MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 4850–4865 doi: 10.1111/mec.13826

The transcriptome response of *Heliconius melpomene* larvae to a novel host plant

QUAN-YOU YU,*† SHOU-MIN FANG,‡ ZE ZHANG* and CHRIS D. JIGGINS†

School of Life Sciences, Chongqing University, Chongqing 401331, China, †Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK, ‡College of Life Science, China West Normal University, Nanchang 637002, China

Abstract

In the warfare between herbivore and host plant, insects have evolved a variety of defensive mechanisms, including allelochemical transformation and excretion. Several studies have explored the transcriptome responses of insects after host plant shifts to understand these mechanisms. We investigated the plastic responses of *Heliconius melpomene* larvae feeding on a native host *Passiflora menispermifolia* and a less strongly defended nonhost species, *Passiflora biflora*. In total, 326 differentially expressed genes were identified, with a greater number upregulated on the more strongly defended native host. Functional annotation showed that detoxifying enzymes, transporters and components of peritrophic membrane were strongly represented. In total, 30 candidate detoxification genes were differentially expressed, with glutathione S-transferases (GSTs) and UDP-glucuronosyltransferases (UGTs) showing the highest proportion of differential expression, 27.3% and 17.3%, respectively. These differentially expressed detoxification genes were shown to evolve mainly under the influence of purifying selection, suggesting that protein-coding evolution has not played a major role in host adaptation. We found only one gene, *GSTe3*, with evidence of adaptive evolution at H40, which is around the G-site and might alter enzyme activity. Based on our transcriptome and molecular evolution analysis, we suggest that transcriptional plasticity of genes in a herbivore may play an important role in adaptation to a new host plant.

Keywords: detoxification, *Heliconius* butterflies, host plant shift, molecular evolution, transcriptome

Received 30 May 2016; revision received 25 August 2016; accepted 26 August 2016

Introduction

Co-evolution between herbivorous insects and host plants is a relatively frequent phenomenon (Ehrlich & Raven 1964). *Heliconius* butterflies are a diverse system in which to investigate theories of co-evolution (*Heliconius* Genome Consortium 2012; Merrill *et al.* 2015). *Passiflora* and related genera are the sole host plants for *Heliconius* larvae (Benson *et al.* 1976), and there is some phylogenetic association between species groups of *Passiflora* and the *Heliconius* species that feed on them (Benson *et al.* 1976; Brower 1997). Furthermore, there is a wide variety of host use strategies among *Heliconius* species and even between different populations. For example, *Heliconius melpomene* (postman butterfly) is a specialist in Central America, typically using only a single species, either *Passiflora menispermifolia* or *P. oerstedii* (Merrill *et al.* 2013), but is more generalist in other parts of its range. Although there are considerable ecological data on host plant use and diversification of *Heliconius* butterflies, little is known about the molecular and genetic basis for host plant adaptation (Benson *et al.* 1976; Brower 1997; Merrill *et al.* 2013).

In the ecological interaction between plants and herbivorous insects, plants defend themselves against herbivores by synthesizing toxic compounds (or allelochemicals) (Despres *et al.* 2007), by physical barriers (Wybouw *et al.* 2015) or by releasing specific attractants to increase predation of herbivores (Turlings *et al.* 1995). Allelochemicals are a subset of secondary

Correspondence: Quan-You Yu, Fax: +86 23 6567 8491; E-mail: yuqy@cqu.edu.cn and Chris D. Jiggins, Fax: +44 1223 336676; E-mail: c.jiggins@zoo.cam.ac.uk

metabolites which are not required for metabolism (i.e. growth, development and reproduction). Based on their production, allelochemicals are divided into phytoanticipins (constitutive chemicals) and phytoalexins (inducible barriers). A general antiherbivore defence is common to *Passiflora* plants. Many *Passiflora* species are cyanogenic, that is they liberate hydrogen cyanide (HCN) when damaged (Olafsdottir *et al.* 1988; Spencer 1988). In addition, glycosyl flavonoid, alkaloid and phenolic compounds have been reported as the major phytoanticipins of *Passiflora* (Dhawan *et al.* 2004).

To counter the toxic effects of plant toxins, herbivorous insects have evolved a variety of mechanisms to adapt to their host plants, such as chemical transformation, xenobiotic excretion and reducing the absorption of ingested allelochemicals (Barbehenn 2001; Despres *et al.* 2007; Li *et al.* 2007). Chemical transformation is the most important detoxification process and has been well studied. Detoxifying enzymes play roles in allelochemical transformation, which are classified into direct metabolism (phase I) and conjugation (phase II) categories. Cytochrome P450 monooxygenases (P450s) are the principal biochemical system for phase I detoxification in insects and decrease (or, less often, increase) the biological activity of a broad range of substrates through oxidation–reduction or hydrolytic reactions (Despres *et al.* 2007). Hydrolysis reactions are primarily carried out by phase I enzyme carboxylesterases (COEs), which catalyse the transformation of an ester linkage into alcohol and carboxylic acid products (Hosokawa 2008). Phase II conjugation reactions generally follow phase I, including glutathione S-transferases (GSTs) and UDP-glucuronosyltransferases (UGTs) (Li *et al.* 2007). Enzymes from both phases work in concert to transform allelochemicals into water-soluble compounds for eventual excretion by transmembrane proteins that are specific for the conjugated toxins, termed phase III transporters (Reddy *et al.* 2012). Mechanisms for reducing the absorption of allelochemicals are less well studied. In insects, peritrophic membranes (PMs) are extracellular matrices composed of chitin microfibrils and proteins that form thin sheaths around the contents of the midgut lumen (Barbehenn 2001). The strengthened PMs may prevent plant secondary metabolites from entering the insect body (Barbehenn 2001; Celorio-Mancera *et al.* 2013).

In order to understand the molecular mechanisms of insect adaptation to plant defence, global analysis of transcriptome responses is one effective approach. For example, RNA-Seq analysis of two instars of *Polygonia c-album* feeding on *Urtica dioica* or *Ribes uva-crispa* identified digestion- and detoxification-related genes and transcripts coding for structural constituents that showed differential expression (Celorio-Mancera *et al.* 2013). Similar studies have been carried out on the moth *Manduca sexta* (Koenig *et al.* 2015), the spider mite, *Tetranychus urticae* (Dermauw *et al.* 2013; Wybouw *et al.* 2015) and *Drosophila mettleri* (Hoang *et al.* 2015). These studies mainly focused on polyphagous insects (or generalists) feeding on different hosts and detected the expression patterns of differentially expressed genes. However, most herbivorous insects are specialized to one or a few host species (Jaenike 1990), yet transcriptome responses of specialist insects after host plant shifts are less well understood (Ragland *et al.* 2015).

Here, we used *H. melpomene* as a specialist insect to study potential mechanisms of host plant adaptation. This species is a host specialist in Central America and the Pacific slopes of the Andes, but is more generalist in the eastern parts of the range including the Guiana Shield, potentially offering insight into mechanisms of recent adaptation to a specialist lifestyle. Paired-end RNA sequencing of larval guts was performed with *H. melpomene* after rearing on its natural host plant from Central America, *P. menispermifolia*, and an acceptable nonhost plant, *Passiflora biflora*. This experiment takes advantage of the fact that *P. biflora* has a much lower cyanogen content and is considered a 'universal donor' plant that is acceptable to a wide variety of *Heliconius* species (Engler-Chaouat & Gilbert 2007). Thus, we aim to detect primary genes upregulated in the natural host which are necessary for detoxification of the natural host defensive compounds. In addition, however, the two host species differ in their main classes of cyanogen compounds (aliphatic cyanogens in *P. menispermifolia*, complex diglucoside cyclopentenyl cyanogens in the case of *P. biflora*), so differential transcriptional responses to the two hosts might also represent differential detoxification mechanisms for these two types of cyanogenic defence (Engler-Chaouat & Gilbert 2007). We specifically investigated the transcriptional responses of the genes encoding detoxifying enzymes. In addition, the genomes of more than 10 closely related *Heliconius* species have been sequenced (*Heliconius* Genome Consortium 2012; Briscoe *et al.* 2013). Some of these species are specialist on only one or a few *Passiflora* species, others feed on particular subgenera within *Passiflora*, while others are generalists (Brown 1981; Briscoe *et al.* 2013). These genomic data sets enable us to detect evolutionary rates and test for positive selection of detoxifying genes among *Heliconius* species. Our work contributes to an understanding the roles of qualitative regulation and transcriptional plasticity of detoxification-related genes in the host plant adaptation of *Heliconius* butterflies.

Materials and methods

Postman butterflies and host plants

Pupae of *Heliconius melpomene rosina* from Costa Rica were purchased from London Pupae Supplies – The Granary Manor Farm (Oxford, UK). We collected eggs from *H. melpomene* females and allowed them to hatch in plastic cups before transferring them to *Passiflora menispermifolia* and *Passiflora biflora*, respectively. All plants were reared in black earth at 25 °C with a relative humidity of 60% (light: night photoperiod of 16:8 h). Larvae from the same mother butterfly were reared with fresh leaves under controlled conditions in a climate chamber (LD 12:12, 25 °C). The guts of larvae were dissected on the third or fourth day of the fifth instar and food debris removed. Each individual gut was used as one sample and preserved in RNAlater (Ambion, Austin, USA). Five replicate samples were taken for each treatment and stored at −80 °C for RNA isolation.

