Widespread Gene Expression Divergence in Butterfly Sensory Tissues Plays a Fundamental Role During Reproductive Isolation and Speciation

Ningning Wu,[1] Elizabeth Evans,[2] Bas van Schooten,[2] Jesyka Meléndez-Rosa,[2] Yadira Ortiz,[2] Silvia M. Planas Soto-Navarro,[2,3] Steven M. Van Belleghem,[2,4] Brian A. Counterman,[5] Riccardo Papa,*[,†2,3] and Wei Zhang *[,†1,6]

[1]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, People's Republic of China

[2]Department of Biology, University of Puerto Rico, Rio Piedras, San Juan 00925, PR, USA

[3]Molecular Sciences and Research Center, University of Puerto Rico, San Juan 00907, PR, USA

[4]Ecology, Evolution and Conservation Biology, Biology Department, KU Leuven, Leuven, Belgium

[5]Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

[6]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, People's Republic of China

[†]These authors jointly supervised this project and contributed equally.

***Corresponding authors:** E-mail: wzhangvv@pku.edu.cn; rpapa.lab@gmail.com.

Associate editor: Fuwen Wei

Abstract

Neotropical *Heliconius* butterflies are well known for their intricate behaviors and multiple instances of incipient speciation. Chemosensing plays a fundamental role in the life history of these groups of butterflies and in the establishment of reproductive isolation. However, chemical communication involves synergistic sensory and accessory functions, and it remains challenging to investigate the molecular mechanisms underlying behavioral differences. Here, we examine the gene expression profiles and genomic divergence of three sensory tissues (antennae, legs, and mouthparts) between sexes (females and males) and life stages (different adult stages) in two hybridizing butterflies, *Heliconius melpomene* and *Heliconius cydno*. By integrating comparative transcriptomic and population genomic approaches, we found evidence of widespread gene expression divergence, supporting a crucial role of sensory tissues in the establishment of species barriers. We also show that sensory diversification increases in a manner consistent with evolutionary divergence based on comparison with the more distantly related species *Heliconius charithonia*. The findings of our study strongly support the unique chemosensory function of antennae in all three species, the importance of the Z chromosome in interspecific divergence, and the nonnegligible role of nonchemosensory genes in the divergence of chemosensory tissues. Collectively, our results provide a genome-wide illustration of diversification in the chemosensory system under incomplete reproductive isolation, revealing strong molecular separation in the early stage of speciation. Here, we provide a unique perspective and relevant view of the genetic architecture (sensory and accessory functions) of chemosensing beyond the classic chemosensory gene families, leading to a better understanding of the magnitude and complexity of molecular changes in sensory tissues that contribute to the establishment of reproductive isolation and speciation.

Key words: **speciation, chemosensation, hybridization, *Heliconius* butterflies, Lepidoptera.**

Introduction

The process of speciation involves divergence in morphology and physiology that influences reproductive isolation. Examples that have radically advanced our understanding of the genomic architecture of speciation include the beak sizes of Darwin's finches (Grant and Grant 2006; Lamichhaney, et al. 2015), the body shapes of sticklebacks (Conte and Schluter 2013; Bay, et al. 2017), and the wing patterns in *Heliconius* butterflies (Jiggins, et al. 2001; Kronforst, et al. 2006). Similarly, major advances have been made in the identification of the so-called "speciation genes", such as those that influence hybrid sterility in *Drosophila* and mice (Ting, et al. 1998; Barbash, et al. 2003; Mihola, et al. 2009; Phadnis and Orr 2009). Perhaps one of the most important advances has been provided by the burgeoning evidence that speciation with gene flow is not only theoretically possible (Gavrilets

Article

2004; van Doorn, et al. 2009) but potentially common across the tree of life and results in substantial evidence of hybridization in nature (Mallet 2005; Rieseberg 2009; Edelman and Mallet 2021). To understand the mechanisms driving speciation under incomplete reproductive isolation, we need to test specific hypotheses about the genetic architecture (i.e., types of genes and genetic changes) that influences reproductive isolation and patterns of genomic divergence in the early stages of speciation (Lawniczak, et al. 2010; Ellegren, et al. 2012; Staubach, et al. 2012; Li, et al. 2015; Martin, et al. 2019; Kautt, et al. 2020; Yamasaki, et al. 2020; Van Belleghem, et al. 2021).

The vast diversity (Seixas, et al. 2021), varying degrees of divergence, and rampant hybridization (Edelman, et al. 2019) observed among *Heliconius* butterflies make this genus a powerful system for addressing questions about speciation associated with gene flow (*Heliconius* Genome Consortium 2012; Kronforst, et al. 2013; Martin, et al. 2013, 2019 Nadeau, et al. 2013; Edelman, et al. 2019). *Heliconius* warning colors are considered one of the best examples of an adaptive radiation (Mallet and Dasmahapatra 2012; Supple, et al. 2015; Merot, et al. 2017). In addition, *Heliconius* species show pronounced physiological and behavioral features related to their sensory ecology, such as alkaloid sequestration from host plants, pupal mating and adult mating, diversification of pheromone compounds, and communal roosting (Merrill, et al. 2015; Jiggins 2017). Several studies have investigated the genes underlying such traits, such as those involved in chemosensation, vision, and mate preference, between *Heliconius* sister species (Schulz, et al. 2008; Briscoe, et al. 2013; van Schooten, et al. 2016, 2020; Darragh, et al. 2017, 2020, 2021; Byers, et al. 2020, 2021). While the importance of chemical ecology in the diversification of *Heliconius* butterflies seems evident from these studies, the role of chemical communication systems in *Heliconius* speciation remains largely unresolved. For example, the molecular architecture of chemical sensing in *Heliconius* has been investigated in only five gene families, including those encoding gustatory receptors, olfactory receptors, ionotropic receptors, olfactory binding proteins and chemosensory proteins, which represent less than 1% of the transcriptome (*Heliconius* Genome Consortium 2012; Briscoe, et al. 2013; van Schooten, et al. 2016, 2020). However, while these studies explored different aspects of chemosensory gene family evolution in *Heliconius*, a pattern of low genetic admixture during speciation has been confidently shown for only 13 genes (van Schooten, et al. 2020). For all of these reasons, given the synergy of sensory and accessory functions in chemosensing, a comprehensive view of whole-transcriptome diversification is needed to disentangle how this complexity contributes to the evolution of the chemosensory system and speciation.

Here, we investigated the evolution of chemosensing during the stage of speciation with incomplete reproductive isolation by exploring the total mRNA differences in sensory tissue in the hybridizing species pair *Heliconius melpomene* and *Heliconius cydno*, which began to diverge approximately 1.4 million years ago (Mya) and have already partially undergone prezygotic and postzygotic isolation. For example, these two species differ in host plant use, wing pattern, and habitat, and although they hybridize occasionally, F1 females are sterile and subject to Haldane's rule, while F1 males can back-cross with either species (Jiggins et al. 2001; Merrill, et al. 2011). Thus, this implies some level of genomic divergence at the level of the non-coding DNA. By integrating genome-wide and transcriptome-wide analyses, we detected the expression of 11,479 genes on average in each sensory tissue, approximately 87 of which belonged to chemosensory gene families. In particular, the *Heliconius* antennae showed markedly different gene expression profiles among the chemosensory organs, suggesting their distinct functions in chemosensing. Furthermore, we identified a greater degree of differential expression between life stages (different adult stages) than between sexes (females and males) in the chemosensory tissues of both species. To examine the robustness of these results, we collected specimens of the more distantly related butterfly species *Heliconius charithonia*, which diverged from the *melpomene-cydno* clade more than 10 Mya, and performed gene expression analyses. With these additional data, we confirmed (a) the uniqueness of the antenna tissue type and (b) the expected increase in the level of overall gene expression divergence when compared to *H. melpomene* and *H. cydno*. Overall, these results suggest that nonchemosensory genes play an indispensable role in shaping unique sensing functions and establishing life-stage differentiation. To further investigate the importance of the chemosensory organs in speciation, we explored the correlation between genome-wide genetic divergence (Martin, et al. 2019) and the expression of genes involved in chemosensing and found that the Z chromosome is a putative hotspot of low admixture, harboring a number of nonintrogressed genes that are differentially expressed in sensory tissues, based on comparisons of the two species. Beyond the Z chromosome, we demonstrated linkage disequilibrium (LD) between low-admixture genes in sensory tissues and color pattern-related molecules, which have been associated with mate preference and shown to be involved in speciation and adaptation (Jiggins, et al. 2017; Westerman, et al. 2018; Merrill, et al. 2019; Rossi, et al. 2020). In conclusion, we attempted to link classic chemosensory and novel genes in the expression networks of chemosensory tissues to better characterize their interactions. These results reinforce the existence of pronounced genome-wide divergence in the chemosensory system between these two butterflies with incomplete reproductive isolation and indicate a nonnegligible role of genes beyond the chemosensory gene families in both chemosensing and speciation.

Results

Transcriptome-wide Phylogeny and Clustering

The sister species *H. melpomene* and *H. cydno* show distinct ecological, morphological, and physiological characteristics.

