsaeus* (hereafter, transcriptome), high-quality reads from all *M. marsaeus* libraries (paired reads and single reads) were assembled *de novo* using the trinity v2.4.0 transcriptome assembler with default parameters (Haas et al. 2013). Completeness of the assembled transcriptome was assessed using the BUSCO v4.0.6 software (Benchmarking Universal Single-Copy Orthologs) (Seppey et al. 2019), which tests the assembly for the presence of 1 367 single-copy genes highly conserved in insects (insect_odb10).

Functional annotation and classification

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

Identification of wing color pattern (WCP) genes

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

in the pathway for melanin synthesis (*yellow*, *yellow_d*, *yellow_h2*, *tan*, *pale*, *black*, *Ddc_dopa_decarboxylase*, *ebony* and *dopamine_N_acetyltransferase* (Hori et al. 1984; Koch et al. 1998; Ferguson et al. 2010; Hines et al. 2012; Daniels et al. 2014; Zhang et al. 2017; Kuwalekar et al. 2020), and in the ommochrome synthesis pathway (*cinnabar* and *kynurenine formamidase,* (Hines et al. 2012; Reed et al. 2008; Daniels et al. 2014). We then extracted the *M. marsaeus* proteins and corresponding transcripts from the Orthofinder Orthogroups.

Annotation of candidate chemosensory genes

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

Differential gene expression analysis (DGE)

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

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

Results

Sequencing and transcriptome statistics

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

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

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

Candidate gene annotation

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

We annotated 51 candidate ORs (125 transcripts), including the coreceptor Orco, 22 candidate IRs (78 transcripts) and 21 candidate GRs (46 transcripts) (Table S3). The large diversity of ORs present in the reference transcriptome of *M. marsaeus* is mirrored by the fact that the ORs (hereafter, MmarORs) were identified within almost every paralogous lineage of the Lepidoptera OR phylogeny, with the notable exception of the so-called pheromone receptor clade (Figure S2). That said, MmarOR35 and MmarOR38 clustered within clades that have recently been shown to also contain sex pheromone receptors (Bastin-Héline et al. 2019; Li et al. 2017). We also identified five members of an OR lineage specific to Papilionoidea ("butterfly-specific expansion" in Figure S2). A similar diversity was found for MmarIRs, as we identified all four coreceptors (IR8a, IR25a, IR76b, IR93a) and all but one of the highly conserved antennal IRs. On the other hand, we identified only four divergent IRs, known to be expressed in gustatory tissues in *Drosophila* (Sánchez-Alcañiz et al. 2018). In regard to GRs, we identified transcripts encoding for candidate CO_2 and sugar receptors as well as homologs of the *Drosophila* fructose receptor GR43a, but annotated only a few MmarGRs belonging to

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

Differential Gene Expression

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

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

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

Patterns of gene Expression during wing disc development

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

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

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

Transcriptomic differences between subspecies

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

The ten most differentially expressed transcripts between the two subspecies in each of the four comparisons, wing discs at 24 and 48h, adult antennae, and the combination of the three libraries, were extracted (Table 2). Of these 40 transcripts, 27 transcripts were unique. Some of them were specific to certain tissues, such as DN21106_c0_g1_i1 and DN49364_c0_g1_i2, which were differentially expressed in adult antennae and DN61874_c0_g1_i1 and DN73911_c7_g1_i2, which were specific to wing discs. By contrast, some were shared between different libraries, found in the comparison of the merged libraries, such as DN74456_c3_g1_i2, or identified in the four comparisons such as DN65831_c0_g1_i1. Two transcripts were identified as differentially expressed by the interaction between subspecies and wing disc conditions (24h and 48h), DN61874_c0_g1_i1 and DN69040_c0_g1_i4. Of these 28 transcripts, 11 proteins have been predicted and 7 had a blast hit on the nr database. Their putative annotation highlights possible functions in traits other than color pattern and odorant detection (Table 2).

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

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

Discussion

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

Reference transcriptome of M. marsaeus

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

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

Annotation of candidate WCP and chemosensory genes

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

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

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

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

Differential expression during wing disc development and across tissues

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

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

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

Differential expression at the early stages of speciation

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

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

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

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

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

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

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

Conclusion

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

Data Availability Statement:

Scripts and methods used to perform RNA-Seq analyses are available on Github: https://github.com/flegeai/Melinaea_marsaeus_askoR. Raw reads are available on the SRA repository BioProject ID PRJNA725991.

Table and figure captions:

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

Table 2. Blast hit results from the seven proteins predicted on the 27 unique transcripts from the Top10 differentially expressed transcripts (top highest FDR) in the four comparisons between the two subspecies: the comparison of the adult antennae from the two subspecies (Mmp_aa/Mmr_aa), the wing discs at 24h (Mmp_1/Mmr_1), the wing discs at 48h (Mmp_2/Mmr_2) and when combining the three previous libraries (Mmp/Mmr). Transcripts without predicted protein or protein without blastp hits are not shown.

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

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

Fig 3. Heatmap of the expression levels (log2 (CPM+1)) of 18 wing color pattern genes. Legend: WD = wing discs, AA = adult antennae, r = *M. marsaeus rileyi*, p = *M. marsaeus phasiana*. The first column of 3 points correspond to the "Tissue comparison" as follows: WD24h vs WD48h, WD24h vs AA, WD48h vs AA. The following columns of four points correspond to the "Subspecies comparison", with separate comparisons of the two subspecies for the three tissue types, lastly followed by a comparison of all pooled tissue types. Green = up-regulated and red = down-regulated, according to the direction of the comparison indicated at the top of the column.

Fig 4. Heatmap for the level of expression (log2 (CPM+1)) of 65 chemosensory receptor genes. Legend: WD = wing discs, AA = adult antennae, r = *M. marsaeus rileyi*, p = *M. marsaeus phasiana*. The first series of points correspond to "Tissue comparison" with the following comparisons: WD24h vs WD48h, WD24h vs AA, WD48h vs AA. The four subsequent points correspond to the "Subspecies comparison", with comparisons between the two subspecies for the three tissues, first individually and then pooled. Green = up-regulated and red = down-regulated, according to the direction of the comparison indicated at the top of the column.

