



HAL
open science

Female behaviour drives expression and evolution of gustatory receptors in butterflies

Adriana D. Briscoe, Aide Macias-Munoz, Krzysztof M. Kozak, James R. Walters, Furong Yuan, Gabriel A. Jamie, Simon H. Martin, Kanchon K. Dasmahapatra, Laura C. Ferguson, James Mallet, et al.

► **To cite this version:**

Adriana D. Briscoe, Aide Macias-Munoz, Krzysztof M. Kozak, James R. Walters, Furong Yuan, et al.. Female behaviour drives expression and evolution of gustatory receptors in butterflies. PLoS Genetics, 2013, 9 (7), 10.1371/journal.pgen.1003620 . hal-01190382

HAL Id: hal-01190382

<https://hal.science/hal-01190382>

Submitted on 29 May 2020

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

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

Female Behaviour Drives Expression and Evolution of Gustatory Receptors in Butterflies

Adriana D. Briscoe^{1*}, Aide Macias-Muñoz¹, Krzysztof M. Kozak², James R. Walters³, Furong Yuan¹, Gabriel A. Jamie², Simon H. Martin², Kanchon K. Dasmahapatra⁴, Laura C. Ferguson⁵, James Mallet^{6,7}, Emmanuelle Jacquin-Joly⁸, Chris D. Jiggins^{2*}

1 Department of Ecology and Evolutionary Biology, University of California, Irvine, California, United States of America, **2** Department of Zoology, University of Cambridge, Cambridge, United Kingdom, **3** Department of Biology, Stanford University, Palo Alto, California, United States of America, **4** Department of Biology, University of York, York, United Kingdom, **5** Department of Zoology, University of Oxford, Oxford, United Kingdom, **6** Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America, **7** Department of Genetics, Evolution and Environment, University College London, London, United Kingdom, **8** INRA, UMR 1272 INRA-UPMC Physiologie de l'Insecte: Signalisation et Communication, Versailles, France

Abstract

Secondary plant compounds are strong deterrents of insect oviposition and feeding, but may also be attractants for specialist herbivores. These insect-plant interactions are mediated by insect gustatory receptors (*Grs*) and olfactory receptors (*Ors*). An analysis of the reference genome of the butterfly *Heliconius melpomene*, which feeds on passion-flower vines (*Passiflora* spp.), together with whole-genome sequencing within the species and across the *Heliconius* phylogeny has permitted an unprecedented opportunity to study the patterns of gene duplication and copy-number variation (CNV) among these key sensory genes. We report *in silico* gene predictions of 73 *Gr* genes in the *H. melpomene* reference genome, including putative CO₂, sugar, sugar alcohol, fructose, and bitter receptors. The majority of these *Grs* are the result of gene duplications since *Heliconius* shared a common ancestor with the monarch butterfly or the silkworm. Among *Grs* but not *Ors*, CNVs are more common within species in those gene lineages that have also duplicated over this evolutionary time-scale, suggesting ongoing rapid gene family evolution. Deep sequencing (~1 billion reads) of transcriptomes from proboscis and labial palps, antennae, and legs of adult *H. melpomene* males and females indicates that 67 of the predicted 73 *Gr* genes and 67 of the 70 predicted *Or* genes are expressed in these three tissues. Intriguingly, we find that one-third of all *Grs* show female-biased gene expression ($n = 26$) and nearly all of these ($n = 21$) are *Heliconius*-specific *Grs*. In fact, a significant excess of *Grs* that are expressed in female legs but not male legs are the result of recent gene duplication. This difference in *Gr* gene expression diversity between the sexes is accompanied by a striking sexual dimorphism in the abundance of gustatory sensilla on the forelegs of *H. melpomene*, suggesting that female oviposition behaviour drives the evolution of new gustatory receptors in butterfly genomes.

Citation: Briscoe AD, Macias-Muñoz A, Kozak KM, Walters JR, Yuan F, et al. (2013) Female Behaviour Drives Expression and Evolution of Gustatory Receptors in Butterflies. *PLoS Genet* 9(7): e1003620. doi:10.1371/journal.pgen.1003620

Editor: Jianzhi Zhang, University of Michigan, United States of America

Received: February 4, 2013; **Accepted:** May 23, 2013; **Published:** July 11, 2013

Copyright: © 2013 Briscoe et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by an Overseas Visiting Scholarship from St. John's College, University of Cambridge, U.K., Faculty Research and Travel Funds from the School of Biological Sciences at the University of California, Irvine, and National Science Foundation grants IOS-1025106 and DBI-0939454 to ADB. Funding for LCF was provided by the John Fell Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: abriscoe@uci.edu (ADB); c.jiggins@zoo.cam.ac.uk (CDJ)

Introduction

Nearly 50 years ago Ehrlich and Raven proposed that butterflies and their host-plants co-evolve [1]. Based on field observations of egg-laying in adult female butterflies, feeding behavior of caterpillars, and studies of systematics and taxonomy of plants and butterflies themselves, they outlined a scenario in which plant lineages evolved novel defensive compounds which then permitted their radiation into novel ecological space. In turn, insect taxa evolved resistance to those chemical defences, permitting the adaptive radiation of insects to exploit the new plant niche. Ehrlich and Raven's theory of an evolutionary arms-race between insects and plants drew primarily from an examination of butterfly species richness and host-plant specialization. It did not specify the sensory mechanisms or genetic loci mediating these adaptive plant-insect interactions.

Insects possess gustatory hairs or contact chemosensilla derived from mechanosensory bristles, scattered along a variety of appendages [2–4]. In adult butterflies and moths, gustatory sensilla are found on the labial palps and proboscis (Figure 1), the legs (Figure 2A) [5], the antennae (Figure 2B) [6,7], and the ovipositor [8,9]. In adult *Heliconius charithonia* legs, the 5 tarsomeres of the male foreleg foretarsus are fused and lack chemosensory sensilla, while female foretarsi bear groups of trichoid sensilla ($n = 70–90$ sensilla/tarsus) associated with pairs of cuticular spines [10]. Each trichoid sensilla contains five receptor neurons. These sensilla are sensitive to compounds that may be broadly classified as phagostimulants (e.g., sugars and amino acids), which promote feeding behavior, or phagodeterrents (secondary plant compounds), which suppress it [11]; in adult females they may also modulate oviposition [12].

Genes for vision, taste and smell are likely to be crucial genomic loci underlying the spectacular diversity of butterfly-plant interac-

Author Summary

Insects and their chemically-defended hostplants engage in a co-evolutionary arms race but the genetic basis by which suitable host plants are identified by insects is poorly understood. Host plant specializations require specialized sensors by the insects to exploit novel ecological niches. Adult male and female *Heliconius* butterflies feed on nectar and, unusually for butterflies, on pollen from flowers while their larvae feed on the leaves of passion-flower vines. We have discovered—between sub-species of butterflies—fixed differences in copy-number variation among several putative sugar receptor genes that are located on different chromosomes, raising the possibility of local adaptation around the detection of sugars. We also show that the legs of adult female butterflies, which are used by females when selecting a host plant on which to lay their eggs, express more gustatory (taste) receptor genes than those of male butterflies. These female-biased taste receptors show a significantly higher level of gene duplication than a set of taste receptors expressed in both sexes. Sex-limited behaviour may therefore influence the long-term evolution of physiologically important gene families resulting in a strong genomic signature of ecological adaptation.

tions. The availability of genomes for two butterfly species, the postman *Heliconius melpomene* (Nymphalidae) [13] and the monarch (*Danaus plexippus*) [14], as well as the silkworm (*Bombyx mori*) [15], enables us to examine the evolutionary diversification of gustatory (*Gr*) and olfactory (*Or*) receptor genes that mediate insect-plant interactions. Each of these species feeds on hosts from different plant families. Silkworm larvae feed on mulberry (*Morus* spp., Moraceae) and monarch larvae feed on milkweed (*Asclepias* spp., Apocynaceae). The larvae of *Heliconius* feed exclusively on passion flower vines, primarily in the genus *Passiflora* (Passifloraceae). In addition, adult *Heliconius* are notable for several derived traits such as augmented UV color vision [16], pollen feeding (Figure 1B) [17,18], and the ability to sequester substances from their host plants that are toxic to vertebrate predators such as birds [19,20].

In *Drosophila melanogaster*, the *Gr* gene family consists of 60 genes [21–24], several of which are alternatively spliced, yielding 68 predicted *Gr* transcripts [24]. One or more of these *Gr* proteins including possibly obligatory co-receptors [25–27] may be expressed in each gustatory receptor neuron [11]. Originally considered members of the G-protein-coupled receptor (GPCR) family, insect *Gr*s have an inverted orientation in the membrane compared to the GPCR family of vertebrate *Grs* [28] and are part of the same superfamily as the insect *Ors* [21]. Signalling pathways for insect *Gr*s may be both G-protein dependent [29,30,31] and G-protein independent [32]. For the vast majority of *Drosophila Gr*s the specific compounds to which they are sensitive remain unknown. Nonetheless, several receptors for sugars [33–35], CO₂ [26,36], bitter substances [37–39] and plant-derived insecticides [25] have been identified in flies.

Knowledge of the *Gr* gene family for insects outside *Drosophila* is sparse and has primarily relied on the analyses of individual reference genomes. Expression studies are challenging, due to the very low expression of *Grs* in gustatory tissues [21,23]. In addition, *Grs* and *Ors* typically have large introns, small exons and undergo fast sequence evolution, making their *in silico* identification using automated gene prediction algorithms from genomic sequences problematic. Thus, the large repertoire of *Grs* (and *Ors*) that have been examined in the reference genomes of the pea aphid [40], the

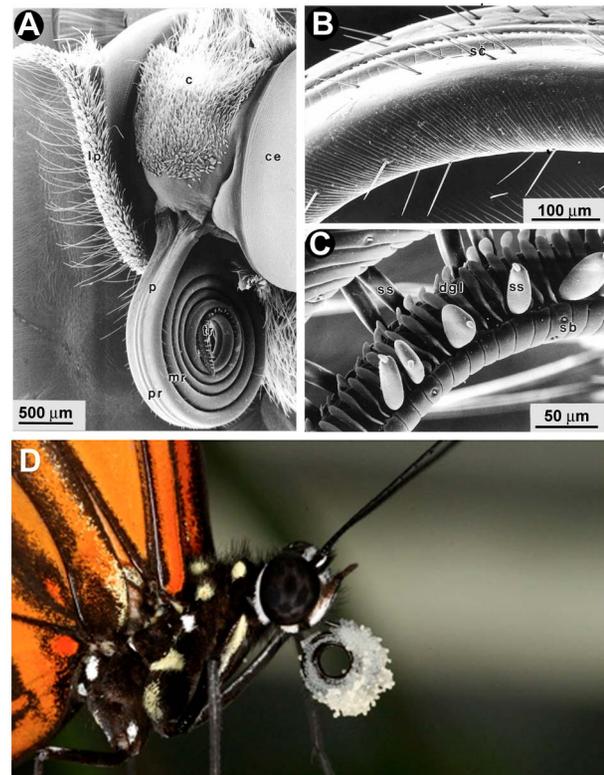


Figure 1. Scanning electron micrographs of the proboscis of *Heliconius* butterflies. (A) The labial palps (lp) and proboscis (p) of the *H. erato* head contain gustatory sensilla. (B) The proximal portion of the *H. melpomene* proboscis has hair-like sensilla chaetica (sc). (C) The tip portion of the proboscis has specialized ridges for pollen collection along with sensilla styloconica (ss). Reproduced with permission [9]. (D) *H. melpomene* with a pollen-load. c, clypeus; ce, compound eye; pr, proximal region; mr, mid region; tr, tip region; dgl, dorsal galeal linking structures; sb, blunt-tipped sensilla. doi:10.1371/journal.pgen.1003620.g001

honey bee [41], the red flour beetle *Tribolium castaneum* [42], the mosquitoes *Aedes aegypti* [43] and *Anopheles gambiae* [44], and several *Drosophila* spp. [45,46] have required extensive manual curation. In Lepidoptera, a large insect group which includes ~175,000 species, completely described *Gr* (and *Or*) gene models from genomes are rare and limited to *B. mori* [47], *D. plexippus* [14] and *H. melpomene* (*Grs*, this study; *Ors*, [13]). In other lepidopteran species, only fragmentary *Gr* data are available: five sequences in *Spodoptera littoralis* [48], three in *Heliothis virescens* [49], two in *Manduca sexta* [50,51] and one in *Papilio xuthus* [52].

Adult females of each *Heliconius* species only lay eggs on a limited number of host plants [53], and therefore need to recognize different species from among the large and diverse Passifloraceae family, which also show a remarkable diversity of chemical defences [54]. The evolutionary arms race between *Heliconius* butterflies and their hosts led us to hypothesize that *Heliconius Gr*s (and *Ors*) might be subject to rapid gene duplication and gene loss as well as copy-number variation (CNV). Recent work taking advantage of published *Drosophila* genomes has shown a relationship between host specialization and/or endemism and an increased rate of gene loss, as well as a positive relationship between genome size and gene duplication [46,55]. Moreover, *Drosophila Gr*s appear to be evolving under weaker purifying selection than *Ors* [55].

