Evolution of Gustatory Receptor Gene Family Provides Insights into Adaptation to Diverse Host Plants in Nymphalid Butterflies

Hiromu C. Suzuki[1], Katsuhisa Ozaki[2], Takashi Makino[1], Hironobu Uchiyama[3], Shunsuke Yajima[3,4], and Masakado Kawata[1,*]

[1]Graduate School of Life Sciences, Tohoku University, Sendai, Japan

[2]JT Biohistory Research Hall, Takatsuki, Japan

[3]NODAI Genome Research Center, Tokyo University of Agriculture, Japan

[4]Department of Bioscience, Tokyo University of Agriculture, Japan

*Corresponding author: E-mail: kawata@m.tohoku.ac.jp.

Abstract

The host plant range of herbivorous insects is a major aspect of insect–plant interaction, but the genetic basis of host range expansion in insects is poorly understood. In butterflies, gustatory receptor genes (GRs) play important roles in host plant selection by ovipositing females. Since several studies have shown associations between the repertoire sizes of chemosensory gene families and the diversity of resource use, we hypothesized that the increase in the number of genes in the GR family is associated with host range expansion in butterflies. Here, we analyzed the evolutionary dynamics of GRs among related species, including the host generalist *Vanessa cardui* and three specialists. Although the increase of the GR repertoire itself was not observed, we found that the gene birth rate of GRs was the highest in the lineage leading to *V. cardui* compared with other specialist lineages. We also identified two taxon-specific subfamilies of GRs, characterized by frequent lineage-specific duplications and higher non-synonymous substitution rates. Together, our results suggest that frequent gene duplications in GRs, which might be involved in the detection of plant secondary metabolites, were associated with host range expansion in the *V. cardui* lineage. These evolutionary patterns imply that the capability to perceive various compounds during host selection was favored during adaptation to diverse host plants.

Key words: gustatory receptor, gene duplication, chemoreception, host range expansion, host plant selection, butterfly.

Introduction

Herbivorous insects comprise more than one-third of described species, and many authors claim that host plant associations have been major forces driving insect species diversification (Ehrlich and Raven 1964; Mitter et al. 1988; Wiens et al. 2015). In particular, the host plant range, which refers to the variety of plant species consumed or used as oviposition sites, is considered an important factor in adaptation of herbivorous insects. Several studies have suggested that host range evolution has facilitated species diversification (Janz et al. 2006; Janz and Nylin 2008; Nylin et al. 2014). According to the "oscillation hypothesis" detailed in Janz and Nylin (2008), expansion of the host plant range in an ancestral species would result in a high speciation rate in the following lineage. However, the evolutionary processes of host range expansion in herbivorous insects are poorly understood, because the genetic basis of host plant adaptation has been largely unexplored in most insect taxa.

Host plant selection by ovipositing females is one of the key determinants of host range (Janz and Nylin 1997). Because the larvae of herbivorous insects generally have low dispersal abilities, adult females have to recognize and lay eggs on plants that are suitable for larval diet. In butterflies, adult females can discriminate their host and non-host plants by touching the surface of the plant with their foreleg tarsi (Renwick and Chew 1994). This drumming behavior allows

butterflies to perceive soluble secondary metabolites on the plant, and adult females decide to oviposit or not based on the chemical blend (Nishida et al. 1987; Pereyra and Bowers 1988; Huang et al. 1993). From these observations, chemosensory gene families have been seen as the biggest candidate genes regulating host selection behavior of butterflies. Although the precise genetic mechanisms underlying the preference for specific hosts for oviposition are yet to be elucidated, Ozaki et al. (2011) showed that a gustatory receptor gene (GR) encoded a receptor for a host plant-specific compound (i.e., synephrine), and regulated the oviposition behavior of the Asian swallowtail butterfly, *Papilio xuthus*. Moreover, Briscoe et al. (2013) revealed that many GR genes had female-biased expression patterns in the postman butterfly, *Heliconius melpomene*, suggesting the importance of GRs for host selection by females. Because drumming behavior is widely observed among butterflies, GRs likely play essential host selection roles in other butterfly taxa, which also influences host range expansion.

GRs comprise one of the biggest gene families in insect genomes. This gene family evolved under the birth-and-death model, in which gene repertoires are shaped by multiple gene gains (duplications) and losses (deletions or pseudogenization), and the gene number varies among lineages (Nei et al. 2008; Sánchez-Gracia et al. 2009). Several studies have reported possible associations between the repertoire sizes of the GR family and the varieties of resource use in insects. In *Drosophila*, for example, two specialist species have fewer numbers of GRs than their non-specialist sister species because of accelerated gene losses, which were apparently caused by relaxed selection after specializations on chemically homogeneous environments (McBride 2007; McBride et al. 2007). Differences in the number of GRs have also been observed in the other direction, that is, host generalists: recent studies on genome sequences of extreme polyphagous herbivores in the family Noctuidae, such as the cotton bollworm moth *Helicoverpa armigera* and the tobacco cutworm *Spodoptera litura*, revealed that they have remarkably larger GR repertoires in the genomes compared with other lepidopteran host specialists (Xu et al. 2016; Cheng et al. 2017; Gouin et al. 2017; Pearce et al. 2017). These observations raise the possibility that evolutionary transition from host specialist to generalist in butterflies is accompanied by the increase in the number of GRs by gene duplications. However, the evolutionary relationships between the GRs and host expansion remain uncertain, because previous works did not present genomes for specialist species in Noctuidae, and specialists used for comparisons were all phylogenetically distant (e.g., *Bombyx mori*). Also, the timings of host range expansion within Noctuidae were not clearly understood.

In this article, we analyzed the evolutionary dynamics of GRs in four closely related butterfly species, including the painted lady *Vanessa cardui*, to test our hypothesis that host range expansion is associated with the increase in the repertoire size of GRs. *V. cardui* is one of the most polyphagous butterfly species, with more than 10 plant orders recorded as its hosts (Robinson et al. 2010). Since it has been estimated that host range expansion in *V. cardui* occurred within the genus *Vanessa* (Nylin et al. 2014), genomic analysis among *V. cardui* and related host specialist species provides an opportunity to address relationships between host plant ranges and GRs in butterflies. Specifically, we investigated the numbers of gene gains and losses of GRs within a monophyletic species group, structures and clusters of the GR gene phylogeny, and amino acid substitution rates across the GR gene family.

Materials and Methods

Butterflies and Host Plants

Four butterfly species from the tribe Nymphalini (*V. cardui*, *Vanessa indica*, *Polygonia c-aureum*, and *Araschnia burejana*) were used in this study. *V. cardui* was chosen as the generalist sample because it is the only butterfly species that is both generalist at the level of plant order (i.e., using more than three plant orders, according to Nylin et al. 2014) and available in Japan. Other three specialist species are also commonly found in Japan, and each has a different phylogenetic distance from *V. cardui*. These specialists rely on a single plant order, the Rosales, as their hosts. *V. cardui* was also recorded to utilize Rosales (family Urticaceae), but its primary host plants are from the order Asterales. Butterfly host use data were obtained from an online database HOSTS (Robinson et al. 2010). Nylin et al. (2014) suggested the host range expansion in the ancestor of *V. cardui* occurred after the divergence of the *V. cardui* lineage and the *V. indica* lineage, which was around 20 million years ago (Wahlberg and Rubinoff 2011).

