

PNAS

www.pnas.org

Supplementary Information for

Divergence of chemosensing during the early stages of speciation

Bas van Schooten^{1, 2, †, *}, Jesyka Meléndez-Rosa^{1, †, *}, Steven M. Van Belleghem¹, Chris Jiggins³, John Tan⁴, W. Owen McMillan² and Riccardo Papa^{1, 2, 5, *}

¹Department of Biology, University of Puerto Rico, Rio Piedras, San Juan, Puerto Rico, United States of America, 00925

²Smithsonian Tropical Research Institution, Balboa Ancón, Panama, Republic of Panama

³Department of Zoology, University of Cambridge, Cambridge, United Kingdom, CB2 8PQ

⁴Roche NimbleGen Inc., Madison, Wisconsin, United States of America, 53719

⁵Molecular Sciences and Research Center, University of Puerto Rico, San Juan 00907, Puerto Rico

* Corresponding author

† These authors contributed equally to this work

Bas van Schooten – Julio García Díaz (JGD) Building, Room 119, University of Puerto Rico, San Juan, PR 00931; 626-421-5648; basvanschooten@gmail.com

Jesyka Meléndez-Rosa – Julio García Díaz (JGD) Building, Room 119, University of Puerto Rico, San Juan, PR 00931; 787-557-3100; jesykamelendez@gmail.com

Riccardo Papa – Julio García Díaz (JGD) Building, Room 119, University of Puerto Rico, San Juan, PR 00931; 626-421-5648; rpapa.lab@gmail.com

This PDF file includes:

Supplementary text

Figures S1 to S7

Tables S1 to S5

Supplementary Information Text

Detailed Materials and Methods

Sampling, RNA extraction, and sequencing

We bred 15 *Heliconius melpomene* and 15 *H. cydno* butterflies in semi-natural conditions from November 2013 to March 2014 in Gamboa, Panama (9°7.4' N, 79°42.2' W, elevation 60 m). Unmated females were two days old at sampling and both males and mated females were five days old; females were mated within the first two days after eclosing. In order to minimize time of day effects in gene expression, all individuals were sampled during the late morning hours (09:00-12:00 hrs) - the period during which butterflies are the most active in the wild. For each individual we collected tissues of three sensory tissue types: antenna, legs (all six legs), and mouthparts (including the proboscis and labial palps). These were submerged in RNAlater and stored for two weeks in a -20 °C freezer at the Smithsonian Tropical Research Institute before being moved to permanent storage at a -80 °C freezer at the University of Puerto Rico, Rio Piedras.

For each individual butterfly we extracted RNA separately from each of the three sensory tissue types (see above) for a total of 90 individual RNA extractions representing three biological groups (5 males, 5 unmated females, and 5 mated females) of two closely related species (15 *H. melpomene* and 15 *H. cydno*) (S1 Figure in SI Appendix). RNA was extracted using a TRIzol™ RNA isolation protocol followed by additional purification using a RNeasy minikit (Qiagen). RNA quality was assessed using RNA 6000 Pico BioAnalyzer chips (Agilent) (1) and concentrations were determined using a NanoDrop spectrophotometer (ThermoFisher) (2). Two Illumina libraries, each representing 45 individually barcoded RNA extractions for either *H. melpomene* or *H. cydno*, were generated using the Illumina TruSeq™ RNA sample preparation kit according to the manufacturer's low-throughput protocol. Each library was sequenced three times on a HiSeq2500 Illumina system for a total of six Illumina sequencing lanes (50bp SR). Sequencing was performed at the Knapp Center for Biomedical Discoveries at the University of Chicago.

Targeted resequencing and chemosensory gene annotation improvement

We first used CrossMap (3) and manual verification to update the annotation of chemosensory gene models on the *Heliconius melpomene* genome v2 Hmel2 genome version (4, 5). For this, we used previously published gene models for 252 chemosensory genes: a) 33 chemosensory proteins (6), b) 43 olfactory binding proteins (6), c) 70 olfactory receptors (6), d) 73 gustatory receptors (7), and e) 31 ionotropic receptors (8).

We further improved the chemosensory gene models by producing a target RNA resequencing data set for the genes of interest (S1 Figure in SI Appendix). For a subset of the sampled *Heliconius melpomene* butterflies (two males, two unmated females, and two mated females) we pooled equal amounts of RNA extracted from the antenna, legs, and mouthparts for each individual using a NanoDrop spectrophotometer (ThermoFisher). The pooled samples (N = 6) were sent to Roche for targeted enrichment of sequencing libraries using the NimbleGen SeqCap protocol (9) followed by sequencing on the Illumina HiSeq2500 platform (100bp PE), generating a total of 168,344,748 (50 bp) reads. We then used the targeted capture data to improve the annotation of the chemosensory genes, in particular, the beginnings and ends of genes, and to find previously unannotated genes in the Hmel2 genome. We used 'Bowtie' (10) to create a BAM file and verified the alignment visually using the Integrative Genomics Viewer (11) for accuracy.

Using a genome guided assembly approach, we aligned our reads to 252 chemosensory genes with a high coverage (237 X) that allowed for a substantial improvement of the chemosensory gene models (S1 Table in SI Appendix). The combined dataset (NimbleGen SeqCap and RNA-seq data) had more power to identify start and stop codons, missing exons, and exon/intron boundaries. Overall, we improved the gene models for 115 (45 ORs, 41 GRs, 22 OBPs, and 7 CSPs) chemosensory genes (S1 Table in SI Appendix).

The major change that emerged from our analysis was related to the number of genes characterizing specific chemosensory gene families. Compared to previous studies (4, 6–8), we reduced the number of Olfactory Receptors (ORs), Gustatory Receptors (GRs) and Chemosensory Proteins (CSPs), while increasing the gene models for olfactory binding proteins (OBPs) (S1 Table in SI Appendix). We detected the erroneous annotation of 6 chemosensory genes in the earlier *Heliconius* genome version (Hmel1): *HmCSP33*, *HmGR37*, *HmGR41*, *HmOR73*, *HmOR48* and *HmOR47*. These 6 genes were identical to *HmCSP34*, *HmGR38*, *HmGR40*, *HmOR43*, *HmOR24*, and *HmOR50*. We found no evidence to indicate that these genes represent real duplications, rather, they appear to be the product of erroneous annotation caused by heterozygosity and were discarded from our gene models and

analyses (4, 6). We confirmed, as suggested by Vogt *et al.* (2015), the existence of 8 additional OBPs. Vogt *et al.* (2015) also suggested the existence of a 9th OBP named *HmelOBP41*. Our analyses, however, could not recover a full gene model for *OBP41*, indicating that it represents a partial copy. To this point, we did not identify - in either our NimbleGen SeqCap or RNA-seq datasets - a single sequence corresponding to the *OBP41* gene. While superficially *OBP41* appears to be a real partial gene duplication, the lack of evidence for the existence of an associated functional gene suggests that it is either the result of errors during genome assembly or a pseudogene. For this reason, we chose to remove *OBP41* from downstream analyses, bringing the total number of OBPs from 51 to 50. Our analyses detected the existence of an additional 29 copies of 23 genes, which we believe likely represent genome assembly errors. These include 9 CSPs, 1 OBP, 4 IRs, and 9 GRs (S1 Table in SI Appendix). Additionally, our data indicate that three previously annotated OR genes (*HmOR56*, *HmOR58*, and *HmOR59*) (6) are actually splice variants of a single larger gene. The three splice variants share four smaller exons totaling 411 bp in length and contain one larger exon, from 790-808 bp, that differs between the three genes. These splice variants were treated as individual genes in downstream analyses.