We previously used the reference genome sequence for *H. melpomene* to annotate three chemosensory gene families, encoding

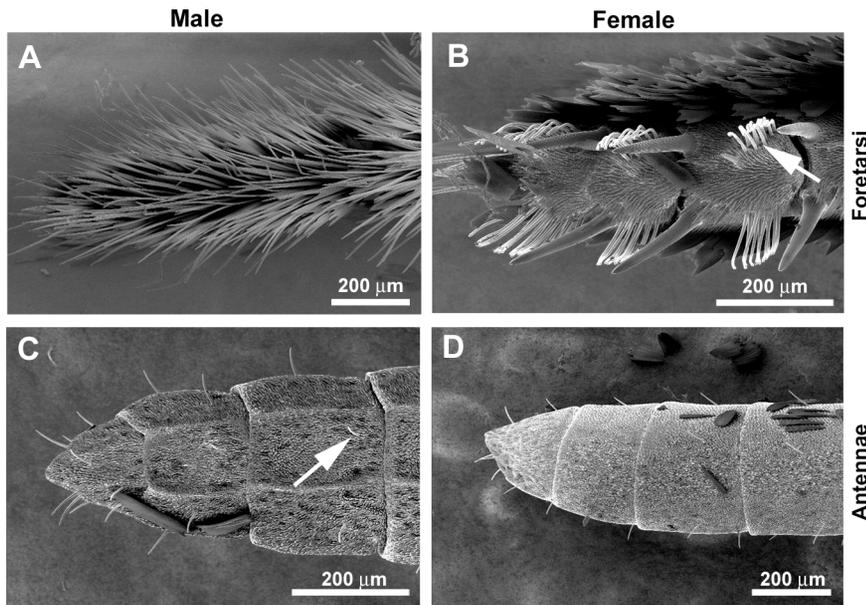


Figure 2. Sexual dimorphism in *H. melpomene* chemosensory tissues. Scanning electron micrographs of adult legs showing a sexual dimorphism in gustatory (trichoid) sensilla. Foreleg foretarsi of a male (A) and a female (B). Four pairs of clumped taste sensilla are each found associated with a pair of cuticular spines on each female foot (only three are shown). Arrow indicates a clump of taste sensilla. Antennae of an adult male (C) and a female (D) showing individual gustatory sensilla (arrow). doi:10.1371/journal.pgen.1003620.g002

the chemosensory proteins (CSPs), the odorant-binding proteins (OBPs), and the olfactory receptors (Ors). This demonstrated a surprising diversity in these gene families. In particular there are more CSPs in the butterfly genomes than in any other insect genome sequenced to date [13]. We build on this work below by characterizing the *Gr* gene family in the reference *H. melpomene* genome and in two other lepidopteran species whose genomes have been sequenced, *B. mori* (Bombycidae) and *D. plexippus* (Nymphalidae), by performing *in silico* gene predictions and phylogenetic analysis. We then analyzed whole-genome sequences of twenty-seven individual butterflies, representing eleven species sampled across all major lineages of the *Heliconius* phylogeny and including sixteen individuals from two species, *H. melpomene* and its sister-species *H. cydno*. We also generated RNA-sequencing expression profiles of the proboscis and labial palps, antennae and legs of individual adult male and female butterflies of the sub-species *H. melpomene rosina* from Costa Rica (~1 billion 100 bp reads). We used these data to address four major questions: Are different chemosensory modalities less prone to duplication and loss than others (e.g., taste vs. olfaction)? Is there evidence of lineage-specific differentiation of *Gr* (and *Or*) repertoires between genera, species and populations? What is the relationship between CNVs and the retention of paralogous genes over long-term evolutionary timescales? Are the life history differences between males and females reflected in the expression of *Grs* and *Ors* as well as in the retention of novel sensory genes in the genome?

We find higher turnover of the *Grs* than the *Ors* over longer evolutionary timescales, and evidence for both gene duplication and loss among a clade of intronless *Grs* between lepidopteran species and within the genus *Heliconius*. We also find for *H. melpomene* and its sister species, *H. cydno*, evidence of copy-number variation (CNVs) within their *Gr* and *Or* repertoires. Lastly, our RNA-sequencing suggests both tissue-specific and sex-specific differences in the diversity of expressed *Grs* and *Ors*, with female legs expressing a more diverse suite of *Grs* than male legs. Our data set revealing the expression of 67 of 73 predicted *Gr* genes and 67

of 70 predicted *Or* genes in adult *H. melpomene* butterflies is the most comprehensive profiling of these chemosensory gene families in Lepidoptera to date, and suggests how female host plant-seeking behaviour shapes the evolution of gustatory receptors in butterflies.

Results

Annotation of *Grs* in the reference genome of *H. melpomene*

In total, we manually annotated 86,870 bp of the *H. melpomene* reference genome (Table S1). Our 73 *Gr* gene models, consisted of 1–11 annotated exons, with the majority having three or four exons; six were intronless. We found genomic evidence (but not RNA-seq evidence) of possible alternative splicing of the last two exons of *HmGr18*, bringing the total number of predicted *Grs* to 74. Alternative splicing has not been previously described in the silkworm *B. mori* [47], but is known to occur in most other insects examined, including *D. melanogaster*, *Anopheles gambiae*, *Aedes aegypti* and *T. castaneum* [24,43,44]. We also identified eleven new putative *Grs* in the monarch butterfly genome, *DpGr48-56*, *DpGr66* and *DpGr68* (Table S1) [14].

All but five of our gene models contained more than 330 encoded amino acids (AAs) while individual gene models ranged from 258–477 AAs. Several *Gr* genes contained internal stop codons (Table S1). In at least one case, we found RNA-seq evidence of an expressed pseudogene—*HmGr61*—with two in-frame stop codons. In other cases, the 5' end of our assembled transcripts was not long enough to verify the internal stop codons in the genome assembly. The *Grs* are located on 33 distinct scaffolds, with 58 forming clusters of 2–8 genes on 18 scaffolds, distributed across 14 chromosomes.

Gene duplication and loss in a clade of putative bitter receptors

To study the patterns of gene duplication and loss more broadly across the Lepidoptera, we next examined the phylogenetic

relationships of *Gr*s from the three lepidopteran reference genomes [13–15]. Across the gene family phylogeny a large number of duplications among the putative ‘bitter’ gustatory receptors of *Heliconius* or *Danaus* have occurred, while the putative CO₂ and sugar receptors are evolving more conservatively, with only single copies in the *H. melpomene* reference genome (see below)(black arcs, Figure 3). A majority (~64%) of *Gr* genes found in the *H. melpomene* genome are the result of gene duplication since *Heliconius* shared a common ancestor with *Danaus* or *Bombyx*. This is in contrast to the more conserved pattern of evolution of the *Or*s (Figure 4) [13] where a majority (37 of 70 or 53%) of genes show a one-to-one orthologous relationship with either a gene in *Danaus*, in *Bombyx* or both.

Within the genus *Heliconius* there is a great diversity of host plant preferences for different *Passiflora* species. To look at the relationship between gene duplication and loss over this shorter timescale, we focussed our efforts on a group of six intronless *Gr*s, *HmGr22–26* and *Gr53*, because it is only feasible to identify single-exon genes with high confidence, given that the Illumina whole-genome sequencing approach leads to poorly assembled genomes (Table S2). These genes are also of interest as some members of this group are very highly expressed. Notably *HmGr22* is one of the

most widely expressed genes in our adult *H. melpomene* transcriptomes, which was verified by reverse-transcriptase (RT)-PCR and sequencing of the PCR products (Figure 5A). In this regard *HmGr22* resembles another intronless *Gr*, the silkworm gene *BmGr53*, which is expressed in adult male and female antennae and larval antennae, maxilla, labrum, mandible, labium, thoracic leg, proleg and gut [32]. The remaining five intronless *Gr*s have much more limited domains of expression in adult *H. melpomene* (see below). We searched for these genes in *de novo* assemblies of whole-genome Illumina sequences from eleven species across the *Heliconius* phylogeny. We investigate whether, as in *Drosophila*, a high turnover in putative bitter receptors is observed in species with host plant specializations or in species which are endemic and thus smaller in effective population size [46].

Although patterns of host plant use are complex within the genus, some notable host-plant shifts have occurred, leading to the prediction that gene loss may have occurred along more specialized lineages [46]. For example, *H. doris* unlike many *Heliconius*, tends to feed on large woody *Passiflora* that can support their highly gregarious larvae [53]. It also probably has a smaller effective population size than most other *Heliconius* species. From the 11 species studied, we identified a total of 44 intact or nearly

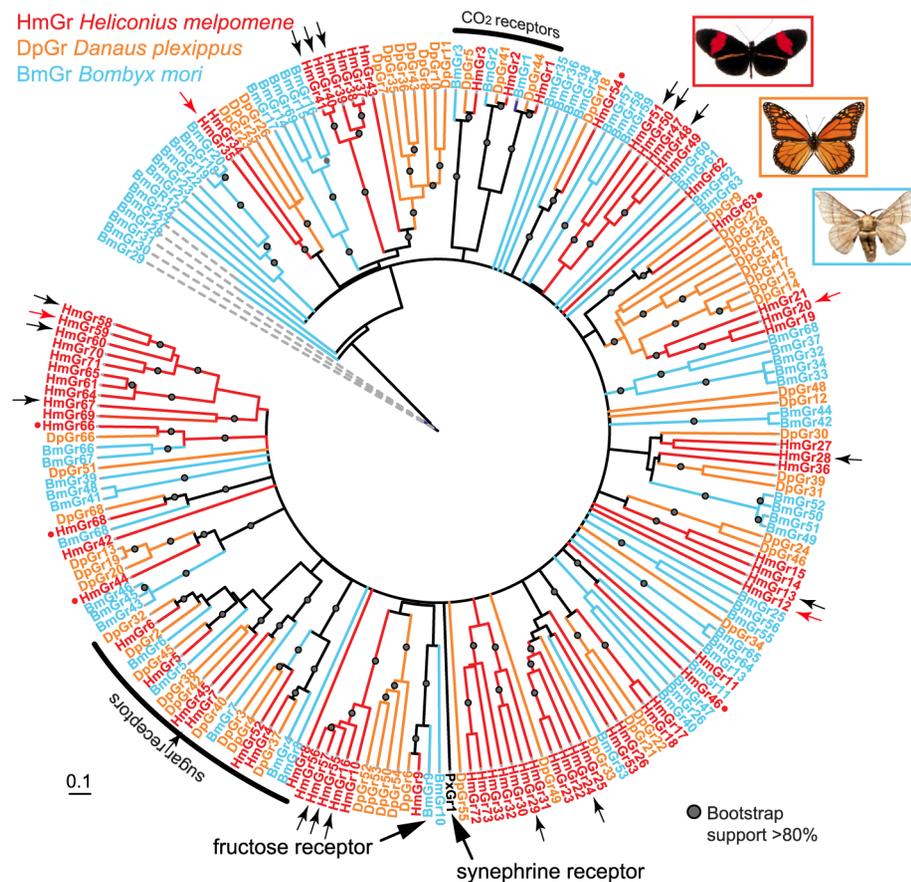


Figure 3. Phylogeny of the *Gr*s identified in three lepidopteran genomes. A maximum likelihood analysis of amino acid sequences was performed. Bootstrap support is out of 500 replicates. Putative CO₂ and fructose receptors show a conserved 1-to-1 orthologous relationship in each of the three lepidopteran genomes, while putative sugar receptors of the monarch butterfly have duplicated twice. By contrast, numerous butterfly- or moth-specific gene duplications are evident among the remaining *Gr*s, which are hypothesized to be bitter receptors. Small red dots indicate single-copy *Heliconius Gr*s classified as conserved genes in the analyses shown in Table 1 and Table 2. Small black arrows indicate female-specific *Gr*s expressed in adult *H. melpomene* legs. Small red arrows indicate *Gr*s expressed in adult *H. melpomene* proboscis only. Bar indicates branch lengths in proportion to amino acid substitutions/site. Synephrine and fructose receptors are described in [52] and [32]. Bm = *Bombyx mori*, Hm = *Heliconius melpomene*, Dp = *Danaus plexippus*, Px = *Papilio xuthus*. doi:10.1371/journal.pgen.1003620.g003