They are known to hybridize, and genetic admixture has been shown to occur via the backcrossing of fertile F1 males (Naisbit, et al. 2002; Kronforst, et al. 2013; Martin, et al. 2013; Jiggins 2017). Both *H. melpomene* and *H. cydno* are polytypic species, and their geographic color pattern morphs form Müllerian mimicry rings with morphs of *H. erato* and *H. sapho*/*H. eleuchia*, respectively (Kronforst and Papa 2015) (fig. 1*A*). In contrast to *H. melpomene*, which is widely distributed in Central and South America, *H. cydno* is mainly found in Central America, where we focused on a sympatric species pair, *H. melpomene rosina* and *H. cydno chioneus*, and obtained RNA-seq datasets for these species from a previous study (van Schooten, et al. 2020). We conducted comparative transcriptome analyses of sensory tissues, including antennae, legs, and mouthparts (fig. 1*A* and *B*, and supplementary tables S1–S2, Supplementary Material online). To ensure the effectiveness of the gene expression analyses, we first estimated the alignment rates of RNA-seq reads from all samples, and the results showed that reads from all tissues were properly mapped to the reference genome of *H. melpomene* (Davey, et al. 2016), with an average alignment rate of 96.7% for *H. melpomene* samples and 96.9% for *H. cydno* samples (supplementary table S1, Supplementary Material online). The average mapping rates of different tissues were comparable, showing no significant differences (Scheffe's test, $P > 0.05$). As all the butterflies used in this study were reared under seminatural conditions in Panama, we first examined their genetic background by constructing a whole-transcriptome phylogeny, which yielded a well-resolved tree with two distinct clades for *H. m. rosina* and *H. c. chioneus*, where tissue samples from the same individuals were clustered as subgroups (fig. 1*C*). Principal component analysis (PCA) further indicated that all the gene expression profiles of different sensory tissues clustered according to species and tissue type (fig. 1*D*). Interestingly, the leg and mouthpart gene clusters of each species partially overlapped, indicating functional correlation between legs and mouthparts in contrast to antennae, which resulted in insufficient resolution to separate leg and mouthpart tissues. This supports the antennae as the primary sensory organs, as previously reported (Gadenne, et al. 2016; Elgar, et al. 2018), and indicates their primary role related to olfaction, while the roles of the legs and mouthparts are more likely related to gustation. We detected the expression of approximately 11,479 genes on average in each sensory tissue, accounting for over 50% of the total 21,661 annotated genes in the *H. melpomene* genome obtained from Lepbase (http://lepbase.org/). However, among the total expressed genes, only approximately 87 chemosensory genes (CSGs; belonging to five gene families, including those encoding olfactory binding proteins, chemosensory proteins, olfactory receptors, gustatory receptors, and ionotropic receptors) were expressed in each tissue on average (supplementary fig. S1 and supplementary table S2, Supplementary Material online). Notably, chemosensory tissues differed in the numbers of expressed CSGs in the two species. The expressed CSGs accounted for 1.06% of the antenna-expressed genes on average, which was significantly higher than the corresponding percentages in the legs (0.60% on average, $P < 0.01$) and mouthparts (0.60% on average, $P < 0.01$), including genes from five chemosensory gene families (supplementary table S2, Supplementary Material online). From the perspective of interspecific comparisons, 259 more total genes and 13 more CSGs were expressed in the chemosensory organs of *H. melpomene* than in those of *H. cydno* (supplementary table S2, Supplementary Material online). Taken together, our results suggest both tissue-specific (3.06%, 1.60%, and 1.08% of the total 21,661 annotated genes in the antennae, legs, and mouthparts of *H. cydno* and 2.59%, 1.55%, and 1.08% in the antennae, legs, and mouthparts of *H. melpomene*, respectively) and species-specific (5.84% in *H. cydno* and 3.91% in *H. melpomene*) patterns of gene expression in the *Heliconius* chemosensory system, where sensory and accessory functions jointly play an integral role in the viability and overall functions of chemosensory organs.

Interspecific Differential Expression and Genetic Divergence in Chemosensing

To investigate the interspecific divergence in chemosensing at both the genome and transcriptome levels, we first compared the in-depth transcription profiles of antenna, leg, and mouthpart tissues (fig. 2). Similar to the PCA results, the gene expression patterns in the antenna tissues were very different from those in the other tissues in both species. Interestingly, the clustering of gene expression patterns showed that the antenna expression patterns of the two species were more similar to each other than to the expression patterns of the other tissue types within the same species. In contrast, the leg and mouthpart tissues showed more similar gene expression patterns to each other within each species than to the pattern in the same tissue in the other species. Although two antenna samples were not clustered with the other antenna samples, they remained as outgroups in the topology of leg and mouthpart samples, which were similar to the other antenna samples in both species (fig. 2). These findings suggest extensive functional divergence between antennae and other chemosensory organs and functional convergence of the antennae in *H. cydno* and *H. melpomene* (fig. 2).

To further characterize the chemosensory system, we identified 484, 340, and 462 unigenes showing significantly different expression patterns between species in antenna, leg, and mouthpart tissues, respectively (supplementary fig. S2, Supplementary Material online, FDR-adjusted $P < 0.05$). Echoing the PCA results, the antennae yielded the most tissue-specific differentially expressed genes (DEGs) among the investigated chemosensory tissues in both species, including 664 genes in *H. cydno* and 561 genes in *H. melpomene* (supplementary fig. S2, Supplementary Material online). A total of 423 genes showed upregulated expression in the antenna tissues of both species and were associated with a variety of sensory-related GO terms, such as odorant binding, olfactory receptor activity, and sensory

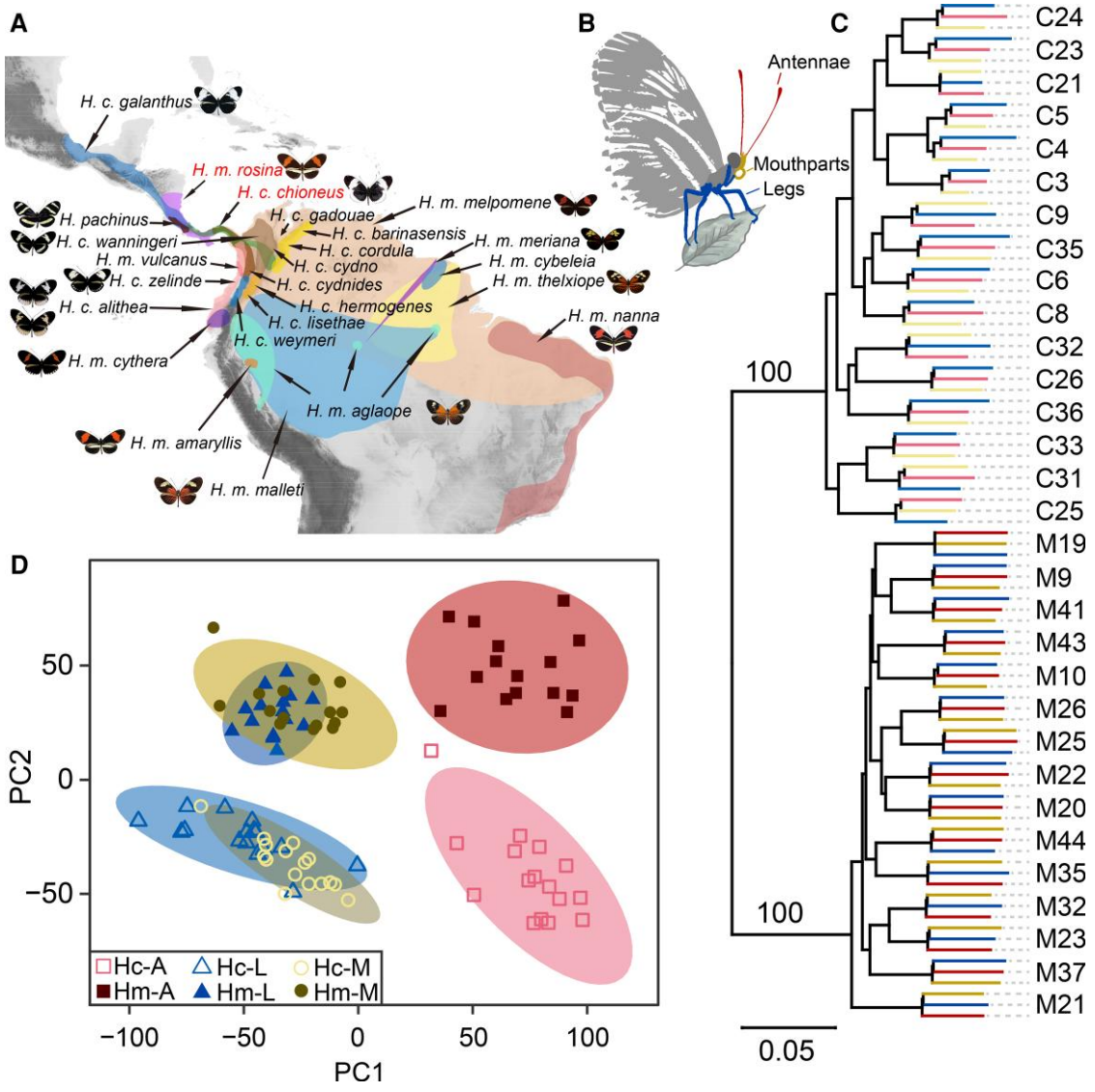


Fig. 1. The geographic distribution, transcriptome-wide phylogeny, and genetic structure of the *Heliconius* species pair analyzed in this study. (*A*) The geographic distributions of *H. melpomene* and *H. cydno*, with the analyzed sympatric species pair highlighted in red. *H. melpomene* is widely distributed in Central and South America, whereas *H. cydno* is mainly found in Central America. (*B*) Diagram of the body structure of a *Heliconius* butterfly with the antennae, legs, and mouthparts labeled, respectively. (*C*) A neighbor-joining phylogeny was constructed for all the transcriptome samples used in this study, showing that different tissue samples from the same individual clustered as a subgroup. (*D*) Principal component analysis (PCA) of all sequenced chemosensory tissues based on their overall gene expression profiles. Solid circles represent the 95% confidence interval regions. Hc and Hm refer to *H. cydno* and *H. melpomene*, whereas A, L and M refer to antennae, legs, and mouthparts, respectively.

perception of smell (supplementary fig. S3 and supplementary table S3, Supplementary Material online). The leg and mouthpart tissues showed far fewer DEGs than the antenna tissues in both species (supplementary fig. S2, Supplementary Material online). The major signatures of the leg-specific genes were enriched for the GO terms related to ion channel activity and ion transmembrane transporter activity (supplementary fig. S4 and supplementary table S4, Supplementary Material online), whereas the signatures of the mouthpart genes were mainly related to metabolic processes and catalytic activities (supplementary fig. S5 and supplementary table S5, Supplementary Material online).