Raw data processing and differential expression analyses

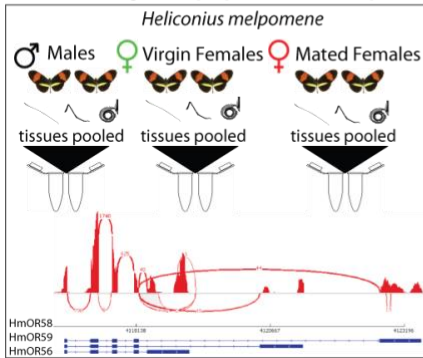
The Illumina RNA-seq data was aligned to the *Hmel2* genome using 'Tophat' (q -value threshold = 0.05; for per-gene q -values see the 'q-values.xlsx' data set) and differential expression analyses were performed with 'Cufflinks' (12). Differential expression analyses were conducted by tissue and biological group as follows. For the tissue-specific expression analyses, we combined data for all 30 butterflies and compared across tissue types. For species-, sex-, and life-stage -specific expression analyses, the data were analyzed for each tissue type independently (antennae, legs, or mouthparts). Species-specific differences were assessed by comparing males, unmated females or mated females of *H. melpomene* to their biological equivalents in *H. cydno*. Here, differences in chemosensory gene expression between closely related species were hypothesized to reflect differences in mate or host-plant choice. Sex-specific differences were assessed by comparing males and unmated females. This comparison was established to identify genes potentially central to both female- and male-choice. Life-stage-specific differences were assessed by comparing unmated and mated females. Here, differences between unmated and mated females were hypothesized to reflect a shift in sensory focus from mate searching (unmated females) to host-plant detection for oviposition (mated female). Expression plots based on FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values were created using the 'heatmap.2' function of the 'gplots' package in R (13).

For *H. melpomene*, we sequenced a total 527,608,789 (50 bp) reads of which 387,543,981 (73.45%) aligned to the *Hmel2* genome (4, 5). For *H. cydno* we sequenced a total of 455,172,761 reads of which 358,702,600 (78.81%) aligned to the *Hmel2* genome. Despite being a different, albeit closely related, species the *H. cydno* genes mapped better to the *Hmel2* genome, alleviating concerns about the use of a *H. melpomene* reference to align and analyze the *H. cydno* data. 7,804,978 (1.05%) of the RNA-seq reads aligned to the 252 chemosensory genes.

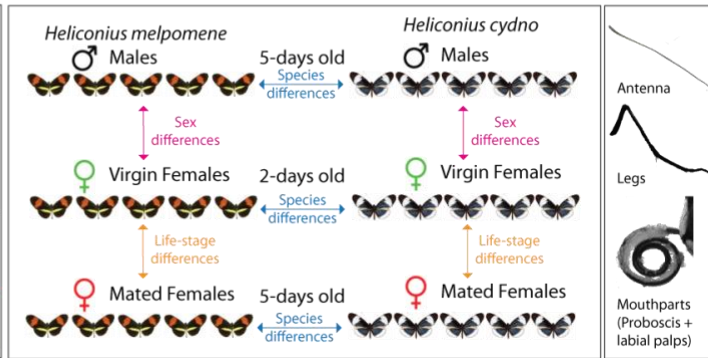
Admixture and Linkage Disequilibrium (LD) analyses

In order to further narrow down genes underlying species-level recognition, we compared results from our species-level differential expression analyses to admixture estimates for the sympatric *H. melpomene* – *H. cydno* pair from Panama (14). The admixture values (f_d) are based on the ABBA-BABA test (15) which measures an excess of derived allele sharing between the sympatric non-sister taxa compared to an allopatric population of *H. melpomene* from French Guiana. We included *H. numata* as an outgroup (Figure 3A). Negative f_d values indicate low admixture or sharing of derived alleles between the sympatric Panama populations of *H. melpomene* and *H. cydno* whereas positive f_d values indicate high admixture. The f_d values were calculated for 100 kb windows with a step size of 20 kb. For chemosensory genes overlapping windows were averaged. With this approach, we were able to identify chemosensory genes that were both significantly differentially expressed (over-expressed) and presented low admixture between the two species. Chemosensory genes at the intercept of these two estimates are likely to be involved in species-specific processes. Finally, because the color pattern locus *optix* has been shown to be associated with mate preference behavior in *H. melpomene* and *H. cydno* (16), and linkage between chemosensory and color pattern genes could facilitate the forming of smell-based prezygotic barriers (17), we tested for the nonrandom association of chemosensory loci (50 kb around transcription start site) at and around the *optix* gene (1,000 kb region) in 50 kb windows. Pairwise comparisons of SNPs within each 50 kb window were averaged to produce a mean r_2 .

A) Chemosensory gene targeted resequencing
Nimblegen SeqCap-Illumina Hiseq



B) Chemosensory gene expression Illumina Hiseq



S1 Figure: Overview of experimental setups. A) Chemosensory gene targeted resequencing: We sampled a total of 6 *H. melpomene* butterflies representing three biological groups (two unmated females, two mated females, and two males). For each individual, we extracted RNA from the antenna, legs, and mouthparts separately. RNA extractions from the three tissues were pooled into one library for each individual (N = 6 libraries). The libraries were sequenced and the targets were enriched as per the Nimblegen SeqCap-Illumina Hiseq protocol. **B)** Chemosensory gene expression: Each biological group contained 5 replicates for both *H. melpomene* and *H. cydno*. For each individual, we extracted RNA from the antenna, legs, and mouthparts separately. We tested for differential expression of chemosensory genes between species, sexes, and life-stages for each tissue type.

S1 Table: Overview of major annotation improvements. For each chemosensory gene category, from left to right: '# of genes' the number of genes remaining after the initial Hmel2 genome annotation; 'Improved annotation' the number of gene models improved at more than 6 base pairs; 'OBPs in Vogt (2015)' (18) the number of OBP genes first published, which also includes one incorrect annotation (*OBP41*) listed in the parenthesis; 'Redundant in Hmel2' the number of redundant genes (the same gene with a different name) mapped to the same location in the Hmel2 genome; 'Identical copies in Hmel2' the number of genes with partial or complete copies (likely generated *in silico*) in the Hmel2 genome, as well as the copy number for each gene provided in parenthesis; 'Splice variants' the number of splice variants identified and the number of genes in parenthesis.

	# of genes	Improved annotation	OBPs in Vogt (2015)	Redundant in Hmel2	Identical copies in Hmel2 (copy number)	# of splice variants (# of genes)
CSP	33	7	-	1	9(13 copies)	0
OBP	50	22	8(1)	0	1	0
GR	71	41	-	2	9(11 copies)	0
IR	31	0	-	0	4	0
OR	67	45	-	3	0	3(1 gene)
Total	252	115	9	6	23(29)	3(1 gene)

S2 Table: Tissue specific expression of chemosensory genes. For each chemosensory gene category, we provide the number of genes presenting tissue specific expression for the various sensory tissue types: antennae, mouthparts, and legs.