Fig 5. Heatmap for the level of expression (log2 (CPM+1)) of 49 OBP and CSP genes. Legend: WD = wing discs, AA = adult antennae, r = *M. marsaeus rileyi*, p = *M. marsaeus phasiana*, first points correspond to the "Tissue comparison" with the comparisons: WD24h vs WD48h, WD24h vs AA, WD48h vs AA, the four points correspond to the "Subspecies comparison" with the comparisons between the two subspecies for the three tissues separated and combined. Green = up-regulated and red = down-regulated, according to the direction of the comparison indicated at the top of the column.

Supplementary material:

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

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

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

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

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

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

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

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

References

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

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

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

Arias M et al. 2016. Crossing fitness valleys: empirical estimation of a fitness landscape 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 neotropicalithomiabutterflies (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 NeotropicalMelinaea (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

differential expression analysis of digital gene expression data. Bioinformatics. 26:139–140.

Saenko SV et al. 2019. Unravelling the genes forming the wing pattern supergene in the polymorphic butterfly Heliconius numata. Evodevo. 10:16.

Sánchez-Alcañiz JA et al. 2018. An expression atlas of variant ionotropic glutamate receptors identifies a molecular basis of carbonation sensing. Nat. Commun. 9:4252.

Sarto i Monteys V, Quero C, Santa-Cruz MC, Rosell G, Guerrero A. 2016. Sexual communication in day-flying Lepidoptera with special reference to castniids or 'butterfly-moths'. Bull. Entomol. Res. 106:421–431.

S. Brown K. 1977. Geographical patterns of evolution in Neotropical Lepidoptera: differentiation of the species of Melinaea and Mechanitis (Nymphalidae, Ithomiinae). Syst. Entomol. 2:161–197.

van Schooten B et al. 2020. Divergence of chemosensing during the early stages of speciation. Proc. Natl. Acad. Sci. U. S. A. 117:16438–16447.

Schulz S et al. 2004. Semiochemicals derived from pyrrolizidine alkaloids in male ithomiine butterflies (Lepidoptera: Nymphalidae: Ithomiinae). Biochem. Syst. Ecol. 32:699–713.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Gene Prediction: Methods and Protocols. Springer New York: New York, NY pp. 227–245.

Smadja C, Butlin RK. 2009. On the scent of speciation: the chemosensory system and its role in premating isolation. Heredity. 102:77–97.

Stamm P, Mann F, McClure M, Elias M, Schulz S. 2019. Chemistry of the Androconial Secretion of the Ithomiine Butterfly Oleria onega. J. Chem. Ecol. 45:768–778.

Vogt RG, Große-Wilde E, Zhou J-J. 2015. The Lepidoptera Odorant Binding Protein gene family: Gene gain and loss within the GOBP/PBP complex of moths and butterflies. Insect Biochem. Mol. Biol. 62:142–153.

Watt WB, Boggs CL. 2003. Butterflies as model systems in ecology and evolution: Present and future. In book: Butterflies: Ecology and Evolution Taking FlightPublisher: University of Chicago Press.

Westerman EL et al. 2018. Aristaless Controls Butterfly Wing Color Variation Used in Mimicry and Mate Choice. Curr. Biol. 28:3469–3474.e4.

Willmott KR, Robinson Willmott JC, Elias M, Jiggins CD. 2017. Maintaining mimicry diversity: optimal warning colour patterns differ among microhabitats in Amazonian clearwing butterflies. Proc. Biol. Sci. 284.

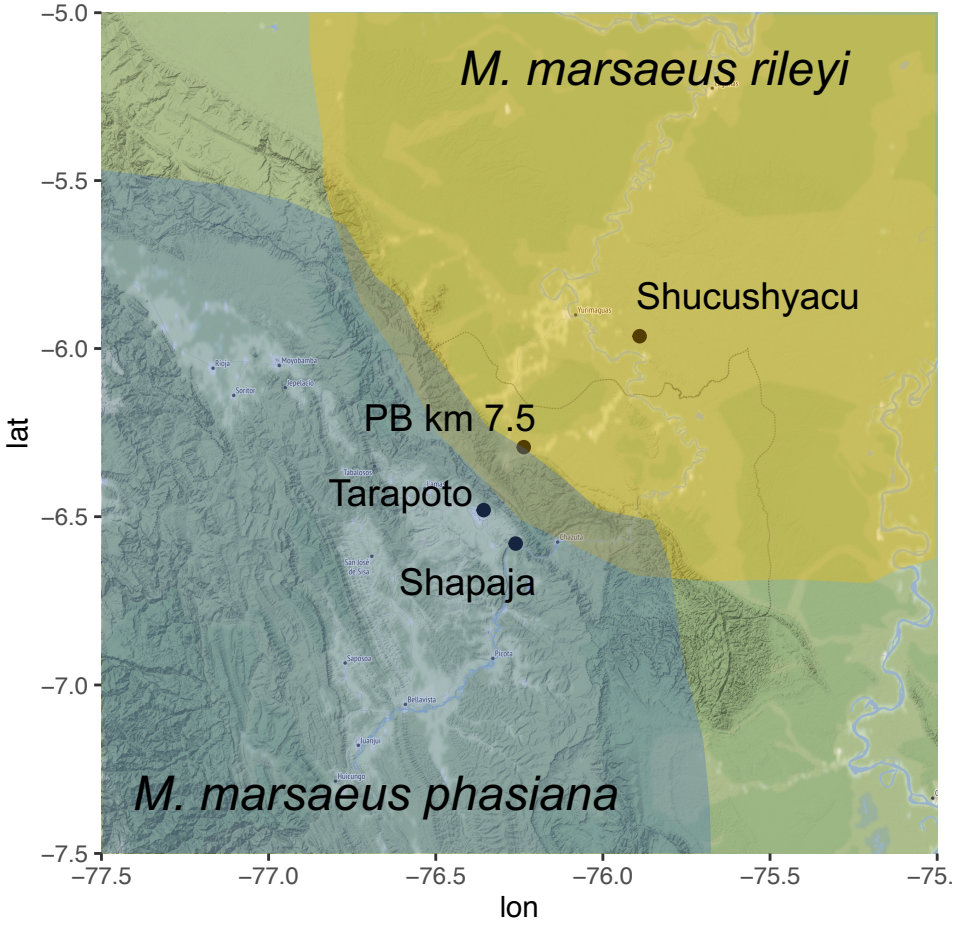
Zhang L et al. 2017. Genetic Basis of Melanin Pigmentation in Butterfly Wings. Genetics. 205:1537–1550.