Moreover, we cross-validated the interspecific diversification of gene expression in sensory tissues by comparing the transcription profiles of antennae between the distantly related species *H. melopmene* and *H. charithonia* and between *H. cydno* and *H. charithonia*, which diverged approximately 12 Mya (fig. 3*A* and supplementary table S1, Supplementary Material online; Kozak, et al. 2015; Catalan, et al. 2019). We again observed different gene expression profiles between the leg and antenna tissues of *H. charithonia* based on the PCA results, further suggesting the uniqueness of antenna tissues (supplementary fig. S6, Supplementary Material online). Then, we found a large number of DEGs between the antenna tissues of *H.*

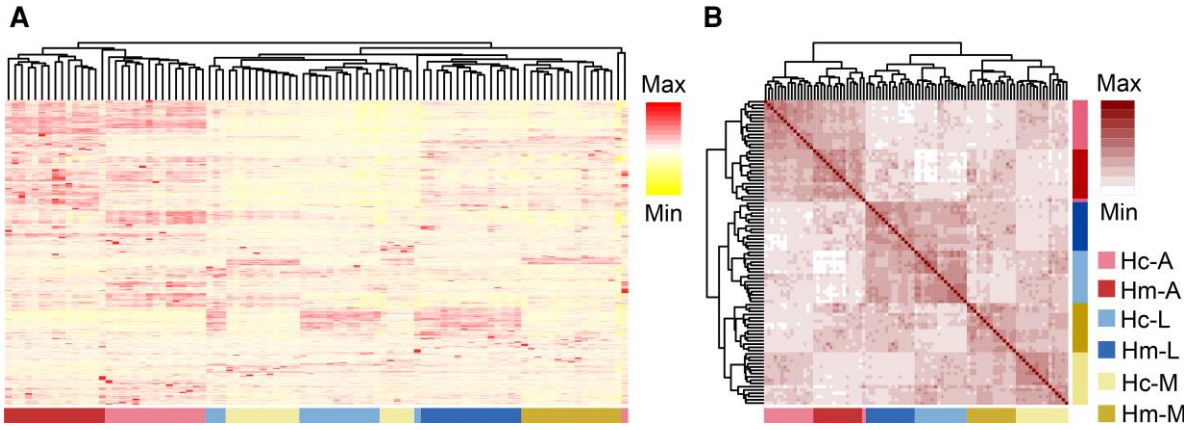


FIG. 2. Differential gene expression patterns in chemosensory tissues of *H. cydno* and *H. melpomene*. (*A*) Overall gene expression profiles of each sample. The columns represent individual samples, and the rows represent individual genes. We transformed the FPKM values of genes into *Z* scores. Genes that were not expressed in any replicate were removed. Samples with similar expression patterns are clustered closely. The antennal expression patterns of *H. cydno* and *H. melpomene* were more similar regardless of sample stage and sex. Replicates of data from legs and mouthparts are clustered by species. Max, 4; Min, −4. The max or min (positive or negative) value indicates that the raw score is above or below the mean average, respectively. (*B*) Correlation heatmaps of RNA-seq samples in all biological replicates of chemosensory tissues. Each box indicates an individual sample. Replicates from the same tissues are clustered closely. The clustering tree shows that the leg and mouthpart tissues are more closely correlated, while the antennae are unique. Max, 1; Min, −0.2. The max or min value represents the correlation coefficient. Hc and Hm refer to *H. cydno* and *H. melpomene*, whereas A, L, and M refer to antennae, legs, and mouthparts, respectively.

melopmene and *H. cydno*, 60.9% and 54.3% of which overlapped with DEGs between the antenna tissues of *H. melopmene* and *H. charithonia* and those of *H. cydno* and *H. charithonia*, respectively, implying that the DEGs identified in antennae were consistently present across the different periods of *Heliconius* speciation (fig. 3*B*–*D*). Given the evidence that conserved protein-coding genes with diverse regulatory elements contribute to the diversification of *Heliconius* wing patterns (Van Belleghem, et al. 2017), we hypothesized that cis-regulatory evolution may also play an important role in driving gene expression divergence in sensory tissues, that is that the different proportions of DEGs between *H. melpomene* and *H. charithonia* and between *H. cydno* and *H. charithonia* may be due to divergence in non-coding regulatory DNA elements. We also considered the top 10% of upregulated DEGs identified in antenna tissues between *H. cyndo* and *H. melpomene*, and we found that 69.4% and 40.8% of these DEGs were differentially expressed between *H. melopmene* and *H. charithonia* and between *H. cydno* and *H. charithonia*, respectively (fig. 3*B*–*D* and supplementary tables S6 and S7, Supplementary Material online), supporting consistent expression differentiation in antennae among *Heliconius* butterflies. These DEGs included an odorant binding gene, *OBP15*, that was differentially expressed in each comparison of pooled antennae of different *Heliconius* butterflies (fig. 3*B*–*D*). We thus speculated that *OBP15* may play an indispensable role in the interspecific diversification of *Heliconius* butterflies. Notably, the greatest differences in the gene expression patterns of the antennae between *H. melpomene* and *H. cydno* were found in 5-day-old females relative to 2-day-old females and males (fig. 3*E*–*M*). The DEGs identified between *H. cydno* and *H. melpomene* in pooled and 5-day-old female antennae were enriched for GO terms involved in odorant binding, primary active transmembrane transporter activity, and ATPase activity (supplementary fig. S7, Supplementary Material online). Therefore, our results demonstrated that the antennal tissue showed a strong pattern of differential expression relative to that in other chemosensory tissues, indicating the unique role of this organ in *Heliconius* chemosensing. Additionally, our findings indicated that the older female stage is a critical period showing interspecific diversification at the gene expression level, which may be driven by host-plant searching and oviposition.

Only a few CSGs were differentially expressed between females at different adult stages and between adult females and males in all the chemosensory tissues of both species (supplementary figs. S8 and S9, Supplementary Material online). In the comparisons of females at different adult stages, more genes with upregulated expression were identified in the antennae and legs of 5-day-old *H. cydno*, whereas more genes with upregulated expression were detected in the mouthparts of 2-day-old *H. melpomene* (supplementary fig. S8 and supplementary tables S8 and S9, Supplementary Material online). In the comparisons between sexes, all the *H. cydno* chemosensory tissues showed a higher proportion of genes with upregulated expression than those of *H. melpomene* in females (supplementary fig. S9 and supplementary tables S10 and S11, Supplementary Material online). In summary, we found greater differentiation between life stages (an average of ~349 DEGs across *H. cydno* tissues and 294 in *H. melpomene*) than between sexes (an average of ~80 DEGs across *H. cydno* tissues and 135 in *H. melpomene*) in both *Heliconius* species. These findings suggest that many nonchemosensory genes likely play an important role in sex and life-stage differentiation.

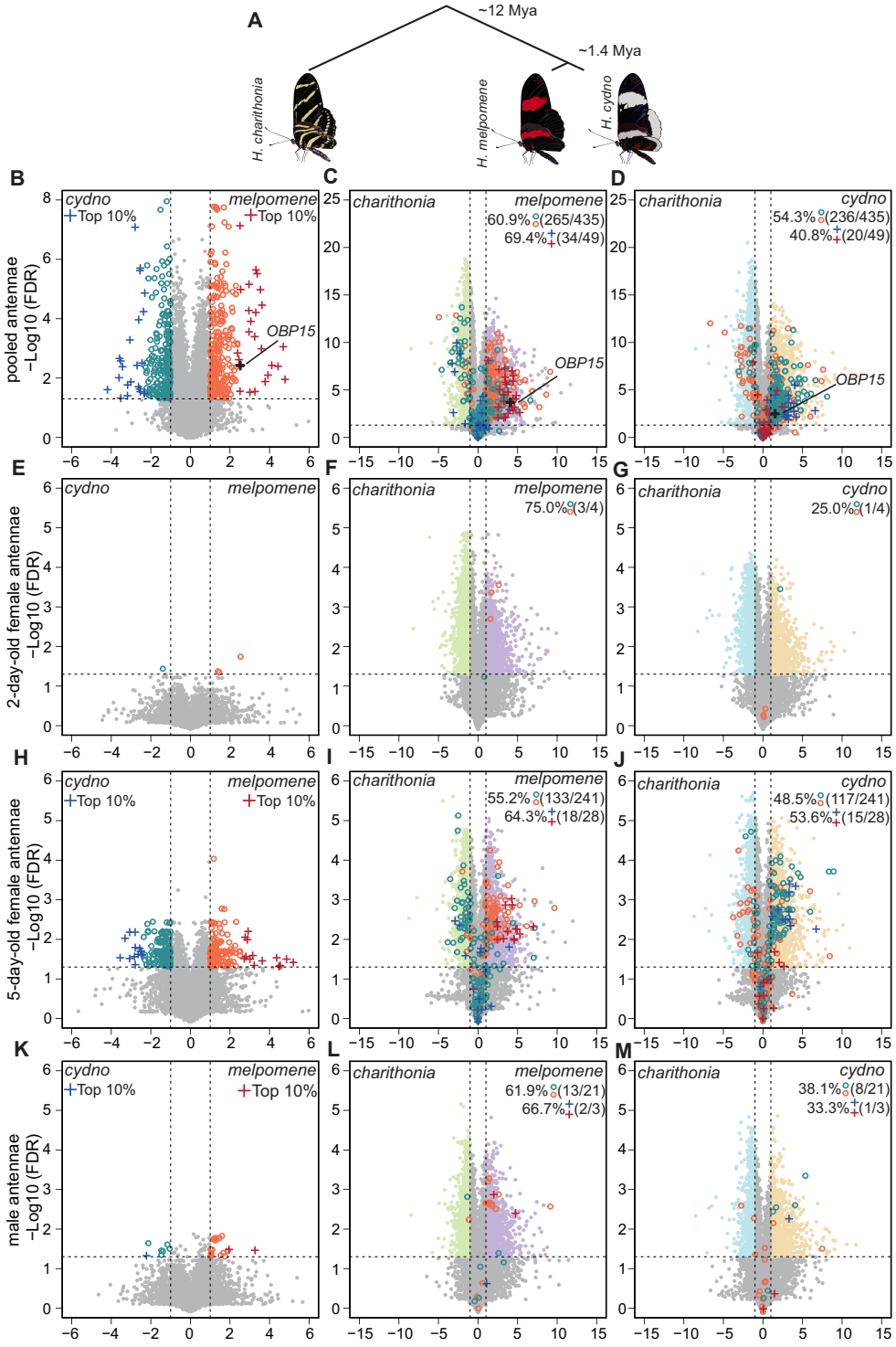


Fig. 3. Convergent differential gene expression pattern in antennae between *Heliconius* butterflies. The phylogenetic relationships of the three species are illustrated schematically in (*A*). Volcano plots showing differentially expressed genes (DEGs) between species in pooled antennae (*B*)–(*D*), 2-day-old female antennae (*E*)–(*G*), 5-day-old female antennae (*H*)–(*J*) and male antennae (*K*)–(*M*). In (*C*), (*F*), (*I*), and (*L*), numbers in brackets indicate the exact number of overlap with those DEGs between *H. cydno* and *H. melpomene* and the total number of DEGs between *H. cydno* and *H. melpomene*, respectively. In (*D*), (*G*), (*J*), and (*M*), numbers in brackets indicate the exact number of overlapping DEGs between *H. cydno* and *H. melpomene* and the total number of DEGs between *H. cydno* and *H. melpomene*, respectively. Blue +, top 10% of upregulated genes in *H. cydno* identified in the comparison of *H. cydno* and *H. melpomene*; light sea-green circles, upregulated genes in *H. cydno* identified in the comparison of *H. cydno* and *H. melpomene*; red +, top 10% of upregulated genes in *H. melpomene* identified in the comparison of *H. cydno* and *H. melpomene*; orange circles, upregulated genes in *H. melpomene* identified in the comparison of *H. cydno* and *H. melpomene*; pale green dots, upregulated genes in *H. charithonia* identified in the comparison of *H. charithonia* and *H. melpomene*; purple dots, upregulated genes in *H. melpomene* identified in the comparison of *H. charithonia* and *H. melpomene*; light sky-blue dots, upregulated genes in *H. charithonia* identified in the comparison of *H. charithonia* and *H. cydno*; light yellow dots, upregulated genes in *H. cydno* identified in the comparison of *H. charithonia* and *H. cydno*.