	CSP	GR	IR	OBP	OR	Total
expressed in all tissues	11	66	23	16	130	147
antenna specific	1	2	5	12	26	46
mouthparts specific	3	2				5
leg specific	1			2		3
mixed expression	17	1	3	20	11	52
total	33	71	31	50	67	252

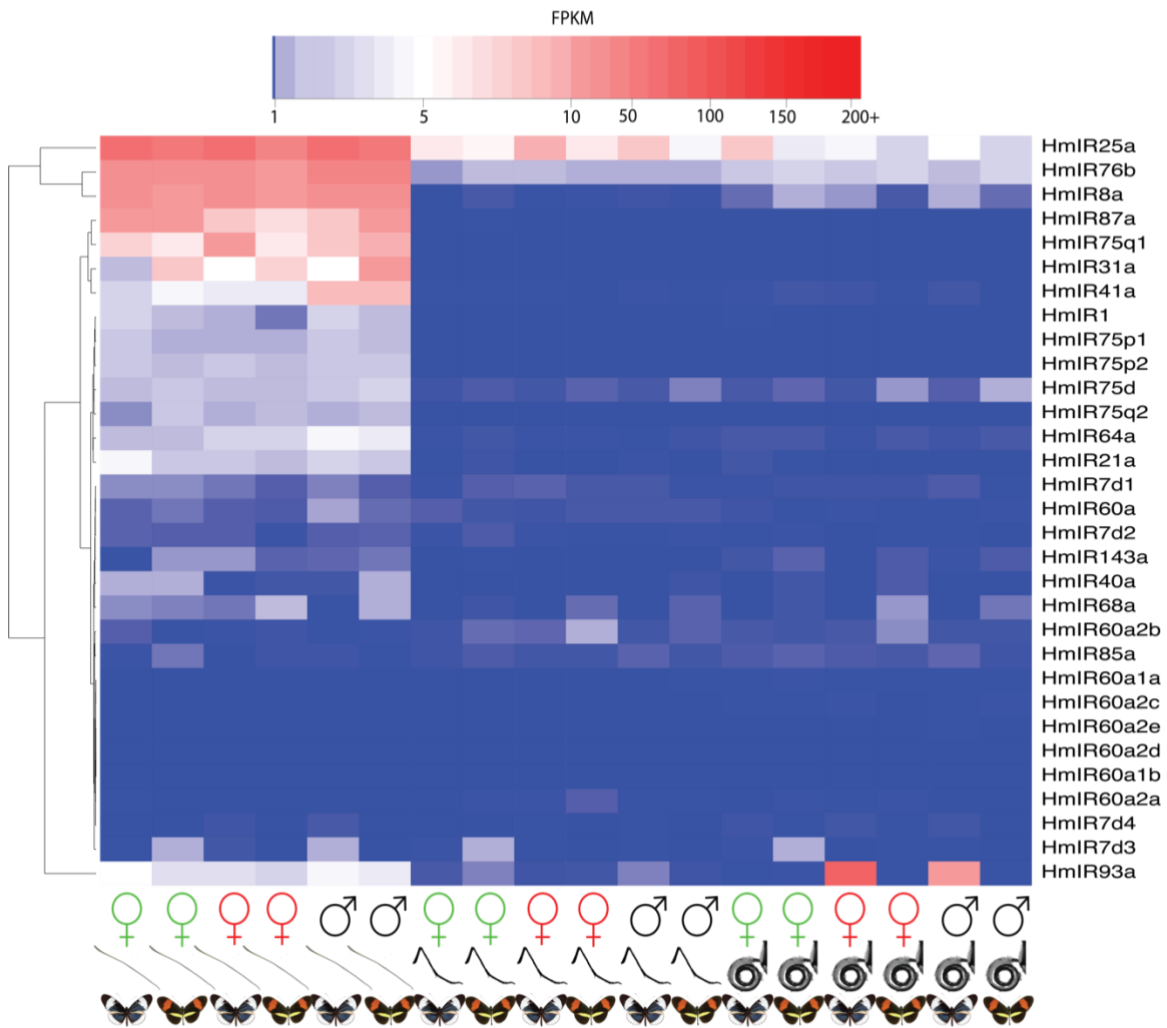
S3 Table: Overview of chemosensory gene (transporters) differential expression patterns. For each tissue type we indicate transporters (OBPs and CSPs) differentially expressed as a function of sex, species, and life-stage. For the species category, we highlight genes (in red and blue) showing consistent differential expression patterns between unmated females, mated females, and males. For the sex and life-stage categories, we highlight genes showing consistent differential expression patterns in both species (*H. cydno* and *H. melpomene*). These highlighted genes correspond with the genes shown in Figure 2 of the manuscript. Bolded and underlined genes are differentially expressed, but don't show consistent expression patterns across sex, life-stage or species. An excel version of this table is available on the Open Science Repository (DOI 10.17605/OSF.IO/2MB38).