Zhang L, Reed RD. 2016. Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns. Nat. Commun. 7:11769.

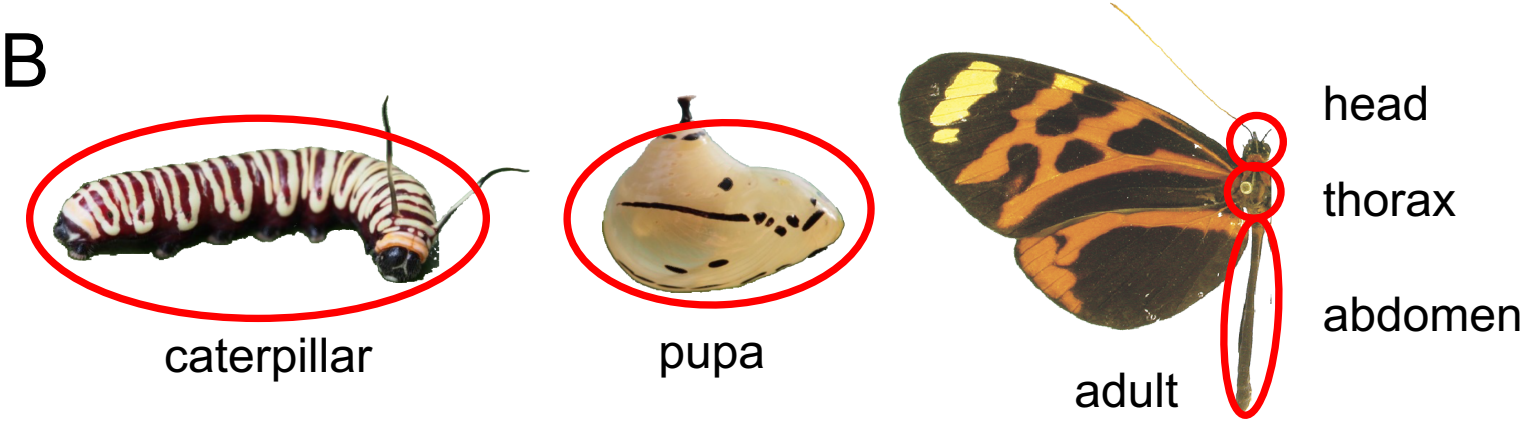
Zhang W et al. 2019. Comparative Transcriptomics Provides Insights into Reticulate and Adaptive Evolution of a Butterfly Radiation. Genome Biol. Evol. 11:2963–2975.

Zhan S, Merlin C, Boore JL, Reppert SM. 2011. The monarch butterfly genome yields insights into long-distance migration. Cell. 147:1171–1185.

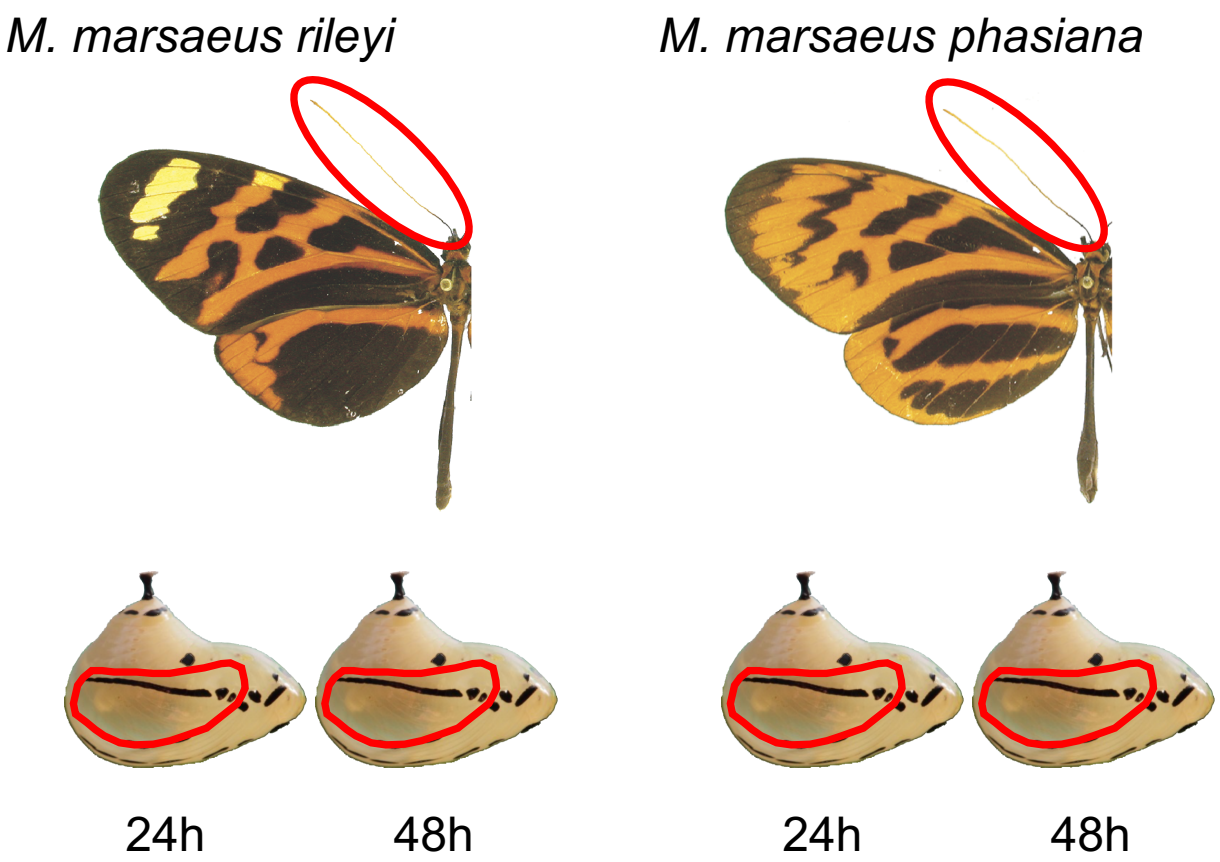
A


M. marsaeus rileyi
Shucushyacu
PB km 7.5
Tarapoto
Shapaja
M. marsaeus phasiana
−5.0
−5.5
−6.0
−6.5
−7.0
−7.5
lat
−77.5
−77.0
−76.5
−76.0
−75.5
−75.
lon