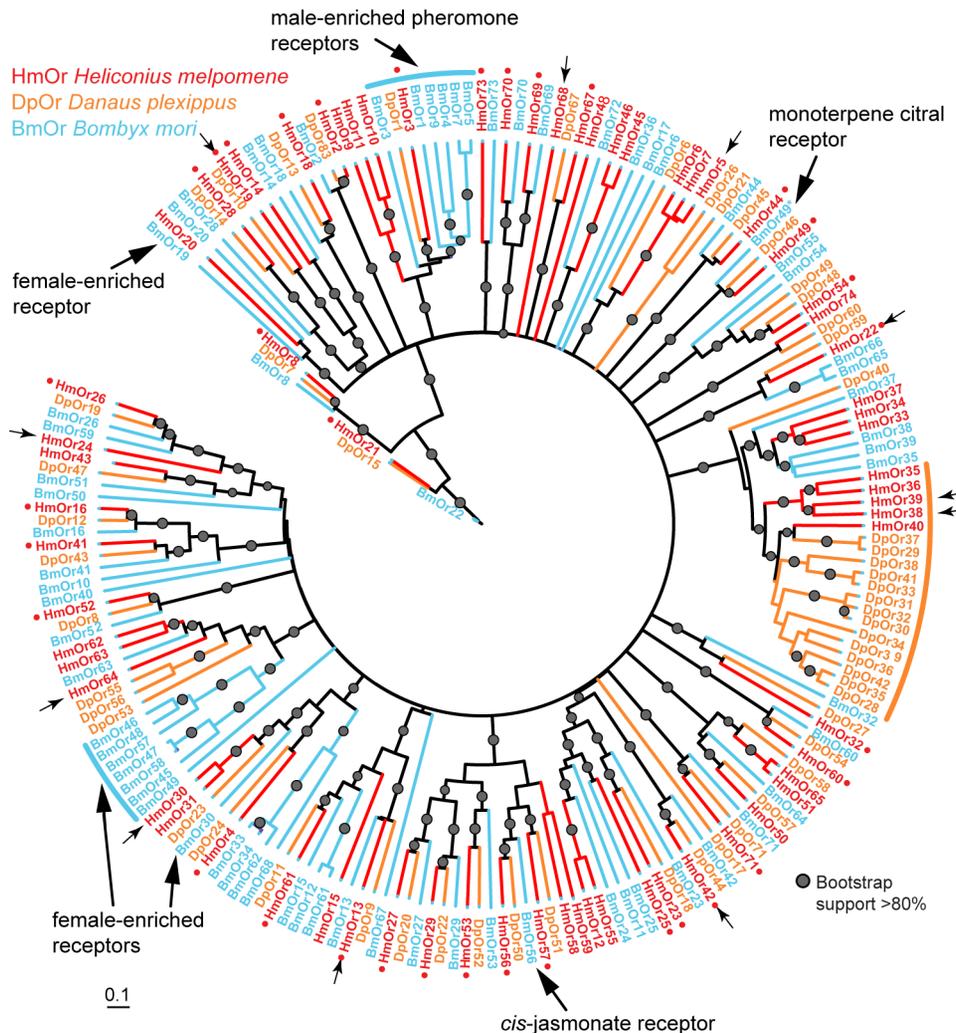


Figure 4. Phylogeny of the Ors identified in three lepidopteran genomes. A maximum likelihood analysis of amino acid sequences was performed. Bootstrap support is out of 500 replicates. Fewer lineage-specific duplications are evident among the Ors compared to the Grs, with the exception of one large butterfly-specific expansion (orange arc). Small red dots indicate single-copy *Heliconius* Ors classified as conserved genes in the analyses shown in Table 1 and Table 2. Ors that are enriched in male or female adult *B. mori* antennae (blue and black arcs) are described in [91]; *cis*-jasmonate and monoterpene citral receptors are described in [92] and [93]. Phylogenetic tree reconstruction details are given in [13]. Bar indicates branch lengths in proportion to amino acid substitutions/site. Small arrows indicate female-specific Ors expressed in adult *H. melpomene* legs. Bm = *Bombyx mori*, Hm = *Heliconius melpomene*, Dp = *Danaus plexippus*. doi:10.1371/journal.pgen.1003620.g004

intact intronless *Grs*, as well as three intronless pseudogenes (Genbank Accession Nos. KC313949-KC313997)(Table S2 and S3). We also identified one intact intronless *Gr* each in monarch and silkworm and one intronless *Gr* pseudogene in monarch. Phylogenetic analysis indicates that six intact intronless *Gr* genes were present at the base of the genus *Heliconius* while the intronless *Gr* pseudogene in monarch was the result of duplication since *Heliconius* and monarch shared a common ancestor (Figure 5B, Figure 6). Subsequent to the radiation of the genus *Heliconius*, there have been a number of gene losses. Whereas all members of the *melpomene* clade (*H. melpomene*, *H. cydno*, *H. timareta*) retained genomic copies of all six genes, members of the *erato* clade (*H. erato*, *H. clysonymus* and *H. telesiphe*) and *sara-sapho* clade (*H. sara* and *H. sapho*) have lost their copies of *Gr22* and *Gr25*. In addition, members of the so-called primitive clade (*H. wallacei*, *H. hecuba*, and *H. doris*) have lost *Gr23*, while *H. doris* and *H. wallacei* have apparently lost *Gr24* independently (Figure 6). The woody plant specialist, *H. doris*, has retained the fewest intronless *Grs*,

apparently also having lost its copy of *Gr53*, a pattern mirrored by *Drosophila* host plant specialists [46]. We have, however, no direct evidence that the intronless *Grs* are in fact involved in host plant discrimination so the observed patterns of loss may be better explained by other variables such as effective population size.

CNVs occur frequently among paralogous gustatory receptor genes

We next tested whether the greater level of diversification of *Grs* as compared to *Ors* over long evolutionary timescales (compare Figure 3 and Figure 4), is similarly reflected in greater population level variation in *Gr* and *Or* duplicate genes. To test this hypothesis, we examined the incidence of CNVs among *Grs* and *Ors* that exist as single-copy genes in the reference *H. melpomene* genome with a one-to-one orthologous relationship with a gene in *Danaus*, *Bombyx* or both (conserved)(red dots, Figure 3 and 4), or as genes that are *Heliconius*-specific where no orthologue exists in either *Danaus* or

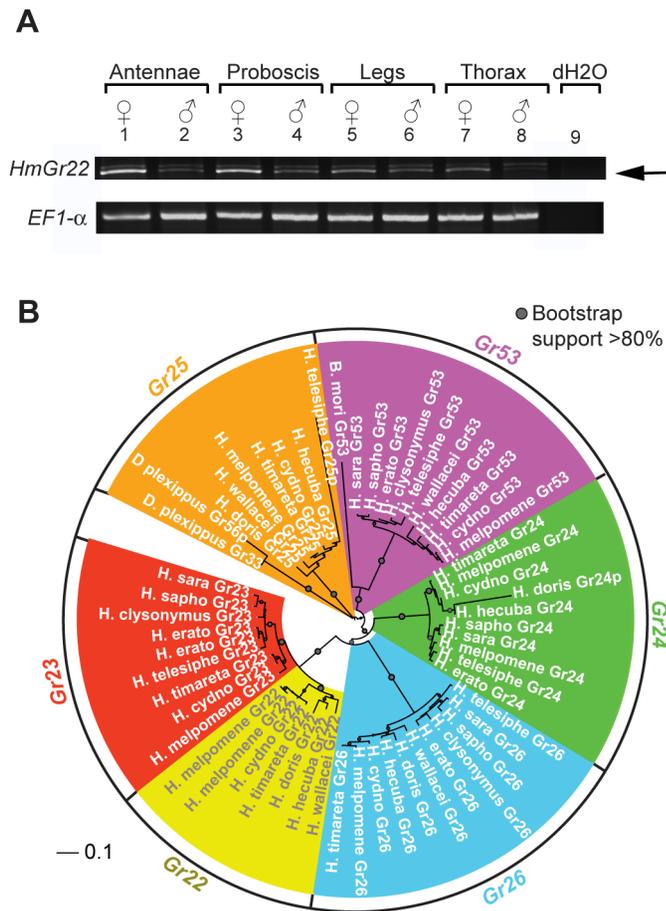


Figure 5. *HmGr22* expression in adults and intronless *Gr*s from whole-genome sequence data across the *Heliconius* phylogeny. (A) Reverse-transcriptase PCR (RT-PCR) of adult *H. melpomene* tissues showing the expression of *HmGr22* and *elongation factor-1 alpha*. Two products are evident from the *Gr22* RT-PCR. The bottom RT-PCR product is *HmGr22* (arrow) and the top RT-PCR product is 18 *s rRNA*, which was verified by Sanger sequencing. (B) Neighbor-joining tree showing the phylogenetic relationship between the forty-six intact *Gr*s and four pseudogenes identified in the 13 lepidopteran genomes. Bootstrap support is out of 500 bootstrap replicates. Pseudogene sequences are indicated by a 'p' after the gene name. doi:10.1371/journal.pgen.1003620.g005

Bombyx (non-conserved). We used whole genome resequence data (12 genomes) for three subspecies of *H. melpomene* (*H. melpomene amaryllis*, $n = 4$; *H. melpomene aglaope*, $n = 4$; and *H. melpomene rosina*, $n = 4$) (Figure 7, inset) and one sub-species of *H. cydno* (*H. cydno chioneus*, $n = 4$) (Table S4). We first mapped genomic resequence reads to the *H. melpomene melpomene* reference genome, and then searched for regions of abnormal coverage using CNVnator [56]. More than half of *Gr* loci showed presence of CNVs (37 out of 68 loci). However, there were noticeably fewer CNVs in *Gr* loci that evolve conservatively over the long-term, such as among the putative CO₂ receptors, while there was an excess of CNVs in loci that show patterns of *Heliconius*-specific duplication (11.1% vs. 54.9%, respectively) (Fisher's Exact Test, two-tailed, $P = 0.0004$) (Table 1) (Figure 7). Intriguingly, many sugar receptor CNVs are sub-species specific; we observed fixed duplications relative to the reference genome in *H. melpomene aglaope* (*HmGr4*, *Gr5*, *Gr6*, *Gr8*, *Gr45*, *Gr52*) and *H. melpomene amaryllis* (*Gr4*, *Gr5*, *Gr6*, *Gr7*, *Gr8*, *Gr45*, *Gr52*), among genes that are found on different chromosomes (Table S5, Figure 7). Although the majority of CNVs are likely to be evolving neutrally, this raises the possibility of local adaptation within the species range around the detection of sugars. As expected given their long-term stability, *Or*s also show a lower incidence of CNVs (12 out of 67 loci), with no association between gene duplication and CNV incidence at least in *H. melpomene*

(Table 1, Table S6). In *H. cydno*, a slight excess of *Or* CNVs was observed in loci that resulted in paralogous genes over longer evolutionary timescales (Fisher's Exact Test, two-tailed, $P = 0.0475$) (Table 1) (Figure 8).

We have not experimentally verified the incidence of copy number variation in any of these genomes, and some of the regions identified as CNVs are likely to be false positives. To investigate the rate of false positives, we analysed resequence data from the reference genome itself and discovered 3 *Gr* and 3 *Or* CNVs, suggesting a false positive rate of around 4%. (We therefore excluded these loci from our statistical tests.) However, the fact that broad patterns of observed CNVs are consistent with the evolutionary patterns at deeper levels supports our conclusion that CNV, in the absence of strong purifying selection, is an important driver of gene family diversification. These results also provide a novel line of evidence that the butterfly *Gr*s have a higher rate of evolutionary turnover as compared to *Or*s.

Sexually dimorphic gustatory sensilla in adult legs mirror *Gr* expression diversity

The life histories of adult male and female butterflies are similar with respect to the need to find food and potential mates except that adult females are under strong selection to identify suitable

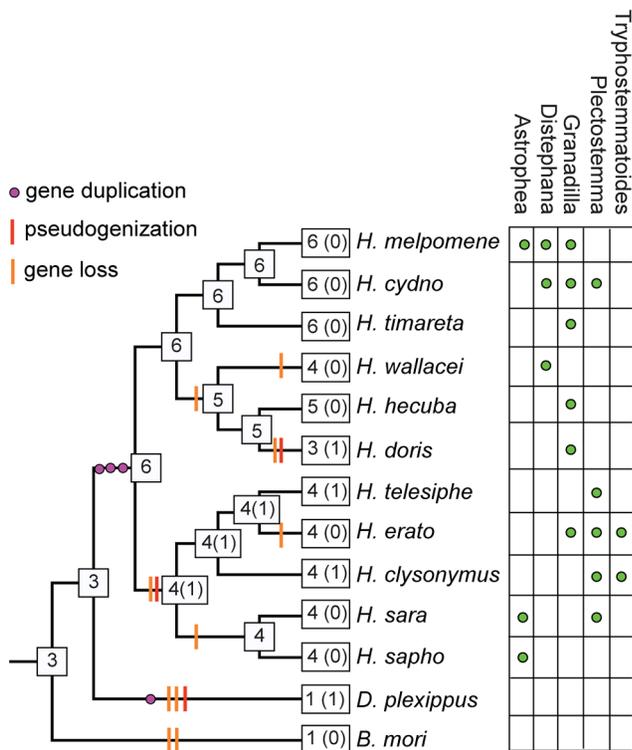


Figure 6. Inferred patterns of intronless *Gr* gene gain and loss across the genus *Heliconius*. Estimates of the number of *Gr* loci (number of pseudogenes is indicated in parentheses) on internal nodes of the lepidopteran phylogeny and gene gain (purple dots), gene loss (orange slashes) and pseudogenization events (red slashes) on each branch. *Heliconius* phylogeny is based on Beltran et al. (2007) [90]. Reconciliation of gene trees onto the species tree was performed in Notung using maximum likelihood gene family trees. Primary *Passiflora* host plant subgenera (green dots) affiliated with each *Heliconius* species [53]. No clear relationship exists between the number of known *Passiflora* subgenera used and the number of intronless *Gr*s in a species, which are presumed to be putative bitter receptors, but whose ligands are not yet identified. The woody vine specialist, *H. doris*, with the smallest effective population size, has the fewest intact intronless *Gr*s. doi:10.1371/journal.pgen.1003620.g006

host plants for oviposition. To ascertain host-plant identity, female butterflies drum with their legs on the surface of leaves before laying eggs [10]. This behaviour presumably allows the female to taste oviposition stimulants. Consistent with this behaviour, adult nymphalid butterfly legs are known to contain gustatory sensilla [57], and it has been reported that while nymphalid butterfly females have clusters of gustatory sensilla on their foreleg foretarsi, males lack these entirely [10,58]. Here we confirm this mostly anecdotal evidence for sexual dimorphism using scanning electron microscopy (SEM). The mid- and hindlegs of both male and female *H. melpomene* have similar numbers of individual gustatory sensilla along their entire lengths, but there is a striking difference in their abundance and distribution on the foretarsi of the female forelegs. Unlike males, females exhibit cuticular spines associated with gustatory (trichoid) sensillae ($n \sim 80$ sensilla/foretarsus for females; $n = 0$ /foretarsus for males) (Figure 2A) [10].