RNA sequencing

Total RNAs were extracted using Trizol (Life Technologies, Grand Island, NY, USA) and RNeasy® Mini Kit (Qiagen, Valencia, CA, USA). DNA contaminations were removed from the samples by DNase I (Ambion, Darmstadt, Germany). Purified total RNA samples were quantified using a Nanodrop spectrophotometer. The quality of the RNA samples was checked using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). RNA-Seq libraries were prepared and sequenced by BGI Hong Kong (China). The transcriptome libraries were generated using Illumina TruSeq™ RNA Sample Preparation Kit (Illumina, San Diego, CA, USA). The mRNA molecules containing polyA were purified using poly-T oligo attached magnetic beads. The resulting cDNA fragments were purified with QIAQuick® PCR extraction kit (Qiagen, Shanghai, China). The cDNA for 10 samples, five from each treatment, were individually barcoded and run for paired-end sequencing on an Illumina HiSeq™ 2000 Genome Analyzer platform (BGI Hong Kong, China).

RNA-Seq read mapping and identifying differentially expressed gene

Raw reads in FASTQ format were filtered by removing reads that contained adapter sequences and low-quality reads containing >10% poly-N or >50% of bases whose Phred quality scores ≤5. The quality of clean reads was checked with FastQC. The paired-end reads of each sample were mapped to the *H. melpomene* reference genome (Hmel1.1) with TopHat (-g 1 -i 20 –mate-inner-dist 50 –mate-std-dev 50) (Trapnell *et al.* 2009). Gene expression levels were estimated using FPKM (fragments per kilobase of transcript per million mapped reads) values using Cufflinks (Trapnell *et al.* 2010). Differential expression analysis was performed using CUFFDIFF2 (Trapnell *et al.* 2012). False discovery rate (FDR) ≤0.05 and the absolute value of $\log_2$-fold-change ($\log_2$FC) ≥ 1 were used as thresholds to judge significant difference in gene expression. To compare, read counts of each gene were calculated with HTSeq-count (Anders *et al.* 2015), and DEGs were also detected with DESEQ2 (Love *et al.* 2014), with the adjusted *P* value ≤ 0.05 and $\log_2$FC ≥ 1 used as thresholds.

Functional annotation of differentially expressed genes

Homology searches of all the genes of *H. melpomene* were performed by INTERPROSCAN (http://www.ebi.ac.uk/interpro/) and BLAST against the nonredundant (nr) sequence database of the National Center for Biotechnology Information (NCBI) with an *e*-value cut-off ≤$1e^{-5}$. Functional annotations by gene ontology (GO) terms were inferred using BLAST2GO software (Conesa *et al.* 2005). Statistical assessments of GO annotations were performed by gene set enrichment analysis, using Fisher's exact test within BLAST2GO, including corrections for multiple testing using false discovery rate (FDR) (FDR < 0.05). To understand the possible physiological functions of DEGs, statistically enriched pathways were investigated using KOBAS 2.0 (Xie *et al.* 2011). Four pathway databases, KEGG, Reactome, BioCyc and PANTHER, were used in our analysis. A combination of the hypergeometric test and Benjamini–Hochberg FDR correction was conducted to screen the enriched pathways.

Validation of differentially expressed genes by real-time quantitative PCR

Real-time quantitative PCR (qPCR) was performed using a CFX96™ Real-Time PCR Detection System with SYBR Green qPCR Mix (Bio-Rad, USA). The cycling parameters were as follows: 95 °C for 3 min, then 40 cycles of 95 °C for 10 s, and annealing for 30 s (the annealing temperature is listed in Table S1, Supporting information). Five replicate samples for each treatment were used for qPCR analysis, and each sample was analysed twice. We used two housekeeping control genes for normalization, *EF1-α* and *RpS3A*. The results were consistent, so *EF1-α* was used for normalization in subsequent analysis. The relative expression levels were analysed using the $R = 2-\Delta\Delta Ct$ method (Livak & Schmittgen 2001).

Annotations of the detoxifying genes in the whole genome

tBLASTn searches (*E*-value < 0.01) were conducted iteratively against the *H. melpomene* genome (version v1.1) and haplotype scaffolds (*Heliconius* Genome Consortium 2012) using amino acid sequences of *B. mori* GSTs (Yu *et al.* 2008) and UGTs (Huang *et al.* 2008; Ahn *et al.* 2012) as queries. Genomic sequences that show even weak sequence similarity to queries and its flanking regions (2 kb or more long) were extracted. The candidate genes were predicted by FGENESH+ or FGENESH (http://www.softberry.com/). Predicted proteins were used to search the NCBI Conserved Domain Database (Marchler-Bauer *et al.* 2002). Proteins containing the corresponding domains were regarded as putative GSTs and UGTs, respectively (Table S2, Supporting information). Otherwise, they were excluded for further analysis. Identical methods were used to identify the detoxification genes in the genome (v3 scaffolds, http://monarchbase.umassmed.edu/) of monarch butterfly, *Danaus plexippus* (Zhan *et al.* 2011). Chromosomal assignments were based on published mapping of scaffolds in the *H. melpomene* reference genome (*Heliconius* Genome Consortium 2012). The validated GSTs and other detoxification genes of *H. melpomene* were used for tBLASTn (*E*-value < 0.01) search against the genome assemblies of other *Heliconius* species, including *H. pachinus*, *H. timareta*, *H. wallacei*, *H. hecuba*, *H. doris*, *H. hierax*, *H. xanthocles*, *H. telesiphe* and *H. clysonymus*. To improve the quality of the sequences found, the coding sequences were confirmed by read mapping in CLC Genomics Workbench v. 6.5.1. Conservative parameters (mismatch, insertion and deletion cost of 3; length fraction and similarity fraction of 0.9) were used to prevent mis-mapping of paralogous sequences. All read-mappings were inspected by eye.

Phylogenetic analysis and gene nomenclature

Multiple sequence alignments of amino acids were aligned using Muscle (Edgar 2004). Phylogenetic trees were reconstructed using the neighbour-joining (NJ) and maximum-likelihood (ML) method implemented in MEGA 5.0 (Tamura *et al.* 2011). The pairwise deletion option was used in the tree reconstruction, and the accuracy of the tree topology was assessed by bootstrap analysis with 500 resampling replicates. Positions that have a high percentage of gaps (>70%) were trimmed. The ML and NJ trees were consistent; only the NJ tree is presented in subsequent analysis. Based on the phylogeny, *H. melpomene* GSTs were named according to silkworm GSTs (Yu *et al.* 2008). According to the current nomenclature guidelines of the UGT Nomenclature Committee (Mackenzie *et al.* 1997), UGT families, indicated by numbers, and subfamilies, indicated by capital letters, are defined at 40% and 60% amino acid identity, respectively.

Molecular evolution of detoxifying genes in Heliconius butterflies

Evolutionary rates. The phylogenetically related members and reciprocal best BLAST hits between species were identified as orthologous genes. N- and C-terminal domains of GSTs were identified using Conserved Domain Search (CD-search) (Marchler-Bauer *et al.* 2002). For further analysis, only genes in which each terminal region of GSTs contained higher than 80% total length of N- or C-terminal domain, or >50% for the P450s, UGTs and COEs genes, were retained (Table S3, Supporting information). Only orthologous genes identified in five or more *Heliconius* species were included. Nucleotide sequences were aligned using PAL2NAL (Suyama *et al.* 2006) to construct a multiple codon alignment from the corresponding aligned protein sequences. All alignments were manually inspected. The tree topology supplied for CODEML in PAML (Yang 2007) followed the species tree in Fig. S6 (Supporting information). To evaluate variation in selective pressure among sites, the site-specific 'discrete' model (M3, $K = 3$) of CODEML was used to estimate three different ω values and its site proportions. The rates of synonymous (d_N) and nonsynonymous (d_S) substitutions and ω values ($\omega = d_N/d_S$) among all pairwise comparisons were calculated by the YN00 program implemented in the PAML 4.5 package (Yang 2007).

Detection of positive selection. To detect site-specific positive selection, the CODEML program as implemented in PAML4.5 (Yang 2007) was used. Model M0 (one ratio), M1a (nearly neutral), M2a (positive selection), M7 (β distribution) and M8 (β distribution and ω) were run for each set of orthologous (Table S3, Supporting information). To avoid being trapped at local optima, three different initial ω-values of 0.5, 1.1 and 2.0 were used in the estimation of the log-likelihood for model7 and model8 (Low *et al.* 2007). Models M2a vs. M1a and M8 vs. M7 were compared using likelihood ratio tests (LRT), in which twice the log-likelihood difference (2ΔlnL) was compared with the distribution of the chi-square statistic to test whether the neutral model should be rejected. Bayes Empirical Bayes was used to calculate the posterior Bayesian probability, sites with probability >95% and $\omega > 1$

inferred to be under positive selection (Yang *et al.* 2005).