ANTENNA									
Sex differences			Species differences				Lifestage differences		
Higher expression	Comparison	Transporter	Higher expression	Comparison	Transporter	Higher expression	Comparison	Transporter	
FEMALE	cydno	<i>HmOBP16, HmOBP17, HmOBP19, HmOBP26, HmOBP31, HmOBP42, HmOBP52, <u>HmOBP9</u>, HmCSP19, HmCSP22, <u>HmCSP29</u></i>	cydno	MALE	<i>HmCSP15, HmCSP16, HmCSP2, HmCSP29, HmCSP5, HmCSP6, HmCSP8, HmCSP9, HmOBP19, HmOBP29, HmOBP39, HmOBP40, HmOBP47</i>	VIRGIN FEMALE	cydno	<i>HmOBP19, HmOBP26, HmOBP31, HmOBP42, HmOBP5, HmOBP9, HmCSP25, <u>HmCSP29</u>, HmCSP6, HmCSP7</i>	
	melpomene	<i>HmOBP31, HmOBP39, HmOBP42, HmOBP5, HmCSP15, HmCSP22, HmCSP28, HmCSP30, HmCSP6, HmCSP7</i>	melpomene		<i>HmCSP12, HmOBP15, HmOBP20, HmOBP38, HmOBP9</i>		melpomene	<i>HmOBP31, HmOBP42, HmOBP5, <u>HmCSP13</u>, HmCSP2, HmCSP28, HmCSP30, HmCSP7</i>	
MALE	cydno	<i>HmOBP10, HmOBP15, HmOBP38, HmOBP40, HmOBP47, HmOBP50, HmOBP51, HmCSP13, HmCSP16, HmCSP3, HmCSP5, HmCSP8</i>	cydno	MATED FEMALE	<i>HmCSP10, HmCSP13, HmCSP2, HmCSP5, HmCSP8, HmCSP9, HmOBP16, HmOBP19, HmOBP29, HmOBP39, HmOBP40, HmOBP42, HmOBP47</i>	MATED FEMALE	cydno	<i>HmOBP38, HmOBP40, HmOBP47, HmOBP50, <u>HmCSP13</u>, HmCSP16, HmCSP8</i>	
	melpomene	<i>HmOBP47, HmOBP9, HmCSP29</i>	melpomene		<i>HmCSP1, HmCSP12, HmCSP7, HmOBP1, HmOBP14, HmOBP15, HmOBP2, HmOBP20, HmOBP38, HmOBP45, HmOBP51, HmOBP9</i>		melpomene	<i>HmOBP47, HmCSP29</i>	
			cydno	VIRGIN FEMALE	<i>HmCSP19, HmCSP29, HmCSP31, HmCSP9, HmOBP16, HmOBP19, HmOBP22, HmOBP26, HmOBP29, HmOBP42</i>		melpomene	<i>HmCSP13, HmCSP28, HmCSP30, HmCSP7, HmOBP15, HmOBP18, HmOBP20, HmOBP38, HmOBP5, HmOBP51</i>	
			melpomene						
LEGS									
Sex differences			Species differences				Lifestage differences		
Higher expression	Comparison	Transporter	Higher expression	Comparison	Transporter	Higher expression	Comparison	Transporter	
FEMALE	cydno	<i>HmOBP25, HmOBP31, HmOBP5, HmOBP52, HmCSP2, HmCSP25, HmCSP9</i>	cydno	Male	<i>HmOBP31, HmOBP39, HmOBP47, HmOB48, HmCSP22, HmCSP29, HmCSP30, HmCSP31, HmCSP6, HmCSP8</i>	VIRGIN FEMALE	cydno	<i>HmOBP25, HmOBP31, HmOBP5, HmCSP2, HmCSP25, HmCSP9</i>	
	melpomene	<i>HmOBP29, HmOBP31, HmOBP39, HmOBP5, HmOBP52, HmCSP10, HmCSP11, HmCSP22, HmCSP30, HmCSP6, HmCSP7, HmCSP8</i>	melpomene	Male	<i>HmCSP25, HmCSP3, HmCSP7</i>		melpomene	<i>HmOBP29, HmOBP31, HmOBP5, HmOBP52, HmCSP10, HmCSP11, HmCSP13, HmCSP2, HmCSP22, HmCSP7</i>	
MALE	cydno	<i>HmOBP45, HmOBP47, HmOBP50, HmCSP1, HmCSP14</i>	cydno	Mated female	<i>HmOBP27, HmOBP29, HmOBP47, HmOB48, HmCSP14, HmCSP15, HmCSP22, HmCSP31, HmCSP6</i>	MATED FEMALE	cydno	<i>HmOBP11, HmOBP12, HmOBP41, HmOBP45, HmOBP47, HmOBP50, HmOBP8, HmOBP8, HmOBP9, HmCSP1, HmCSP13, HmCSP23, HmCSP25, HmCSP7</i>	
	melpomene	<i>HmOBP50, HmCSP5</i>	melpomene	Mated female	<i>HmOBP11, HmOBP14, HmOBP16, HmOBP18, HmOBP21, HmOBP30, HmOBP35, HmOBP8, HmOBP9, HmCSP1, HmCSP13, HmCSP23, HmCSP25, HmCSP7</i>		melpomene	<i>HmOBP11, HmOBP15, HmOBP16, HmOBP50, HmOBP8, HmOBP9, HmCSP29</i>	
			cydno	virgin female	<i>HmOBP31, HmOBP48, HmCSP2, HmCSP29, HmCSP9</i>		melpomene	<i>HmOBP14, HmOBP30, HmOBP5, HmOBP8, HmCSP13, HmCSP14, HmCSP30, HmCSP7</i>	
			melpomene	virgin female	<i>HmOBP14, HmOBP30, HmOBP5, HmOBP8, HmCSP13, HmCSP14, HmCSP30, HmCSP7</i>				
MOUTHPARTS									
Sex differences			Species differences				Lifestage differences		
Higher expression	Comparison	Transporter	Higher expression	Comparison	Transporter	Higher expression	Comparison	Transporter	
FEMALE	cydno	<i>HmOBP31, HmOBP39, HmOBP5, HmOBP52, HmOBP8, HmCSP10, HmCSP13, HmCSP15, HmCSP16, HmCSP19, HmCSP25, HmCSP5, HmCSP6, HmCSP8</i>	cydno	Male	<i>HmOBP16, HmOBP17, HmOBP31, HmOBP39, HmOBP40, HmOBP53, HmCSP1, HmCSP28, HmCSP29, HmCSP9</i>	VIRGIN FEMALE	cydno	<i>HmOBP31, HmOBP39, HmOBP43, HmOBP52, HmOBP53, HmCSP1, HmCSP25</i>	
	melpomene	<i>HmOBP31, HmOBP39, HmOBP5, HmOBP52, HmCSP10, HmCSP15, HmCSP16, HmCSP19, HmCSP22, HmCSP28, HmCSP6, HmCSP8</i>	melpomene	Male	<i>HmOBP14, HmOBP43, HmOBP45, HmOBP8, HmCSP14, HmCSP2, HmCSP25, HmCSP7</i>		melpomene	<i>HmOBP31, HmOBP5, HmCSP10, HmCSP16, HmCSP19, HmCSP22, HmCSP25, <u>HmCSP28</u>, HmCSP30</i>	
MALE	cydno	<i>HmOBP17, HmOBP18, HmOBP50, HmCSP23, HmCSP29, HmCSP30, HmCSP7</i>	cydno	Mated female	<i>HmOBP12, HmOBP16, HmOBP17, HmOBP5, HmCSP13, HmCSP19, HmCSP22, HmCSP28, HmCSP30, HmCSP8</i>	MATED FEMALE	cydno	<i>HmOBP12, HmOBP16, HmOBP17, HmOBP18, HmOBP50, HmOBP8, HmCSP23, HmCSP28, HmCSP30, HmCSP7</i>	
	melpomene	<i>HmOBP15, HmOBP18, HmOBP43, HmOBP50, HmCSP13, HmCSP2</i>	melpomene	Mated female	<i>HmOBP19, HmOBP45, HmOBP50, HmOBP52, HmOBP8, HmCSP1, HmCSP14, HmCSP7</i>		melpomene	<i>HmOBP12, HmOBP15, HmOBP18, HmOBP19, HmOBP45, HmOBP50, HmOBP8, HmCSP7</i>	
			cydno	virgin female	<i>HmOBP16, HmOBP22, HmOBP26, HmOBP31, HmOBP39, HmOBP43, HmCSP1, HmCSP13, HmCSP25, HmCSP9</i>		melpomene	<i>HmOBP14, HmOBP18, HmOBP50, HmCSP11, HmCSP14, <u>HmCSP28</u>, HmCSP30, HmCSP7</i>	
			melpomene	virgin female	<i>HmOBP14, HmOBP18, HmOBP50, HmCSP11, HmCSP14, <u>HmCSP28</u>, HmCSP30, HmCSP7</i>				

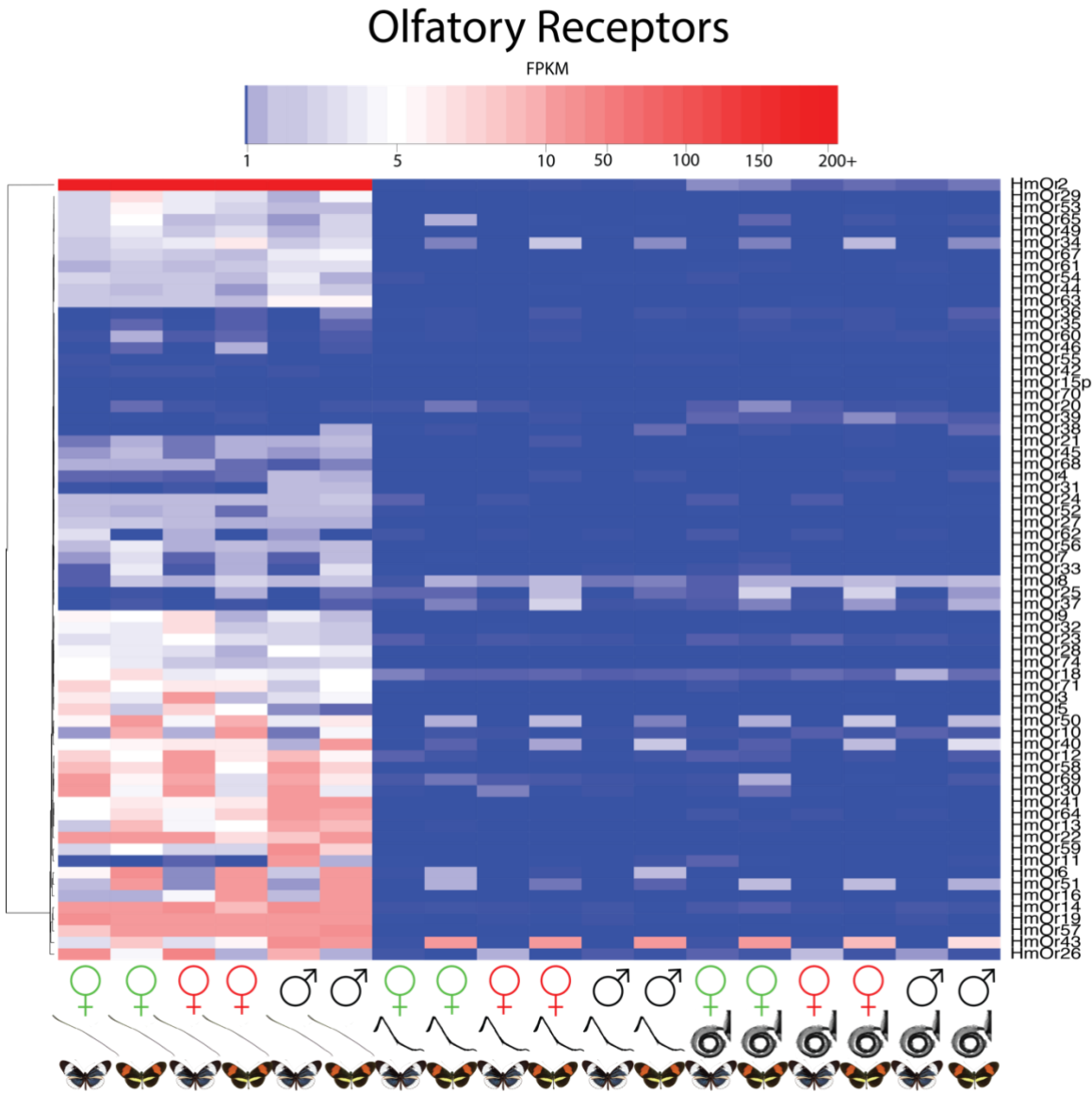
S4 Table: Overview of chemosensory gene (receptors) differential expression patterns. For each tissue type we indicate receptors (OR, GRs, and IRs) differentially expressed as a function of sex, species, and life-stage. For the species category, we highlight genes (in red and blue) showing consistent differential expression patterns between unmated females, mated females, and males. For the sex and life-stage categories, we highlight genes showing consistent differential expression patterns in both species (*H. cydno* and *H. melpomene*). These highlighted genes correspond with the genes shown in Figure 2 of the manuscript. Note the increase in the number of differentially expressed genes in the species comparison (particularly for antennae) relative to the sex or life-stage comparisons. An excel version of this table is available on the Open Science Repository (DOI 10.17605/OSF.IO/2MB38).