We therefore hypothesized that the repertoire of expressed *Gr* and *Or* genes in *H. melpomene* legs might be more diverse in females as compared to males. Furthermore, if female-specific genes are used for assessment of potential host plants, then fast-evolving insect-host interactions might produce rapid duplication of these genes over evolutionary timescales. Accordingly, we examined the

expression profiles of *Gr*s and *Or*s in adult *H. melpomene* by RNA-sequencing of libraries prepared from mRNAs expressed in adult antennae, labial palps and proboscis, and legs from one deeply-sequenced male and female each of *H. melpomene* (6 libraries total)(Table S7 and S8). The number of 100 bp reads per individual library ranged from 17.4 to 25.9 million for paired-end sequencing or 74.8–103.9 million for single-end sequencing (Table S8). To confirm these findings, we subsequently made 12 individual libraries from two more males and two more females (Table S7). As coverage was uneven across these libraries, we analysed them by merging biological replicates by sex and tissue type, and then downsampling so that an equal number of reads was analyzed for each treatment. The number of 100 bp reads analyzed for paired-end sequencing ranged from 19.4 to 49.6 million (Table S8). After downsampling, we examined the expression levels of the widely-expressed *elongation factor-1 alpha* gene in each of the libraries as a control, and found a comparable level of expression between sexes within each tissue type (Table S8). By careful visual examination of the uniquely-mapped reads to our 143 reference *Gr* and *Or* sequences, we found evidence of *Gr* and *Or* expression in all three adult tissue-types, with both tissue-specific and sex-specific differences as detailed below (Figure 9, Tables S9, S10, S11, S12, S13, S14). In total, we found evidence for expression of 67 of 73 *Gr*s and 67 of 70 *Or*s identified in the *H. melpomene* reference genome.

Strikingly, the sexual dimorphism of gustatory sensilla we observed among the foreleg foretarsi is reflected in *Gr* gene expression patterns. A total of thirty-two *Gr*s are expressed in both male and female *H. melpomene* leg transcriptomes including three CO₂ receptors, *HmGr1-3*, four putative sugar receptors *HmGr4*, *Gr6*, *Gr45* and *Gr52* and a fructose receptor, *HmGr9* (Figure 9A, Table S9, Supplementary Text). Many *Gr*s showed sex-specific expression, however, with many more *Gr*s in female ($n = 46$) as compared to male leg transcriptomes ($n = 33$)(Figure 9B, C).

In total 15 of these *Gr*s expressed in female legs, *HmGr10*, *Gr24*, *Gr26*, *Gr29*, *Gr40*, *Gr41*, *Gr48*, *Gr50*, *Gr51*, *Gr16*, *Gr55*, *Gr57*, *Gr58*, *Gr60* and *Gr67*, are the result of duplications since *Heliconius* and *Danaus* shared a common ancestor (Figure 3 small arrows, Figure 9B, Table S9). By contrast, only one of the three male-biased *Gr*s, *HmGr19*, evolved as a result of recent duplication. There is an excess of *Heliconius*-specific *Gr*s but not *Or*s (see below) that are expressed in female legs (Fisher's Exact Test, two-tailed, $p = 0.019$)(Table 2). Since male *H. melpomene* do not need to identify host-plants for oviposition, it seems likely that the 17 female-specific *Gr*s in our leg transcriptomes are candidate receptors involved in mediating oviposition (Figure S1).

Female *Gr* expression is more diverse in antennae than male *Gr* expression

Besides using their antennae for olfaction, female nymphalid butterflies also taste a host plant by antennal tapping before oviposition. This tapping behaviour presumably allows the host plant chemicals to come into physical contact with gustatory sensilla on the antennae. We therefore examined whether there was any difference in the abundance of gustatory sensilla on the antennae of male and female *H. melpomene*. Using scanning electron microscopy, we found individual gustatory sensilla scattered along each antennae of both male and female *H. melpomene* but no obvious sexual dimorphism in their abundance or distribution (Figure 2B). We found 28 *Gr*s expressed in both male and female *H. melpomene* antennae (Figure 9A, Table S10), including two sugar receptors, *HmGr4* and *HmGr52*, a putative fructose receptor *HmGr9* and two CO₂ receptors, *HmGr1* and *Gr3*. Besides the sugar and CO₂ receptors noted, other conserved genes

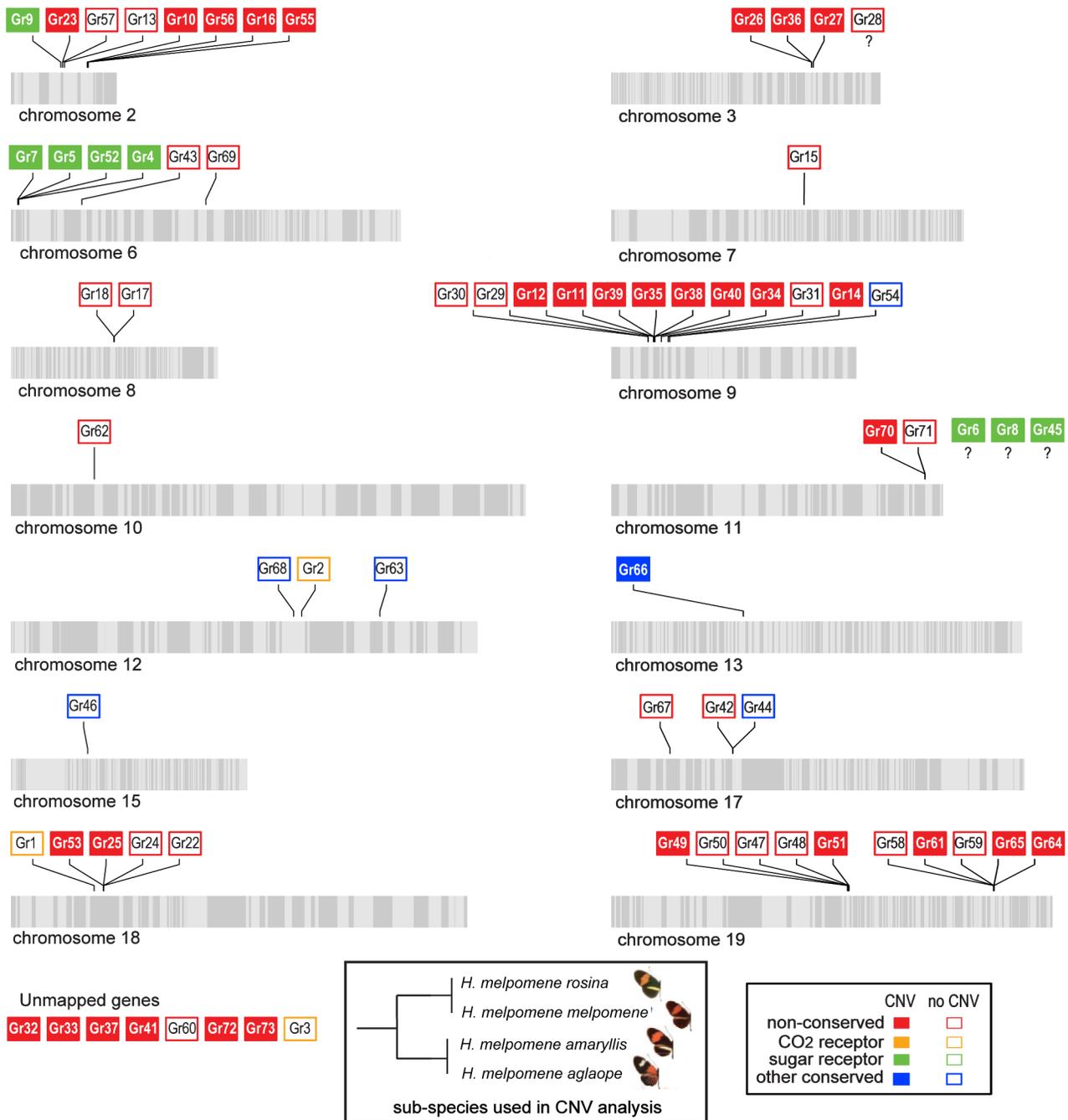


Figure 7. Copy-number variant (CNV) analysis of *Grs* in the *H. melpomene* genome. Scaffolds comprising each chromosome are indicated by alternating light and grey stripes. *Grs* without CNVs are indicated by open boxes and *Grs* with CNVs are indicated by closed boxes. *Grs* are classified as conserved if, in the *H. melpomene* reference genome, they have a one-to-one orthologous relationship with either a gene in *Danaus*, *Bombyx* or both (red dots, Figure 3). *Grs* are classified as non-conserved if they are duplicated in the *H. melpomene* reference genome or have no orthologue in either *Danaus*, *Bombyx* or both. Genes mapped to chromosomes but without precise locations are indicated by question marks. Scaffold arrangement is based on the published linkage map [13]. doi:10.1371/journal.pgen.1003620.g007

that are expressed in both male and female antennae include *HmGr63*, a candidate Gr co-receptor (see Text S1), and *HmGr66*, a candidate bitter receptor.

We also found 11 *Grs* expressed in female *H. melpomene* antennae that did not appear to be expressed in male antennae. Two of these, *HmGr47* and *Gr68*, appeared in the top one-third of the most abundant female antennal *Grs* in terms of number of

reads recovered from the individual butterfly transcriptome. In contrast, just four *Grs* were expressed in male antennae *HmGr11*, *Gr25*, *Gr31*, and *Gr69* but not female antennae (Figure 9B, C, Table S10). Six of the female-biased *Grs* and two of the male-biased *Grs* (*Gr31*, *Gr69*) expressed in antennae are the result of duplication events since *Heliconius* and *Danaus* shared a common ancestor.