Results

Transcriptome characteristics in the larval gut of Heliconius melpomene

The lepidopteran gut plays key roles in nutrient digestion and allocation. The guts of the fifth-instar larvae reared with the host plants *Passiflora menispermifolia* and *Passiflora biflora* were dissected and used for RNA sequencing. After filtering low-quality reads and trimming adapters, the clean reads of 5 replicate samples for each treatment were mapped to the *Heliconius melpomene* reference genome with TopHat (Trapnell *et al.* 2009). The expression level of each gene was estimated by Cufflinks (Trapnell *et al.* 2010). To discard transcript models that had no read coverage or low coverage, a threshold was set and only transcripts with FPKM ≥ 1 were considered for expression. In total, 7993 genes were expressed in the gut samples, of which 600 genes were expressed at high levels (hereafter 'highly expressed') (FPKM ≥ 100). GO assignments were used to classify the functions of the genes expressed and highly expressed in the gut. Most of the genes expressed in the gut (GEG) and genes highly expressed in the gut (GHG, FPKM ≥ 100) have a similar distribution of GO terms (Fig. S1, Supporting information). Based on the molecular function GO annotation, the most abundant GO terms were 'catalytic activity' (47.1% for GEG, 57.6% for GHG) and 'binding' (57.7% for GEG, 43.8% for GHG). The functional categories enriched in the GEG and GHG subsets were related to nutrient digestion and allocation, and xenobiotic detoxification in the gut.

Transcriptome response of Heliconius melpomene larvae after host plant shift

To explore the physiology of the plastic response of *H. melpomene* to a novel host environment, newly hatched caterpillars were reared on either their native host plant, *P. menispermifolia*, or the nonhost plant, *P. biflora*, and their guts dissected in the fifth instar. A previous study indicated that larval growth rates of *H. melpomene* are similar on five species of *Passiflora*, including *P. biflora* (Smiley 1978). Extracted RNA from each individual gut represented one sample, and five replicate samples for each treatment were used for RNA sequencing, respectively. After quality control, a total of ~338 million, 100-bp clean reads were obtained (Table S4, Supporting information) from ten guts of the fifth-instar caterpillars reared on host and nonhost plants. We mapped the clean reads to the *H. melpomene* reference genome and identified significantly differentially expressed genes (DEGs) (absolute value of $\log_2$ FC ≥ 1; FDR-corrected $P < 0.05$) using CUFFDIFF2. In total, 326 genes were characterized as being significantly differentially regulated in the guts after host-feeding treatment (Table S5, Supporting information), of which 173 genes were upregulated in larvae reared on the native host. To validate the RNA-Seq data, qPCR was performed for 14 DEGs. The relative expression levels of the selected DEGs are shown in Fig. 1A. There was a strong positive correlation (Pearson correlation coefficient $r = 0.939$, $P < 0.001$) between RNA-Seq and qPCR data (Fig. 1B), supporting comparisons based on RNA-Seq data alone.

To analyse the functions of the DEGs, Gene Ontology (GO) enrichment analysis was performed using Fisher's exact test in BLAST2GO (Conesa *et al.* 2005). Based on GO classification, 269 DEGs (82.5%) could be automatically annotated with GO terms. A total of 27 GO categories were over-represented when compared to the GOslim categories of all genes in the *H. melpomene* genome (Table S6, Supporting information). The GO terms, catalytic activity (GO:0003824; $n = 153$) and metabolic process (GO:0008152; $n = 149$), contained the highest number of genes. Typical stress-induced GO terms were identified, such as 'response to stress' and 'response to external stimulus'. To obtain more insight into the possible physiological functions of the DEGs, statistically enriched pathways were investigated using KOBAS 2.0 (Xie *et al.* 2011). The enriched pathways with gene number ≥6 were shown in Fig. S2 (Supporting information). This shows that the pathways related to xenobiotic detoxification metabolism, transmembrane transport and nutrient metabolism were over-represented. The significantly enriched detoxification pathways mainly included 'metabolism of xenobiotics by cytochrome P450', 'drug metabolism-cytochrome P450', 'drug metabolism-other enzymes' and 'glutathione metabolism' (Fig. S2, Supporting information).

To identify the most important mechanisms involved in the response, differentially expressed genes were used for a tBLASTn search (*E*-value < 1e-5) against the nonredundant sequence database of National Center for Biotechnology Information (NCBI). Some important functional categories were manually grouped, such as nutrient digestion, detoxification enzymes, transporter, peritrophic matrix (PM) biosynthesis (Table S5, Supporting information). Based on the functional annotations for the DEGs, 16 genes encoding nutrient digestion enzymes were found, of which 12 were trypsin-like proteinases. These digestion enzymes might play important roles in phenotypic adaptation to the nonhost plant. We found 30 detoxifying genes that were

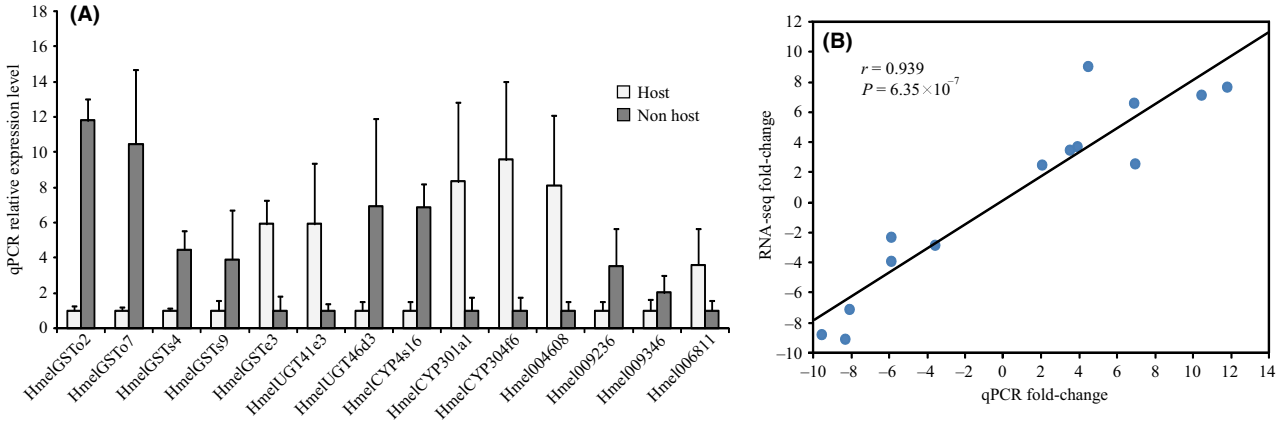


Fig. 1 Quantitative real-time PCR validation of the differentially expressed genes and comparison with RNA-Seq data. The relative expression of a candidate gene was normalized against *EF1-α*. The data are the average ± standard error of five independent replicated qPCR experiments.

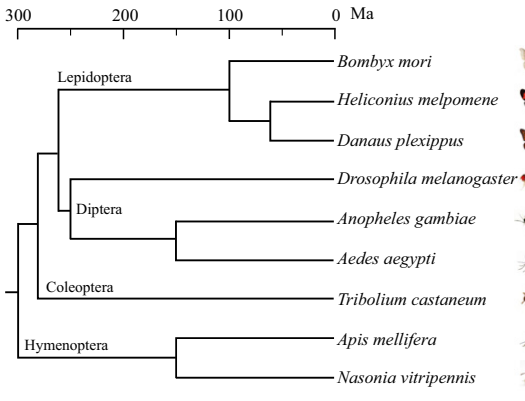
significantly differentially regulated following the host plant shift, which belonged to four major detoxification supergene families, P450s, GSTs, UGTs and COEs. This shows that 10 of 12 phase I detoxifying (P450s and COEs) DEGs were upregulated in larvae reared on native host plant. In contrast, only one (*HmelGSTe3*) of 9 differentially expressed GSTs was upregulated on the native host (Table S5, Supporting information). Twenty-nine transporter genes were identified as DEGs, in which more than half were upregulated in native host-reared larvae. Functional annotation showed that 12 of the 29 transporters belonged to the major facilitator superfamily (MFS). Previous studies suggested that MFS might be key players facilitating insect adaptation and survival in response to new diets (Dermauw *et al.* 2013). Chitin and peritrophin are the major components of the peritrophic matrix (Lehane 1997). One chitin synthase and eight chitin-degrading enzymes (chitin deacetylase and chitinase) responded to the host plant shift. Six genes encoding peritrophin also showed differential expression. This regulation of genes related to PM biosynthesis might affect the penetration of allelochemicals in the gut. These results suggest that multiple gene categories are involved in the defence network for plant secondary metabolites in *H. melpomene*.

Specific responses of detoxifying enzymes to plant allelochemicals

Allelochemicals are the major defence chemicals in plants and are also therefore the selective agents on the detoxification systems of insect herbivores (Li *et al.* 2007; Ragland *et al.* 2015). Accordingly, we observed major plastic changes in expression of detoxification-related genes after the host plant shift (Fig. S2, Table S5, Supporting information). Based on functional annotations of the DEGs, we found 9, 9, 8 and 4 members of GSTs, UGTs, P450s and COEs, respectively, that were significantly differentially expressed (Table S5, Supporting information). Generally, GSTs and UGTs families contained fewer genes than P450s and COEs in insect genomes (Ranson *et al.* 2002; Ahn *et al.* 2012). However, GST and UGT gene families have more members differentially regulated by the host plant shift in *H. melpomene*. Thus, it would be interesting to understand whether the two families were expanded in the genome of *Heliconius* butterflies and the role of duplicated genes in host plant adaptation.

Using the amino acid sequences of silkworm GSTs (Yu *et al.* 2008) and UGTs (Huang *et al.* 2008; Ahn *et al.* 2012) as queries, we manually annotated the detoxification genes through a local tBLASTn search in the *H. melpomene* reference genome (*Heliconius* Genome Consortium 2012). In total, 33 GSTs and 52 UGTs were identified in the *H. melpomene* genome (Fig. 2). GSTs had the highest percentage (9 of 33, 27.3%) of differential expression, followed by UGTs (9 of 52, 17.3%). For P450s, one hundred genes have been found in *H. melpomene* (Chauhan *et al.* 2013). One new candidate (*HMEL004608*) was found in our analysis (Table S5, Supporting information). Thus, only 7.9% of P450 genes showed differential expression after the host plant shift.