Sampling and RNA Extraction

Adult females were captured in the wild around Sendai, Japan, from 2015 to 2016. They were placed in cages with their host plants for oviposition. After oviposition, the eggs were collected and reared under constant environmental conditions of 25 °C, 16L/8D until eclosion. Three female individuals of each species had all their legs dissected within 2 days of eclosion. The legs of each individual were separately homogenized as soon as possible after dissection, and the total RNA was extracted from the homogenates using a Maxwell 16 LEV Plant RNA kit (Promega Corporation, WI, USA) following the manufacturer's protocol. The qualities of total RNA samples were measured using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA).

RNA-Sequencing and *De Novo* Assembly

Although detection and comparison of the gene repertoires should be based on whole genome sequences, they were not

available for our study species. Instead, we applied RNA-sequencing and *de novo* assembly to identify candidate genes. As insect GRs generally have very low expression levels (Clyne et al. 2000; Ozaki et al. 2011), we focused our sequencing efforts on female legs, in which the most diverse GRs are expressed among body parts of the butterfly (Briscoe et al. 2013), rather than sequencing all tissues evenly at low-sequencing depths. Our strategy for RNA-seq was to combine 100 bp and 300 bp paired-end reads obtained using HiSeq 2500 and MiSeq sequencers (Illumina, CA, USA), respectively, to obtain the longest possible gene sequences. Briefly, paired-end cDNA libraries were constructed using 500 ng of total RNA from each butterfly obtained using the TruSeq RNA Sample Prep Kit v2 (Illumina) according to the manufacturer's protocol. Three libraries were obtained for the HiSeq system and one library for the MiSeq system from each species (i.e., RNA samples from one individual of each species was used for both HiSeq and MiSeq libraries (see supplementary fig. 1, Supplementary Material online for a schematic explanation). All RNA-seq runs were conducted at NODAI genome research center (Tokyo, Japan) between 2015 and 2016, except for a MiSeq library of *A. burejana* that was sequenced at JT Biohistory Research Hall (Osaka, Japan) in 2016.

Low-quality raw RNA-seq reads were removed using the fastq_quality_filter of FASTX-toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/; last accessed May 19, 2018), with a setting of -v -Q 33 -q 20 -p 30. Processed reads were assembled using Trinity 2.1.1 (Grabherr et al. 2011) with the –trimmomatic option on the default settings. All the reads from both HiSeq and MiSeq data were collectively assembled (supplementary fig. 1, Supplementary Material online). Qualities of *de novo* assemblies were listed in supplementary table 1, Supplementary Material online.

Gene Homology Search, Phylogenetic Analysis, and Gene Model Construction

To detect candidate GR genes from the Trinity assemblies, TBLASTN searches (e-value = 1e-05) (https://blast.ncbi.nlm.nih.gov/Blast.cgi; last accessed May 19, 2018) were performed using the amino acid sequences of GRs of *B. mori* (Guo et al. 2017), *Danaus plexippus* (Zhan et al. 2011), and *H. melpomene* (Briscoe et al. 2013) as input queries and our Trinity assemblies as databases. Contig profiles selected in TBLASTN were further tested with BLASTX (e-value = 1e-02) against the NCBI non-redundant protein database (nr). Contigs that matched GRs of other insect species in BLASTX were considered to be candidate GR genes of the focal species. For contigs having multiple isoforms, we chose the longest contig as the representative sequence. Amino acid sequences of candidate GR contigs were predicted using TransDecoder (https://transdecoder.github.io; last accessed May 19, 2018) and Sequence Manipulation Suite (Stothard 2000). Subsequently, we repeated the same procedure using amino acid sequences of the candidate GR contigs

as the query for TBLASTN to search for contigs that were not detected in the first round.

The amino acid sequences of the GR genes of our study species (*V. cardui*, *V. indica*, *P. c-aureum*, and *A. burejana*) and those of five other insect species (*H. melpomene*, Briscoe et al. 2013; *D. plexippus*, Zhan et al. 2011; *P. xuthus*, Ozaki et al. 2011; *B. mori*, Guo et al. 2017; and *Drosophila melanogaster*, Robertson et al. 2003) were aligned using MAFFT 7.273 (Katoh and Standley 2013) with the L-INS-i algorithm. An initial gene phylogeny was constructed using the maximum likelihood method in RAxML 8.2.8 (Stamatakis 2014), followed by bootstrap analyses with 100 replications. We utilized "-m PROTGAMMAAUTO" option to identify the most suitable protein substitution model, and JTT model was assigned.

At this point, however, we observed many cases where one GR gene of the nymphalid species seemed to be fragmented into several contigs in the gene phylogeny. To avoid overestimation of the number of GRs, we searched for contigs that could be integrated into a single gene model. We first identified sets of orthologous genes occupying close positions in the phylogeny. Nucleotide sequences of each gene set were then aligned using MAFFT. Here, we set specific criteria to determine whether the multiple contigs originated from a single gene: 1) If multiple contigs of one species aligned to their orthologous genes without overlaps between contigs, they were combined into a single gene model. The contig identities were later confirmed by PCR and electrophoresis (Supplementary Material online). For a few gene models, sequence gaps were determined by Sanger sequencing. 2) If multiple contigs of the species overlapped in alignments, we calculated synonymous substitution rates (dS) in the overlapping regions, following the method used in Duncan et al. (2014). If dS was <0.19, we collapsed the contigs into a single gene model (i.e., contigs were considered as alleles). Otherwise, we counted those contigs as paralogs. As dS is often considered as an index for divergence times between homologs (Lynch and Conery 2000), what we tried here was collapsing contigs with shorter divergence times. The cut-off value for dS (0.19) was average pairwise dS between *V. cardui* and *V. indica*, calculated from 30 randomly chosen BUSCO genes using codeml in PAML v4.8 (Yang 2007) (supplementary table 2, Supplementary Material online). Because of this methodology, we had to repeat the phylogeny construction process and gene alignment tests until no contigs remained that could be combined with others. Hereafter, we considered each gene model as a unit of GR gene.

After obtaining the final set of GRs, a final gene phylogeny was constructed using the maximum likelihood method in RAxML, with JTT substitution model assigned by "PROTGAMMAAUTO" option. Bootstrap analyses were further carried out with 500 replications. The phylogeny was visualized using iTOL v3 (Letunic and Bork 2016). Based on ligand information of several insect GRs from previous studies, we inferred the ligand affinities of each phylogenetic subfamily.