ANTENNA								
Sex differences			Species differences			Lifestage differences		
Higher expression	Comparison	Receptor	Higher expression	Comparison	Receptor	Higher expression	Comparison	Receptor
FEMALE	cydno	<i>HmOR19, HmOR26, HmOR5, HmOR71</i>	cydno	MALE	<i>HmOR11, HmOR14, HmOR26, HmOR30, HmOR43, HmOR54, HmOR59</i>	UNMATED FEMALE	cydno	<i>HmGR56, HmOR6</i>
	melpomene	<i>HmOR53</i>	melpomene		<i>HmGR64, HmIR31a, HmOR10, HmOR16, HmOR33, HmOR40, HmOR51, HmOR6, HmOR71</i>		melpomene	
MALE	cydno	<i>HmIR41a, HmOR13, HmOR41, HmOR43, HmOR57</i>	cydno	MATED FEMALE	<i>HmGR22, HmIR25a, HmIR75q1, HmIR76b, HmIR8a, HmOR12, HmOR14, HmOR19, HmOR26, HmOR3, HmOR30, HmOR32, HmOR58, HmOR69, HmOR9</i>	MATED FEMALE	cydno	<i>HmOR12</i>
	melpomene	<i>HmOR16</i>	melpomene		<i>HmGR1, HmIR68a, HmOR16, HmOR51, HmOR6, HmOR26, HmOR62</i>		melpomene	<i>HmOR16</i>
			cydno	UNMATED FEMALE				
			melpomene		<i>HmIR31a, HmOR13, HmOR51, HmOR57, HmOR6</i>			
LEGS								
Sex differences			Species differences			Lifestage differences		
Higher expression	Comparison	Receptor	Higher expression	Comparison	Receptor	Higher expression	Comparison	Receptor
FEMALE	cydno	<i>HmGR22</i>	cydno	Male	<i>HmIR25a</i>	UNMATED FEMALE	cydno	<i>HmGR22</i>
	melpomene		melpomene	Male	<i>HmGR63</i>		melpomene	
MALE	cydno		cydno	Mated female	<i>HmGR56</i>	MATED FEMALE	cydno	
	melpomene		melpomene	Mated female			melpomene	<i>HmOR37</i>
			cydno	Unmated female	<i>HmGR56</i>			
			melpomene	Unmated female				
MOUTHPARTS								
Sex differences			Species differences			Lifestage differences		
Higher expression	Comparison	Receptor	Higher expression	Comparison	Receptor	Higher expression	Comparison	Receptor
FEMALE	cydno	<i>HmIR25a, HmGR22, HmGR3, HmGR5, HmGR56,</i>	cydno	Male	<i>HmIR25a, HmIR93a, HmGR52, HmGR56</i>	UNMATED FEMALE	cydno	<i>HmIR25a, HmGR22, HmGR3, HmGR5</i>
	melpomene		melpomene	Male	<i>HmGR63</i>		melpomene	
MALE	cydno		cydno	Mated female	<i>HmIR93a</i>	MATED FEMALE	cydno	
	melpomene	<i>HmOR40</i>	melpomene	Mated female	<i>HmGR63</i>		melpomene	
			cydno	Unmated female	<i>HmIR25a, HmGR3, HmGR5, HmGR52</i>			
			melpomene	Unmated female				