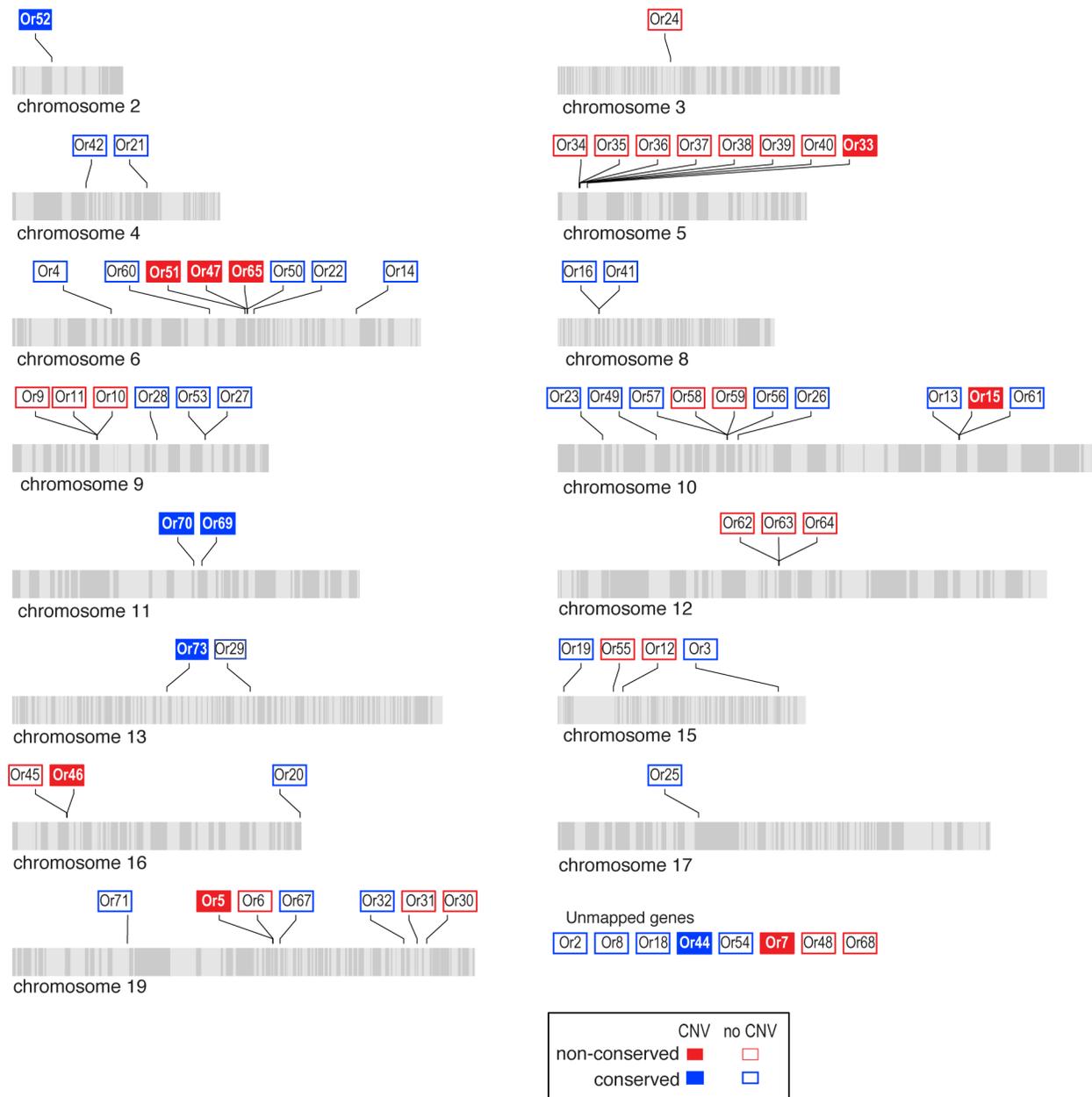


Figure 8. Copy-number variant (CNV) analysis of *Ors* in the *H. melpomene* genome. Scaffolds comprising each chromosome are indicated by alternating light and grey stripes. *Ors* without CNVs are indicated by open boxes and *Ors* with CNVs are indicated by closed boxes. The classification of *Ors* as being either conserved or non-conserved follows the same criteria as for the *Gr*s. The eight genes for which the chromosome locality is not known are shown at the bottom. doi:10.1371/journal.pgen.1003620.g008

Candidate *Heliconius* gustatory receptors for nectar- and pollen-feeding

By contrast with the leg and antennal tissue, where more *Gr*s are expressed in females compared to males, the labial palps and proboscis (Figure 1) transcriptomes contained the largest number of *Gr*s ($n = 35$) expressed in both sexes (Figure 9A, C, Table S11). Five of the six candidate sugar receptors in the *H. melpomene* genome are expressed in both the male and the female transcriptomes along with two of the three conserved CO₂ receptors, which may be used to assess floral quality [59] (Figure 3, Table S11). A majority (21 of 35) of *Heliconius* *Gr*s expressed in both male and female labial palps and proboscis

libraries have no existing ortholog in the silkworm genome, apparently the result of gene loss in *B. mori* or gene duplication along the lineage leading to *Heliconius* (Figure 3). This may in part reflect the fact that adult silkworms have lost the ability to feed. Interestingly, four *Gr*s expressed in both male and female labial palps and proboscis transcriptomes could not be detected in male and female antennae and legs (*HmGr12*, *Gr20*, *Gr35*, and *Gr59*) (Figure 3, red arrows, Figure 9B). Some of these *Gr*s might play a role in the pollen-feeding behaviour that is specific to *Heliconius*, and which involves preferences for particular species of flowers in the plant families Rubiaceae, Cucurbitaceae and Verbenaceae (see Discussion).

Table 1. Relationship between evolutionarily-conserved genes and copy-number variation (CNV).

| Species | Gene family | Gene classification | Number of genes with | | P value [§] |
|---------------------|-------------------------|---|----------------------|--------|----------------------|
| | | | CNV | No-CNV | |
| <i>H. melpomene</i> | <i>Grs</i> [†] | <i>Heliconius</i> -specific | 28 | 23 | 0.0004 |
| | | CO ₂ receptors+other conserved <i>Grs</i> [*] | 1 | 8 | |
| | | Sugar receptors | 8 | 0 | |
| <i>H. cydno</i> | <i>Grs</i> [†] | <i>Heliconius</i> -specific | 10 | 41 | 0.247 |
| | | CO ₂ receptors + other conserved <i>Grs</i> | 0 | 9 | |
| | | Sugar receptors | 0 | 8 | |
| <i>H. melpomene</i> | <i>Ors</i> [‡] | <i>Heliconius</i> -specific | 7 | 24 | 0.527 |
| | | Conserved <i>Ors</i> | 5 | 29 | |
| <i>H. cydno</i> | <i>Ors</i> [‡] | <i>Heliconius</i> -specific | 6 | 25 | 0.0475 |
| | | Conserved <i>Ors</i> | 1 | 33 | |

*Consists of single-copy genes in *H. melpomene*; in the monarch or *Bombyx* genomes, homologues are either single-copy or duplicate genes with bootstrap support $\geq 80\%$.

[†]Fisher's exact test, two-tailed.

[‡]Excludes 3 *Ors* where read-mapping of the reference genome reads back to the reference assembly indicated areas of poor assembly: *Or37*, *Or39* and *Or49*.

[§]Excludes 3 *Ors* where read-mapping of the reference genome reads back to the reference assembly indicated areas of poor assembly: *Or20*, *Or24*, *Or43*, *Or50* and *Or74*. doi:10.1371/journal.pgen.1003620.t001

Widespread expression of *Ors* in *H. melpomene* antennae, proboscis and labial palps and legs

In addition to the *Gr* gene expression described above, we examined *Or* expression in the three adult tissues. The expression of *Ors* in antennal tissue has been widely studied in a variety of insects including *Drosophila* and some Lepidoptera [50,60]. As expected, we observed that *Or* gene expression was high in the antennae. Unexpectedly, *Or* expression was about as prevalent as *Gr* expression in the proboscis and labial palps and leg transcriptomes (Figure 9D, E, F). In total across all three tissues profiled, we found evidence for the expression of nearly all predicted *Or* genes (67 of 70 genes)(Table S12, S13, S14) in the *H. melpomene* reference genome [13].

Discussion

Outside *Drosophila*, the study of sensory gene family evolution in insects has generally been limited to the comparison of a small number of phylogenetically distant reference genomes. Such studies have commonly involved a comparison of the size of gene families between taxa in order to document lineage-specific expansions (Figure 10), and the comparison of dN/dS ratios to identify branches subject to rapid evolution [61]. Here we have used a similar approach to annotate 73 *Grs* in the *Heliconius melpomene* reference genome. However, we have also demonstrated the power of next-generation sequencing to elucidate patterns of evolution and expression of these genes. These data have offered exciting new insights into a set of genes that show both rapid evolution and sex-specific expression patterns, suggesting that female oviposition behaviour drives the evolution of butterfly gustatory receptors.

Previous work in other insects indicates that *Grs* are an important target for gene duplication and loss between species. Most notably, *D. sechellia* and *D. erecta* are host specialists, on *Morinda citrifolia* and *Pandanus candelabrum* respectively, while *D. simulans* is a generalist fly exploiting a broad array of rotting fruit [46]. Host specialization in the former species is associated with an acceleration of gene loss and increased rates of amino acid evolution at receptors that remain intact. Here we have used whole-genome Illumina sequencing of single diploid individuals to

similarly document patterns of gene gain and loss across *Heliconius*. This method yields highly fragmented genome assemblies, but such assemblies have proven very informative, most notably for studying the evolution of the clade of single-exon bitter receptor genes. We identified three gene duplication events along the lineage leading to *Heliconius*, followed by eight independent instances of clade-specific pseudogenizations or losses of different members of the intronless *Grs*, *Gr22-26* and *Gr53*, within *Heliconius* and one instance within *Danaus plexippus* (Figure 5 and Figure 6). In both *Heliconius* and *Drosophila* gene gain and loss appear to primarily affect *Grs* that are presumed to respond to bitter compounds (Figure 3). To verify whether this pattern holds within the genus *Heliconius* for the remaining gene family members with more complex intron-exon structure will require better genome assemblies for multiple *Heliconius* species (Table S2).

These patterns of rapid gene gain and loss are mirrored by within-population variation in copy number. From 16 resequenced genomes for *H. melpomene* and its sister species *H. cydno*, we have shown that CNVs occur more commonly among the *Grs* than the *Ors* (Figure 7, 8, Table 1). Within the *Grs*, the bitter receptors of *H. melpomene* represent a class of genes that are both highly prone to lineage-specific duplication and commonly subject to population-level copy number variation. These putative bitter receptor genes are also more likely to show female-specific expression, especially in the legs, which suggests a role in insect-host chemical interactions (Table 2, Figure 3, Figure S1).

In human genomes, a tendency for CNV-rich areas to display higher dN/dS ratios and yield paralogous genes has been noted [62], along with an enrichment of CNVs in genes involved in immune function and in the senses (specifically in *Ors* which are unrelated to the insect *Ors*) [63,64]. It is also widely known that copy-number variation is an important source of disease-causing mutations in humans [64]. With the exception of insecticide resistance in insects [65,66], the spectrum of naturally-occurring copy-number variants is only just starting to be explored in *Drosophila* [67,68] and non-model systems. Our results demonstrate the great utility of high throughput sequencing to reveal the naturally-occurring spectrum of CNVs that underlie gene family expansions in non-model systems, in traits of ecological relevance.

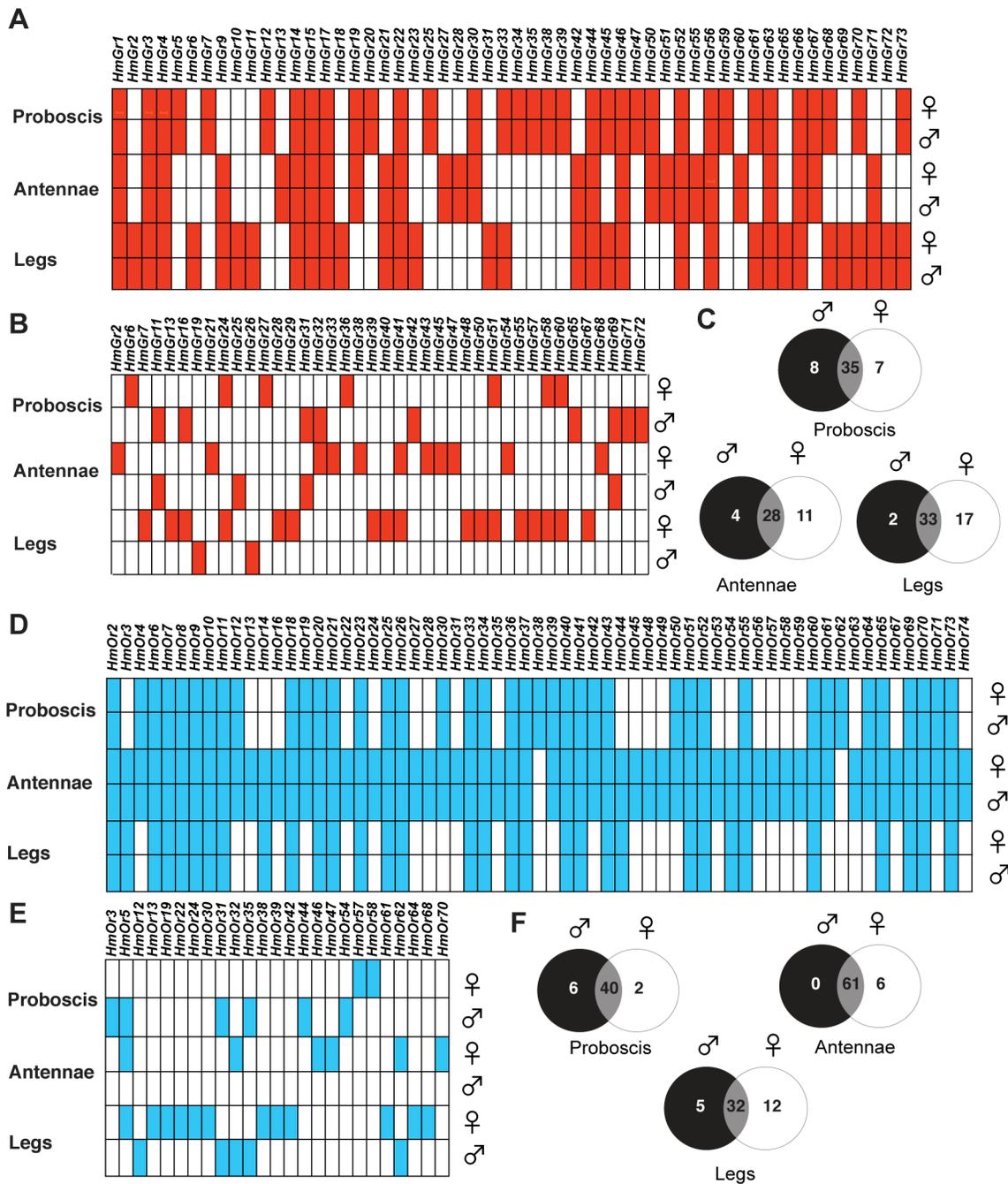


Figure 9. Comparison of *Grand Or* expression in male and female adult *H. melpomene* chemosensory tissues. (A) The common set of *Grs* expressed in each tissue in both males and females. Red box indicates the presence of reads uniquely mapping to the coding region of each *Gr* gene model. To facilitate the visualization of tissue-specific expression found in both males and females, only *Grs* where both sexes show expression are indicated. Where only one sex or neither sex shows expression, the box is empty. (B) *Grs* showing sex-specific expression. To facilitate the visualization of sex-specific *Grs*, only *Grs* where one sex shows expression are indicated by a filled box. *Grs* which are expressed in both sexes or no sex are indicated by an empty box. (C) Venn diagram showing the number of uniquely expressed gustatory receptors in each transcriptome. (D) The common set of *Ors* expressed in each tissue in both males and females. Blue box indicates the presence of reads uniquely mapping to the coding region of each *Or* gene model. As above, only *Ors* where both sexes show expression are indicated. Where only one sex or neither sex show expression, the box is empty. (E) *Ors* showing sex-specific expression are indicated by a filled box. *Ors* which are expressed in both sexes or no sex are indicated by an empty box. (F) Venn diagram showing the number of uniquely expressed gustatory receptors in each transcriptome. The proboscis libraries also included both labial palps, the antennal libraries included both antennae, and the leg libraries included all six legs. doi:10.1371/journal.pgen.1003620.g009