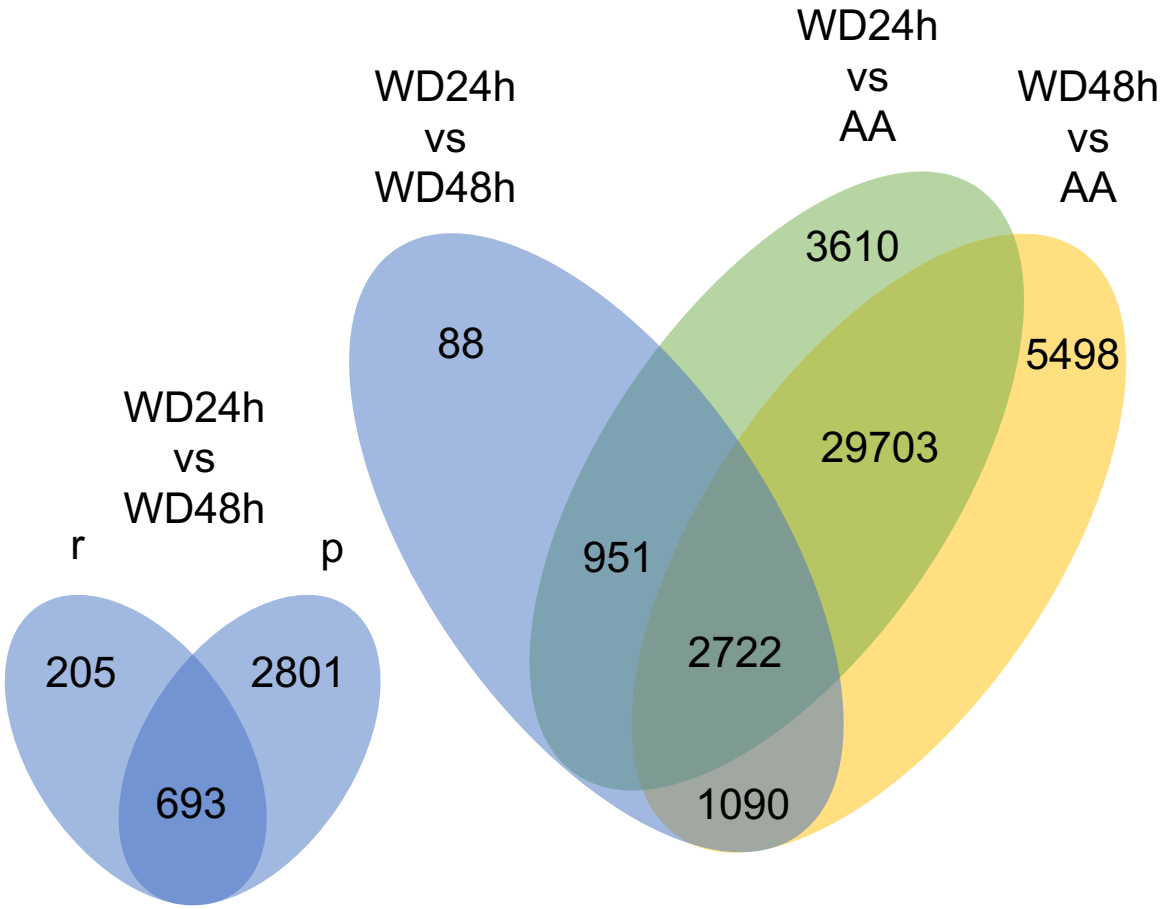
B


caterpillar
pupa
adult
head
thorax
abdomen

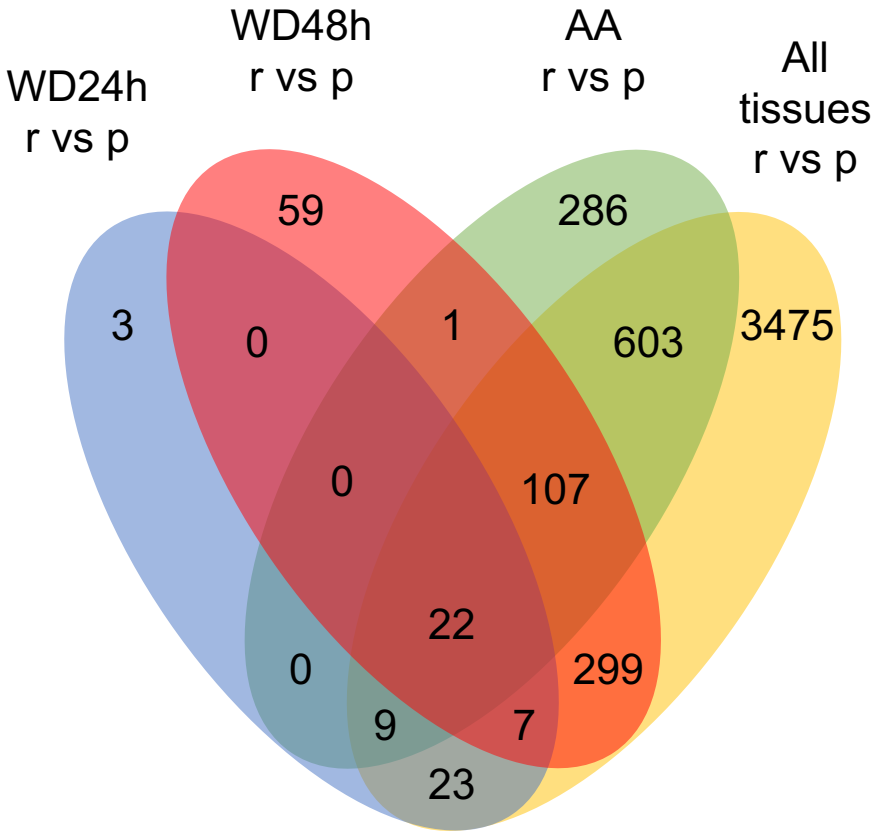
C

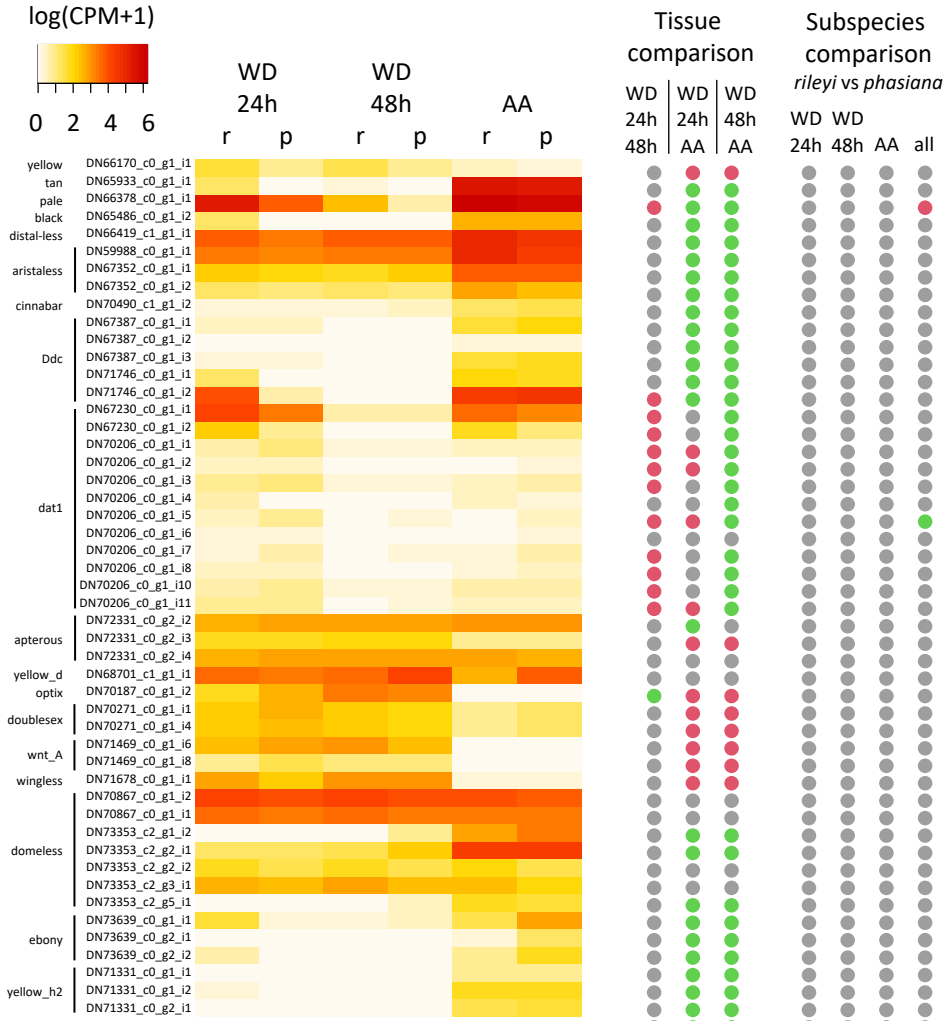

M. marsaeus rileyi
M. marsaeus phasiana
24h
48h
24h
48h

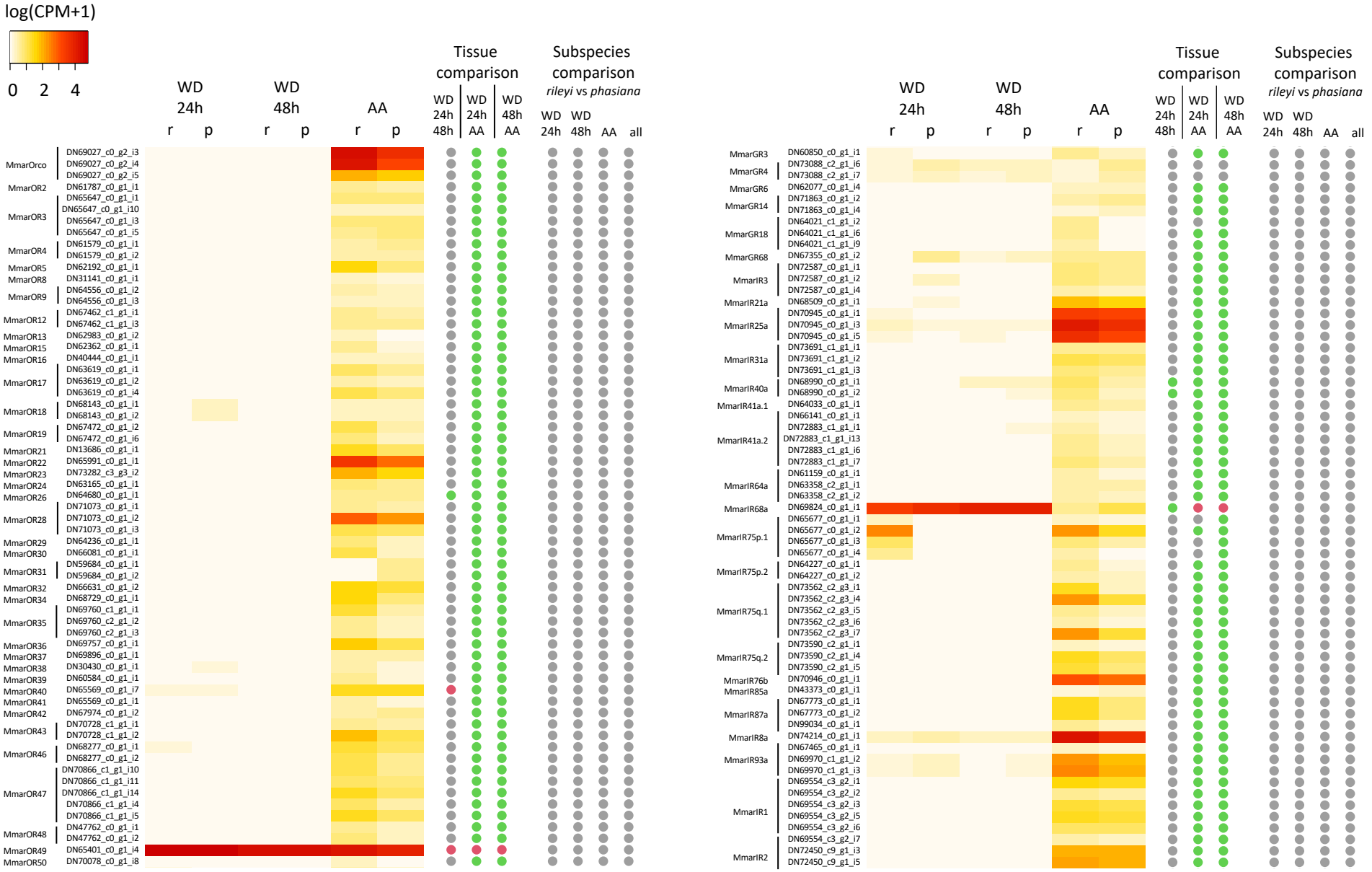
Tissue comparison

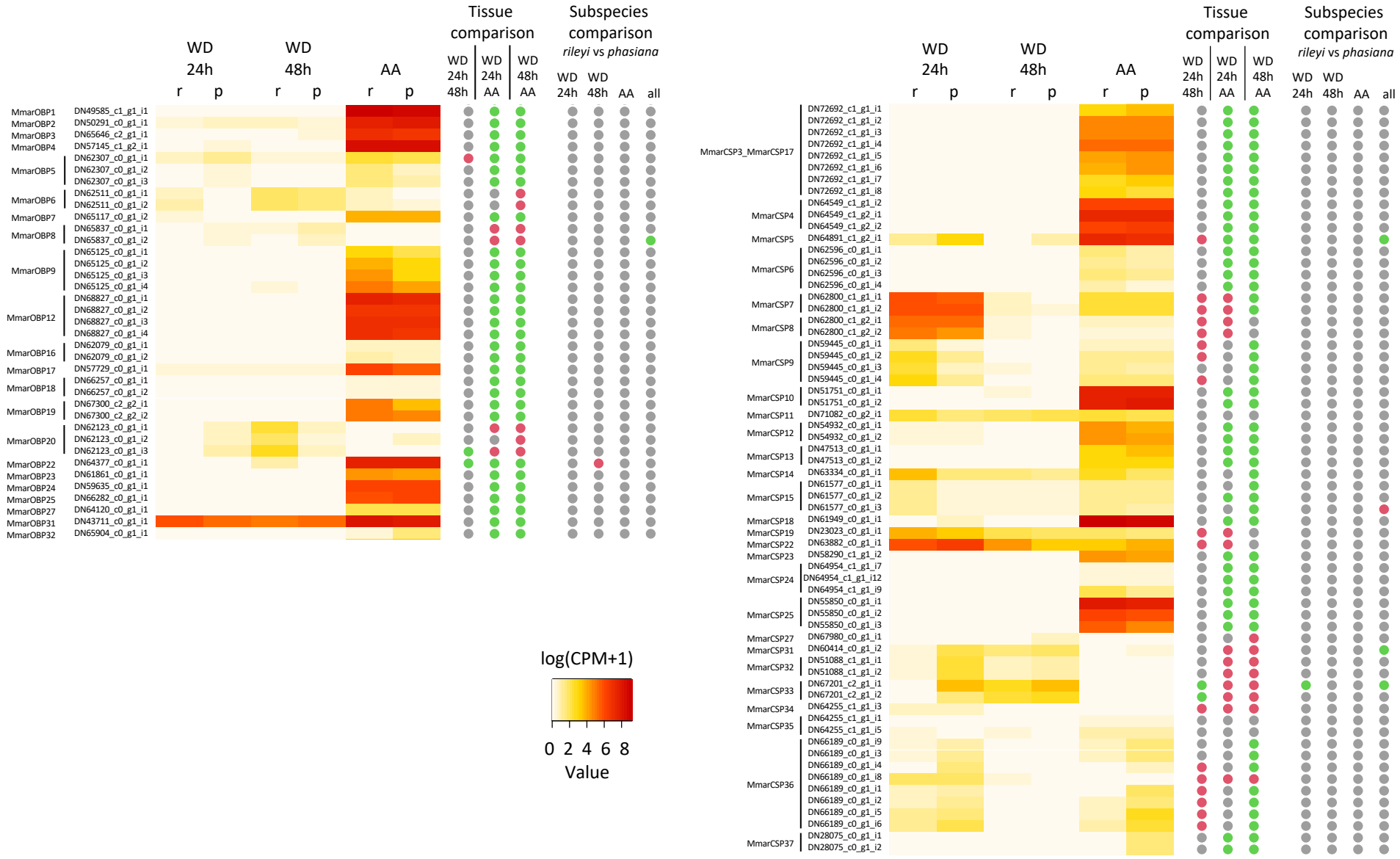


Subspecies comparison




log(CPM+1)
0 2 4 6
WD 24h
WD 48h
AA
r p r p r p
yellow DN66170_c0_g1_i1
tan DN65933_c0_g1_i1
pale DN66378_c0_g1_i1
black DN65486_c0_g1_i2
distal-less DN66419_c1_g1_i1
aristaless DN59988_c0_g1_i1
DN67352_c0_g1_i1
DN67352_c0_g1_i2
cinnabar DN70490_c1_g1_i2
Ddc DN67387_c0_g1_i1
DN67387_c0_g1_i2
DN67387_c0_g1_i3
DN71746_c0_g1_i1
DN71746_c0_g1_i2
dat1 DN67230_c0_g1_i1
DN67230_c0_g1_i2
DN70206_c0_g1_i1
DN70206_c0_g1_i2
DN70206_c0_g1_i3
DN70206_c0_g1_i4
DN70206_c0_g1_i5
DN70206_c0_g1_i6
DN70206_c0_g1_i7
DN70206_c0_g1_i8
DN70206_c0_g1_i10
DN70206_c0_g1_i11
apterous DN72331_c0_g2_i2
DN72331_c0_g2_i3
DN72331_c0_g2_i4
yellow_d DN68701_c1_g1_i1
optix DN70187_c0_g1_i2
doublesex DN70271_c0_g1_i1
DN70271_c0_g1_i4
wnt_A DN71469_c0_g1_i6