Genomic Admixture of Chemosensory Tissue Genes

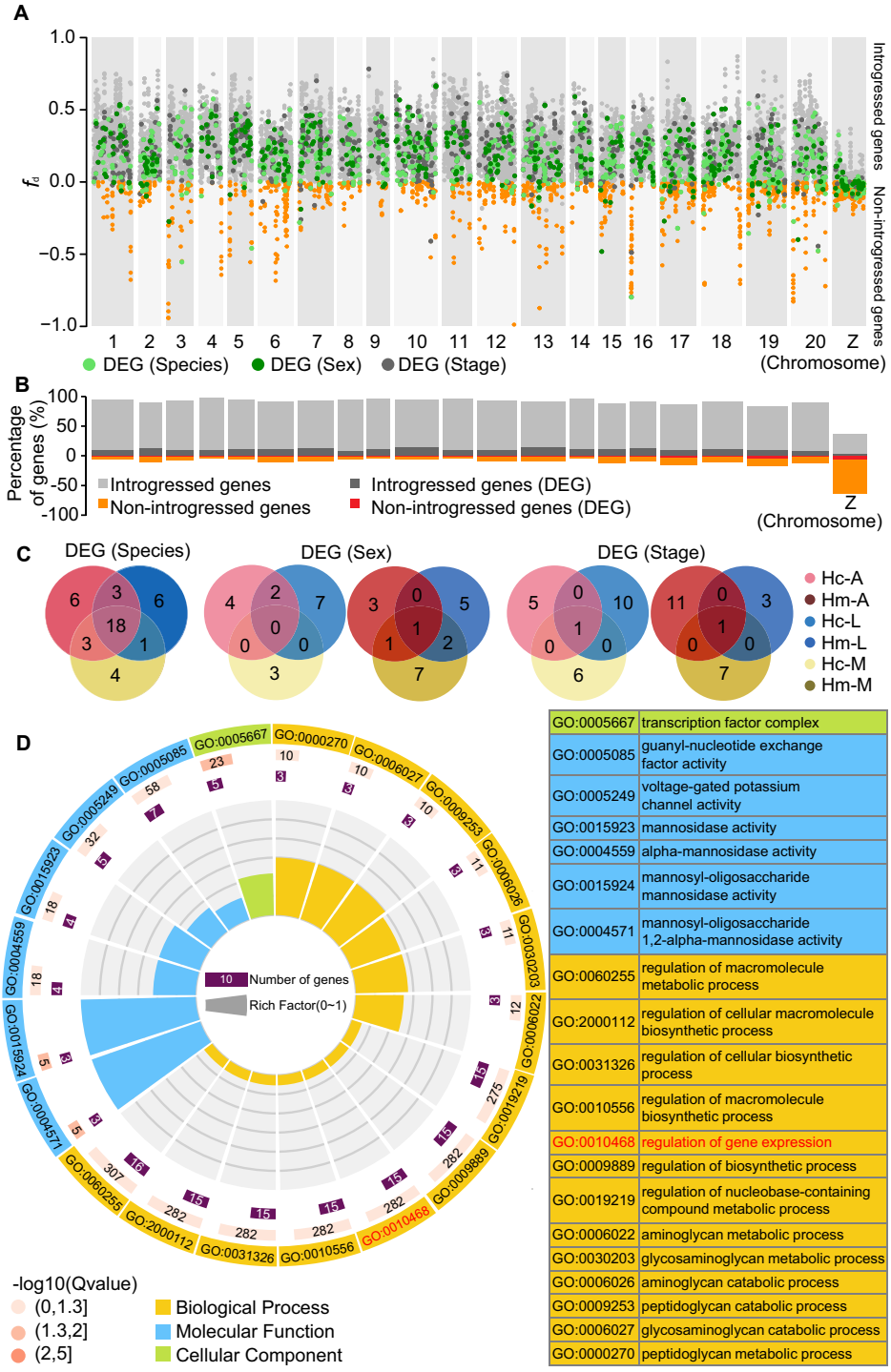
The abovementioned patterns of differential expression could be a consequence of regulatory changes in the focal genes. Therefore, these DEGs may not necessarily exhibit low genetic admixture or directly contribute to reproductive isolation. To test this hypothesis, we examined the level of genetic admixture related to chemosensing between the hybridizing species pair, and our results showed that the DEGs between species exhibited low admixture, which differed significantly from the findings at the genome level (supplementary table S12, Supplementary Material online), suggesting that many DEGs directly functionally contribute to reproductive isolation or ultimately act as genetic barriers due to genetic drift. For the whole-genome except the Z chromosome, 8.4% of the total 21,661 annotated genes showed a signature of low admixture, whereas over 63% of genes on the Z chromosome were categorized as nonintrogressed genes (fig. 4*A* and *B*), similar to the findings of previous studies on divergence and admixture of the Z chromosome in *Heliconius* (Van Belleghem, et al. 2018; Martin, et al. 2019). We noted that the DEGs identified in chemosensory tissues exhibited similar trends and were particularly enriched on the Z chromosome (fig. 4*A*). There were more nonintrogressed than introgressed genes on the Z chromosome (720 of 2,412 total nonintrogressed genes), many of which were differentially expressed (59 of 226 genome-wide nonintrogressed DEGs). Among these genes, species-specific DEGs accounted for the largest proportion, and many of them were shared by all the chemosensory tissues, suggesting that the Z chromosome plays an important role in determining gene expression levels related to chemosensory diversification (fig. 4*C*). However, the DEGs identified between species, stages, and sexes as well as the nonintrogressed genes were largely nonchemosensory genes. We found that the nonintrogressed genes on the Z chromosome were enriched for GO terms related to the regulation of gene expression as well as multiple metabolic and biosynthetic processes, suggesting their possible roles in regulating the expression of downstream genes on the Z chromosome or autosomes, indirectly promoting diversification (fig. 4*D*). Taken together, our results showed concordant patterns of interspecific genetic divergence and differential expression enriched on the Z chromosome in particular, indicating a disproportionately large role of the Z chromosome in modulating changes in chemosensing that may impact premating isolation.

Linkage Disequilibrium Related to Chemosensing

To further understand the role of chemosensing-related genes in speciation, we also investigated the possible linkage between these genes or their regulatory elements and other genetic loci that contribute to reproductive isolation and are subject to divergent selection. Previous studies have revealed that wing patterns and mate preference may facilitate assortative mating behavior in incipient *Heliconius* species (Jiggins, et al. 2001; Kronforst and Papa 2015) and have identified associated genetic loci, such as the wing-pattern loci *B*/*D*, including the *optix* gene; *Yb*, including the *cortex* gene; *Ac*, including *WntA*; and *K*, including *aristaless* (Jiggins, et al. 2017; Westerman, et al. 2018) (fig. 5*A*). Additionally, three genome-wide quantitative trait loci (QTLs) have been associated with mate preference between *H. melpomene* and *H. cydno* (Merrill, et al. 2019; Rossi, et al. 2020). These findings enabled us to investigate the LD between genes related to chemosensing, wing patterns, and mate preference. Although a previous study (van Schooten, et al. 2020) reported that no CSGs located close to the wing-pattern genes showed signatures of differential expression or divergence and that the total CSGs did not show elevated LD with *optix*, we identified several nonchemosensory genes with signatures of differential expression or nonintrogression located near wing-pattern genes and mate preference QTLs (fig. 5*B* and supplementary figs. S10–S14, Supplementary Material online). We noted an overall nonrandom excess of interspecifically nonintrogressed genes in the peak region of the mate preference QTL on chromosome 18, which is located near the *optix* gene (fig. 5*B*, and supplementary table S13, Supplementary Material online). Coincidently, this peak region showed elevated LD with genome-wide nonintrogressed genes (fig. 5*C* and *D*). We also found that some genes expressed in chemosensory tissues were located in this region and showed patterns of differential expression between sexes (four genes), life stages (12 genes, including the CSG *OBP42*), and species (13 genes, including the CSG *OBP40*). However, we did not observe similar patterns among other mate preference and wing-pattern loci, and we consistently observed weak LD with genome-wide nonintrogressed genes in *H. cydno*, likely because of the fewer *H. cydno* single-nucleotide polymorphisms (SNPs) mapped to the *H. melpomene* reference genome. Nevertheless, our findings reveal that nonchemosensory genes in chemosensory tissues can be linked to wing patterns and mate preference, indicating that chemosensing, wing patterns, and mate preference may synergistically contribute to reproductive isolation.

Integration of Chemosensory and Nonchemosensory Genes

The abovementioned results demonstrate that in chemosensory tissues, nonchemosensory genes show some of the greatest interspecific differences in expression and the lowest genetic admixture relative to chemosensory-related gene families. In regard to the functional importance of chemosensation in speciation, our *a priori* hypothesis was that nonchemosensory DEGs would show higher connectivity and a stronger correlation with CSGs than with other, nonchemosensory non-DEGs. To test this hypothesis, we constructed gene coexpression (GC) networks for *H. c. chioneus* and *H. m. rosina* chemosensory tissues and explored the integration and relationships between chemosensory and nonchemosensory genes (fig. 6 and supplementary fig. S15, Supplementary Material online).