Ionotropic Receptors

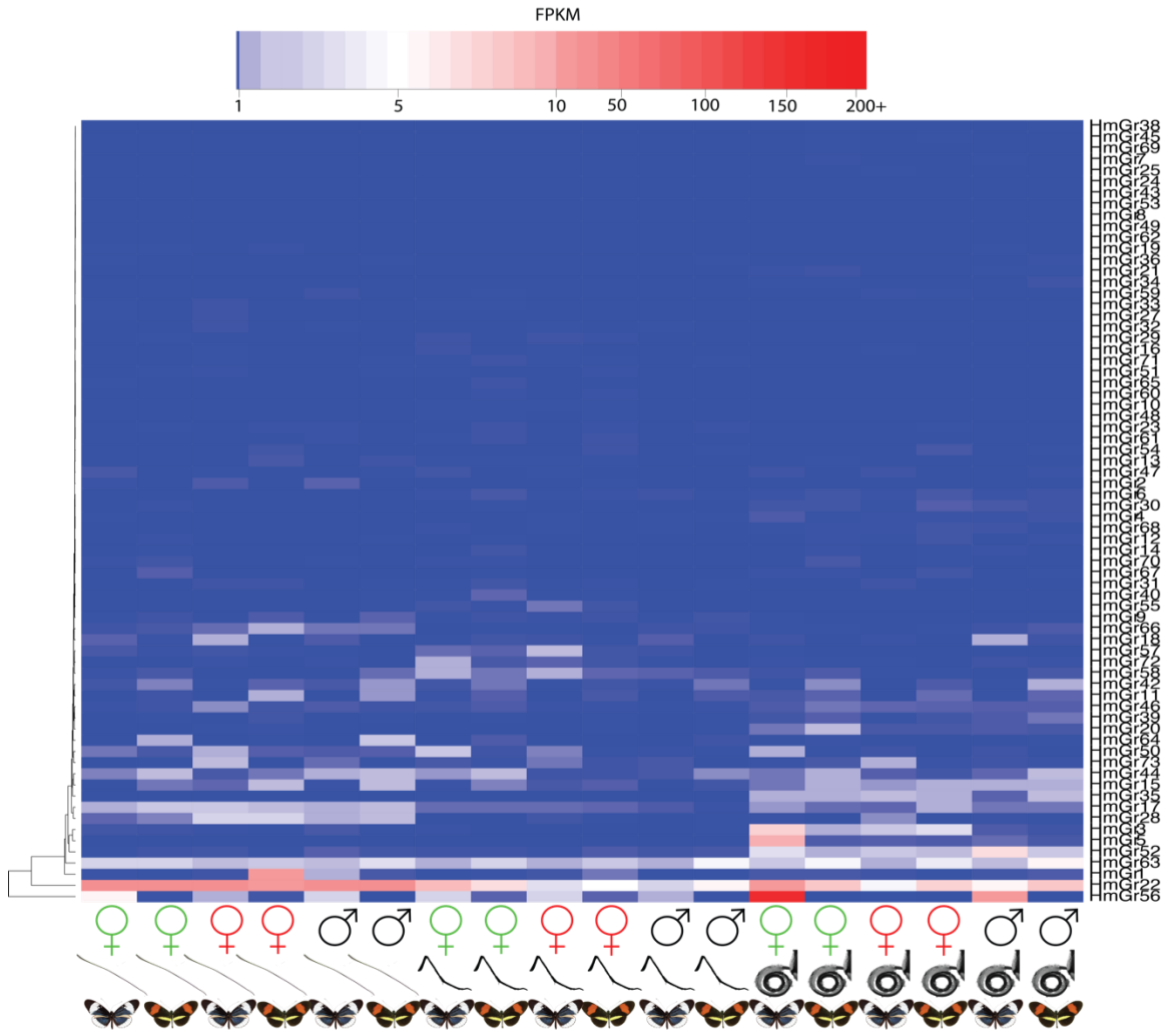


S2 Figure: Expression values (FPKM) for ionotropic receptors. The heatmap illustrates gene expression patterns for ionotropic receptors (IRs). On the y-axis, genes (one per row) are clustered based on expression. On the x-axis, individuals are identified by tissue, species, sex, and life-stage as indicated in the Figure S1 legend. Measures of expression are reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM); the scale for expression counts is shown under the header of each heatmap. Note how IRs are expressed predominantly in the antenna.

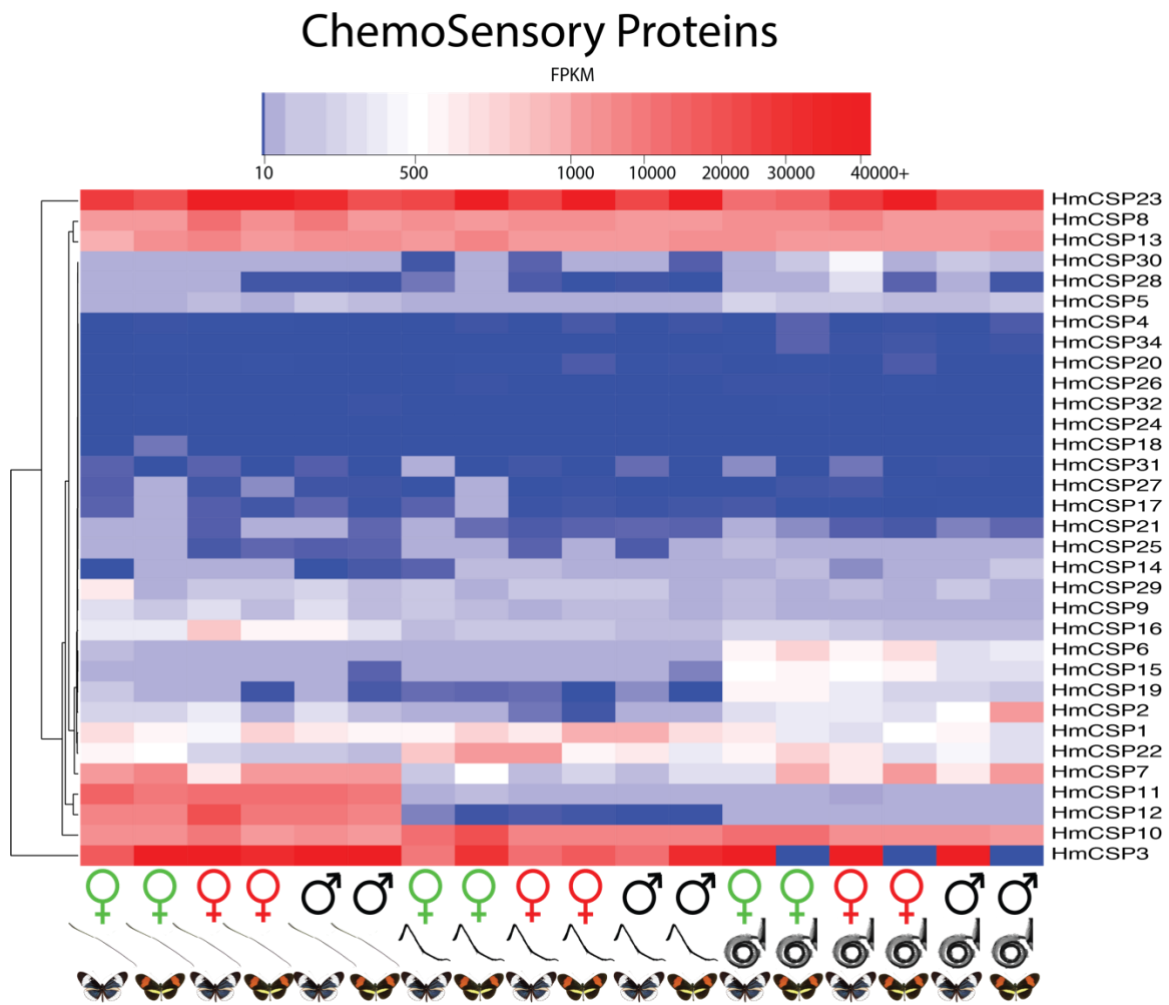


S3 Figure: Expression values (FPKM) for olfactory receptors. The heatmap illustrates gene expression patterns for olfactory receptors (ORs). On the y-axis, genes (one per row) are clustered based on expression. On the x-axis, individuals are identified by tissue, species, sex, and life-stage as indicated in the Figure S1 legend. Measures of expression are reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM); the scale for expression counts is shown under the header of each heatmap. Note how ORs are expressed predominantly in the antenna.