DN71469_c0_g1_i8
wingless DN71678_c0_g1_i1
domeless DN70867_c0_g1_i2
DN70867_c0_g1_i1
DN73353_c2_g1_i2
DN73353_c2_g2_i1
DN73353_c2_g2_i2
DN73353_c2_g3_i1
DN73353_c2_g5_i1
ebony DN73639_c0_g1_i1
DN73639_c0_g2_i1
DN73639_c0_g2_i2
yellow_h2 DN71331_c0_g1_i1
DN71331_c0_g1_i2
DN71331_c0_g2_i1
Tissue comparison
WD 24h 48h
WD 24h AA
WD 48h AA
Subspecies comparison
rileyi vs phasiana
WD 24h
WD 48h
AA
all


log(CPM+1)
0 2 4
WD 24h
WD 48h
AA
r p
Tissue comparison
WD 24h 48h
WD 24h AA
WD 48h AA
Subspecies comparison
rileyi vs phasiana
WD 24h
WD 48h
AA
all
MmarOrco
DN69027_c0_g2_i3
DN69027_c0_g2_i4
DN69027_c0_g2_i5
MmarOR2
DN61787_c0_g1_i1
MmarOR3
DN65647_c0_g1_i1
DN65647_c0_g1_i10
DN65647_c0_g1_i3
DN65647_c0_g1_i5
MmarOR4
DN61579_c0_g1_i1
DN61579_c0_g1_i2
MmarOR5
DN62192_c0_g1_i1
MmarOR8
DN31141_c0_g1_i1
MmarOR9
DN64556_c0_g1_i2
DN64556_c0_g1_i3
MmarOR12
DN67462_c1_g1_i1
DN67462_c1_g1_i3
MmarOR13
DN62983_c0_g1_i2
MmarOR15
DN62362_c0_g1_i1
MmarOR16
DN40444_c0_g1_i1
MmarOR17
DN63619_c0_g1_i1
DN63619_c0_g1_i2
DN63619_c0_g1_i4
MmarOR18
DN68143_c0_g1_i1
DN68143_c0_g1_i2
MmarOR19
DN67472_c0_g1_i2