As chemoreception plays a major role in host selection of butterflies, we also explored other chemosensory gene families expressed in female legs. Homology searches and constructions of gene models were also performed for olfactory receptor (OR), ionotropic receptor (IR), odorant-binding protein (OBP), and chemosensory protein (CSP) gene families in the same manner as GRs.

Comparisons of the Number of GRs

The difference in the number of GRs (i.e., the number of GR gene models) among species was determined. First, we counted the numbers of insect-specific BUSCO (single-copy orthologs conserved among the lineage) genes (v3, Insecta) (Simão et al. 2015) in each assembly to consider the variation in the number of genes due to the *de novo* assembly procedure (supplementary table 3, Supplementary Material online). We calculated the ratio of the number of GRs to that of BUSCO in each assembly and tested the heterogeneity of the frequency among species using the pairwise Fisher test with the Benjamini–Hochberg correction (Benjamini and Hochberg 1995). Other chemosensory genes detected were also analyzed in the same procedure.

Estimation of GR Gain/Loss Events and Birth/Death Rates

We estimated the numbers of gains and losses of GRs along the species tree using Notung 2.8.1.7 (Stolzer et al. 2012), which reconciles a gene phylogeny onto a species tree. Input species tree was manually drawn based on phylogenetic information of species reported in previous studies (Wahlberg 2006; Wahlberg et al. 2009, 2013; Wahlberg and Rubinoff 2011). We used the "–phylogenomics" command implemented in Notung to estimate counts of gains (duplications) and losses (deletions or pseudogenization) of GRs occurring in each phylogenetic branch. We then tested the difference in the ratio of the number of gains and losses among the most recent branches for each species using the pairwise Fisher test with the Benjamini–Hochberg correction (Benjamini and Hochberg 1995). The same analysis was also performed for OR and IR gene family.

We then estimated the gene birth rate (ß) and the gene death rate (∂), defined as the number of gene gains or losses per million years per gene, for each phylogenetic branch after the common ancestor of Nymphalini. Divergence times of lineages were inferred using BEAST 2.4.7. (Bouckaert et al. 2014), based on the nucleotide sequences of five gene regions (COI, EF-a, wingless, RpS5, and NADPH) of 34 species obtained from Genbank (supplementary table 4, supplementary fig. 2, Supplementary Material online). We calculated the gene birth and death rate using the formula developed by Niimura et al. (2014) (supplementary table 5, Supplementary Material online).

However, we admit our analysis of gene gains and losses based on leg transcriptomic data may be inaccurate, because

genes exclusively expressed in other tissues or at other developmental stages would have been ignored from the analysis. In this case, the number of gene gains would be underestimated, and the number of losses would be overestimated. Nevertheless, we believe that the overall pattern of gains and losses was reliable for several reasons. First, legs are one of the primary gustatory-sensing organs for butterflies, thus we assumed that the numbers of GRs expressed in female legs could be used as substitutes for the numbers of the whole GR repertoires. In fact, it has been shown that female legs of *H. melpomene* expressed the most diverse set of GRs among its body parts (50 out of 73 total GRs, 68%) (Briscoe et al. 2013). The positive relationships between the number of expressed genes at a tissue and the total repertoire size have been reported in ORs. For example, *D. melanogaster* expressed 44 out of 62 total ORs (70%), and female *H. melpomene* expressed 67 out of 70 total ORs (96%) only in the antennae, the primary olfactory organ of insects (Couto et al. 2005; Briscoe et al. 2013). Although phylogenetically distant from insects, similar patterns were found for vertebrate ORs expressed at olfactory epithelium, the most important olfactory organ for mammals: 295 out of 366 ORs (80%) and 817 out of 1,154 ORs (70%) were detected from human and mouse, respectively (Zhang et al. 2004, 2007). In contrast, the relationship may not be straightforward for GRs. In *S. litura*, the most GR-rich tissue was larval maxilla, but it expressed only 84 out of 237 total GRs (35%) (Cheng et al. 2017). Adult legs possessed only 20 GRs (8%) (Cheng et al. 2017). However, these percentages might have been underestimated, because only 123 GRs were detected in their transcriptome analysis across all tissues of larvae and adults (Cheng et al. 2017). Taken together, these observations imply that the number of expressed chemosensory genes at an organ is generally influenced by the size of total gene repertoires, but the extent to which it reflects the total repertoires would depend on the importance of chemoreception at the organ. At least, we expect the numbers of expressed GRs in butterfly legs reflect high proportions of the whole repertoire sizes, as leg chemoreception is commonly important among butterflies. Second, the gene gain/loss estimation by Notung is based on the topology of the input gene phylogeny (Stolzer et al. 2012). Since we collected complete GR repertoires from several species, our final GR gene phylogeny, in general, was highly supported by bootstrap analysis. Therefore, we assume that the topology of the phylogeny would not dramatically change even if missing GRs of Nymphalini were included, and the overall pattern of gene gains and losses would not be affected. Further analysis using complete GR gene repertoires needs to be performed to verify this assumption.

Evolutionary Rate Analysis

We took two approaches to infer d_N/d_S ratios across the GR family. First, we chose pairs of orthologous GRs between

Table 1

Number of chemosensory genes (i.e., gene models) and BUSCO genes (v3, Insecta) detected from female leg transcriptomes.

Species	Vanessa cardui	Vanessa indica	Polygonia c-aureum	Araschnia burejana
Host Range	Generalist	Specialist	Specialist	Specialist
GR	50	28	17	45
OR	25	15	17	24
IR	20	21	27	35
OBP	29	27	32	30
CSP	31	36	33	39
BUSCO	1618	1586	1609	1642

V. cardui and *V. indica*, and between *V. cardui* and *A. burejana*, based on phylogenetic positions on the GR gene phylogeny. If a GR gene had a one-to-many relationship with its orthologs, we considered each different combination as one pair. In total, we obtained 24 orthologous pairs between *V. cardui* and *V. indica* (supplementary table 6, Supplementary Material online), and 26 pairs between *V. cardui* and *A. burejana* (supplementary table 7, Supplementary Material online). After nucleotide alignments using MAFFT, we calculated pairwise d_N/d_S ratios with codeml in PAML 4.8 (Yang 2007). Second, we searched for orthologous gene groups consisting three species (*V. cardui*, *V. indica*, and *A. burejana* as the outgroup), and aligned those nucleotide sequences within groups. When one-to-many genes were included in the group, each different combination was considered as an original sample. We obtained 14 orthologous groups (OGs) in total (supplementary table 8, Supplementary Material online). d_N/d_S ratios for each branch in the unrooted three-species tree were calculated using codeml under branch model (Yang 2007). The estimation based on branch model is likely to be more accurate than pairwise, but a problem in our data set was that phylogenetic relationships of Nymphalini GRs were so complicated it was difficult to find clear ortholog groups among three species. Therefore, we integrated pairwise dn/ds estimation to cover across the GR phylogeny. Most GRs in our data set were partial sequences, and consequently, the analysis was performed only in aligned regions. We then tested the variation of mean d_N/d_S ratios using one-way ANOVA for phylogenetic subfamilies and using *t*-test between one-to-one and one-to-many pairs, respectively. For three-species OGs, we also tested the difference in selective pressures between the *V. cardui* lineage and the *V. indica* lineage using paired Wilcoxon test. Any genes showing dS < 0.01 were excluded from the analysis.