For comparison, detoxification-related genes have also been characterized in another butterfly, *D. plexippus* (Fig. 2). Phylogenetic analysis was performed for GSTs and UGTs of lepidopteran insect genomes (Figs 3 and 4). The GSTs and UGTs of *H. melpomene* and *D. plexippus* were named using published nomenclature (Yu *et al.* 2008; Ahn *et al.* 2012). Interestingly, it was not insect-specific delta and epsilon classes, which are



	UGT	GST							
		Total	De	Ep	Om	Si	Th	Ze	Un
Bombyx mori	45	23	4	8	4	2	1	2	2
Heliconius melpomene	52	33	2	7	8	11	1	2	2
Danaus plexippus	49	23	2	8	3	5	1	2	2
Drosophila melanogaster	34	37	11	14	5	1	4	2	0
Anopheles gambiae	26	28	12	8	1	1	2	1	3
Aedes aegypti	34	27	8	8	1	1	4	1	3
Tribolium castaneum	43	35	26	2	2	7	2	1	1
Apis mellifera	12	10	1	0	1	4	1	1	0
Nasonia vitripennis	58	19	5	0	2	8	3	1	0

Fig. 2 Gene numbers in multigene families related to detoxification in the genomes of insects. The species tree was modified from Honeybee Genome Sequencing Consortium (2006). The divergence time of *Bombyx mori*, *Danaus plexippus* and *Heliconius melpomene* was from Pringle *et al*. (2007). Gene numbers of the other detoxification enzymes were obtained from previous literature (Claudianos *et al*. 2006; Strode *et al*. 2008; Tribolium Genome Sequencing Consortium 2008; Yu *et al*. 2008; Oakeshott *et al*. 2010; Ahn *et al*. 2012). De, delta; Ep, epsilon; Om, omega; Si, sigma; Th, theta; Ze, zeta; Un, unclassified class.

known to be related to insecticide resistance and detoxification in insects (Li *et al*. 2007), but rather the omega and sigma classes that are more ubiquitously distributed in all organisms, that were most commonly duplicated in the *H. melpomene* genome (Figs 2 and 3). *H. melpomene* is the first species in which omega and sigma classes both showed expansion (Fig. 2). The intron positions of the members of the ubiquitous sigma and omega classes are highly conserved and show very strong class specificity (Fig. S3, Supporting information). Furthermore, omega and sigma classes were mainly clustered on chromosome 7 and scaffold HE669239, respectively (Fig. S4, Supporting information). RNA-Seq data showed that most of the significantly differentially regulated GSTs belonged to the sigma (4) and omega (3) classes (Table S5, Supporting information).

Compared with Diptera, the UGTs were more commonly duplicated in Lepidoptera (Fig. 2). Phylogenetic analysis showed that most of them have duplicated through species-specific expansion (Fig. 4). In *H. melpomene*, UGT members were clustered and mainly located on chromosomes 9, 10, 16, 17 and 19 (Fig. S4, Supporting information). UGT33 is the largest family and contained 24 members (Fig. 4). Based on the analysis of segmental and tandem duplication by MCScanX (Wang *et al*. 2012), segmental duplication mainly occurred in the early expansions, which resulted in spread of UGT33 among different chromosomes or regions (Fig. S5A, Supporting information). Tandem duplication produced more recent copies (Fig. S5A, Supporting information). These duplicated UGT33s share a common signature motif (FhTQhGLQST xExxxxxVPhhxxPhxxDQ) (Mackenzie *et al*. 1997) with UGT33 members from other lepidopteran species (Fig. S5B, Supporting information). Transcriptome data indicated that five of nine differentially expressed UGTs belonged to the UGT33 subfamily (Table S5, Supporting information).

Evolutionary rate and positive selection of detoxification genes among Heliconius butterflies

Molecular evolution of the differentially expressed detoxifying genes was studied in 10 closely related *Heliconius* species. In total, 24 orthologous gene sets of 30 differentially expressed detoxifying genes were analysed (Table S3, Supporting information). The ω values were calculated by the YN00 program in the PAML (Yang 2007), and a plot of d_N/d_S was shown in Fig. 5. All the genes have evolved mainly under the influence of purifying selection ($\omega < 1$, Fig. 5). There was no evidence for different evolutionary rates between differentially expressed GSTs and the GSTs (oGSTs) unaffected by the host plant shift (Fig. 5). In addition, GST proteins consist of two well-defined domains, the N-terminal domain that binds reduced glutathione (GSH) and the C-terminal domain that binds the hydrophobic substrates (Ranson & Hemingway 2005). The comparison indicated that the selective constraint on the C-terminus of oGSTs was more relaxed than on the N-terminus (*t*-test, $P \ll 0.01$), while the selective pressure on the C-terminus and N-terminus was similar among the differentially expressed GSTs (Fig. 5).

To detect the evolutionary rates of different sites, CODEML (model 3 with $K = 3$) implemented in PAML was used to calculated three different ω values and the corresponding proportions of sites. The results

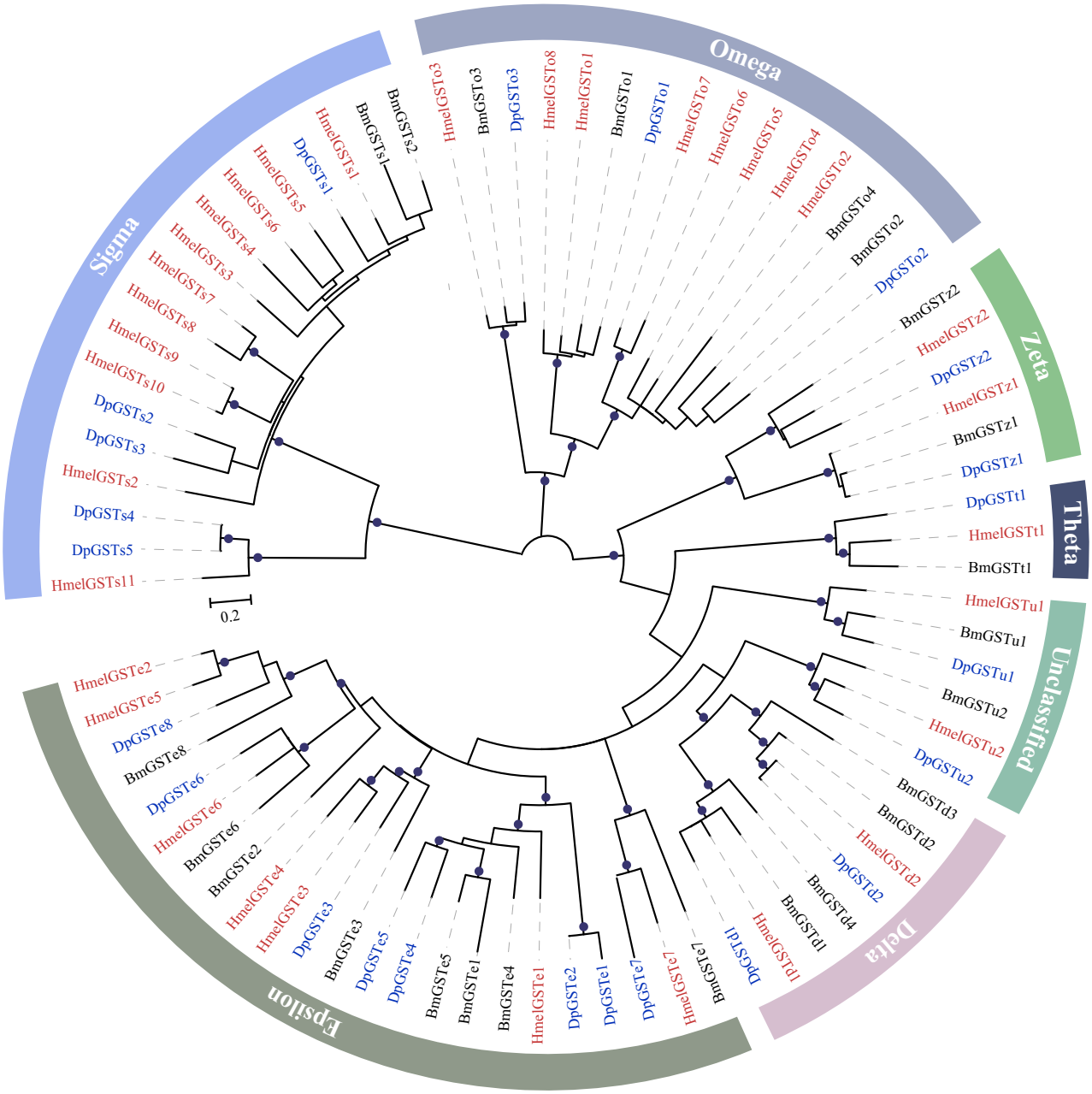


Fig. 3 Phylogenetic tree of GSTs in *Heliconius melpomene*, *Bombyx mori* and *Danaus plexippus*. The tree was reconstructed with MEGA5.0 using neighbour-joining (NJ), Jones–Taylor–Thornton (JTT) substitution model and 500 replicates. The dots on the nodes of phylogenetic tree indicate bootstrap values higher than 0.8. GST sequences of *Bombyx mori* were retrieved from Yu *et al.* (2008). GST sequences of *H. melpomene* and *D. plexippus* are in Table S2 (Supporting information).

indicated that some of the detoxification genes might have positively selected sites ($\omega > 1$) (Table S7, Supporting information). We used two standard PAML comparisons (model M2a vs. M1a and M8 vs.M7) to test whether an adaptive model fits the data better than the neutral model. Only *GSTe3* was found to be significant with Bayes Empirical Bayes (BEB) method and $P < 0.01$ (Table 1). The amino acid state of the positively selected site in N-terminal region of GSTE3 was shown in Fig. 6A. As shown in Fig. 6A, a histidine/alanine/glycine substitution at site 40 has occurred across the six *Heliconius* species. The tertiary protein structure of HmelGSTE3 was predicted by homology modelling (Fig. 6B). The positively selected site (H40) was located near the glutathione-binding site (G-site) in N-terminal domain.