DN67472_c0_g1_i6
MmarOR21
DN13686_c0_g1_i1
MmarOR22
DN65991_c0_g1_i1
MmarOR23
DN73282_c3_g3_i2
MmarOR24
DN63165_c0_g1_i1
MmarOR26
DN64680_c0_g1_i1
MmarOR28
DN71073_c0_g1_i1
DN71073_c0_g1_i2
DN71073_c0_g1_i3
MmarOR29
DN64236_c0_g1_i1
MmarOR30
DN66081_c0_g1_i1
MmarOR31
DN59684_c0_g1_i1
DN59684_c0_g1_i2
MmarOR32
DN66631_c0_g1_i2
MmarOR34
DN68729_c0_g1_i1
MmarOR35
DN69760_c1_g1_i1
DN69760_c2_g1_i2
DN69760_c2_g1_i3
MmarOR36
DN69757_c0_g1_i1
MmarOR37
DN69896_c0_g1_i1
MmarOR38
DN30430_c0_g1_i1
MmarOR39
DN60584_c0_g1_i1
MmarOR40
DN65569_c0_g1_i7
MmarOR41
DN65569_c0_g1_i1
MmarOR42
DN67974_c0_g1_i2
MmarOR43
DN70728_c1_g1_i1
DN70728_c1_g1_i2
MmarOR46
DN68277_c0_g1_i1
DN68277_c0_g1_i2
MmarOR47
DN70866_c1_g1_i10
DN70866_c1_g1_i11
DN70866_c1_g1_i14
DN70866_c1_g1_i4
DN70866_c1_g1_i5
MmarOR48
DN47762_c0_g1_i1
DN47762_c0_g1_i2
MmarOR49
DN65401_c0_g1_i4
MmarOR50
DN70078_c0_g1_i8
MmarGR3
DN60850_c0_g1_i1
MmarGR4
DN73088_c2_g1_i6
DN73088_c2_g1_i7
MmarGR6
DN62077_c0_g1_i4
MmarGR14
DN71863_c0_g1_i2
DN71863_c0_g1_i4
MmarGR18
DN64021_c1_g1_i2
DN64021_c1_g1_i6
DN64021_c1_g1_i9
MmarGR68
DN67355_c0_g1_i2
MmarIR3