Additionally, we searched for signatures of gene conversion among homologous genes of Nymphalini using GENECONV (Sawyer 1989), as gene conversion can modify the estimation of evolutionary rates. Protein sequences alignments of more than three genes (paralogs or homologs in the close positions) were created by MAFFT 2.723 (Katoh and Standley 2013) and used as input data. For genes in which

Table 2

P-values for the Pairwise Fisher's Test Using Benjamini–Hochberg Correction, Testing Differences in the Ratio of the Number of GRs to that of BUSCOs among four species

	V. cardui	V. indica	P. c-aureum
V. indica	0.04275	—	—
P. c-aureum	0.00037	0.12034	—
A. burejana	0.6034	0.11378	0.00217

gene conversion were detected, we re-estimated pairwise d_N/d_S ratios after removing the region of putative conversion from the alignments.

Results

Detection and Comparisons of the Numbers of GRs

Although a variation of total assembled bases was observed, our *de novo* assemblies exhibited generally similar qualities among four species in terms of N50, average contig length, and the numbers of BUSCO (supplementary tables 1 and 3, Supplementary Material online). The number of GRs found in *V. cardui*, *V. indica*, *P. c-aureum*, and *A. burejana* were 50, 27, 17, and 45, respectively (tables 1 and 2). Most of the annotated GRs were partial sequences. The frequencies of GRs in the assembly (i.e., GR/BUSCO ratio) were significantly different between *V. cardui* and *V. indica* (pairwise Fisher test; $P < 0.05$), and between *V. cardui* and *P. c-aureum* (pairwise Fisher test; $P < 0.001$). However, the difference was not observed between *V. cardui* and *A. burejana* (pairwise Fisher test; $P = 0.603$) (tables 1 and 2). For the other chemosensory gene families, we did not observe any significant differences in the number of genes between *V. cardui* and other specialists (tables 1 and 2).

GR Gains and Losses along the Nymphalini Phylogeny

It was estimated that losses of GRs occurred more frequently than gains in most branches in the Nymphalini lineage (fig. 2 and table 3). In contrast, the most recent branch leading to

Table 3

The Estimated Numbers of Gains and Losses in the GR Family, Along With Gene Birth Rates (β) and Death Rates (∂), Among the Most Recent Branches of Four Species.

| Species | Vanessa cardui | Vanessa indica | Polygonia c-aureum | Araschnia burejana |
Host Range	Generalist	Specialist	Specialist	Specialist
No. of Gains	7	1	0	3
No. of Losses	9	25	40	24
Gain/Loss ratio	0.778	0.04	0	0.125
Birth rate (β)	0.00536	0.00101	0	0.00099
Death rate (δ)	0.00689	0.02519	0.02853	0.00793

Table 4

P-values for the Pairwise Fisher's Test Using Benjamini-Hochberg Correction, Testing Differences in Gain/Loss Ratios of GRs.

	V. cardui	V. indica	P. c-aureum
V. indica	0.0079	—	—
P. c-aureum	0.0003	0.4727	—
A. burejana	0.0487	0.6104	0.0916

V. cardui exhibited a relatively high GR gain/loss ratio (seven gains/nine losses) (fig. 2 and table 3). Comparison of GR gain/loss ratios among the most recent branches for each species revealed that the gain/loss ratio in *V. cardui* was significantly higher than those of the other species (pairwise Fisher test; $P < 0.01$ for *V. cardui*–*V. indica*; $P < 0.001$ for *V. cardui*–*P. c-aureum*; $P < 0.05$ for *V. cardui*–*A. burejana*) (table 4). We then estimated the gene birth and death rate at each phylogenetic branch. The branch leading to *V. cardui* showed the highest birth rate ($\beta = 0.00536$), whereas the death rate was similar to that of the branch leading to *A. burejana* (*V. cardui*: $\partial = 0.00689$, *A. burejana*: $\partial = 0.00793$) (table 3 and supplementary table 5, Supplementary Material online). However, expansion of the repertoire itself during host range expansion was not observed (from 52 to 50 GRs, fig. 1). The same analysis for OR and IR families did not indicate any lineage-specific accelerations of gene gains or losses (supplementary fig. 3, Supplementary Material online).

Phylogenetic Analysis

Ligand information of several GRs could be estimated from the gene phylogeny because they were included in the same subfamilies as GRs whose ligands had been reported in past studies, namely, CO_2 (Kwon et al. 2007), sugar (Dahanukar et al. 2007), and fructose receptors (Sato et al. 2011) (fig. 2). These subfamilies were mainly characterized by one-to-one orthologous relationships among distantly related species (fig. 2).

Two major taxon-specific GR subfamilies were observed in the gene phylogeny (fig. 2). One comprised nearly half of all the genes, exhibiting frequent lineage-specific expansions. This subfamily only included GRs of Lepidopteran species (i.e., no *Drosophila* GRs), and was thus named the "Lepidoptera-specific" (LS) subfamily. The other smaller subfamily only contained GRs of butterfly species and was designated the "butterfly-specific" (BS) subfamily (fig. 2). This subfamily originated from the fructose receptor clade, but, unlike the sugar and fructose subfamilies, it was characterized by several lineage-specific gene expansions. Putative species-specific gene duplications in the most recent branches of Nymphalini, including seven putatively duplicated genes in the *V. cardui* lineage, were mapped on either LS or BS subfamilies (fig. 2).