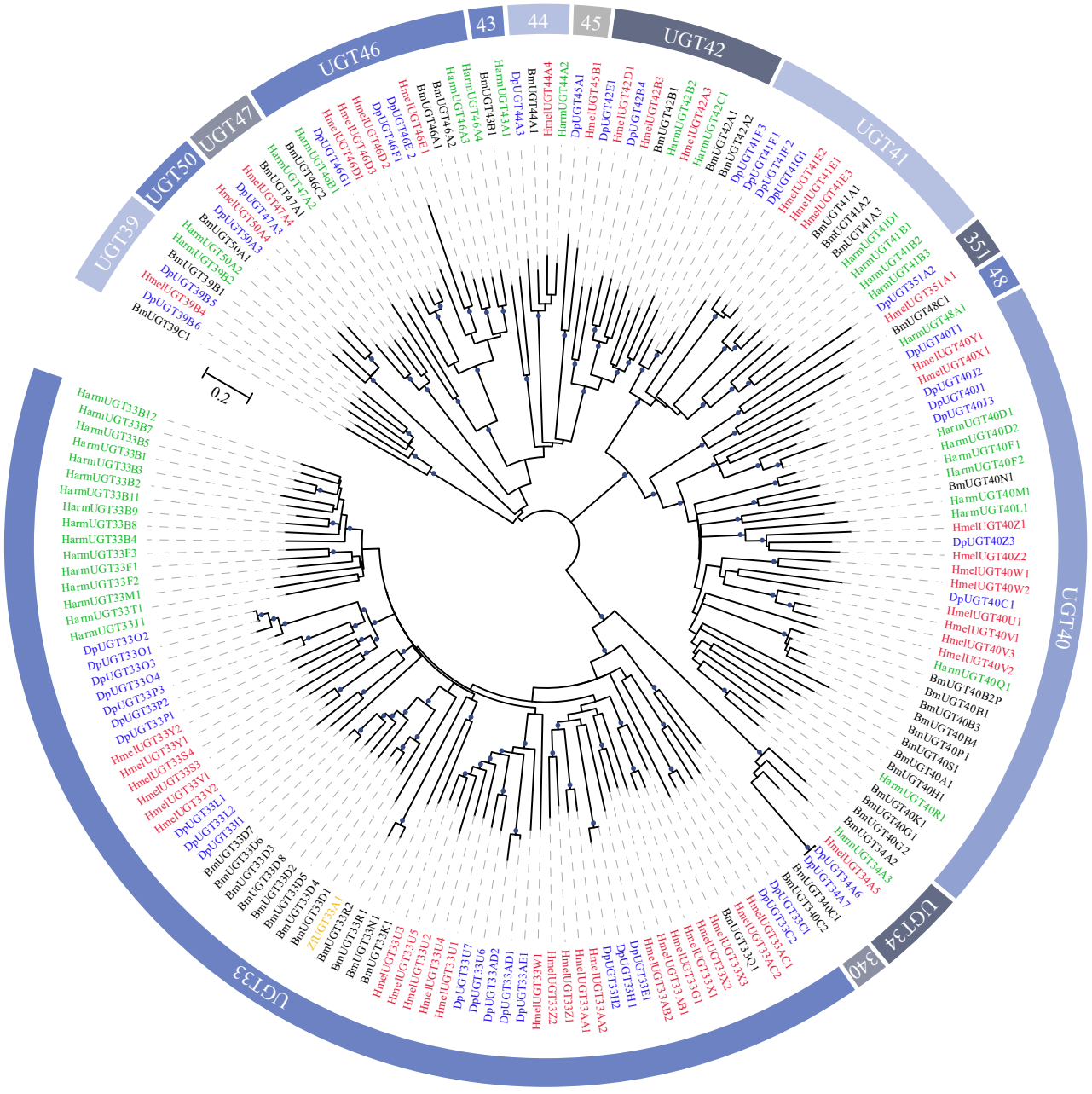


Fig. 4 Phylogenetic tree of UGTs in *Heliconius melpomene*, *Danaus plexippus*, *Bombyx mori* and *Helicoverpa armigera*. The dots on the nodes of phylogenetic tree indicate bootstrap values higher than 0.8. UGT sequences of *B. mori* and *H. armigera* were retrieved from Huang *et al.* (2008) and Ahn *et al.* (2012). UGT sequences of *H. melpomene* and *D. plexippus* were from Table S2 (Supporting information).

Discussion

Differentially expressed genes

Digestive enzymes. We anticipated that digestion-related genes of herbivorous insects might respond to exposure to different hosts. Previous studies suggested that serine proteases are key enzymes allowing larvae to adapt to many different diets (Celorio-Mancera *et al.* 2013; Chikate *et al.* 2013). In this study, 16 digestion-related genes showed plastic responses in *Heliconius melpomene* larvae after the host plant shift (Table S5, Supporting information). Pfam domain searches indicated that 12 of these genes contain a trypsin domain (PF00089.21), and correspond to serine proteases. In addition, another three genes that correspond to carboxypeptidase, alpha-glucosidase and gamma-glutamyl hydrolase (Table S5, Supporting information) were identified. These

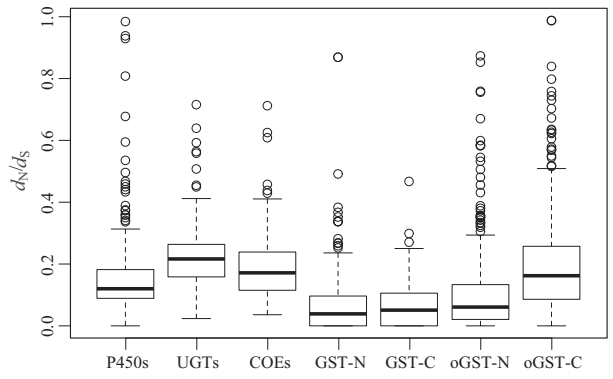


Fig. 5 Box plot of the d_N/d_S values of the detoxification genes in *Heliconius* butterflies. In total, 24 orthologous gene sets of differentially expressed detoxifying genes were used for evolutionary rate analysis. To compare the evolutionary rates between differentially and nondifferentially expressed detoxifying genes, 21 nondifferentially expressed GST (oGST) genes were characterized in a further nine *Heliconius* species (Table S3, Supporting information). GSTs were divided into N-terminus (GST-N) and C-terminus (GST-C) regions to compare evolutionary rates in different functional domains.

digestion-related genes might play important roles in adaptation to a novel host plant.

Detoxification genes. RNA-Seq is an effective means to identify candidate detoxification genes related to host plant adaptation (Celorio-Mancera *et al.* 2013; Dermauw *et al.* 2013; Ragland *et al.* 2015; Wybouw *et al.* 2015). In the spider mite *Tetranychus urticae*, 8.1% of all detoxification genes were differentially expressed following adaptation to tomato or bean plants over 30 generations, including 10 P450s, 4 COEs, 5 GSTs and 9 UGTs (Wybouw *et al.* 2015). For the four detoxification gene families, GSTs had the highest percentage of differential expression (15.6%), followed by P450s (13.3%), UGTs (12.0%) and COEs (6.0%). In contrast, a recent study of *Manduca sexta* showed limited responses of detoxification genes to host plant use (Koenig *et al.* 2015). In our study, 30 detoxifying genes responded significantly to the host plant shift (Table S5, Supporting information). GSTs had the highest percentage (27.3%), followed by UGTs (17.3%) and P450s (7.9%). Compared with the spider mite, GSTs and UGTs have a higher percentage of differential regulation in *Heliconius melpomene*.

Although we might expect detoxification genes to be upregulated in order to deal with the allelochemicals from a nonhost plant, in this case *Passiflora biflora* contains lower concentrations of cyanogenic compounds and is therefore often considered a less toxic host plant (Engler-Chaouat & Gilbert 2007). It is therefore perhaps not surprising that almost all of the differentially expressed P450s (7 of 8) and COEs (3 of 4) were upregulated in larvae reared on the native host plant (Table S5, Supporting information). For some species including *H. melpomene*, *Passiflora* natural product diversity does not cause major deleterious effects on *Heliconius* growth suggesting that larval feeding specialization is not primarily driven by biochemical co-evolution (Smiley 1978; Smiley & Wisdom 1985). *Heliconius–Passiflora* interactions are therefore mediated primarily by adult female butterflies correctly identifying suitable host plants for oviposition (Briscoe *et al.* 2013). So, these differentially expressed phase I detoxifying genes showed higher constitutive expressions and might be responsible for routine detoxification of allelochemicals in *H. melpomene* on its natural host in Central America.