GO:0005667	transcription factor complex
GO:0005085	guanyl-nucleotide exchange factor activity
GO:0005249	voltage-gated potassium channel activity
GO:0015923	mannosidase activity
GO:0004559	alpha-mannosidase activity
GO:0015924	mannosyl-oligosaccharide mannosidase activity
GO:0004571	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity
GO:0060255	regulation of macromolecule metabolic process
GO:2000112	regulation of cellular macromolecule biosynthetic process
GO:0031326	regulation of cellular biosynthetic process
GO:0010556	regulation of macromolecule biosynthetic process
GO:0010468	regulation of gene expression
GO:0009889	regulation of biosynthetic process
GO:0019219	regulation of nucleobase-containing compound metabolic process
GO:0006022	aminoglycan metabolic process
GO:0030203	glycosaminoglycan metabolic process
GO:0006026	aminoglycan catabolic process
GO:0009253	peptidoglycan catabolic process
GO:0006027	glycosaminoglycan catabolic process
GO:0000270	peptidoglycan metabolic process

Fig. 4. Transcriptome-wide admixture pattern in chemosensory tissues of *H. cydno* and *H. melpomene*. (*A*) The admixture pattern of DEGs between *H. cydno* and *H. melpomene*. The admixture pattern between *H. cydno* and *H. melpomene* was plotted according to the admixture levels (f_d) calculated with the topology ((*H. m. melpomene*, *H. m. rosina*), *H. c. chioneus*, *H. numata*) for annotated genes across the genome. Among the total genes, those with $f_d > 0$ were categorized as introgressed genes between *H. m. rosina* and *H. c. chioneus*, and the others were categorized as nonintrogressed genes. The DEGs identified in chemosensory tissues between species, sexes, and life stages are also categorized and labeled in different colors. DEGs (species), DEGs between *H. cydno* and *H. melpomene*; DEGs (sex), DEGs between 5-day-old females and males of *H. cydno* or *H. melpomene*; DEGs (stage), DEGs between 2-day-old females and 5-day-old females of *H. cydno* or *H. melpomene*. (*B*) Percentages of introgressed and nonintrogressed genes on each chromosome. The percentage of DEGs is also indicated in each category. (*C*) Venn diagrams of DEGs on the Z chromosome. Hc and Hm refer to *H. cydno* and *H. melpomene*, whereas A, L, and M refer to antennae, legs, and mouthparts, respectively. (*D*) Top 20 GO terms and descriptions in the GO enrichment analysis of nonintrogressed genes on the Z chromosome. From the outer to inner circles, the first circle represents the classes of enriched GO terms; the second circle represents the background gene numbers, with the color indicating the threshold values of −log10 (Q value); the third circle represents the numbers of genes enriched for each GO term; and the fourth circle represents the enrichment fold-change.

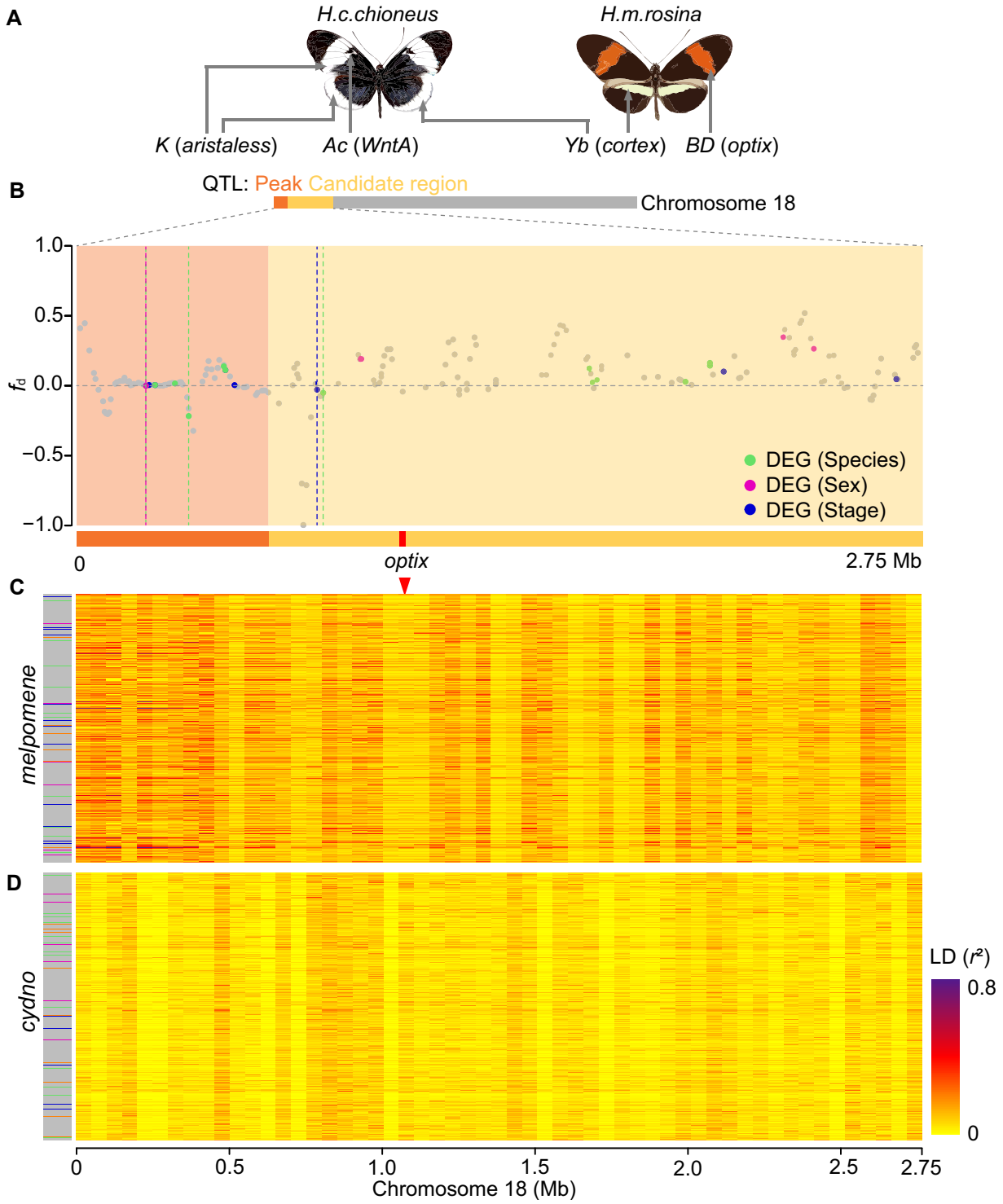


FIG. 5. Genomic divergence and differential expression in chemosensory tissues associated with divergent mate preference and color patterns. (*A*) The diagram includes the wing-pattern loci and their associated wing patterns in *H. melpomene* and *H. cydno*. (*B*) Nonintrogressed and DEGs at the mate preference QTL and the color pattern gene *optix* on chromosome 18, including the QTL peak region and the nearby 1.5-log-odds-ratio candidate region. Gray points indicate non-DEGs. Green, purple, and blue points indicate DEGs. Points above the horizontal dotted line are introgressed genes ($f_d > 0$), while the others are nonintrogressed genes ($f_d \leq 0$). Vertical dashed lines indicate nonintrogressed DEGs. Linkage disequilibrium (LD) map of genome-wide nonintrogressed genes and the focal region containing the mate preference QTL and *optix* in *H. melpomene* (*C*) and *H. cydno* (*D*). Nonintrogressed genes are ordered according to their f_d values. The *x*-axis indicates the 2.75 Mb region corresponding to the physical positions shown in (*B*). The LD values (r^2) were calculated in 50 kb windows.

Comparisons of the network topologies of the two species revealed similar topological characteristics, with three major clusters, suggesting general conservation of the GC network in *Heliconius* chemosensory tissues (fig. 6*A*, *D* and supplementary fig. S15, Supplementary Material online). We found that CSGs were clustered with many other