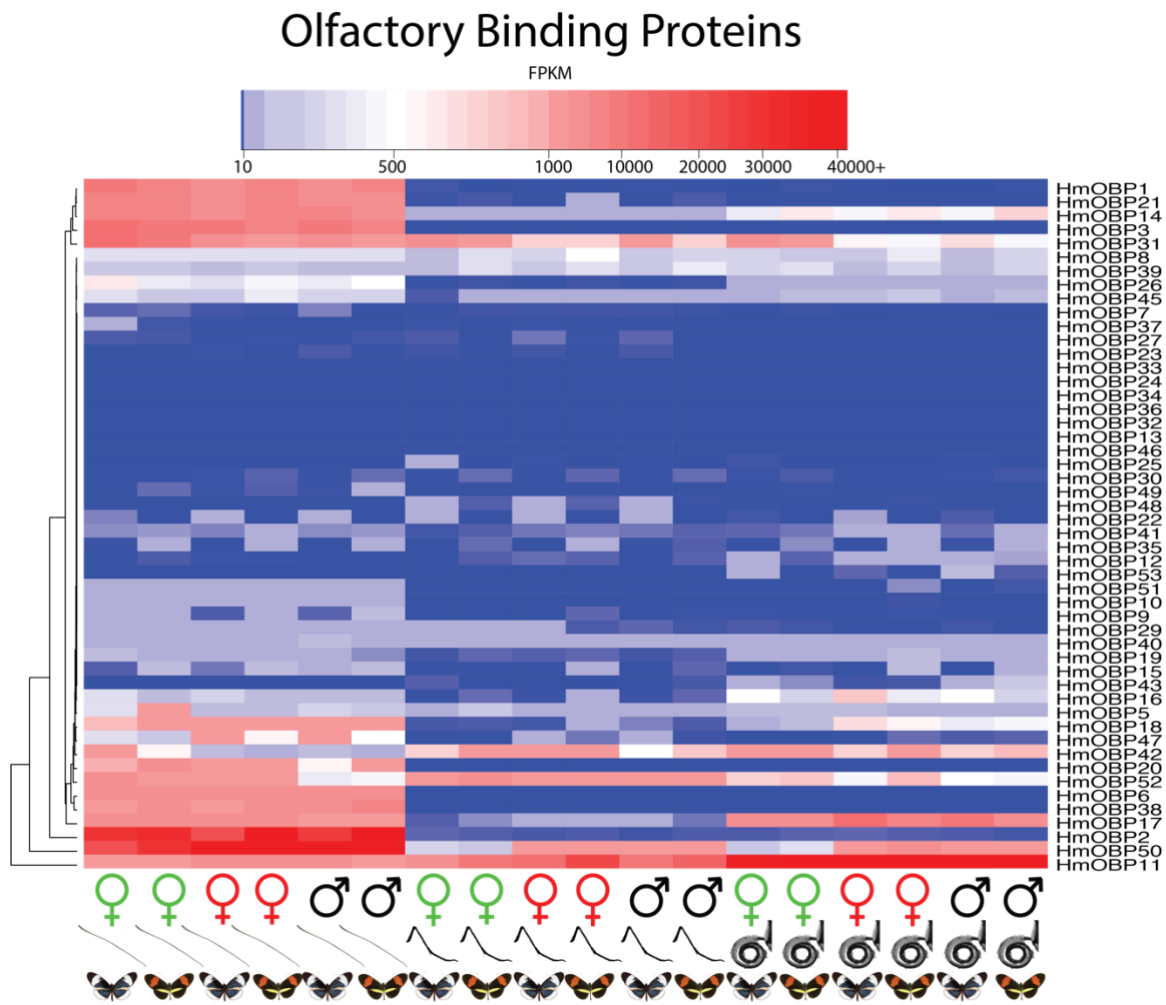
Gustatory Receptors



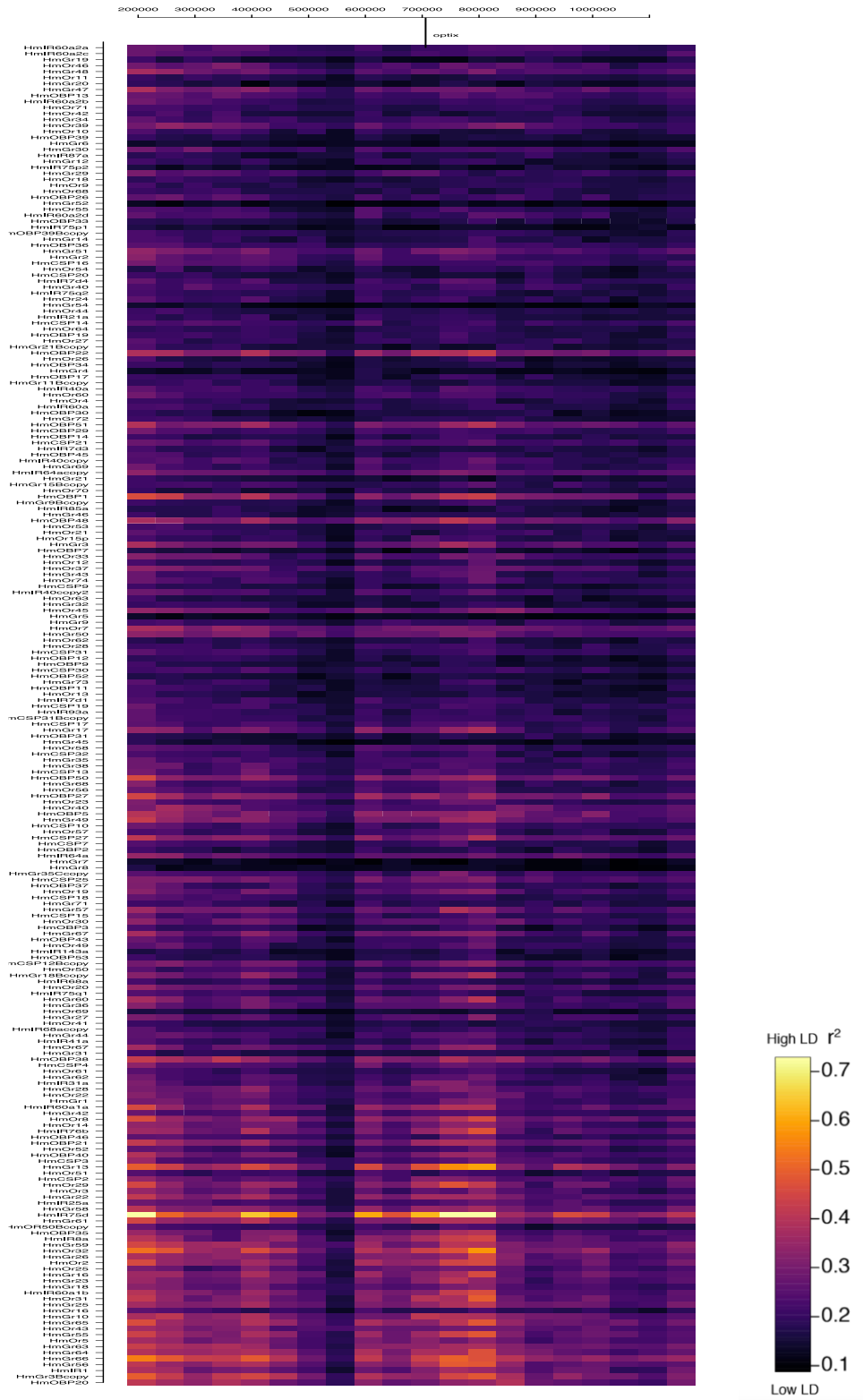
S4 Figure: Expression values (FPKM) for gustatory receptors. The heatmap illustrates gene expression patterns for gustatory receptors (GRs). On the y-axis, genes (one per row) are clustered based on expression. On the x-axis, individuals are identified by tissue, species, sex, and life-stage as indicated in the Figure S1 legend. Measures of expression are reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM); the scale for expression counts is shown under the header of each heatmap.



S5 Figure: Expression values (FPKM) for chemosensory proteins. The heatmap illustrates gene expression patterns for chemosensory proteins (CSPs). On the y-axis, genes (one per row) are clustered based on expression. On the x-axis, individuals are identified by tissue, species, sex, and life-stage as indicated in the Figure S1 legend. Measures of expression are reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM); the scale for expression counts is shown under the header of each heatmap. Note that the FPKM scale is different from Figures S6; S7, and S8 - expression counts are higher for transporters relative to receptors. Generally, CSPs show a homogeneous expression pattern throughout tissues, with some individual genes showing comparatively higher expression levels.



S6 Figure: Expression values (FPKM) for olfactory binding proteins. The heatmap illustrates gene expression patterns for olfactory binding proteins (OBPs). On the y-axis, genes (one per row) are clustered based on expression. On the x-axis, individuals are identified by tissue, species, sex, and life-stage as indicated in the Figure S1 legend. Measures of expression are reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM); the scale for expression counts is shown under the header of each heatmap. Note that the FPKM scale is different from Figures S6; S7, and S8 - expression counts are higher for transporters relative to receptors. Generally, OBPs are characterized by increased expression levels in the antennae, many of these genes are also found to be expressed in the legs and labial palps.