Heliconius butterflies have complex relationships with their *Passifloraceae* host plants. Some species are host-specialist, feeding on only one or a few *Passiflora* species, others specialise on

particular sub-genera within *Passiflora*, while others are generalists, albeit within this one host plant family (Figure 6) [53]. The *Passifloraceae* is extremely chemically diverse, most notably in

Table 2. An overabundance of *Gr*s expressed in female legs are the result of *Heliconius*-specific duplication.

| Gene Family | Gene duplication | Gene Expression | | |
|------------------------|-----------------------------|-----------------|------------|-----------------------------|
| | | Female-specific | Both sexes | <i>P</i> value [†] |
| <i>Gr</i> [‡] | <i>Heliconius</i> -specific | 15 | 20 | 0.019 |
| | Conserved* | 1 | 13 | |
| <i>Or</i> [§] | <i>Heliconius</i> -specific | 6 | 12 | 0.483 |
| | Conserved* | 5 | 19 | |

*Consists of single-copy genes in *H. melpomene*; in the monarch or *Bombyx* genomes, homologues are either single-copy or duplicate genes with bootstrap support $\geq 80\%$.

[†]Fisher's exact test, two-tailed, d.f. = 1.

[‡]Excludes *Gr39* because of poor coverage in the reference genome read-mapping.

[§]Excludes *Or20* and *Or24* because of coverage in the reference genome.

doi:10.1371/journal.pgen.1003620.t002

their diversity of cyanogenic glycosides that protect the plant from herbivores. It seems likely that coevolution of the butterfly chemosensory and detoxification system on the one hand, with the plant biochemical defense on the other, has played an important role in the evolution of this chemical arsenal. In contrast to the research already carried out on the chemistry of the host plants [54], until recently almost nothing was known about the chemosensory system of *Heliconius* butterflies. All of these insect host-plant interactions are mediated primarily by adult female butterflies, which must correctly identify suitable host plants for oviposition [69,70], or risk the survival of their offspring.

Expression data for *Gr*s in the Lepidoptera have been limited until now—especially for adults—due to their low expression level. The largest previous study identified 14 *Gr*s profiled in larval *B. mori* [32]. We have found evidence for adult expression for most (~91%) of the 73 predicted *Gr* genes. This provides a marked contrast to the handful of gustatory receptors that have been identified from traditional expressed sequence tag (EST) projects

in other Lepidoptera. Our methods may provide a greatly improved yield of expressed genes because we now have a set of well-annotated target *Gr* genes against which RNA-seq data can be mapped, together with a greater diversity of transcripts afforded by deep sequencing. Such methods have also permitted us to find widespread expression of their sister gene family, the *Or*s, in the adult chemosensory tissues examined (68 of 70 or 97% of predicted genes) (Figure 9).

Many of these *Gr* genes are likely to be involved in the detection of host plant attractants as well as toxic secondary metabolites and thus allow the discrimination of suitable hosts. Most notably, there were a large number of *Heliconius*-specific *Gr*s with female-biased expression in both legs and antennae (Figure 9). As mentioned previously, these female-biased leg *Gr*s (but not *Or*s) are also more likely to represent unique duplicates on the *Heliconius* lineage (Table 2). Female-biased *Or* expression, as quantified using RNA-seq data, has been reported for *Or*s expressed in the antennae of the adult mosquito, *Anopheles gambiae* [71]. Specifically, 22 *Or*s displayed enhanced expression in mosquito female antennae but not in male antennae. Since adult mosquito females but not males need to find hosts for a blood-meal, and adult butterfly females but not males need to find host plants for egg-laying, this suggests that host-seeking behaviour of female insects may be an important general driver of sensory gene evolution. Indirect evidence for the possible role of some of these *Gr*s in *Heliconius* host plant detection comes from comparative studies of *Gr*s mediating oviposition behaviour in swallowtail butterflies (Papilionidae). *Papilio xuthus* PxGr1 a member of the *Gr* subgroup that contains *D. melanogaster* Gr43a and HmGr9, has been characterized as a receptor for synephrine, which is an alkaloid found in citrus trees [52]. It is expressed in female *P. xuthus* tarsi and is necessary for the correct oviposition behavior of swallowtail butterflies [52]. Within the two clades most closely-related to PxGr1, are 9 butterfly-specific *Gr*s: HmGr10, Gr16, Gr55, Gr56 and Gr57, and the newly-described DpGr16, Gr50, Gr52, and Gr54 (Figure 3). Four these *Gr*s, HmGr16, Gr55, Gr56 and Gr57, result from *Heliconius*-specific gene duplications (i.e., no *Danaus* or *Bombyx* homologs). Gr55-57 are also in the top ten most highly expressed *Gr*s in female legs. The identification of these sex-biased leg *Gr*s has provided an important starting point

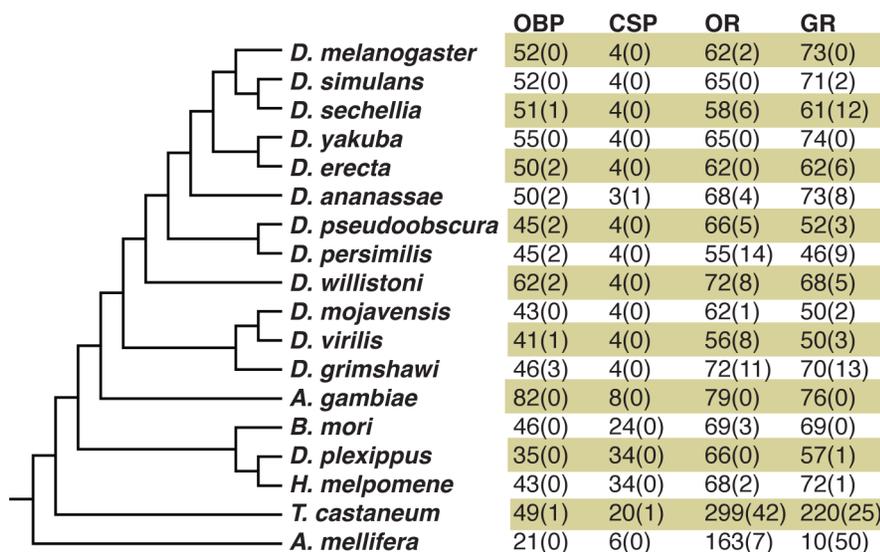


Figure 10. Insect chemosensory gene family repertoires. Numbers indicate intact genes and numbers in parentheses indicate pseudogenes. References are given in [13,55,94]. OBP = odorant binding protein; CSP = chemosensory protein; OR = olfactory receptor; GR = gustatory receptor. doi:10.1371/journal.pgen.1003620.g010

for future ligand specificity studies combining heterologous expression, electrophysiology, RNAi [51], assays of the proboscis-extension reflex, and female oviposition behavior.

Lastly, the patterns of *Gr* gene expression among different tissues and sexes has permitted us to identify a number of *Gr*s that are strong candidates for mediating the remarkable pollen feeding behaviour that is unique to *Heliconius*, among the butterflies. The *Heliconius* proboscis contains at least two types of gustatory sensilla, hair-like *sensilla chaetica*, and *sensilla styloconica* (Figure 1). Like other butterflies, *Heliconius* respond to varying amounts of sugars including sucrose present in floral nectar [72]. Unlike other moths and butterflies, *Heliconius* actively collect pollen with their proboscides, preferentially from *Psychotria* (Rubiaceae), *Psiguria/Gurania* (Cucurbitaceae) and *Lantana* (Verbenaceae) flowers [17,18,73]. Once a pollen load is collected (Figure 1D), the butterflies use a combination of mechanical shearing (coiling and uncoiling of the proboscis) and enzymatic activity (using proteases found in saliva) to release amino acids from the pollen [74]. The RNA-seq data we have collected for *H. melpomene* proboscis and labial palps should provide a useful resource for future studies examining the molecular basis of this unique digestive trait.

Pollen feeding in adult *Heliconius* has an important ecological function. Amino acids obtained from pollen are key resources used in male nuptial gifts and egg allocation [18,75–77]. They also permit *Heliconius* adults to have exceptionally long lifespans. Pollen feeding behavior is not found outside the genus *Heliconius*, even in the sister genus *Eueides*, whose larvae share a preference for *Passiflora* host-plants with *Heliconius*. In the present study we have identified four *Heliconius*-specific *Gr*s that are only expressed in the proboscis (*HmGr12*, *Gr20*, *Gr35* and *Gr59*) but not in antennae or legs (Figure 9B), suggesting a role for these genes in pollen-feeding behaviour.

Taken together, the whole-genome and whole-transcriptome data suggest that *Gr* genes in particular are highly evolutionarily labile both on short and long evolutionary timescales, and begin to offer an insight into the likely molecular basis for the rapid coevolution observed between these butterflies and their host plants. Understanding the remarkable diversity underlying this ecological interaction at a molecular level has remained a challenge (but see [32,52,78,79]). Thanks to technological innovations in sequencing, the genetic basis of taste and olfaction involved in host-plant adaptation in *Heliconius* is beginning to be uncovered.

Conclusions

We have shown that like the opsin visual receptors [80], the chemosensory superfamily composed of constituent *Gr* and *Or* families in Lepidoptera show rapid gene family evolution, with higher rates of copy-number variation and gene duplication among the *Gr*s than the *Or*s, as well as gene losses in the *Gr*s. In particular, there is a group of putative bitter receptors that show female-specific expression in the legs and that are especially prone to gene duplication, providing new material for sensory diversification in the insect-host plant arms race. We have also shown, for the first time, widespread expression of *Or*s in non-antennal tissues in a lepidopteran. With the most comprehensive data set on *Gr* and *Or* expression in butterflies to date we are one step closer to identifying the sensory and molecular genetic basis of the *Heliconius*-*Passiflora* co-evolutionary race that inspired Ehrlich and Raven in 1964.

Materials and Methods

Genome annotation

tBLASTn searches were conducted iteratively against the *H. melpomene melpomene* genome (version v1.1) and haplotype scaffolds

[13] using *B. mori* [28,47] and *D. plexippus Grs* [14] as input sequences. For these *in silico* gene predictions, intron-exon boundaries were identified by first translating the scaffold nucleotides in MEGA version 5 [81], searching for exons identified in the tBLASTn searches, then back translating to identify splice junctions. Intron sequences were then excised to verify that the remaining exonic sequences formed an in-frame coding sequence. Insect *Gr*s are defined by a conserved C-terminal motif TYhHHHhQF, where ‘h’ is any hydrophobic amino acid [21]. We inspected our predicted protein sequences for this motif or variants thereof, specifically ‘S’, ‘M’ or ‘K’ instead of a ‘T’ or ‘L’, ‘T’ or ‘I’ instead of ‘F’. In the handful of cases where we were unable to find the last short exon that contains this motif, final assignment to the *Gr* gene family was based on using the predicted amino acid sequence as a search string for either tBLASTn or BLASTp against the nr/nt Genbank database. Gene annotations were submitted to the EnsemblMetazoa database http://metazoa.ensembl.org/Heliconius_melpomene/Info/Index as part of the *H. melpomene* v. 2 genome release (for GeneIDs see Table S1). Chromosomal assignments were based on published mapping of scaffolds in the *H. melpomene melpomene* reference genome [13].

Following amino acid alignment using ClustalW, preliminary phylogenetic trees were constructed in MEGA using neighbor-joining and pair-wise deletion to identify orthologous relationships with *B. mori* and *D. plexippus Grs*. Reciprocal tBLASTn searches against the *B. mori* and *D. plexippus* genomes as well as searches using the protein2genome module in EXONERATE [82] were then performed in order to search for ‘missing’ *Gr*s in those genomes. Final phylogenetic analysis was performed using a maximum-likelihood (ML) algorithm and JTT model on an amino acid alignment that was inspected by eye and manually adjusted. These results were compared to a ML tree made from a Clustal-Omega alignment [83] and were found to be nearly identical. Once the initial *H. melpomene Gr* gene predictions were obtained, EXONERATE, Perl scripts and manual annotations in Apollo [84] were used to produce gff3 files for submission of the annotated *H. melpomene* genome scaffolds to EMBL-EBI.