Evolutionary Rate Analysis

We estimated pairwise d_N/d_S ratios between orthologous genes, and those in lineages of *V. cardui* and *V. indica* based on the three-species tree, to examine patterns of selective pressures acting on GRs among phylogenetic subfamilies (i.e., sugar, fructose, CO_2, LS, and BS). For the pairwise analysis, significant differences in d_N/d_S were observed among subfamilies in both combinations of species (one-way ANOVA; $P < 0.05$ for *V. cardui*–*V. indica*; $P < 0.01$ for *V. cardui*–*A. burejana*) (fig. 3). Particularly, LS and BS subfamilies showed higher d_N/d_S ratios, whereas sugar, fructose and CO_2 receptor genes exhibited very low evolutionary rates (fig. 3). Moreover, GRs in one-to-many relationships had higher d_N/d_S ratios compared with one-to-one genes in both combinations (t-test; $P < 0.01$ for *V. cardui* and *V. indica*; $P < 0.05$ for *V. cardui* and *A. burejana*) (fig. 3). The variation of d_N/d_S estimated for lineages of *V. cardui* and *V. indica* showed similar patterns in terms of mean values (fig. 3). However, the differences in d_N/d_S among subfamilies and between one-to-one and one-to-many genes were not significant for the *V. cardui* lineage, mainly because of low sample coverage and a few outlier genes with $d_N/d_S > 2$ (fig. 3). Although the mean d_N/d_S was higher in the *V. cardui* lineage than in *V. indica*, there was no significant difference in evolutionary rates
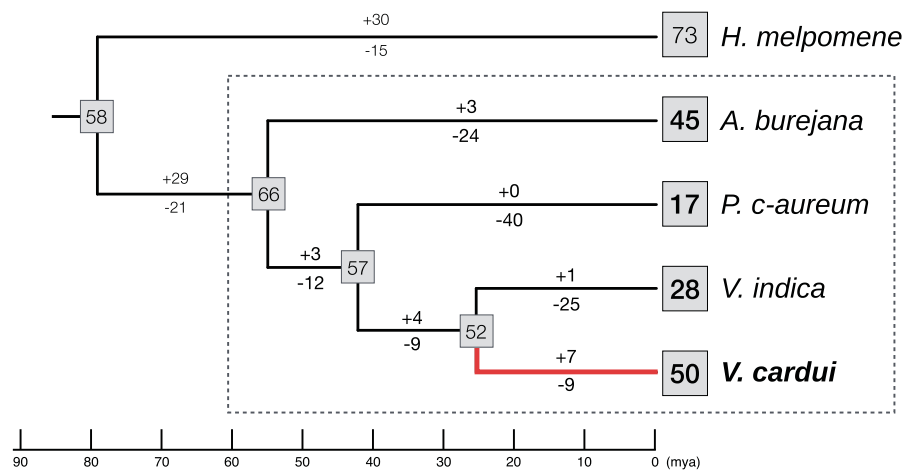
Fig. 1.—Estimation of gene gains and losses in the GR family. The results below the *H. melpomene* lineage are shown in the figure (see supplementary fig. 3, Supplementary Material online for the results on the whole phylogeny). Nymphalini lineage is boxed in dashed line. Numbers at each tip show the current repertoire size of GRs, and numbers at each node indicate an inferred GR repertoire size of a common ancestor. Estimated numbers of gene gains (+) and losses (−) in GRs are shown on branches. The generalist species (*V. cardui*) is labeled in bold. The branch where host range expansion occurred (according to Nylin et al. 2014) was colored in red. Divergence times were estimated with BEAST 2.4.7.

between orthologs of these species (supplementary table 8, Supplementary Material online).

We detected a possible gene conversion between VindGR23a and VindGR23b (supplementary table 9, Supplementary Material online). To examine the effect of gene conversion, we calculated pairwise d_N/d_S ratios for these genes after removing regions of the putative gene conversion. As a result, estimated d_N/d_S ratios were slightly higher without conversion than original values (supplementary table 9, Supplementary Material online).

Discussion

The genetic basis underlying the transition from specialist to generalist has been unexplored in herbivorous insects. In butterflies, GRs play important roles in host plant selection by ovipositing females (Ozaki et al. 2011). It has been suggested that the repertoire sizes of this gene family are associated with the diversity of resource use (McBride 2007; McBride et al. 2007; Xu et al. 2016; Cheng et al. 2017; Gouin et al. 2017; Pearce et al. 2017). Thus, we hypothesized that evolutionary events in the GR family, particularly the increase of the repertoire by gene duplications, were associated with adaptation to diverse host plants by butterflies. To test this hypothesis, we investigated characteristics of GRs among four closely related species from the tribe Nymphalini, including the generalist *V. cardui* and three specialists.

Frequent Gene Duplications of GRs Were Associated With Host Range Expansion

We observed that the gain/loss ratio was particularly higher at the branch leading to *V. cardui* than the other branches within Nymphalini (fig. 1). Moreover, the gene birth rate was highest at the *V. cardui* branch, whereas the death rate was not clearly different from those in the other branches (table 3). These findings imply that gene duplications in the GR family occurred frequently in the course of host range expansion, which is consistent with our hypothesis. However, as Notung simply considers clusters of conspecific genes in the phylogeny (i.e., genes in one-to-many relationships) to be species-specific gene duplication events, it was uncertain whether those one-to-many genes represented actual duplications. Nevertheless, the results showed that one-to-many orthologous pairs had higher d_N/d_S ratios than one-to-one pairs (fig. 3), which is consistent with the observation that duplicated chemoreceptor genes have higher evolutionary rates than non-duplicated homologs (Gardiner et al. 2008; Almeida et al. 2014). Therefore, we expect one-to-many genes in our data set repersented recent duplication events in the GR family.

In contrast, the expansion of the GR repertoire size itself was not observed along with the increased rate of gene duplication (fig. 1). Although this result was not consistent with our hypothesis, we note that it was largely influenced by our study design using partial GR repertoires, in which many GRs were likely to be missing. The algorithm of Notung predicts the number of genes at each node as the number of OGs retained in at least one species below the node (Stolzer et al. 2012). For example, the estimated gene number at the common ancestor of *V. cardui* and *V. indica* is the sum of the numbers of OGs found only in *V. cardui*, those found only in *V. indica*, and those found in both species. Therefore, if partial gene repertoires are placed at the tips, there would be a strong bias that the gene numbers at nodes tend to be larger than those at tips, regardless of the actual changes in
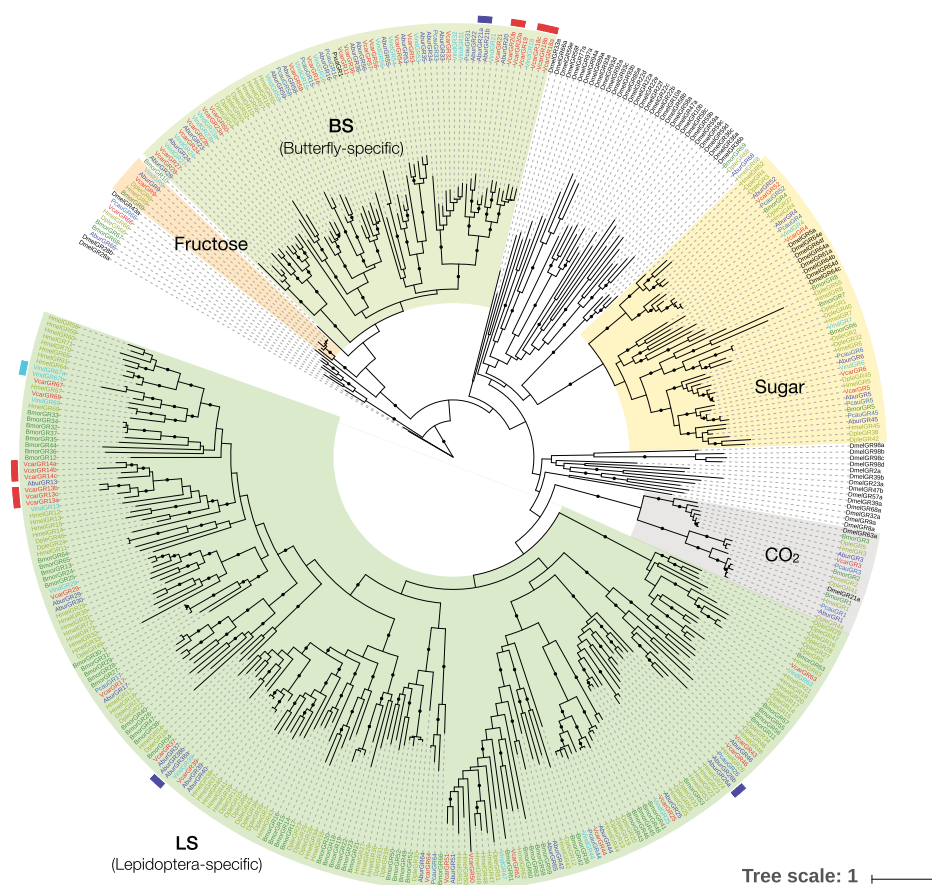
FIG. 2.—Phylogenetic relationships of GRs from nine insect species. The maximum likelihood tree was constructed with RAxML based on amino acid sequences of GRs. Bootstrap analysis was carried out with 500 replicates. Black dots indicate bootstrap support >80%. Subfamilies with putative ligand information are colored in yellow (sugar), orange (fructose), and gray (CO_2). Taxon-specific subfamilies are colored in green (Lepidoptera-specific, LS) and light green (Butterfly-specific, BS). Putative species-specific gene duplications are labeled with colored bars. Vcar, *V. cardui*; Vind, *V. indica*; Pcau, *P. c-aureum*; Abur, *A. burejana*; Hmel, *H. melpomene*; Dple, *D. plexippus*; Pxut, *P. xuthus*; Bmor, *B. mori*; Dmel, *D. melanogaster*.