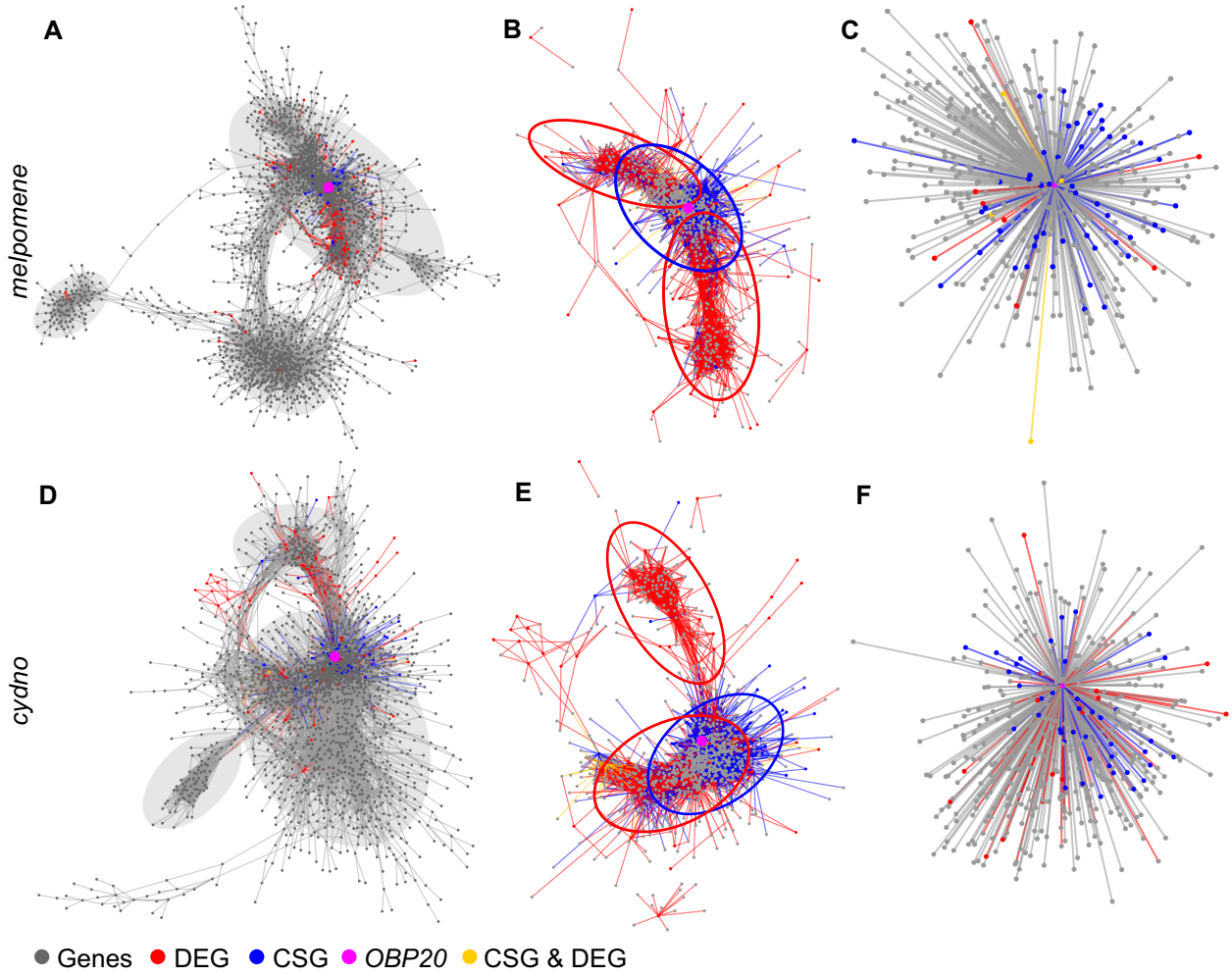


Fig. 6. Weighted gene coexpression (GC) networks of chemosensory genes (CSGs) and DEGs. The main networks were constructed for *H. melpomene* (*A*) and *H. cydno* (*D*) chemosensory tissues, showing three clusters interacting with each other, indicated by solid circles. The filtered networks included CSGs, DEGs and their primary network partners in *H. melpomene* (*B*) and *H. cydno* (*E*), where DEGs interact closely and extensively with CSGs from two directions. *OBP20* and its primary network partners are illustrated in *H. melpomene* (*C*) and *H. cydno* (*F*). Red dots, genes differentially expressed in antennae, legs or mouthparts between species; blue dots, CSGs; pink dot, *OBP20*; yellow dots, CSGs with differential expression between species; gray dots, genes involved in networks. Red edges, interactions involving DEGs; blue edges, interactions involving CSGs; yellow edges, interactions involving CSGs and DEGs; gray edges, interactions with other genes.

genes, which led us to speculate that the three major clusters exhibit a cascade of amplified regulatory effects. Notably, most hub genes in the three major clusters were nonchemosensory genes (supplementary table S14, Supplementary Material online), suggesting that these genes could be regulators of downstream chemosensory and other genes in the network clusters. The CSGs showed a strong correlation with nonchemosensory DEGs in both species and formed major clusters with dense connections. However, the nonintrogressed, nonchemosensory genes were generally distributed across the three major clusters and were not densely connected with the CSGs (fig. 6*B*, *E* and supplementary fig. S15, Supplementary Material online). Based on previous evidence indicating *OBP20* involvement in speciation and suggesting that its divergence was driven by gene regulatory elements (van Schooten, et al. 2020), we investigated *OPB20*'s connectivity in the GC networks of the two species and found that *OBP20* showed very high connectivity with CSGs and nonchemosensory genes as a hub gene (410 adjacent nodes in *H. melpomene* and 441 in *H. cydno*) (fig. 6*C*, *F*, supplementary fig. S15 and supplementary table S15, Supplementary Material online). Collectively, our findings highlight the importance of not only CSGs (i.e., *OBP20*) but also nonchemosensory DEGs in the incipient stages of *Heliconius* speciation.

Discussion

Chemosensing is a fundamental function in multicellular organisms, as it allows the recognition of chemical signals and the expression of specific behaviors (Yan, et al. 2020). Chemical communication is essential to butterfly life history activities such as foraging, mate preference, and oviposition (Jiggins 2017). From a broader perspective,

chemosensory organs are sites of expression for both CSGs and nonCSGs, which are subject to selection as a whole. Both types of genes are necessary to generate distinct sensory organs (i.e., antennae, legs, and mouthparts) with specific structures and molecular characteristics suited to their unique functions. In this study, we used *Heliconius* butterflies, including both partially and completely reproductively isolated species, to obtain molecular evidence of the distinctive features of antennae relative to other chemosensory organs, implying a unique functional role of these organs in chemosensing. As the main chemosensory organs, *Heliconius* antennae contribute to multiple life activities, acting either synergistically with other organs or independently. For example, the female antennae, proboscis, and forelegs contribute to host plant recognition during oviposition (Benson, et al. 1975), whereas the antennae alone can detect complex floral scents involving multiple compounds that can serve as cues in foraging or mating (Andersson and Dobson 2003; Estrada and Gilbert 2010); these functions correspond to antenna-enriched GO terms such as odorant binding, olfactory receptor activity, and sensory perception of smell. In addition, insect antennae may assume more complex roles, such as auditory, wind-sensing, gravity-sensing, and circadian oscillator functions, as reported for *Drosophila* antennae (Tanoue, et al. 2004; Todi, et al. 2004; Kamikouchi, et al. 2009; Fuller, et al. 2014). Consistent with the expected functions of antennae in insects, our data suggest a unique molecular architecture relative to those of the other chemosensory organs and possibly a far more important role in reproductive isolation. Additional experiments will be needed to functionally investigate the strongest candidate genes for these functions using CRISPR/Cas9 to determine the corresponding behavioral phenotype. However, this type of experiment is still not easy to perform due to the mosaic nature of CRISPR/Cas9 effects and the difficulty of screening F0 mutants in real time. Another possibility is the generation of a stable F2 mutant line, but this approach presents its own challenges, which currently limits the comprehensive understanding of the roles of candidate CSGs. Finally, these chemosensory phenotypes cannot be validated using silkworms or *Drosophila* because of their quite different life histories compared to those of *Heliconius* butterflies.

Our results provide transcriptomic and genetic perspectives on the consequences of reproductive isolation for gene expression in sensory tissues. A previously noted (Van Belleghem, et al. 2018; Martin, et al. 2019) role of sex chromosomes in speciation with incomplete reproductive isolation emerged from our findings. However, by mapping the DEGs to the patterns of admixture, we observed that the Z chromosome showed clear signatures of both genetic divergence (normal distribution, P [percentage of nonintrogressed genes on the Z chromosome] < 0.01) and functional differentiation (normal distribution, P [percentage of DEGs among nonintrogressed genes on the Z chromosome] < 0.01). The enriched GO terms of the nonintrogressed genes on the Z chromosome can be used to not only investigate the functional categories in which interspecific divergence directly occurs but also help understand how genetic divergence promotes expression differences through regulation. As mentioned above, the nonintrogressed genes on the Z chromosome enriched in the regulation of gene expression GO terms may serve as candidates involved in bridging interspecific divergence and functional differentiation through regulation. Divergence in the gene expression of such Z-linked regulators can impact the expression of genes on other chromosomes and generate profound changes in spatiotemporal complexity across tissues and life stages. Our study supports the role of sex chromosomes in speciation and further demonstrates their link to chemosensory differentiation. The confirmation of the above hypothesis regarding the Z chromosome and the patterns identified based on our data in additional tissues (i.e., head, thorax, and wings) deserve further investigation to better understand functional differentiation at the genome-wide level during the stages of speciation with incomplete reproductive isolation.

Network analysis further suggested the important roles of nonchemosensory genes with differential expression patterns in both species. Interestingly, these genes seem to function as connecting hubs in complex networks, as they show relatively high connectivity, whereas the other nonCSGs and CSGs are located at the network periphery. Therefore, these nonchemosensory hub genes likely show pleiotropic biological interactions and possibly present greater potential to promote reproductive isolation. Our study demonstrates the great complexity underlying chemosensation in *Heliconius* butterflies and highlights a group of genes involved in chemosensation whose roles remained unrecognized until now. Such interactions between chemosensory and nonchemosensory genes provide a more complete view of the molecular mechanism underlying chemosensation and a better understanding of the routes of evolutionary adaptations underlying changes in chemical communication during the stages of speciation with incomplete reproductive isolation. Such integrative research enables us to investigate the roles of chemosensory organs in a holistic manner, which will ultimately lay the foundation for an in-depth understanding of the functions of different sensory tissues.

Materials and Methods

Data Collection

We reared *H. charithonia* individuals under seminatural conditions in Gamboa, Panama. Females were collected 2 days or 5 days after emergence, while males were sampled at 5 days of age. The antennae, forelegs, and hindlegs of each individual were dissected for RNA extraction using the TRIzol RNA isolation protocol, and additional purification steps were performed using the RNeasy mini kit (Qiagen). The Illumina libraries were constructed using the Illumina TruSeq RNA sample preparation kit and

sequenced on a NextSeq 500 system. We obtained the RNA-seq dataset for *H. melpomene* and *H. cydno* from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) repository with BioProject accession number PRJNA577441 (van Schooten, et al. 2020). Briefly, *H. melpomene* and *H. cydno* individuals were also reared under seminatural conditions in Gamboa, Panama. Females were collected 2 days or 5 days after emergence, while males were sampled at 5 days of age. The antennae, legs, and mouthparts of each individual were dissected for RNA extraction. Detailed sample information and sequencing statistics are provided in supplementary table S1, Supplementary Material online.

Phylogenetic Analysis

Raw RNA-seq reads were first trimmed using Seqtk v1.3 (https://github.com/lh3/seqtk) to filter out low-quality bases, and the qualified reads were aligned to the reference genome of *H. melpomene* v2 (Davey, et al. 2016) obtained from Lepbase (http://lepbase.org/) using STAR v2.7.6a (Dobin, et al. 2013) with the default parameters. We performed a statistical comparison of filtered and trimmed RNA-seq alignment rates between identical tissues from different species and found no significant differences between *H. cydno* and *H. melpomene* (Scheffe's test, $P > 0.05$). Genotype calling was performed using Genome Analysis Toolkit (GATK) following a best-practice workflow (McKenna, et al. 2010). Briefly, after removing PCR duplicates, we used HaplotypeCaller in GATK v4.1.2.0 to call genotypes for all sequenced samples, combined GVCF files using CombineGVCFs, and performed joint variant calling using GenotypeGVCFs (McKenna, et al. 2010). We then retained SNPs with quality > 30 and constructed a phylogenetic tree of all sequenced samples based on the genetic distance matrix described previously (Xia, et al. 2009; Xiang, et al. 2018). Briefly, we used a simple model to calculate pairwise genetic distances for each locus, where 0 represented different homozygous genotypes, 0.5 represented heterozygous genotypes, and 1 represented identical genotypes. We performed 100 bootstrap replicates by randomly selecting SNPs with 10 M replacements. The neighbor-joining phylogeny was inferred on the basis of the obtained distance matrix using MEGA X (Kumar, et al. 2018).