RNA-sequencing

Butterfly pupae of *H. melpomene rosina* were obtained from Suministros Entomológicos Costarricenses, S.A., Costa Rica. Adult males and females were sexed and frozen at -80°C . Total RNAs were extracted separately from antennae, proboscis together with labial palps, and all six legs of three males and three females of *H. melpomene* using Trizol (Life Technologies, Grand Island, NY). A NucleoSpin RNA II kit (Macherey-Nagel, Bethlehem, PA) was used to purify total RNAs. Each total RNA sample was purified through one NucleoSpin RNA II column. Purified total RNA samples were quantified using a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY). The quality of the RNA samples was checked using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). 0.3–4.0 μg of purified total RNAs were used to make cDNA libraries. A TruSeq RNA sample prep kit (Illumina, San Diego, CA) was used to prepare 18 individual cDNA libraries. After being normalized according to their concentrations, the enriched individual libraries were pooled and then run on a 2% agarose gel. cDNA products ranging from 280 to 340 bp with an average of 310 bp were cut out and purified using a GeneClean III kit (MP Biomedicals, Solon, OH) to facilitate post-sequencing assembly. After being re-purified using Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics, Danvers, MA), the cDNA pool was quantified using the Qubit 2.0 Fluorometer, and quality control-checked using the Agilent Bioanalyzer 2100. The cDNA pools were then normalized

to 10 nM and run as either two paired-end or three single-end 100 bp runs on a HiSeq 2000 (Illumina, San Diego, CA) by the UCI Genomics High-Throughput Facility.

RNA-seq assembly and read mapping

mRNA sequences were demultiplexed, trimmed and sorted using Python and Perl scripts. A single *de novo* assembly of the combined libraries was performed using CLC Genomics Workbench 5 to check for missing exons in our gene models. The 73 corrected *Gr* gene models and 70 *Or* gene models were then used as an alignment reference to perform unique read mapping of each individual chemosensory transcriptome. To determine if an individual *Gr* or *Or* was expressed in a given tissue, each of the 1716 individual *Gr* and *Or* mapping alignments was inspected by eye for uniquely mapped reads, and any spuriously-mapped reads (i.e., reads <70 bp in length with indels or sequence mismatches at the ends) were discarded. As a control for potential differences in RNA preparation between samples, we also quantified the number of uniquely mapped fragments to the widely-expressed *elongation factor 1-alpha (EF1 α)* gene transcript and calculated the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) [85]. Illumina reads for each of the libraries were deposited as fastq files in the ArrayExpress archive under the accession number: E-TAB-1500 (Table S7).

Scanning electron microscopy

One week old adult *H. melpomene rosina* butterflies were sexed, frozen at -80°C , then dissected and mounted for imaging on an FEI/Philips XL30 FEG scanning electron microscope at UCI's Materials Characterization User Facility. Forelegs, middle legs, hindlegs and antennae were examined for the presence of gustatory sensilla.

Copy number variation analysis

We also examined resequenced genomes of twelve *H. melpomene* and four *H. cydno* individuals, including *H. melpomene aglaope*, *H. melpomene amaryllis* and *H. melpomene rosina* (Table S4), sequenced by The GenePool, University of Edinburgh, U.K. and the FAS Center, Harvard University, U.S.A., for evidence of copy-number variation (CNV) in the *Grs* and *Ors* using CNVnator [56]. These sequences were deposited in the European Nucleotide Archive (ENA) under accession number: ERP002440. The Illumina resequenced genomes were first mapped to the *H. melpomene* reference genome and the average read depth was calculated along a 100 bp sliding window. The output of CNVnator was parsed for candidate insertion and deletion variants, and those with estimated copy number of $>2\times$ were counted as potential duplications and $<0.5\times$ as potential deletions.

Whole-genome sequence assembly

The GenePool, University of Edinburgh, and the Oxford Genomics Centre, University of Oxford, U.K., produced whole genome 100 bp sequences from *H. cydno*, *H. timareta*, *H. wallacei*, *H. doris*, *H. clysonymus*, *H. telesiphe*, *H. erato petiverana*, *H. sara* and *H. sapho* using the Illumina Pipeline v. 1.5–1.7 with insert sizes ranging from 300 to 400 bp. We deposited sequences for *H. sapho* and *H. sara* in the Sequence Read Archive (SRA) under accession number ERP002444. We performed *de novo* assembly of the short reads using Abyss v. 1.2 [86] implemented in parallel at the School of Life Sciences, University of Cambridge, U.K. Based on previous results [87], recommendations estimated by the software, and comparison of N50 values in preliminary experiments, we chose a k-mer size of 31, a minimum number of pairs required $n = 5$ and

the minimum mean k-mer coverage of a unitig $c = 2$ (full command: `abyss-pe n = 5 k = 31 c = 2 in = 'for.fastq rev.fastq'`). In all assemblies, at least 96% of reads mapped back to the contigs. We created BLAST databases of these whole genome sequence assembly contigs (Table S2) in Geneious Pro v. 5.5.6. The lack of introns in the putative bitter receptor genes *Gr22-26* and *Gr53* permitted us to easily retrieve them from these BLAST databases. To confirm the identity and improve the quality of the sequences found, we mapped the reads to the assembled exon sequences in CLC Genomics Workbench v. 5.5.1, using the following conservative settings to prevent mis-mapping of paralogous sequences: mismatch, insertion and deletion cost of 3; length fraction and similarity fraction of 0.9. We then inspected all read-mappings by eye. Because the intronless *Grs* are closely related, we aligned the translated nucleotide sequences in MEGA using the ClustalW algorithm, and also inspected the alignment by eye. For all intronless *Gr* sequences except for the pseudogenes, sequence length was highly conserved (i.e., there were few indels). To illustrate the high substitution rate of the retrieved pseudogene sequences, we selected the neighbor-joining method for tree reconstruction and performed 500 bootstrap replicates.

Inferring gene duplications and losses

To infer the number of intronless *Gr* gene duplications and losses, we used the program Notung v. 2.6 [88,89], which reconciles gene trees onto the species tree. The gene tree was made by a maximum likelihood analysis of 1074 nucleotide sites, aligned by Clustal-Omega, and 500 bootstrap replications. The species tree was derived from a phylogeny based on independent nuclear and mitochondrial DNA sequences [90].

RT-PCR

We verified the presence of *HmGr22* in several adult tissues using reverse-transcriptase PCR and primers for *HmGr22* (5'-CCATAATTTTGTTCATCCT-3' and 5'-GATTTGAAA-TAAGGTCTGT-3') and *EF1alpha* (5'-CGTTTTGAGGAAAT-CAAGAAGG-3' and 5'-GACATCTTGTAAGGGAAGACG-CAG 3'). RNA was extracted from fresh frozen specimens using Trizol and purified using the Nucleospin RNA II kit, which contains a DNAase-treatment step. RNA concentration was diluted to 12.5 $\mu\text{g}/\text{ml}$. Each 25 μl reaction had 2.5 μl 10 \times BD Advantage 2 PCR buffer, 2.5 μl dNTPs (2 mM), 0.5 μl (100 μM) forward and 0.5 μl reverse primer, 0.5 μl (1:20 diluted) Stratagene Affinity Script Reverse Transcriptase, 0.5 μl 50 \times Advantage 2 Polymerase Mix, 17 μl H₂O and 1 μl RNA. The PCR reaction consisted of 38 cycles of 95 $^{\circ}\text{C}$ for 30 s, 55 $^{\circ}\text{C}$ for 30 s, and 68 $^{\circ}\text{C}$ for 55 s. The identity of the RT-PCR products was confirmed by Sanger sequencing.

Supporting Information

Figure S1 “For Bitter or Worse: A Tale of Sexual Dimorphism and Good Taste”, an original cartoon by author and illustrator of science-oriented comics, Jay S. Hosler.

(PDF)

Table S1 *Heliconius melpomene* genome gustatory receptor annotations. Gene name, EnsemblMetazoa GeneID, amino acid sequence, nucleotide sequence, number of exons, top BLAST hit. (XLS)

Table S2 Whole genome Illumina sequencing *de novo* assembly statistics.

(DOC)

Table S3 Intronless gustatory receptor genes retrieved from whole-genome Illumina assemblies.
(DOC)

Table S4 CNV sample data and whole-genome resequencing statistics.
(DOC)

Table S5 CNVs in *H. melpomene* and *H. cydno* gustatory receptors.
(XLS)

Table S6 CNVs in *H. melpomene* and *H. cydno* olfactory receptors.
(XLS)

Table S7 List of specimens and localities used in RNA-seq.
(DOC)

Table S8 Number of 100 bp Illumina reads sequenced per RNA-seq library.
(DOC)

Table S9 Gustatory receptor mRNAs expressed in adult *H. melpomene* legs.
(DOC)

Table S10 Gustatory receptor mRNAs expressed in adult *H. melpomene* antennae.
(DOC)

Table S11 Gustatory receptor mRNAs expressed in adult *H. melpomene* labial palps and proboscis.
(DOC)

Table S12 Olfactory receptor mRNAs expressed in adult *H. melpomene* antennae.
(DOC)

Table S13 Olfactory receptor mRNAs expressed in adult *H. melpomene* legs.
(DOC)

Table S14 Olfactory receptor mRNAs expressed in adult *H. melpomene* proboscis and labial palps.
(DOC)

Text S1 Identification of *H. melpomene* homologs of all described insect Gr subfamilies.
(DOC)

Acknowledgments

Thanks to Jenny Barna for remote computing support; Harald Krenn and Carla Penz for permission to use the proboscis SEMs; Francesca Frentiu, Kyle McCulloch, Arnaud Martin and Susan Finkbeiner for feedback on the manuscript; Dan Hughes and Peter Andolfatto for bioinformatics assistance.

Author Contributions

Conceived and designed the experiments: ADB CDJ. Performed the experiments: ADB AMM FY. Analyzed the data: ADB AMM KMK JRW GAJ SHM EJJ CDJ. Wrote the paper: ADB EJJ CDJ. Sequenced the *H. melpomene* aglaope and *H. melpomene* amaryllis genomes: KKD JM. Sequenced the *H. sara* and *H. sapho* genomes: LCF.