the GR gene repertoires. Rather, we would like to focus on the fact that the increase in gene birth rate was observed despite the bias toward gene loss. Changes in the GR repertoire sizes before and after host range expansion should be further addressed in whole genome-based studies.

Another concern is that the variation of the GR repertoire sizes between specialist and generalist was not entirely consistent with previous studies (Xu et al. 2016; Cheng et al. 2017; Gouin et al. 2017; Pearce et al. 2017). Specifically, the number of GRs detected from the generalist *V. cardui* was significantly larger than *V. indica* and *P. c-aureum*, but not different from *A. burejana*. Again, this could be caused by the use of partial GR repertoires: Even if there is a significant difference in the repertoire sizes at the whole genome scale, that difference could be reduced when we extracted only genes expressed in female legs. Another explanation is that *A. burejana* lineage, the most distant branch from *V. cardui*, was influenced by other factors that maintained its GR repertoire relatively large. Besides host range, genome size was

proposed as a variable explaining the variation of chemoreceptor gene repertoire sizes among *Drosophila* (Gardiner et al. 2008), which could also be affecting nymphalid butterflies. To our knowledge, however, there is no evidence showing that genome size is significantly different among our study species, thus we could not reach any conclusion about the relatively large GR repertoire for *A. burejana*. At least, the fact that lineages for *V. cardui* and *A. burejana* can be clearly separated by the gene birth rates supported the importance of GRs during host range expansion.

Because we used only a single generalist species, it was difficult to identify whether the higher rate of gene duplications in GRs was specifically associated with host plant range. However, several studies suggest that there could be associations between gene duplications and adaptation to diverse environments in various organisms. For example, species that survive under a variety of climatic conditions tend to have higher proportions of duplicated genes in the genomes among *Drosophila* (Makino and Kawata 2012) and mammals
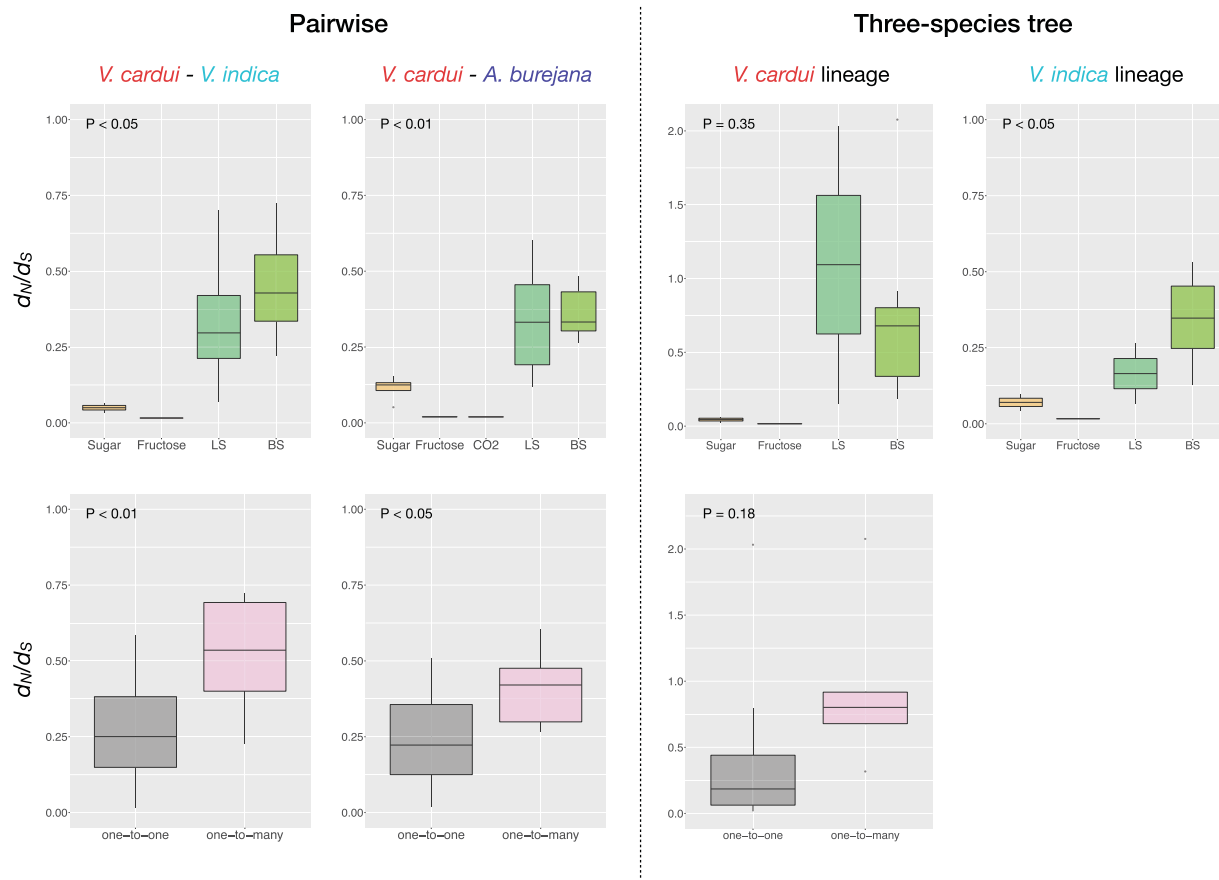
FIG. 3.—Variations of d_N/d_S ratios across the GR family, estimated in two ways. The upper section shows variations among phylogenetic subfamilies, analyzed with one-way ANOVA. The lower section shows comparisons between one-to-one and one-to-many genes, analyzed with t-test. One-to-many genes were not found in the *V. indica* lineage among ortholog sets for the three-species branch model estimation.