Cytosolic GSTs are important enzymes involved in detoxification of various plant xenobiotics, which can be induced by xanthotoxin, indoles and flavonoids (Li *et al.* 2007). Previous studies indicated that insect-specific delta and epsilon classes primarily mediated allelochemical tolerance and insecticide resistance (Li *et al.* 2007). In our study, nine GST genes showed plastic responses to the host plant shift in *H. melpomene*, two of

Table 1 Parameters and maximum-likelihood values under model estimates for *GSTe3* gene in *Heliconius* butterflies

Model	No. of parameters	Estimates of parameters	ln *L*	*P*-value	BEB*
M0: one ratio	1	$\omega_0 = 0.29855$	−446.40		
M1a: neutral	1	$p_0 = 0.82404$, $p_1 = 0.17596$ $\omega_0 = 0.05580$, $\omega_1 = 1.00000$	−443.02		
M2a: selection	3	$p_0 = 0.98698$, $p_1 = 0.00000$, $p_2 = 0.01302$ $\omega_0 = 0.20073$, $\omega_1 = 1.00000$, $\omega_2 = 87.94816$	−437.75	0.0051	40
M7: beta	2	$p = 0.01427$, $q = 0.04601$	−442.80		
M8: beta and ω	4	$p_0 = 0.98698$, $p = 24.66788$ $q = 99.00000$, $p_1 = 0.01302$, $\omega = 84.22935$	−437.47	0.0049	40

*The amino acid site under positive selection was determined with the Bayes Empirical Bayes (BEB) method ($P > 99\%$). Parameters p and q are the shape parameters of the beta distribution which underlies M7 and M8.

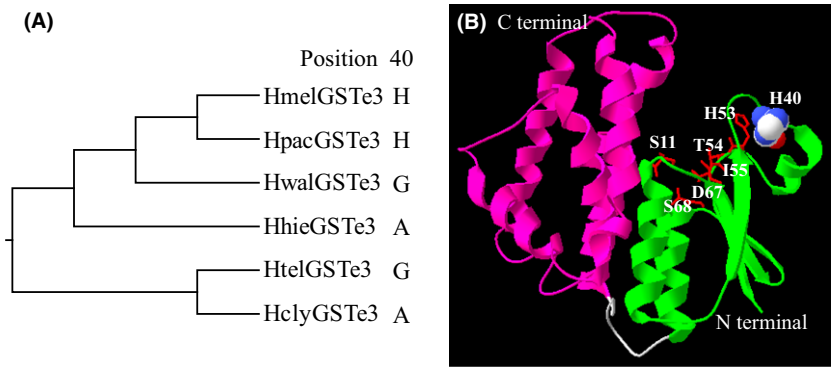


Fig. 6 Homology modelling and character state of a putatively selected site in the HmelGSTE3 protein. (A) The character state of position 40 of HmelGSTE3 in six species. (B) Tertiary structure modelling of HmelGSTE3. The folding pattern of HmelGSTE3 protein sequence was predicted by the SWISS-MODEL workspace web server (Arnold *et al.* 2006). The epsilon GST of *Musca domestica* was used as a template for protein modelling (Nakamura *et al.* 2013). The model was generated with SWISS-PDBVIEWER v4.1 (Guex & Peitsch 1997). The glutathione-binding site (G-site) was predicted by CD-search (Marchler-Bauer *et al.* 2002) and shown with red sticks. The putative positively selected site H40 is represented in a sphere. The N-terminal, C-terminal and linker regions are shown in green, red and grey, respectively.

which (*HmelGSTd2* and *HmelGSTe3*) belonged to insect-specific classes (Table S5, Supporting information). In addition, only *HmelGSTe3* showed a similar response to phase I detoxifying genes and was upregulated on the native host plant. Thus, it is surprising that a large number of ubiquitous omega and sigma GSTs were upregulated in non-host-feeding larvae. The GSTs offer passive protection against pyrethroid insecticides in *Tenebrio molitor* (Kostaropoulos *et al.* 2001), and in *Petunia,* AN9 is a flavonoid-binding GST protein that acts as a cytoplasmic flavonoid carrier protein (Mueller *et al.* 2000). It may be that upregulated omega and sigma GSTs might mainly act as allelochemical binding proteins in the non-host-feeding larvae of *H. melpomene*.

Transporter genes. Transporters play an important role in excretion, conjugation or sequestration of xenobiotic compounds from cells. Combinations of enzymes and transporters work together as a detoxification 'system'. In total, 29 transporter genes were differentially expressed after host plant shift in *H. melpomene*. ATP binding cassette (ABC) transporters have traditionally been regarded as major components in the excretion process (Glavinas *et al.* 2004), but only two ABC transporters showed upregulation in non-host-feeding larvae. Similar responses have been seen in previous studies (Dermauw *et al.* 2013; Wybouw *et al.* 2015). Major facilitator superfamily (MFS, Transporter Classification DataBase, TC# 2.A.1) is one of the largest families of membrane transporters along with ABC transporters (Reddy *et al.* 2012). The potential role for MFS transporters in detoxification of herbivores has only recently been demonstrated (Celorio-Mancera *et al.* 2013; Dermauw *et al.* 2013; Wybouw *et al.* 2015). In our study, twelve MFS members were differentially regulated in *H. melpomene* after the host plant shift (Table S5, Supporting information). In addition, other transporters were also found, such as zinc transporter, proton-coupled amino acid transporter, organic anion transporter. These membrane proteins might function as efflux transporters and facilitate host plant adaptation.

Peritrophic matrix biosynthesis-related genes. During long-term adaptive evolution the peritrophic matrix (PM) can decrease penetration of allelochemicals to enhance adaptation of insects to their host plants (Barbehenn 2001; Celorio-Mancera *et al.* 2013). It has been shown that the PM can decrease or prevent the permeability of plant allelochemicals, such as rutin, tannic acid, digitoxin and chlorogenic acid (Barbehenn 2001). Peritrophic matrix is a proteoglycan matrix formed of proteins, glycoproteins, and chitin microfibrils, which may protect insects from insult by pathogens and toxins (Lehane 1997). Peritrophins are a major fraction of the PM proteins and noncovalently bonded to chitin. Chitin, a homopolymer of β-(1-4)-linked *N*-acetyl-D-glucosamine, can be hydrolysed by chitinase and chitin deacetylase, which breaks down glycosidic bonds and acetamido groups, respectively (Lehane 1997; Arakane *et al.* 2009). After host plant shift, six genes encoding peritrophin proteins and one chitin synthase were differentially expressed, all of which were upregulated in non-host-feeding larvae. In addition, 3 of 8 differentially expressed chitin-degrading enzymes were upregulated in nonhost treatment. Taken together, more peritrophins

and chitins might be synthesized in the non-host-feeding larval gut. The strengthened cuticular component of PM might be another important mechanism to adapt to host plants in *H. melpomene*.

Detoxification of cyanogenic glucosides. Cyanogenic glycosides (CNglc, cyanogen) are important defensive compounds in *Passiflora* plants (Olafsdottir *et al.* 1988; Spencer 1988). After long-term co-evolution between *Heliconius* butterflies and *Passiflora* hosts, some *Heliconius* not only specifically sequester simple monoglycoside cyclopentenyl (SMC) cyanogens, but also most species synthesize CNglc linamarin and lotaustralin *de novo* for their own defence against predators (Nahrstedt & Davis 1981; Engler-Chaouat & Gilbert 2007). In this study, *H. melpomene* was exposed to two non-SMC plants, *Passiflora menispermifolia* and *P. biflora*, with the former having a higher cyanogen content (Engler-Chaouat & Gilbert 2007). The ingested non-SMC cyanogens might be degraded by β-glucosidases and α-hydroxynitrile lyases (Zagrobelny *et al.* 2004; Ketudat Cairns & Esen 2010). Although no specific β-glucosidase involved in CNglc catabolism has been characterized in insects yet, a β-glucosidase (ZfBGD1) has been identified in CNglc-containing droplets of *Zygaena filipendulae* larvae (Pentzold *et al.* 2016). Among the 326 DEGs in the *H. melpomene* gut after host plant shift, three homologs of β-glucosidase enzymes (HMEL013290, HMEL014474 and HMEL014483) were identified. All of these were significantly upregulated in larvae reared on the native host, *P. menispermifolia* (Table S5, Supporting information). Except for sequestration and degradation of cyanogens, insects have evolved to cope with the toxic effects of HCN, through conversion of HCN into β-cyanoalanine by β-cyanoalanine synthase (CAS) (Despres *et al.* 2007; Wybouw *et al.* 2014). Three candidates of CAS (HMEL013489, HMEL016300 and HMEL002400) were found in *H. melpomene* (Wybouw *et al.* 2014). Although none of them showed significantly differential expression after the host plant shift, HMEL002400 was expressed higher in larvae reared on the host (FPKM = 46.62) as compared to nonhost plant (FPKM = 20.12), with an FDR value of 0.052. Our results suggested that the differentially expressed β-glucosidases and elevated CAS might play important roles in detoxifying cyanogens and its toxic product, HCN.