S7 Figure: Linkage disequilibrium (LD) map of chemosensory genes and the color pattern gene *optix*. On the y-axis chemosensory genes are annotated. The x-axis indicates the 1,000kb window around *optix*, the position of which is noted with a line, analyzed in 50kb windows. Please see S5 table for LD values and gene names.

hmc7z26	0.4483	0.3811	0.2790	0.3423	0.4208	0.3212	0.2944	0.1953	0.3746	0.3091	0.3302	0.3984	0.4237	0.3749	0.2429	0.3035	0.3028	0.2328	0.2117	0.2702	-0.0139
hmc7z1	0.5322	0.4242	0.3273	0.3323	0.4141	0.3427	0.2762	0.2091	0.4188	0.3145	0.3966	0.4692	0.5332	0.3280	0.2249	0.3185	0.3076	0.2535	0.2319	0.2917	-0.0139
hmc7z5	0.3922	0.3346	0.2903	0.3167	0.3483	0.3075	0.2482	0.2162	0.3218	0.2985	0.3577	0.4224	0.4011	0.3126	0.2132	0.2694	0.2714	0.2562	0.2013	0.2745	-0.0151
hmc7z16	0.4706	0.3971	0.3211	0.3478	0.3767	0.3322	0.3022	0.1904	0.3558	0.3057	0.4133	0.4380	0.4448	0.3209	0.1895	0.2911	0.2912	0.2586	0.2263	0.2725	-0.0164
hmc7z3	0.4068	0.3529	0.2639	0.2814	0.3456	0.2777	0.2488	0.1781	0.3473	0.2730	0.3119	0.3961	0.3659	0.3050	0.1888	0.2484	0.2713	0.2386	0.2271	0.2345	-0.0167
hmc7z13	0.3924	0.3202	0.2689	0.2927	0.3310	0.2548	0.2300	0.1875	0.2940	0.2872	0.3497	0.3733	0.3934	0.2869	0.1888	0.2362	0.2520	0.2649	0.1983	0.2500	-0.0173
hmc7z10a10	0.4574	0.3602	0.3019	0.3125	0.3727	0.3131	0.2776	0.1643	0.3891	0.2686	0.3386	0.3986	0.4650	0.3028	0.2054	0.2697	0.2684	0.2564	0.2587	0.2619	-0.0174
hmc7z31	0.3856	0.3528	0.2832	0.3720	0.3678	0.2828	0.2840	0.1971	0.3265	0.2789	0.3782	0.4329	0.3949	0.3375	0.1896	0.2487	0.3004	0.2835	0.2244	0.2894	-0.0199
hmc7z5	0.4555	0.3911	0.3062	0.3477	0.3799	0.3578	0.2739	0.1979	0.3704	0.3019	0.3633	0.3958	0.4468	0.3394	0.2187	0.2787	0.3157	0.2969	0.2377	0.2694	-0.0206
hmc7z16	0.2297	0.1999	0.1897	0.2014	0.2148	0.2052	0.1737	0.1557	0.2195	0.1729	0.2226	0.2629	0.2234	0.1832	0.1396	0.1835	0.1907	0.1643	0.1519	0.1873	-0.0244
hmc7z10	0.4796	0.3950	0.2905	0.3021	0.3817	0.2903	0.2543	0.1854	0.3529	0.2904	0.3791	0.4275	0.4271	0.3207	0.2143	0.2475	0.2894	0.2592	0.2465	0.2862	-0.0357
hmc7z5	0.4482	0.3875	0.3072	0.3448	0.4168	0.3622	0.2758	0.1941	0.3587	0.3038	0.3552	0.4540	0.4241	0.2947	0.2296	0.3093	0.2992	0.2689	0.2275	0.2917	-0.0362
hmc7z43	0.3742	0.2905	0.2671	0.3021	0.2921	0.2852	0.2670	0.1863	0.3024	0.2636	0.3194	0.3482	0.3623	0.2957	0.1992	0.2740	0.2836	0.2417	0.1956	0.2285	-0.0363
hmc7z3	0.4610	0.3775	0.3135	0.2990	0.4064	0.3061	0.2585	0.1831	0.3502	0.2803	0.3320	0.4024	0.4138	0.2999	0.1999	0.2707	0.2901	0.2423	0.2375	0.2713	-0.0808
hmc7z4	0.4556	0.3783	0.3130	0.3445	0.3886	0.3162	0.2773	0.1938	0.3358	0.3094	0.3764	0.4026	0.4035	0.3301	0.2000	0.2689	0.3322	0.2803	0.2308	0.2903	-0.0864
hmc7z6	0.5122	0.4049	0.3301	0.3925	0.4505	0.3558	0.3398	0.2163	0.4489	0.3589	0.4634	0.5151	0.5209	0.3667	0.2570	0.3747	0.3453	0.2940	0.2940	0.3391	-0.1051
hmc7z6	0.4568	0.3926	0.3430	0.3408	0.4309	0.3561	0.2743	0.2138	0.3381	0.2879	0.3192	0.4132	0.4450	0.3554	0.2147	0.2816	0.2901	0.2706	0.2581	0.2536	-0.1093
hmc7z5	0.4407	0.3608	0.2792	0.3434	0.4362	0.3355	0.2872	0.1604	0.3847	0.2958	0.4604	0.5064	0.4715	0.3330	0.2177	0.2803	0.3083	0.2833	0.2147	0.2813	-0.1149
hmc7z20	0.3401	0.2767	0.2706	0.2523	0.2970	0.2531	0.2219	0.2028	0.3048	0.2634	0.2817	0.2887	0.3425	0.2585	0.1710	0.2175	0.2374	0.2106	0.1932	0.2552	-0.3562

SI References

1. N. J. Panaro, *et al.*, Evaluation of DNA fragment sizing and quantification by the Agilent 2100 bioanalyzer. *Clinical Chemistry* **46**, 1851–1853 (2000).
2. P. R. Desjardins, D. S. Conklin, Microvolume quantitation of nucleic acids. *Current Protocols in Molecular Biology* **45**, e2565 (2011).
3. H. Zhao, *et al.*, CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
4. J. W. Davey, *et al.*, Major Improvements to the *Heliconius melpomene* Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. 14 (2016).
5. J. W. Davey, *et al.*, Data from: Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution (2016) <https://doi.org/10.5061/dryad.3s795> (October 13, 2019).
6. K. K. Dasmahapatra, *et al.*, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
7. A. D. Briscoe, *et al.*, Female Behaviour Drives Expression and Evolution of Gustatory Receptors in Butterflies. *PLoS Genetics* **9**, e1003620 (2013).
8. B. van Schooten, C. D. Jiggins, A. D. Briscoe, R. Papa, Genome-wide analysis of ionotropic receptors provides insight into their evolution in *Heliconius* butterflies. *BMC Genomics* **17** (2016).
9. E. Samorodnitsky, *et al.*, Comparison of custom capture for targeted next-generation DNA sequencing. *Journal of Molecular Diagnostics* **17**, 64–75 (2015).
10. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25 (2009).
11. M. J. P. Robinson, J. T. Thorvaldsdóttir, H. Winckler, W. Guttman, M. Lander, E. S. Getz, G. Integrative genomics viewer. - PubMed - NCBI. *Nat Biotechnol* **29**, 29 (1): 24–26 (2011).
12. C. Trapnell, *et al.*, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562–578 (2013).
13. G. R. Warnes, *et al.*, *gplots: various R programming tools for plotting data*. R package version 2.17.0. Computer software. (2015).
14. S. H. Martin, J. W. Davey, C. Salazar, C. D. Jiggins, Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biology* **17**, e2006288 (2019).
15. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution* **28**, 2239–2252 (2011).
16. R. M. Merrill, *et al.*, Genetic dissection of assortative mating behavior. *PLoS Biology* **17**, e2005902 (2019).

17. R. M. Merrill, B. Van Schooten, J. A. Scott, C. D. Jiggins, Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. *Proceedings of the Royal Society B: Biological Sciences* **278**, 511–518 (2011).
18. R. G. Vogt, E. Große-Wilde, J. J. Zhou, The Lepidoptera Odorant Binding Protein gene family: Gene gain and loss within the GOBP/PBP complex of moths and butterflies. *Insect Biochemistry and Molecular Biology* **62**, 142–153 (2015).