(Tamate et al. 2014). In the soil bacteria genus *Frankia*, metabolism-related genes were expanded by gene duplications in a strain infecting diverse plant orders (Normand et al. 2007). Collectively, these observations raise the possibility that gene duplications and gene family expansions are common processes facilitating adaptation to diverse ecological niches across various organisms. Therefore, it would be reasonable to assume that the gene duplication of GRs in the *V. cardui* lineage are associated with adaptation to diverse host plants. However, a more comprehensive study including multiple generalist lineages will be necessary to investigate whether the pattern we observed is peculiar to generalist lineages or specific to *V. cardui*.

Taxon-Specific GR Subfamilies and Perception of Plant Secondary Metabolites

The phylogenetic analysis identified two taxon-specific GR subfamilies, LS and BS (fig. 2). These subfamilies were characterized by frequent lineage-specific gene gains and losses, including genes duplicated in the recent *V. cardui* lineage

(fig. 2), and higher evolutionary rates compared with sugar and CO_2 clades (fig. 3). Based on these results, we assume that genes in LS and BS are involved in the perception of plant secondary metabolites by nymphalid butterflies. In *Drosophila*, the GR subfamily responding to bitter tastants, such as plant-derived secondary metabolites, evolved faster than sugar and CO_2 subfamilies (McBride et al. 2007; Weiss et al. 2011; Ling et al. 2014; Delventhal and Carlson 2016; supplementary fig. 4 and supplementary table 10, Supplementary Material online), which resembles the estimation for LS and BS in our data. Moreover, more than half of lost genes during host specialization in *Drosophila* belonged to the putative bitter receptor clade (McBride et al. 2007; supplementary table 10, Supplementary Material online), which is consistent with the frequent turnover of GR gene repertoires in LS/BS. As secondary metabolite repertoires on plants are likely to be quite variable at both intra- and inter-species scales, divergent evolution of GRs in LS/BS in terms of sequences and gene repertoires might reflect their role as receptors for secondary metabolites. The fact that GRs for synephrine receptor in *P. xuthus* (Ozaki et al. 2011) and those

expressed exclusively in female legs in *H. melpomene* (Briscoe et al. 2013) were included in BS also imply that the subfamily plays some roles in host selection. Although GRs in LS do not have empirical knowledge to support their functions, three GRs of *H. armigera*, which could be classified in LS in our gene phylogeny, responded to extracts of cotton leaves, suggesting their functions as bitter taste receptors (Xu et al. 2016). Overall, these facts imply that GRs for secondary metabolite receptors of nymphalid butterflies are most likely to exist in the taxon-specific subfamilies. We acknowledge that we cannot ultimately verify their roles without functional analysis, thus candidate bitter GRs should be deorphanized in the following studies.

Expansion of Detectable Secondary Metabolites May Be an Adaptation to a Wider Host Plant Range

Our findings together suggest that frequent gene duplications in GRs, which might be responsible for secondary metabolite detection, were associated with host range expansion in Nymphalini. Given the variation of ligand affinities among GRs, the increase in the number of GRs is likely to reflect the repertoire expansion of detectable compounds during host plant selection by ovipositing females. For herbivorous insects, especially butterflies, plant secondary metabolites are classified either as a stimulant or deterrent, which respectively induces or prevents oviposition (Renwick and Chew 1994). Thus, the increase in GRs can be interpreted as the repertoire expansion of detectable stimulants or deterrents, which is possibly an adaptation to diverse host plants.

If the increase in GRs represents an expansion of detectable deterrents, duplication in GRs would not be the direct cause of the host range expansion because adding new deterrents into the repertoire would prevent colonization of novel plant taxa. However, once the host range is expanded by other mechanisms, insects may lay eggs on less suitable species for newly included host plant taxa, because they may not sense the compounds that represent toxicity of the new taxa. In this situation, the increase in GRs would be favored in generalists because it can enhance the accuracy of host selection by expanding repertoires of detectable deterrents. In contrast, GRs would not increase as long as the lineage remains specialized on specific plant taxa. This concept is consistent with an explanation for herbivorous vertebrates tending to have larger numbers of bitter taste receptor genes: recognition of various compounds would be adaptive for accurate discrimination of better plants for food (Li and Zhang 2014).

On the other hand, it is also possible that the increase in GRs represents an expansion of detectable oviposition stimulants. When butterflies include new compounds into their stimulant repertoire, they may start to recognize some plants that were previously avoided as oviposition sites. Thus, in this case, duplication of GRs can trigger host range expansion. Under this assumption, the number of GRs would be increased under selection in species or populations that favor a more extensive host plant range, whereas the number would be stable as long as specialized hosts are favored and maintained.

We are unable to determine which scenario is more plausible at this point. In fact, the two scenarios of adaptation may have occurred simultaneously in the course of host range expansion. To understand the evolutionary relationships between the number of GRs and host range expansion in detail, we have to answer more questions, such as the timings for duplication and fixation of GRs, ligand affinities and behavioral influences of each receptor, and distribution of ligands on candidate host plant taxa in nature.

Conclusion

Although a substantial number of studies have focused on oviposition behavior of butterflies, few studies revealed mechanisms of evolutionary transition between specialists and generalists. Our analysis of nymphalid butterflies, including the generalist *V. cardui*, showed that rapid GR gene duplications have occurred during host range expansion and that there were two taxon-specific GR subfamilies that may be related to secondary metabolite detection. Together, the results imply that the expansion of detectable secondary metabolites during host selection was an adaptation to various host plants by generalist butterflies. These findings would establish the foundation for understanding the evolution of plant–insect interactions, and how it has facilitated species diversification.

Supplementary Material

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

Acknowledgments

Literature Cited

Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. 2014. Family size evolution in drosophila chemosensory gene families: a comparative analysis with a critical appraisal of methods. Genome Biol Evol. 6(7):1669–1682.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300.

Bouckaert R, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 10(4):e1003537–e1003536.

Briscoe AD, et al. 2013. Female behaviour drives expression and evolution of gustatory receptors in butterflies. PLoS Genet. 9(7):e1003620.

Cheng T, et al. 2017. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. Nat Ecol Evol. 1(11):1747–1756.