Comparison of DEGs detected by CUFFDIFF*2 and* DESEQ*2.* For comparison, we compared the DEGs detected by CUFFDIFF2 to those detected using DESEQ2. In total 217 DEGs were identified by DESEQ2 (Table S8, Supporting information), of which 147 were found in the differentially expressed gene lists detected by CUFFDIFF2 (Table S5, Supporting information). Although the total number of DEGs detected by DESEQ2 decreased, the number of genes related to detoxification, peritrophic matrix biosynthesis, transporter and nutrient digestion represented more than half of the corresponding genes detected by CUFFDIFF2. For the detoxification gene group, 7 of 9 differentially expressed GST genes (Table S5, Supporting information) were characterized as DEGs using DESEQ2, meanwhile *HmelGSTs10* and *HmelGSTs11* were identified as new candidate DEGs. Although *HmelGSTe3* was not identified as differentially expressed using DESEQ2, it was a differentially expressed gene validated by qPCR (Fig. 1A). Among UGT genes only 2 of 9 genes were detected as DEGs using DESEQ2 (Tables S5 and S8, Supporting information). In typical high-throughput sequencing experiments, as few as two or three replicates per condition were used (Love *et al.* 2014). In contrast, five replicates were performed in our study. Nonetheless, while most of our results were robust, slightly different gene sets were identified using the two methods of differential expression analysis. Further validation would therefore be needed to definitively confirm differential expression of the genes identified here.

Molecular evolution of detoxifying genes in Heliconius butterflies

Heliconius butterflies have complex relationships with their Passifloraceae host plants. Some species are specialist herbivores, while others are generalists (Brown 1981; Briscoe *et al.* 2013). It seems likely that the detoxification system has played an important role in host adaptation among these butterflies. To understand the roles of detoxification enzymes in herbivore adaptation, we studied the molecular evolution of detoxifying genes. We focused on gene gain and loss of GSTs and showed that there has been broad conservation of gene number subsequent to the *Heliconius* radiation (Fig. S6, Supporting information), while omega and sigma classes expanded before the radiation of *Heliconius* (Fig. S6, Supporting information).

After the host plant shift in *H. melpomene*, 30 detoxification genes showed significant differential expression (Table S5, Supporting information). Of these, 24 orthologous gene sets had enough coverage across 10 *Heliconius* species and were used to estimate ω values (Fig. 5). Broadly, GSTs, UGTs, COEs and P450s (Fig. 5) were all under strong purifying selection. Using the 'site-specific' model in PAML software, one positively selected site (H40) was found in GSTE3, located near the glutathione-binding pocket in N-terminal domain (Fig. 6). In *Pinus tabuliformis*, two positively selected sites (corresponding to Arg-12 and Asn-37 of PtaGSTU17) located close to the G-site

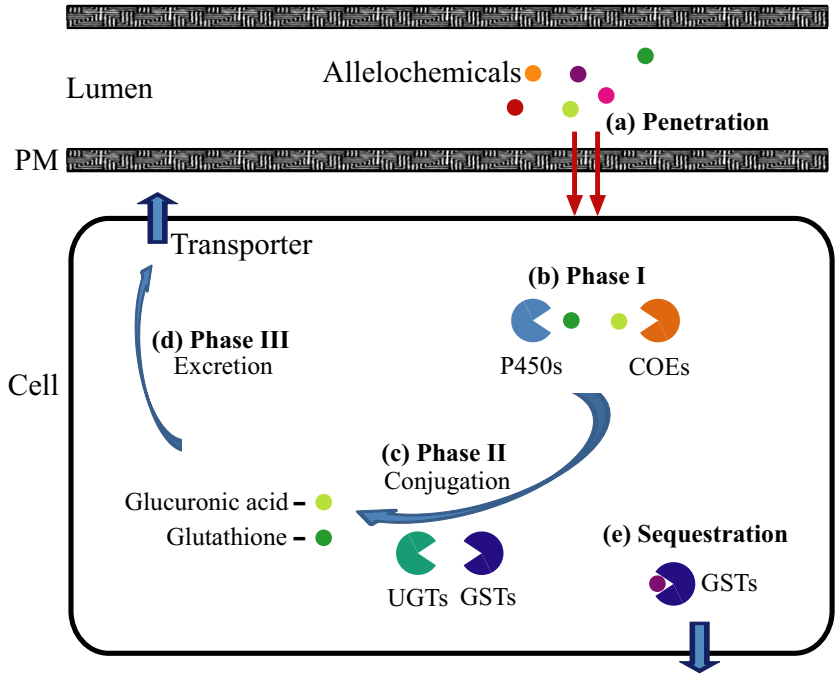


Fig. 7 Schematic models of proposed mechanisms for resistance to host secondary metabolites in *Heliconius* butterflies. Based on Despres *et al.* (2007), the detoxification model was reconstructed in the larval gut of *Heliconius* butterflies. The first mechanism is decreasing penetration (a) through regulating the components of peritrophic matrix (PM). The ingested allelochemicals can be partially detoxified by the classical processes, including phase I direct metabolism (b), phase II conjugation (c) and phase III excretion (d). The duplicated sigma and omega GSTs may sequester (e) some of the allelochemicals and help to excrete them. Detoxification of cyanogenic glucosides is not shown here.

can dramatically alter enzyme activities (Lan *et al.* 2013). In *Anopheles dirus*, E25Q residue of adGST1-1 is located between helix 1 and sheet 2 and appears to affect tertiary structure around the hydrophobic core in N-terminal domain and further change the enzyme activity (Ketterman *et al.* 2001). It seems plausible therefore that the site of positive selection in GSTE3 might alter enzyme activity. Thus, *GSTe3* shows both a transcriptional response and evidence for positive selection, so could be considered a potential candidate for host adaptation among *Heliconius* butterflies.

Conclusion

Previous studies suggested that larval feeding specialization of *Heliconius* is mediated primarily by adult female butterflies (Briscoe *et al.* 2013), not primarily driven by biochemical co-evolution (Smiley 1978; Smiley & Wisdom 1985). In insects, the gut is the first barrier for penetration and detoxification of plant allelochemicals (Krieger *et al.* 1971). We have analysed the transcriptome response of guts in *H. melpomene* after a host plant shift. Based on the functional annotations of DEGs, we suggest that *Heliconius* have a broad-based defence mechanism for detoxifying host secondary metabolites (Fig. 7). Our results demonstrate that phase I detoxifying genes and β-glucosidases might play important roles in routine detoxification of allelochemicals, as most were upregulated in larvae reared on the more strongly defended host species, *P. menispermifolia* (Table S5, Supporting information). Most of the differentially expressed UGTs/GSTs belonged to lineage-specific duplications (Figs 3 and 4) and were upregulated in larvae reared on the nonhost plant. The high degree of plasticity may help to partly cope with natural shifts between hosts in *H. melpomene* (Smiley 1978).

Acknowledgements

Thanks to Jenny Barna for remote computing support; Nicola Nadeau, Richard Wallbank, Richard Merrill and Carolina Pardo for providing much help during the larval rearing and experiment; and John Davey, Simon H. Martin, Krzysztof M. Kozak and Martin Thompson for bioinformatics assistance. This work was supported by the National High Technology Research and Development Program of China (2013AA102507-2 to ZZ), the Fundamental Research Funds for the Central Universities (106112015CDJXY290002 to YQY), the National Natural Science Foundation of China (31201851 to YQY). CDJ is supported by an ERC Advanced Grant Speciation Genetics 339873.

References

Ahn SJ, Vogel H, Heckel DG (2012) Comparative analysis of the UDP-glycosyltransferase multigene family in insects. *Insect Biochemistry and Molecular Biology*, **42**, 133–147.

Anders S, Pyl PT, Huber W (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

Arakane Y, Dixit R, Begum K *et al.* (2009) Analysis of functions of the chitin deacetylase gene family in *Tribolium castaneum*. *Insect Biochemistry and Molecular Biology*, **39**, 355–365.

Arnold K, Bordoli L, Kopp J *et al.* (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201.

Barbehenn RV (2001) Roles of peritrophic membranes in protecting herbivorous insects from ingested plant allelochemicals. *Archives of Insect Biochemistry and Physiology*, **47**, 86–99.

Benson WW, Brown KS, Gilbert LE (1976) Coevolution of plants and herbivores: passion flower butterflies. *Evolution*, **29**, 659–680.

Briscoe AD, Macias-Munoz A, Kozak KM *et al.* (2013) Female behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS Genetics*, **9**, e1003620.

Brower AVZ (1997) The evolution of ecologically important characters in *Heliconius* butterflies (Lepidoptera: Nymphalidae): a cladistic review. *Zoological Journal of the Linnean Society*, **119**, 457–472.

Brown KS (1981) The biology of *Heliconius* and related genera. *Annual Review of Entomology*, **26**, 427–456.

Celorio-Mancera MDLP, Wheat CW, Vogel H *et al.* (2013) Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-Seq. *Molecular Ecology*, **22**, 4884–4895.

Chauhan R, Jones R, Wilkinson P *et al.* (2013) Cytochrome P450-encoding genes from the *Heliconius* genome as candidates for cyanogenesis. *Insect Molecular Biology*, **22**, 532–540.

Chikate YR, Tamhane VA, Joshi RS *et al.* (2013) Differential protease activity augments polyphagy in *Helicoverpa armigera*. *Insect Molecular Biology*, **22**, 258–272.

Claudianos C, Ranson H, Johnson RM *et al.* (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Molecular Biology*, **15**, 615–636.

Conesa A, Gotz S, Garcia-Gomez JM *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

Dermauw W, Wybouw N, Rombauts S *et al.* (2013) A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E113–E122.

Despres L, David JP, Gallet C (2007) The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution*, **22**, 298–307.