Differential Gene Expression Analysis and Functional Enrichment

We used the well-annotated chemosensory gene set (van Schooten, et al. 2020) to improve the gene set of *H. melpomene* v2 (Davey, et al. 2016) and used RSEM v1.3.3 (Li and Dewey 2011) to determine gene expression profiles based on the fragments per kilobase of transcript per million mapped reads (FPKM) method. To efficiently detect the expression pattern of specific isoforms in different tissues, we treated each isoform as an independent unit, which also reflects the total differential expression levels of the focal gene loci between species. We used the built-in R function "prcomp" (https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prcomp) to normalize the gene expression data and performed PCA based on the expression profile. The expression profiles based on overall FPKM values were generated with the "pheatmap" package in R (https://CRAN.R-project.org/package=pheatmap). The correlation coefficients of all samples were calculated using the R functions "cormat" (https://www.rdocumentation.org/packages/plsRglm/versions/1.3.0/topics/CorMat) and "pheatmap" (https://CRAN.R-project.org/package=pheatmap).

Genes that were expressed in all replicates were identified as reliably expressed genes. We applied *t*-tests to compare the patterns of genome-wide differential expression between different samples. For species-specific expression analyses, we performed comparisons across tissue types. For life-stage-specific and sex-specific expression analyses, comparisons were conducted within species for each tissue type. We compared the antenna, leg and mouthpart tissues in a pairwise manner to determine tissue-specific expression within species. On the basis of the overall distribution, we identified significant DEGs between species with a cutoff of an FDR-adjusted $P < 0.05$. A cutoff of $P < 0.05$ was used to filter significant DEGs within species in a specific tissue.

Functional enrichment was examined based on GO analysis and further mapping to the identified GO terms using the OMICSHARE cloud platform (https://www.omicshare.com/tools/Home/Soft/gogseasenior). We identified significantly enriched pathways with a cutoff of an FDR-adjusted $P < 0.05$ and processed significantly DEGs for functional enrichment analysis. Enriched ontology terms were categorized by class.

Detecting Gene Flow Between *H. c. chioneus* and *H. m. rosina*

To reveal the roles of the significant DEGs identified in chemosensory tissues in speciation, our further analyses focused on characterizing their admixture patterns in the genome. We analyzed genome-wide introgression using *H. m. melpomene*, *H. c. chioneus* and *H. m. rosina* as ingroups and *H. numata* as an outgroup. Among these focal species, *H. c. chioneus* and *H. m. rosina* are sympatric species from Panama, while *H. m. melpomene* is an allopatric race of *H. m. rosina* from French Guiana. We performed a modified *f*-statistic (f_d) analysis (Martin, et al. 2015, 2019) to estimate genome-wide admixture and then integrated Patterson's *D*-statistic (Durand, et al. 2011) and f_d (Martin, et al. 2015, 2019) to identify potential nonintrogressed loci across the genome, both of which measure an excess of derived allele sharing between the sympatric species *H. c. chioneus* and *H. m. rosina* compared to allopatric *H. m. melpomene*. Both the f_d and *D*-statistics were calculated for 20 kb sliding windows across the whole genome with a 5 kb step size. We averaged the overlapping windows for each gene and identified introgressed genes according to cutoffs of $f_d > 0$ and *D*-statistic > 0. We considered the remaining genes to be nonintrogressed genes.

Linkage Disequilibrium Analysis

To investigate the roles of nonintrogressed genes in speciation, we calculated LD between nonintrogressed genes and three QTLs, including the *optix* color pattern locus (Merrill, et al. 2019), that were previously reported to be associated with preference behavior and contributed to species differences between *H. c. chioneus* and *H. m. rosina*. Additionally, we calculated LD between nonintrogressed genes and other known color pattern loci, including the *WntA*, *cortex* and *aristaless* genes (Kronforst and Papa 2015; Jiggins, et al. 2017; Westerman, et al. 2018), which were maintained or accumulated divergence during *Heliconius* speciation. We randomly selected 10 SNPs of each nonintrogressed gene (or all SNPs if there were fewer than 10 SNPs) and 100 SNPs near the focal region (1 Mb). We performed LD analysis using PLINK v1.9 (Purcell, et al. 2007) for pairwise comparisons of these SNPs. The r^2 values in each 50 kb window were averaged to obtain the mean r^2.

To exclude the random effect of reduced recombination, we randomly selected 20,000 SNPs at the genome-wide level, divided them into 10 groups, and then calculated LD within each group. The LD between nonintrogressed genes and regions located near *optix* was significantly higher than that at the genome-wide level (supplementary table S13, Supplementary Material online). To estimate the possible effect of reduced recombination toward the chromosome end, we further calculated LD between nonintrogressed genes and corresponding regions on chromosomes 1, 12, and 19, which have a length similar to that of chromosome 18. The LD between nonintrogressed genes and regions on chromosome 18 was significantly higher than that between nonintrogressed genes and regions on chromosomes 1 and 12. The average LD between nonintrogressed genes and regions on chromosome 18 was also higher than that between nonintrogressed genes and regions on chromosome 19, but the difference was not significant (supplementary table S13, Supplementary Material online). Our results suggested an overall nonrandom association of genomic divergence and differential expression in chemosensory tissues with *optix* in *H. melpomene*.

Network Analysis

Gene expression data of *H. melpomene* and *H. cydno* were used for coexpression network analysis. GC networks were constructed using the weighted correlation network analysis (WGCNA) R package (Langfelder and Horvath 2008). We used the PickSoftThreshold function to select an appropriate soft threshold (signed R^2 threshold > 0.85) to construct the network. We referred to genes as nodes, and the edges between nodes were determined based on pairwise correlations between gene expression values. We used Cytoscape v3.8.2 (Su, et al. 2014) to visualize and modify the network association data.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgements

This project was supported by grants from the Beijing Natural Science Foundation (JQ19021 to W.Z.), the National Natural Science Foundation of China (31871271 and 32170420 to W.Z., 32100330 to N.W.), the Peking-Tsinghua Center for Life Sciences (to W.Z.), the State Key Laboratory of Protein and Plant Gene Research and Qidong-SLS Innovation Fund (to W.Z.), and the China Postdoctoral Science Foundation (2020M680215 to N.W.). This study was also supported by an Institutional Development Award Networks of Biomedical Research Excellence (INBRE) Grant (P20GM103475) from the National Institute of General Medical Sciences, a component of the NIH, the Bioinformatics Research Core of the INBRE. R.P. and B.C. were funded by NSF EPSCoR RII Track-2 FEC (grant no. OIA 1736026). R.P. was also supported by NSF IOS 1656389, the Fondo Institucional para la Investigación (FIPI), Universidad de Puerto Rico—Recinto de Rı´o Piedras, Decanato de Estudios Graduados e Investigación, and the Puerto Rico Science, Technology, and Research Trust under Agreement ARG 2020-00138. S.M.V.B. was also supported by a Puerto Rico Science, Technology & Research Trust catalyzer award (#2020-00142) We thank Sequencing and Genomics Facility staff for their assistance with RNA sequencing and the Bioinformatics Research Core of the INBRE for providing the computational infrastructure to run the analyses. We thank the Computing Platforms of the Center for Life Sciences and the School of Life Sciences at Peking University for assistance with computation.

Author Contributions

W.Z. and R.P. conceived and designed the study; N.W., S.M.V.B., E.E., J.M.-R., B.V.S., B.A.C., Y.O. and S.M.P.S.-N. analyzed the data; W.Z. and N.W. wrote the manuscript with input from R.P. and S.M.V.B., and all the authors proofread and approved the manuscript.

Data Availability

We deposited RNA-seq data of *H. charithonia* into the NCBI SRA under accession numbers SAMN30448756–SAMN30448800 (https://trace.ncbi.nlm.nih.gov/Traces/sra/). RNA-seq data of *H. melpomene* and *H. cydno* are available from the NCBI SRA with accession number PRJNA577441 (https://trace.ncbi.nlm.nih.gov/Traces/sra/). The reference genome of *H. melpomene* v2 is available from Lepbase (http://lepbase.org/).

References

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. **487**:94–98.

Andersson S, Dobson HE. 2003. Antennal responses to floral scents in the butterfly *Heliconius melpomene*. *J Chem Ecol*. **29**:2319–2330.

Barbash DA, Siino DF, Tarone AM, Roote J. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc Natl Acad Sci U S A*. **100**:5302–5307.

Bay RA, Arnegard ME, Conte GL, Best J, Bedford NL, McCann SR, Dubin ME, Chan YF, Jones FC, Kingsley DM, *et al.* 2017. Genetic coupling of female mate choice with polygenic ecological divergence facilitates stickleback speciation. *Curr Biol*. **27**: 3344–3349.e3344.

Benson WW, Brown KS Jr., Gilbert LE. 1975. Coevolution of plants and herbivores: passion flower butterflies. *Evolution*. **29**:659–680.

Briscoe AD, Macias-Munoz A, Kozak KM, Walters JR, Yuan F, Jamie GA, Martin SH, Dasmahapatra KK, Ferguson LC, Mallet J, *et al.* 2013. Female behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS Genet*. **9**:e1003620.

Byers K, Darragh K, Fernanda Garza S, Abondano Almeida D, Warren IA, Rastas PMA, Merrill RM, Schulz S, McMillan WO, Jiggins CD. 2021. Clustering of loci controlling species differences in male chemical bouquets of sympatric *Heliconius* butterflies. *Ecol Evol*. **11**:89–107.

Byers K, Darragh K, Musgrove J, Almeida DA, Garza SF, Warren IA, Rastas PM, Kucka M, Chan YF, Merrill RM, *et al.* 2020. A major locus controls a biologically active pheromone component in *Heliconius melpomene*. *Evolution*. **74**:349–364.

Catalan A, Briscoe AD, Hohna S. 2019. Drift and directional selection are the evolutionary forces driving gene expression divergence in eye and brain tissue of *Heliconius* butterflies. *Genetics*. **213**: 581–594.

Conte GL, Schluter D. 2013. Experimental confirmation that body size determines mate preference via phenotype matching in a stickleback species pair. *Evolution*. **67**:1477–1484.

Darragh K, Montejo-Kovacevich G, Kozak KM, Morrison CR, Figueiredo CME, Ready JS, Salazar C, Linares M, Byers K, Merrill RM, *et al.* 2020. Species specificity and intraspecific variation in the chemical profiles of *Heliconius* butterflies across a large geographic range. *Ecol Evol*. **10**:3895–3918.

Darragh K, Orteu A, Black D, Byers KJRP, Szczerbowski D, Warren IA, Rastas P, Pinharanda A, Davey JW, Fernanda Garza SF, *et al.* 2021. A novel terpene synthase controls differences in anti-aphrodisiac pheromone production between closely related *Heliconius* butterflies. *PLoS Biol*. **19**:e3001022.