Clyne PJ, Warr CG, Carlson JR. 2000. Candidate taste receptors in Drosophila. Science 287(5459):1830–1834.

Couto A, Alenius M, Dickson BJ. 2005. Molecular, anatomical, and functional organization of the Drosophila olfactory system. Curr Biol. 15(17):1535–1547.

Dahanukar A, Lei YT, Kwon JY, Carlson JR. 2007. Two Gr genes underlie sugar reception in Drosophila. Neuron 56(3):503–516.

Delventhal R, Carlson JR. 2016. Bitter taste receptors confer diverse functions to neurons. Elife 5:1–23.

Duncan RP, et al. 2014. Dynamic recruitment of amino acid transporters to the insect/symbiont interface. Mol Ecol. 23(6):1608–1623.

Ehrlich PR, Raven PH. 1964. Butterflies and plants: a study in coevolution. Evolution 18(4):586–608.

Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. 2008. Drosophila chemoreceptor gene evolution: selection, specialization and genome size. Mol Ecol. 17(7):1648–1657.

Gouin A, et al. 2017. Two genomes of highly polyphagous lepidopteran pests (Spodoptera frugiperda, Noctuidae) with different host-plant ranges. Sci Rep. 7(1):1–12.

Grabherr MG, et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 29(7):644–652.

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

Huang X, Renwick JAA, Sachdev-Gupta K. 1993. Oviposition stimulants and deterrents regulating differential acceptance of Iberis amara by Pieris rapae and P. napi oleracea. J Chem Ecol. 19(8):1645–1663.

Janz N, Nylin S. 1997. The role of female search behaviour in determining host plant range in plant feeding insects: a test of the information processing hypothesis. Proc R Soc B Biol Sci. 264(1382):701–707.

Janz N, Nylin S. 2008. The oscillation hypothesis of host-plant range and speciation. Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects. University of California Press. p. 203–215.

Janz N, Nylin S, Wahlberg N. 2006. Diversity begets diversity: host expansions and the diversification of plant-feeding insects. BMC Evol Biol. 6:4.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Kwon JY, Dahanukar A, Weiss LA, Carlson JR. 2007. The molecular basis of CO2 reception in Drosophila. Proc Natl Acad Sci U S A. 104(9):3574–3578.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44(W1):W242–W245.

Li D, Zhang J. 2014. Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. Mol Biol Evol. 31(2):303–309.

Ling F, Dahanukar A, Weiss LA, Kwon JY, Carlson JR. 2014. The molecular and cellular basis of taste coding in the legs of drosophila. J Neurosci. 34(21):7148–7164.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290(5494):1151–1155.

Makino T, Kawata M. 2012. Habitat variability correlates with duplicate content of drosophila genomes. Mol Biol Evol. 29(10):3169–3179.

McBride CS. 2007. Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proc Natl Acad Sci U S A. 104(12):4996–5001.

McBride CS, Arguello JR, O'Meara BC. 2007. Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics 177(3):1395–1416.

Mitter C, Farrell B, Wiegmann B. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? Am Nat. 132(1):107–128.

Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet. 9(12):951–963.

Niimura Y, Matsui A, Touhara K. 2014. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. Genome Res. 24(9):1485–1496.

Nishida R, Ohsugi T, Kokubo S, Fukami H. 1987. Oviposition stimulants of a Citrus-feeding swallowtail butterfly, Papilio xuthus L. Experientia 43(3):342–344.

Normand P, et al. 2006. Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography. Genome Res. 17(1):7–15.

Nylin S, Slove J, Janz N. 2014. Host plant utilization, host range oscillations and diversification in nymphalid butterflies: a phylogenetic investigation. Evolution 68(1):105–124.

Ozaki K, et al. 2011. A gustatory receptor involved in host plant recognition for oviposition of a swallowtail butterfly. Nat Commun. 2:542.

Pearce SL, et al. 2017. Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive Helicoverpa pest species. BMC Biol. 15(1):1–30.

Pereyra PC, Bowers MD. 1988. Iridoid glycosides as oviposition stimulants for the buckeye butterfly, Junonia coenia (Nymphalidae). J Chem Ecol. 14(3):917–928.

Renwick JAA, Chew FS. 1994. Oviposition behavior in Lepidoptera. Annu Rev Entomol. 39(1):377–400.

Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. Proc Natl Acad Sci U S A. 100(Suppl 2):14537–14542.

Robinson GS, Ackery PR, Kitching IJ, Beccaloni GW, Hernández LM. 2010. HOSTS - A database of the world's Lepidopteran Hostplants. London: Natural History Museum.

Sánchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major chemosensory gene families in insects. Heredity 103(3):208–216.

Sato K, Tanaka K, Touhara K. 2011. Sugar-regulated cation channel formed by an insect gustatory receptor. Proc Natl Acad Sci U S A. 108(28):11680–11685.

Sawyer SA. 1989. Statistical tests for detecting gene conversion. Mol Biol Evol. 6(5):526–538.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.

Stolzer M, et al. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics 28(18):i409–i415.

Stothard P. 2000. The sequence manipulation suite: javaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102–1104.

Tamate SC, Kawata M, Makino T. 2014. Contribution of nonohnologous duplicated genes to high habitat variability in mammals. Mol Biol Evol. 31(7):1779–1786.

Wahlberg N. 2006. That Awkward Age for Butterflies: insights from the Age of the Butterfly Subfamily Nymphalinae (Lepidoptera: nymphalidae). Syst Biol. 55(5):703–714.

Wahlberg N, et al. 2009. Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. Proc R Soc B Biol Sci. 276(1677):4295–4302.

Wahlberg N, Rubinoff D. 2011. Vagility across Vanessa (Lepidoptera: nymphalidae): Mobility in butterfly species does not inhibit the formation and persistence of isolated sister taxa. Syst Entomol. 36(2):362–370.

Wahlberg N, Wheat CW, Peña C. 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLoS One 8(11):e80875–e80878.

Weiss LA, Dahanukar A, Kwon JY, Banerjee D, Carlson JR. 2011. The molecular and cellular basis of bitter taste in Drosophila. Neuron 69(2):258–272.

Wiens JJ, Lapoint RT, Whiteman NK. 2015. Herbivory increases diversification across insect clades. Nat Commun. 6(1):8370.

Xu W, Papanicolaou A, Zhang H-J, Anderson A. 2016. Expansion of a bitter taste receptor family in a polyphagous insect herbivore. Sci Rep. 6(1):23666.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

Zhan S, Merlin C, Boore JL, Reppert SM. 2011. The monarch butterfly genome yields insights into long-distance migration. Cell 147(5):1171–1185.

Zhang X, et al. 2007. Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. Genome Biol. 8(5):r86.

Zhang X, et al. 2004. High-throughput microarray detection of olfactory receptor gene expression in the mouse. Proc Natl Acad Sci U S A. 101(39):14168–14173.

Associate editor: Belinda Chang