DN72587_c0_g1_i1
DN72587_c0_g1_i2
DN72587_c0_g1_i4
MmarIR21a
DN68509_c0_g1_i1
MmarIR25a
DN70945_c0_g1_i1
DN70945_c0_g1_i3
DN70945_c0_g1_i5
MmarIR31a
DN73691_c1_g1_i1
DN73691_c1_g1_i2
DN73691_c1_g1_i3
MmarIR40a
DN68990_c0_g1_i1
DN68990_c0_g1_i2
MmarIR41a.1
DN64033_c0_g1_i1
MmarIR41a.2
DN66141_c0_g1_i1
DN72883_c1_g1_i1
DN72883_c1_g1_i13
DN72883_c1_g1_i6
DN72883_c1_g1_i7
MmarIR64a
DN61159_c0_g1_i1
DN63358_c2_g1_i1
DN63358_c2_g1_i2
MmarIR68a
DN69824_c0_g1_i1
MmarIR75p.1
DN65677_c0_g1_i1
DN65677_c0_g1_i2
DN65677_c0_g1_i3
DN65677_c0_g1_i4
MmarIR75p.2
DN64227_c0_g1_i1
DN64227_c0_g1_i2
MmarIR75q.1
DN73562_c2_g3_i1
DN73562_c2_g3_i4
DN73562_c2_g3_i5
DN73562_c2_g3_i6
DN73562_c2_g3_i7
MmarIR75q.2
DN73590_c2_g1_i1
DN73590_c2_g1_i4
DN73590_c2_g1_i5
MmarIR76b
DN70946_c0_g1_i1
MmarIR85a
DN43373_c0_g1_i1
MmarIR87a
DN67773_c0_g1_i1
DN67773_c0_g1_i2
DN99034_c0_g1_i1
MmarIR8a
DN74214_c0_g1_i1
MmarIR93a
DN67465_c0_g1_i1
DN69970_c1_g1_i2
DN69970_c1_g1_i3
MmarIR1
DN69554_c3_g2_i1
DN69554_c3_g2_i2
DN69554_c3_g2_i3
DN69554_c3_g2_i5
DN69554_c3_g2_i6
DN69554_c3_g2_i7
MmarIR2
DN72450_c9_g1_i3
DN72450_c9_g1_i5