Dhawan K, Dhawan S, Sharma A (2004) *Passiflora*: a review update. *Journal of Ethnopharmacology*, **94**, 1–23.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

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

Engler-Chaouat HS, Gilbert LE (2007) De novo synthesis vs. sequestration: negatively correlated metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. *Journal of Chemical Ecology*, **33**, 25–42.

Glavinas H, Krajcsi P, Cserepes J *et al.* (2004) The role of ABC transporters in drug resistance, metabolism and toxicity. *Current Drug Delivery*, **1**, 27–42.

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.

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

Hoang K, Matzkin LM, Bono JM (2015) Transcriptional variation associated with cactus host plant adaptation in *Drosophila mettleri* populations. *Molecular Ecology*, **24**, 5186–5199.

Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*, **444**, 512.

Hosokawa M (2008) Structure and catalytic properties of carboxylesterase isozymes involved in metabolic activation of prodrugs. *Molecules*, **13**, 412–431.

Huang FF, Chai CL, Zhang Z *et al.* (2008) The UDP-glucosyltransferase multigene family in *Bombyx mori*. *BMC Genomics*, **9**, 563.

Jaenike J (1990) Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics*, **21**, 243–273.

Ketterman AJ, Prommeenate P, Boonchauy C *et al.* (2001) Single amino acid changes outside the active site significantly affect activity of glutathione S-transferases. *Insect Biochemistry and Molecular Biology*, **31**, 65–74.

Ketudat Cairns JR, Esen A (2010) β-Glucosidases. *Cellular and Molecular Life Sciences*, **67**, 3389–3405.

Koenig C, Bretschneider A, Heckel DG *et al.* (2015) The plastic response of *Manduca sexta* to host and non-host plants. *Insect Biochemistry and Molecular Biology*, **63**, 72–85.

Kostaropoulos I, Papadopoulos AI, Metaxakis A *et al.* (2001) Glutathione S-transferase in the defence against pyrethroids in insects. *Insect Biochemistry and Molecular Biology*, **31**, 313–319.

Krieger RI, Feeny PP, Wilkinson CF (1971) Detoxication enzymes in the guts of caterpillars: an evolutionary answer to plant defenses? *Science*, **172**, 579–581.

Lan T, Wang XR, Zeng QY (2013) Structural and functional evolution of positively selected sites in pine glutathione S-transferase enzyme family. *The Journal of Biological Chemistry*, **288**, 24441–24451.

Lehane MJ (1997) Peritrophic matrix structure and function. *Annual Review of Entomology*, **42**, 525–550.

Li X, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, **52**, 231–253.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2 (-Delta Delta C(T)) method. *Methods*, **25**, 402–408.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Low WY, Ng HL, Morton CJ *et al.* (2007) Molecular evolution of glutathione S-transferases in the genus *Drosophila*. *Genetics*, **177**, 1363–1375.

Mackenzie PI, Owens IS, Burchell B *et al.* (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics*, **7**, 255–269.

Marchler-Bauer A, Panchenko AR, Shoemaker BA *et al.* (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Research*, **30**, 281–283.

Merrill RM, Naisbit RE, Mallet J *et al.* (2013) Ecological and genetic factors influencing the transition between host-use strategies in sympatric *Heliconius* butterflies. *Journal of Evolutionary Biology*, **26**, 1959–1967.

Merrill RM, Dasmahapatra KK, Davey JW *et al.* (2015) The diversification of *Heliconius* butterflies: what have we learned in 150 years? *Journal of Evolutionary Biology*, **28**, 1417–1438.

Mueller LA, Goodman CD, Silady RA *et al.* (2000) AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiology*, **123**, 1561–1570.

Nahrstedt A, Davis RH (1981) Cyanogenic glycosides in butterflies: detection and synthesis of linamarin and lotaustralin in the heliconiinae. *Planta Medica*, **42**, 124–125.

Nakamura C, Yajima S, Miyamoto T *et al.* (2013) Structural analysis of an epsilon-class glutathione transferase from housefly, *Musca domestica*. *Biochemical and Biophysical Research Communications*, **430**, 1206–1211.

Oakeshott JG, Johnson RM, Berenbaum MR *et al.* (2010) Metabolic enzymes associated with xenobiotic and chemosensory responses in *Nasonia vitripennis*. *Insect Molecular Biology*, **19**, 147–163.

Olafsdottir ES, Andersen JV, Jaroszewski JW (1988) Cyanohydrin glycosides of Passifloraceae. *Phytochemistry*, **28**, 127–132.

Pentzold S, Zagrobelny M, Khakimov B *et al.* (2016) Lepidopteran defence droplets – a composite physical and chemical weapon against potential predators. *Scientific Reports*, **6**, 22407.

Pringle EG, Baxter SW, Webster CL *et al.* (2007) Synteny and chromosome evolution in the lepidoptera: evidence from mapping in *Heliconius melpomene*. *Genetics*, **177**, 417–426.

Ragland GJ, Almskaar K, Vertacnik KL *et al.* (2015) Differences in performance and transcriptome-wide gene expression associated with *Rhagoletis* (Diptera: Tephritidae) larvae feeding in alternate host fruit environments. *Molecular Ecology*, **24**, 2759–2776.

Ranson H, Hemingway J (2005) Glutathione transferases. In: *Comprehensive Molecular Insect Science-Pharmacology Volume 5: Glutathione Transferases* (eds Gilbert LI, Iatrou K, Gill SS), pp. 383–398. Elsevier, Oxford, UK.

Ranson H, Claudianos C, Ortelli F *et al.* (2002) Evolution of supergene families associated with insecticide resistance. *Science*, **298**, 179–181.

Reddy VS, Shlykov MA, Castillo R *et al.* (2012) The major facilitator superfamily (MFS) revisited. *FEBS Journal*, **279**, 2022–2035.

Smiley J (1978) Plant chemistry and the evolution of host specificity: new evidence from *Heliconius* and *Passiflora*. *Science*, **201**, 745–747.

Smiley JT, Wisdom CS (1985) Determinants of growth rate on chemically heterogeneous host plants by specialist insects. *Biochemical Systematics and Ecology*, **13**, 305–312.

Spencer KC (1988) Chemical mediation of coevolution in the *Passiflora-Heliconius* interaction. In: *Chemical Mediation of Coevolution* (ed Spencer KC), pp. 167–240. Academic Press, London, UK.

Strode C, Wondji CS, David JP *et al.* (2008) Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. *Insect Biochemistry and Molecular Biology*, **38**, 113–123.

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, W609–W612.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.

Trapnell C, Williams BA, Pertea G *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, **28**, 511–515.

Trapnell C, Roberts A, Goff L *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, **7**, 562–578.

Tribolium Genome Sequencing Consortium (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, **452**, 949–955.

Turlings TC, Loughrin JH, McCall PJ *et al.* (1995) How caterpillar-damaged plants protect themselves by attracting parasitic wasps. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 4169–4174.

Wang Y, Tang H, Debarry JD *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49.

Wybouw N, Dermauw W, Tirry L *et al.* (2014) A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. *Elife*, **3**, e02365.

Wybouw N, Zhurov V, Martel C *et al.* (2015) Adaptation of a polyphagous herbivore to a novel host plant extensively shapes the transcriptome of herbivore and host. *Molecular Ecology*, **24**, 4647–4663.

Xie C, Mao X, Huang J *et al.* (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, **39**, W316–W322.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang Z, Wong WS, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107–1118.

Yu Q, Lu C, Li B *et al.* (2008) Identification, genomic organization and expression pattern of glutathione S-transferase in the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, **38**, 1158–1164.

Zagrobelny M, Bak S, Rasmussen AV *et al.* (2004) Cyanogenic glucosides and plant-insect interactions. *Phytochemistry*, **65**, 293–306.

Zhan S, Merlin C, Boore JL *et al.* (2011) The monarch butterfly genome yields insights into long-distance migration. *Cell*, **147**, 1171–1185.

Y.Q.Y. and C.D.J. conceived and designed the study. Y.Q.Y. performed the experiments, analysed the data and drafted and revised manuscript. C.D.J. and Z.Z. revised the manuscript. S.M.F. performed the qPCR experiment and revised the manuscript.

Data accessibility

Illumina reads for each of the libraries were deposited in Sequence Read Archive (SRA) database under the Accession no.: SRP074338. Supplementary data are deposited in the Dryad repository: doi: 10.5061/dryad.65k4n.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Gene ontology classification of the postman butterfly genes expressed in the gut.

Fig. S2 Scatterplot of enriched pathways for DEGs in larval gut after host plant shift.

Fig. S3 NJ tree and intron positions of HmelGST genes.

Fig. S4 Distributions of *H. melpomene* GST and UGT genes on chromosomes.

Fig. S5 Duplication mechanisms and conserved motif of UGT33 family in *H. melpomene*.

Fig. S6 Gene gain and loss of GSTs superfamily in *Heliconius* butterflies.

Table S1 Primer sequences used for quantitative PCR validation.

Table S2 Annotation and sequences of GST and UGT multigene families in the *H. melpomene* and *Danaus plexippus* genomes.

Table S3 Sequences of detoxifying genes used for PAML analysis.

Table S4 Summary of reads and assembly of the gut transcriptomes in *H. melpomene*.

Table S5 Annotations and expression signals of 326 DEGs in larval gut after host plant shift.

Table S6 GO enrichment analysis of the differentially expressed genes.

Table S7 Orthologous divergence of detoxification genes among *Heliconius* butterflies.

Table S8 Differentially expressed genes detected by DESEQ2 package.