Darragh K, Vanjari S, Mann F, Gonzalez-Rojas MF, Morrison CR, Salazar C, Pardo-Diaz C, Merrill RM, McMillan WO, Schulz S, *et al.* 2017. Male sex pheromone components in *Heliconius* butterflies released by the androconia affect female choice. *PeerJ*. **5**:e3953.

Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, Merrill RM, Joron M, Mallet J, Dasmahapatra KK, *et al.* 2016. Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3 (Bethesda)*. **6**:695–708.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**:15–21.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol*. **28**: 2239–2252.

Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, Garcia-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, *et al.* 2019. Genomic architecture and introgression shape a butterfly radiation. *Science*. **366**:594–599.

Edelman NB, Mallet J. 2021. Prevalence and adaptive impact of introgression. *Annu Rev Genet*. **55**:265–283.

Elgar MA, Zhang D, Wang Q, Wittwer B, Pham HT, Johnson TL, Freelance CB, Coquilleau M. 2018. Insect antennal morphology: the evolution of diverse solutions to odorant perception. *Yale J Biol Med*. **91**:457–469.

Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, Kawakami T, Kunstner A, Makinen H, Nadachowska-Brzyska K, Qvarnstrom A, *et al.* 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*. **491**:756–760.

Estrada C, Gilbert LE. 2010. Host plants and immatures as mate-searching cues in *Heliconius* butterflies. *Anim Behav*. **80**:231–239.

Fuller SB, Straw AD, Peek MY, Murray RM, Dickinson MH. 2014. Flying *Drosophila* stabilize their vision-based velocity controller by sensing wind with their antennae. *Proc Natl Acad Sci U S A*. **111**:E1182–E1191.

Gadenne C, Barrozo RB, Anton S. 2016. Plasticity in insect olfaction: to smell or not to smell? *Annu Rev Entomol*. **61**:317–333.

Gavrilets S. 2004. *Fitness landscapes and the origin of species*. Princeton (NJ): Princetion University Press.

Grant PR, Grant BR. 2006. Evolution of character displacement in Darwin's Finches. *Science*. **313**:224–226.

Jiggins CD. 2017. *The ecology and evolution of* Heliconius *butterflies*. Oxford: Oxford University Press.

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

Jiggins CD, Wallbank RW, Hanly JJ. 2017. Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns? *Philos Trans R Soc Lond B Biol Sci*. **372**: 1713.

Kamikouchi A, Inagaki HK, Effertz T, Hendrich O, Fiala A, Gopfert MC, Ito K. 2009. The neural basis of *Drosophila* gravity-sensing and hearing. *Nature*. **458**:165–171.

Kautt AF, Kratochwil CF, Nater A, Machado-Schiaffino G, Olave M, Henning F, Torres-Dowdall J, Harer A, Hulsey CD, Franchini P, *et al.* 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature*. **588**:106–111.

Kozak KM, Wahlberg N, Neild AF, Dasmahapatra KK, Mallet J, Jiggins CD. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. *Syst Biol*. **64**: 505–524.

Kronforst MR, Hansen ME, Crawford NG, Gallant JR, Zhang W, Kulathinal RJ, Kapan DD, Mullen SP. 2013. Hybridization reveals the evolving genomic architecture of speciation. *Cell Rep*. **5**: 666–677.

Kronforst MR, Papa R. 2015. The functional basis of wing patterning in *Heliconius* butterflies: the molecules behind mimicry. *Genetics*. **200**:1–19.

Kronforst MR, Young LG, Kapan DD, McNeely C, O'Neill RJ, Gilbert LE. 2006. Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proc Natl Acad Sci U S A*. **103**:6575–6580.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. **35**:1547–1549.

Lamichhaney S, Berglund J, Almen MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerova M, Rubin CJ, Wang C, Zamani N, *et al.* 2015. Evolution of Darwin's Finches and their beaks revealed by genome sequencing. *Nature*. **518**:371–375.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf*. **9**:559.

Lawniczak MK, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, Redmond S, Fulton L, Appelbaum E, Godfrey J, *et al.* 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*. **330**: 512–514.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf*. **12**:323.

Li K, Hong W, Jiao H, Wang GD, Rodriguez KA, Buffenstein R, Zhao Y, Nevo E, Zhao H. 2015. Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. *Proc Natl Acad Sci U S A*. **112**:11905–11910.

Mallet J. 2005. Hybridization as an invasion of the genome. *Trends Ecol Evol*. **20**:229–237.

Mallet J, Dasmahapatra KK. 2012. Hybrid zones and the speciation continuum in *Heliconius* butterflies. *Mol Ecol*. **21**:5643–5645.
Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res*. **23**:1817–1828.
Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol Biol Evol*. **32**:244–257.
Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol*. **17**:e2006288.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **20**: 1297–1303.
Merot C, Salazar C, Merrill RM, Jiggins CD, Joron M. 2017. What shapes the continuum of reproductive isolation? Lessons from *Heliconius* butterflies. *Proc Biol Sci*. **284**:1856.
Merrill RM, Dasmahapatra KK, Davey JW, Dell'Aglio DD, Hanly JJ, Huber B, Jiggins CD, Joron M, Kozak KM, Llaurens V, *et al.* 2015. The diversification of *Heliconius* butterflies: what have we learned in 150 years? *J Evol Biol*. **28**:1417–1438.
Merrill RM, Rastas P, Martin SH, Melo MC, Barker S, Davey J, McMillan WO, Jiggins CD. 2019. Genetic dissection of assortative mating behavior. *PLoS Biol*. **17**:e2005902.
Merrill RM, Van Schooten B, Scott JA, Jiggins CD. 2011. Pervasive genetic associations between traits causing reproductive isolation in *Heliconius* butterflies. *Proc R. Soc. B*. **278**:511–518.
Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*. **323**:373–375.
Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, Baxter SW, Blaxter ML, Mallet J, Jiggins CD. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Mol Ecol*. **22**:814–826.
Naisbit RE, Jiggins CD, Linares M, Salazar C, Mallet J. 2002. Hybrid sterility, Haldane's Rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics*. **161**:1517–1526.
Phadnis N, Orr HA. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science*. **323**: 376–379.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, *et al.* 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. **81**:559–575.
Rieseberg LH. 2009. Evolution: replacing genes and traits through hybridization. *Curr Biol*. **19**:R119–R122.
Rossi M, Hausmann AE, Thurman TJ, Montgomery SH, Papa R, Jiggins CD, McMillan WO, Merrill RM. 2020. Visual mate preference evolution during butterfly speciation is linked to neural processing genes. *Nat Commun*. **11**:4763.
Schulz S, Estrada C, Yildizhan S, Boppre M, Gilbert LE. 2008. An antiaphrodisiac in *Heliconius melpomene* butterflies. *J Chem Ecol*. **34**: 82–93.
Seixas FA, Edelman NB, Mallet J. 2021. Synteny-based genome assembly for 16 species of *Heliconius* butterflies, and an assessment of structural variation across the genus. *Genome Biol Evol*. **13**: evab069.
Staubach F, Lorenc A, Messer PW, Tang K, Petrov DA, Tautz D. 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genet*. **8**:e1002891.
Su G, Morris JH, Demchak B, Bader GD. 2014. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinform*. **47**:8.13.1–24.
Supple MA, Papa R, Hines HM, McMillan WO, Counterman BA. 2015. Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evol Biol*. **15**:204.
Tanoue S, Krishnan P, Krishnan B, Dryer SE, Hardin PE. 2004. Circadian clocks in antennal neurons are necessary and sufficient for olfaction rhythms in *Drosophila*. *Curr Biol*. **14**:638–649.
Ting CT, Tsaur SC, Wu ML, Wu CI. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science*. **282**:1501–1504.
Todi SV, Sharma Y, Eberl DF. 2004. Anatomical and molecular design of the *Drosophila* antenna as a flagellar auditory organ. *Microsc Res Tech*. **63**:388–399.
Van Belleghem SM, Baquero M, Papa R, Salazar C, McMillan WO, Counterman BA, Jiggins CD, Martin SH. 2018. Patterns of Z chromosome divergence among *Heliconius* species highlight the importance of historical demography. *Mol Ecol*. **27**: 3852–3872.
Van Belleghem SM, Cole JM, Montejo-Kovacevich G, Bacquet CN, McMillan WO, Papa R, Counterman BA. 2021. Selection and isolation define a heterogeneous divergence landscape between hybridizing *Heliconius* butterflies. *Evolution*. **75**: 2251–2268.
Van Belleghem SM, Rastas P, Papanicolaou A, Martin SH, Arias CF, Supple MA, Hanly JJ, Mallet J, Lewis JJ, Hines HM, *et al.* 2017. Complex modular architecture around a simple toolkit of wing pattern genes. *Nat Ecol Evol*. **1**:52.
van Doorn GS, Edelaar P, Weissing FJ. 2009. On the origin of species by natural and sexual selection. *Science*. **326**:1704–1707.
van Schooten B, Jiggins CD, Briscoe AD, Papa R. 2016. Genome-wide analysis of ionotropic receptors provides insight into their evolution in *Heliconius* butterflies. *BMC Genomics*. **17**:254.
van Schooten B, Melendez-Rosa J, Van Belleghem SM, Jiggins CD, Tan JD, McMillan WO, Papa R. 2020. Divergence of chemosensing during the early stages of speciation. *Proc Natl Acad Sci U S A*. **117**:16438–16447.
Westerman EL, VanKuren NW, Massardo D, Tenger-Trolander A, Zhang W, Hill RI, Perry M, Bayala E, Barr K, Chamberlain N, *et al.* 2018. *Aristaless* controls butterfly wing color variation used in mimicry and mate choice. *Curr Biol*. **28**: 3469–3474.e3464.
Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R, *et al.* 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science*. **326**:433–436.
Xiang H, Liu X, Li M, Zhu Y, Wang L, Cui Y, Liu L, Fang G, Qian H, Xu A, *et al.* 2018. The evolutionary road from wild moth to domestic silkworm. *Nat Ecol Evol*. **2**:1268–1279.
Yamasaki YY, Kakioka R, Takahashi H, Toyoda A, Nagano AJ, Machida Y, Moller PR, Kitano J. 2020. Genome-wide patterns of divergence and introgression after secondary contact between *Pungitius* sticklebacks. *Philos Trans R Soc Lond B Biol Sci*. **375**: 20190548.
Yan H, Jafari S, Pask G, Zhou X, Reinberg D, Desplan C. 2020. Evolution, developmental expression and function of odorant receptors in insects. *J Exp Biol*. **223**:jeb208215.