References

- Ehrlich PR, Raven PH (1964) Butterflies and plants: A study in coevolution. *Evolution* 18: 586–608.
- Dethier VG (1937) Gustation and olfaction in lepidopterous larvae. *Biol Bull* 72: 7–23.
- Schneider D (1964) Insect antennae. *Annu Rev Entomol* 9: 103–122.
- Schoonhoven LM, Dethier VM (1966) Sensory aspects of host-plant discrimination by lepidopterous larvae. *Arch Neerl Zool* 16: 497–530.
- Anderson AL (1932) The sensitivity of the legs of common butterflies to sugars. *J Exp Zool B Mol Dev Evol* 63: 235–259.
- Calatayud P-A, Chintawi M, Tauban D, Marion-Poll R, Rü BL, et al. (2006) Sexual dimorphism of antennal, tarsal and ovipositor chemosensilla in the African stemborer, *Busseola fusca* (Fuller) (Lepidoptera: Noctuidae). *An Soc Entomol Fr* 42: 403–412.
- Jørgensen K, Almaas TJ, Marion-Poll F, Mustaparta H (2007) Electrophysiological characterization of responses from gustatory receptor neurons of sensilla chaetica in the moth *Heliothis virescens*. *Chem Senses* 32: 863–879.
- Marion-Poll FC, Guillaumin D, Masson C (1992) Sexual dimorphism of tarsal receptors and sensory equipment of the ovipositor in the European corn borer, *Ostrinia nubilalis*. *Cell Tissue Res* 267: 507–518.
- Krenn HW, Penz CM (1998) Mouthparts of *Heliconius* butterflies (Lepidoptera: Nymphalidae): A search for anatomical adaptations to pollen feeding behavior. *Int J Insect Morphol Embryol* 27: 301–309.
- Renou M (1983) Les récepteurs gustatifs du tarse antérieur de la femelle d'*Heliconius charitonius* (Lep.: Heliconiidae). *Annls Soc Ent Fr (NS)* 19: 101–106.
- Chapman RF (2003) Contact chemoreception in feeding by phytophagous insects. *Annu Rev Entomol* 48: 455–484.
- Myers J (1969) Distribution of foodplant chemoreceptors on the female Florida Queen butterfly, *Danaus gilippus berenice* (Nymphalidae). *J Lepid Soc* 23: 196–198.
- Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94–98.
- Zhan S, Merlin C, Boore JL, Reppert SM (2011) The monarch butterfly genome yields insights into long-distance migration. *Cell* 147: 1171–1185.
- International Silkworm Genome Consortium (2008) The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* 38: 1036–1045.
- Briscoe AD, Bybee SM, Bernard GD, Yuan F, Sison-Mangus MP, et al. (2010) Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. *Proc Natl Acad Sci USA* 107: 3628–3633.
- Boggs CL, Smiley JT, Gilbert LE (1981) Patterns of pollen exploitation by *Heliconius* butterflies. *Oecologia* 48: 284–289.
- Gilbert LE (1972) Pollen feeding and reproductive biology of *Heliconius* butterflies. *Proc Natl Acad Sci USA* 69: 1403–1407.
- Engler-Chaouat HS, Gilbert LE (2007) *De novo* synthesis vs. sequestration: Negatively correlated metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. *J Chem Ecol* 33: 25–42.
- Benson WW, Brown KS, Gilbert LE (1975) Coevolution of plants and herbivores: Passion flower butterflies. *Evolution* 29: 659–680.
- Clyne PJ, Warr CG, Carlson JR (2000) Candidate taste receptors in *Drosophila*. *Science* 287: 1830–1834.
- Dunipace L, Meister S, McNealy C, Amrein H (2001) Spatially restricted expression of candidate taste receptors in the *Drosophila* gustatory system. *Curr Biol* 11: 822–835.
- Scott K, Brady R, Jr., Cravchik A, Morozov P, Rzhetsky A, et al. (2001) A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell* 104: 661–673.
- Robertson HM, Warr CG, Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 100: 14537–14542.
- Lee Y, Kang MJ, Shim J, Cheong CU, Moon SJ, et al. (2012) Gustatory receptors required for avoiding the insecticide L-canavanine. *J Neurosci* 32: 1429–1435.
- Jones WD, Cayirlioglu P, Kadow IG, Voshall LB (2007) Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* 445: 86–90.
- Montell C (2009) A taste of the *Drosophila* gustatory receptors. *Curr Opin Neurobiol* 19: 345–353.
- Zhang HJ, Anderson AR, Trowell SC, Luo AR, Xiang ZH, et al. (2011) Topological and functional characterization of an insect gustatory receptor. *PLoS ONE* 6: e24111.
- Ishimoto H, Takahashi K, Ueda R, Tanimura T (2005) G-protein gamma subunit 1 is required for sugar reception in *Drosophila*. *EMBO J* 24: 3259–3265.
- Kain P, Badsha F, Hussain SM, Nair A, Hasan G, et al. (2010) Mutants in phospholipid signaling attenuate the behavioral response of adult *Drosophila* to trehalose. *Chem Senses* 35: 663–673.
- Yao CA, Carlson JR (2010) Role of G-proteins in odor-sensing and CO₂-sensing neurons in *Drosophila*. *J Neurosci* 30: 4562–4572.
- Sato K, Tanaka K, Touhara K (2011) Sugar-regulated cation channel formed by an insect gustatory receptor. *Proc Natl Acad Sci USA* 108: 11680–11685.
- Dahanukar A, Foster K, van der Goes van Naters WM, Carlson JR (2001) A Gr receptor is required for response to the sugar trehalose in taste neurons of *Drosophila*. *Nat Neurosci* 4: 1182–1186.
- Chyb S, Dahanukar A, Wickens A, Carlson JR (2003) *Drosophila Gr5a* encodes a taste receptor tuned to trehalose. *Proc Natl Acad Sci USA* 100 Suppl 2: 14526–14530.
- Slone J, Daniels J, Amrein H (2007) Sugar receptors in *Drosophila*. *Curr Biol* 17: 1809–1816.

36. Kwon JY, Dahanukar A, Weiss LA, Carlson JR (2007) The molecular basis of CO₂ reception in *Drosophila*. *Proc Natl Acad Sci USA* 104: 3574–3578.
37. Moon SJ, Kottgen M, Jiao Y, Xu H, Montell C (2006) A taste receptor required for the caffeine response *in vivo*. *Curr Biol* 16: 1812–1817.
38. Lee Y, Moon SJ, Montell C (2009) Multiple gustatory receptors required for the caffeine response in *Drosophila*. *Proc Natl Acad Sci USA* 106: 4495–4500.
39. Weiss LA, Dahanukar A, Kwon JY, Banerjee D, Carlson JR (2011) The molecular and cellular basis of bitter taste in *Drosophila*. *Neuron* 69: 258–272.
40. Smadja C, Shi P, Butlin RK, Robertson HM (2009) Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Mol Biol Evol* 26: 2073–2086.
41. Robertson HM, Wanner KW (2006) The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res* 16: 1395–1403.
42. Abdel-Latif M (2007) A family of chemoreceptors in *Tribolium castaneum* (Tenebrionidae: Coleoptera). *PLoS ONE* 2: e1319.
43. Kent LB, Walden KK, Robertson HM (2008) The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. *Chem Senses* 33: 79–93.
44. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, et al. (2002) G protein-coupled receptors in *Anopheles gambiae*. *Science* 298: 176–178.
45. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.
46. McBride CS, Arguello JR (2007) Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177: 1395–1416.
47. Wanner KW, Robertson HM (2008) The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. *Insect Mol Biol* 17:621–629.
48. Jacquin-Joly E, Legeai F, Montagne N, Monsempe C, Francois MC, et al. (2012) Candidate chemosensory genes in female antennae of the noctuid moth *Spodoptera littoralis*. *Int J Biol Sci* 8: 1036–1050.
49. Krieger J, Raming K, Dewar YM, Bette S, Conzelmann S, et al. (2002) A divergent gene family encoding candidate olfactory receptors of the moth *Heliothis virescens*. *Eur J Neurosci* 16: 619–628.
50. Grosse-Wilde E, Kuebler LS, Bucks S, Vogel H, Wicher D, et al. (2011) Antennal transcriptome of *Manduca sexta*. *Proc Natl Acad Sci USA* 108: 7449–7454.
51. Howlett N, Dauber KL, Shukla A, Morton B, Glendinning JI, et al. (2012) Identification of chemosensory receptor genes in *Manduca sexta* and knockdown by RNA interference. *BMC Genomics* 13: 211.
52. Ozaki K, Ryuda M, Yamada A, Utoguchi A, Ishimoto H, et al. (2011) A gustatory receptor involved in host plant recognition for oviposition of a swallowtail butterfly. *Nat Commun* 2: 542.
53. Brown KS (1981) The biology of *Heliconius* and related genera. *Annu Rev Entomol* 26: 427–456.
54. Spencer KC (1988) Chemical mediation of coevolution in the *Passiflora-Heliconius* interaction. In: Spencer KC, editor. *Chemical mediation of coevolution*. London: Academic Press. pp. 167–240.
55. Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG (2008) *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol Ecol* 17: 1648–1657.
56. Abyzov A, Urban AE, Snyder M, Gerstein M (2011) CNVator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21: 974–984.
57. Omura H, Honda K, Asaoka K, Inoue TA (2011) Divergent behavioral and electrophysiological taste responses in the mid-legs of adult butterflies, *Vanessa indica* and *Argyreus hyperbius*. *J Insect Physiol* 57: 118–126.
58. Fox RM (1966) Forelegs of butterflies. I. Introduction: chemoreception. *J Res Lepid* 5: 1–12.
59. Thom C, Guerenstein PG, Mechaber WL, Hildebrand JG (2004) Floral CO₂ reveals flower profitability to moths. *J Chem Ecol* 30: 1285–1288.
60. Liu Y, Gu S, Zhang Y, Guo Y, Wang G (2012) Candidate olfaction genes identified within the *Helicoverpa armiga* antennae transcriptome. *PLoS One* 7: e48260.
61. Lavagnino N, Serra F, Arbiza L, Dopazo H, Hasson E (2012) Evolutionary genomics of genes involved in olfactory behavior in the *Drosophila melanogaster* species subgroup. *Evol Bioinform Online* 8: 89–104.
62. Schuster-Böckler B, Conrad D, Bateman A (2010) Dosage sensitivity shapes the evolution of copy-number varied regions. *PLoS One* 5: e9474.
63. Benton R, Sachse S, Michnick SW, Vosshall LB (2006) Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors *in vivo*. *PLoS Biol* 4: e20.
64. Cooper GM, Nickerson DA, Eichler EE (2007) Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* 39: S22–S29.
65. Puinean AM, Foster SP, Opiphant L, Denholm I, Field LM, et al. (2010) Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLoS Genet* 6: e1000999.
66. Bariami V, Jones CM, Poupardin R, Vontas J, Ranson H (2012) Gene amplification, ABC transporters and cytochrome P450s: Unraveling the molecular basis of pyrethroid resistance in the dengue vector, *Aedes aegypti*. *PLoS Negl Trop Dis* 6: e1692.
67. Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, et al. (2010) Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS Genet* 6: e1000998.
68. Schrider DR, Stevens K, Cardeno CM, Langley CH, Hahn MW (2011) Genome-wide analysis of regurgene polymorphism in *Drosophila melanogaster*. *Genome Res* 21: 2087–2095.
69. Baur R, Haribal M, Renwick JAA, Städler E (1998) Contact chemoreception related to host selection and oviposition behavior in the monarch butterfly, *Danaus plexippus*. *Physiol Entomol* 23: 7–19.
70. Zalucki MP, Brower LP, Malcolm SB (1990) Oviposition by *Danaus plexippus* in relation to cardenolide content of three *Asclepias* species in the southeastern U.S.A. *Ecol Entomol* 15: 231–240.
71. Pitts RJ, Rinker DC, Jones PL, Rokas A, Zwiebel IJ (2011) Transcriptome profiling of chemosensory appendages in the malaria vector *Anopheles gambiae* reveals tissue- and sex-specific signatures of odor coding. *BMC Genomics* 12: 271.
72. Barp EA, Soares GLG, Giani EJM, Rodrigues D, Moreira GRP (2011) Variation in nectar and pollen availability, sucrose preference and daily response in the use of flowers by *Heliconius erato phyllis*. *J Insect Behav* 24: 200–219.
73. Estrada C, Jiggins CD (2008) Interspecific sexual attraction because of convergence in warning coloration: is there a conflict between natural and sexual selection in mimetic species? *J Evol Biol* 21: 749–760.
74. Krenn HW, Eberhard MJB, Eberhard SH, Hiki A-L, Huber W, et al. (2009) Mechanical damage to pollen aids nutrient acquisition in *Heliconius* butterflies (Nymphalidae). *Arthropod-Plant Interact* 3: 203–208.
75. Dunlap-Pianka H, Boggs CL, Gilbert LE (1977) Ovarian dynamics in Heliconiine butterflies: programmed senescence versus eternal youth. *Science* 197: 487–490.
76. Boggs CL, Gilbert LE (1979) Male contribution to egg production in butterflies: evidence for transfer of nutrients at mating. *Science* 206: 83–84.
77. Cardoso MZ, Gilbert LE (2007) A male gift to its partner? Cyanogenic glycosides in the spermatophore of long wing butterflies (*Heliconius*). *Naturwissenschaften* 94: 39–42.
78. Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P (2012) Parallel molecular evolution in an herbivore community. *Science* 337: 1634–1637.
79. Cohen MB, Schuler MA, Berenbaum MR (1992) A host-inducible cytochrome P450 from a host-specific caterpillar—molecular cloning and evolution. *Proc Natl Acad Sci USA* 89: 10920–10924.
80. Frentiu FD, Bernard GD, Sison-Mangus MP, Brower AV, Briscoe AD (2007) Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. *Mol Biol Evol* 24: 2016–2028.
81. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
82. Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.
83. Sievers F, Wilm A, Dineen D, Gibson TJ, Karpus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7: Article Number 539.
84. Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, et al. (2002) Apollo: a sequence annotation editor. *Genome Biol* 3: RESEARCH0082.
85. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28: 511–515.
86. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, et al. (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res* 19: 1117–1123.
87. Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, et al. (2012) GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Res* 22: 557–567.
88. Vernot B, Stolzer M, Goldman A, Durand D (2008) Reconciliation with non-binary species trees. *J Comput Biol* 15: 981–1006.
89. Durand D, Halldorsson BV, Vernot B (2006) A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J Comput Biol* 13: 320–335.
90. Beltran M, Jiggins CD, Brower AVZ, Bermingham E, Mallet J (2007) Do pollen feeding, pupal-mating and larval gregariousness have a single origin in *Heliconius* butterflies? Inferences from multilocus DNA sequence data. *Biol J Linn Soc Lond* 92: 221–239.
91. Wanner KW, Anderson AR, Trowell SC, Theilmann DA, Robertson HM, et al. (2007) Female-biased expression of odorant receptor genes in the adult antennae of the silkworm, *Bombyx mori*. *Insect Mol Biol* 16: 107–119.
92. Tanaka K, Uda Y, Ono Y, Tatsuura N, Suwa M, et al. (2009) Highly selective tuning of a silkworm olfactory receptor to a key mulberry leaf volatile. *Curr Biol* 19: 881–890.
93. Jordan MD, Anderson A, Begum D, Carraher C, Authier A, et al. (2009) Odorant receptors from the light brown apple moth (*Epiphyas postvittana*) recognize important volatile compounds produced by plants. *Chem Senses* 34: 383–394.
94. Sánchez-Gracia A, Vieira FG, Rozas J (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103: 208–216.