WD 24h
WD 48h
AA
r
p
Tissue comparison
WD 24h 48h
WD 24h AA
WD 48h AA
Subspecies comparison
rileyi vs phasiana
WD 24h
WD 48h
AA
all
MmarOBP1
MmarOBP2
MmarOBP3
MmarOBP4
MmarOBP5
MmarOBP6
MmarOBP7
MmarOBP8
MmarOBP9
MmarOBP12
MmarOBP16
MmarOBP17
MmarOBP18
MmarOBP19
MmarOBP20
MmarOBP22
MmarOBP23
MmarOBP24
MmarOBP25
MmarOBP27
MmarOBP31
MmarOBP32
DN49585_c1_g1_i1
DN50291_c0_g1_i1
DN65646_c2_g1_i1
DN57145_c1_g2_i1
DN62307_c0_g1_i1
DN62307_c0_g1_i2
DN62307_c0_g1_i3
DN62511_c0_g1_i1
DN62511_c0_g1_i2
DN65117_c0_g1_i2
DN65837_c0_g1_i1
DN65837_c0_g1_i2
DN65125_c0_g1_i1
DN65125_c0_g1_i2
DN65125_c0_g1_i3
DN65125_c0_g1_i4
DN68827_c0_g1_i1
DN68827_c0_g1_i2
DN68827_c0_g1_i3
DN68827_c0_g1_i4
DN62079_c0_g1_i1
DN62079_c0_g1_i2
DN57729_c0_g1_i1
DN66257_c0_g1_i1
DN66257_c0_g1_i2
DN67300_c2_g2_i1
DN67300_c2_g2_i2
DN62123_c0_g1_i1
DN62123_c0_g1_i2
DN62123_c0_g1_i3
DN64377_c0_g1_i1
DN61861_c0_g1_i1
DN59635_c0_g1_i1
DN66282_c0_g1_i1
DN64120_c0_g1_i1
DN43711_c0_g1_i1
DN65904_c0_g1_i1
log(CPM+1)
0 2 4 6 8
Value
MmarCSP3_MmarCSP17
MmarCSP4
MmarCSP5
MmarCSP6
MmarCSP7
MmarCSP8
MmarCSP9
MmarCSP10
MmarCSP11
MmarCSP12
MmarCSP13
MmarCSP14
MmarCSP15
MmarCSP18
MmarCSP19
MmarCSP22
MmarCSP23
MmarCSP24
MmarCSP25
MmarCSP27
MmarCSP31
MmarCSP32
MmarCSP33
MmarCSP34
MmarCSP35
MmarCSP36
MmarCSP37
DN72692_c1_g1_i1
DN72692_c1_g1_i2
DN72692_c1_g1_i3
DN72692_c1_g1_i4
DN72692_c1_g1_i5
DN72692_c1_g1_i6
DN72692_c1_g1_i7
DN72692_c1_g1_i8
DN64549_c1_g1_i2
DN64549_c1_g2_i1
DN64549_c1_g2_i2
DN64891_c1_g2_i1
DN62596_c0_g1_i1
DN62596_c0_g1_i2
DN62596_c0_g1_i3
DN62596_c0_g1_i4
DN62800_c1_g1_i1
DN62800_c1_g2_i2
DN62800_c1_g2_i1
DN62800_c1_g2_i2
DN59445_c0_g1_i1
DN59445_c0_g1_i2
DN59445_c0_g1_i3
DN59445_c0_g1_i4
DN51751_c0_g1_i1
DN51751_c0_g1_i2
DN71082_c0_g2_i1
DN54932_c0_g1_i1
DN54932_c0_g1_i2
DN47513_c0_g1_i1
DN47513_c0_g1_i2
DN63334_c0_g1_i1
DN61577_c0_g1_i1
DN61577_c0_g1_i2
DN61577_c0_g1_i3
DN61949_c0_g1_i1
DN23023_c0_g1_i1
DN63882_c0_g1_i1
DN58290_c1_g1_i2
DN64954_c1_g1_i7
DN64954_c1_g1_i12
DN64954_c1_g1_i9
DN55850_c0_g1_i1
DN55850_c0_g1_i2
DN55850_c0_g1_i3
DN67980_c0_g1_i1
DN60414_c0_g1_i2
DN51088_c1_g1_i1
DN51088_c1_g1_i2
DN67201_c2_g1_i1
DN67201_c2_g1_i2
DN64255_c1_g1_i3
DN64255_c1_g1_i1
DN64255_c1_g1_i5
DN66189_c0_g1_i9
DN66189_c0_g1_i3
DN66189_c0_g1_i4
DN66189_c0_g1_i8
DN66189_c0_g1_i1
DN66189_c0_g1_i2
DN66189_c0_g1_i5
DN66189_c0_g1_i6
DN28075_c0_g1_i1
DN28075_c0_g1_i2

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 lepbase	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