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

Quan-You Yu[1,2*], Shou-Min Fang[3], Ze Zhang[1], Chris D. Jiggins[2*]

[1]School of Life Sciences, Chongqing University, Chongqing 401331, China.
[2]Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, U.K.
[3]College of Life Science, China West Normal University, Nanchong 637002, China.

[*]Correspondence to:
Quan-You Yu
School of Life Sciences
Chongqing University
Shapingba District
Chonging 401331
Tel. +86-23-65678491
Email: yuqy@cqu.edu.cn

Chris D. Jiggins
Department of Zoology
University of Cambridge
Downing Street
Cambridge CB2 3EJ
Tel. +44-1223-769021
Email: c.jiggins@zoo.cam.ac.uk

Running title

Host plant adaptation and detoxification

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 in order to understand these mechanisms. We investigated the plastic responses of *H. melpomene* larvae feeding on a native host *Passiflora menispermifolia* and a less strongly defended non-host species, *P. 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 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.

Key words: Heliconius butterflies; host plant shift; transcriptome; detoxification; molecular evolution

Introduction

Coevolution 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 coevolution (*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, *H. 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 is 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 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)

65 when damaged (Olafsdottir *et al.* 1988; Spencer 1988). In addition, glycosyl flavonoid, alkaloid
66 and phenolic compounds have been reported as the major phytoanticipins of *Passiflora* (Dhawan
67 *et al.* 2004).

68 To counter the toxic effects of plant toxins, herbivorous insects have evolved a variety of
69 mechanisms to adapt to their host plants, such as chemical transformation, xenobiotic excretion
70 and reducing the absorption of ingested allelochemicals (Barbehenn 2001; Despres *et al.* 2007; Li
71 *et al.* 2007). Chemical transformation is the most important detoxification process and has been
72 well studied. Detoxifying enzymes play roles in allelochemical transformation, which are
73 classified into direct metabolism (phase I) and conjugation (phase II) categories. Cytochrome
74 P450 monooxygenases (P450s) are the principal biochemical system for phase I detoxification in
75 insects, and decrease (or, less often, increase) the biological activity of a broad range of substrates
76 through oxidation-reduction or hydrolytic reactions (Despres *et al.* 2007). Hydrolysis reactions are
77 primarily carried out by phase I enzyme carboxylesterases (COEs), which catalyze the
78 transformation of an ester linkage into alcohol and carboxylic acid products (Hosokawa 2008).
79 Phase II conjugation reactions generally follow phase I, including glutathione S-transferases
80 (GSTs) and UDP-glucuronosyltransferases (UGTs) (Li *et al.* 2007). Enzymes from both phases
81 work in concert to transform allelochemicals into water-soluble compounds for eventual excretion
82 by transmembrane proteins that are specific for the conjugated toxins, termed phase III
83 transporters (Reddy *et al.* 2012). Mechanisms for reducing the absorption of allelochemicals are
84 less well studied. In insects, peritrophic membranes (PMs) are extracellular matrices composed of
85 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 defense, 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 regulation (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 non-host plant, *P. 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 diglycoside 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 (Briscoe *et al.* 2013; *Heliconius* Genome Consortium 2012). Some of these species are specialist on only one or a few *Passiflora* species, others feed on particular sub-genera within *Passiflora*, while others are generalists (Briscoe *et al.* 2013; Brown 1981). These genomic datasets 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 *H. 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 *P. menispermifolia* and *P. 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 Hongkong (China). The transcriptome libraries were generated using Illumina TruSeq™ RNA Sample Preparation Kit (Illumina, San Diego, CA). 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 Hongkong, 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% ploy-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), which the adjusted *P* value ≤ 0.05 and $\log_2$FC ≥ 1 were 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 non-redundant (nr) sequence database of the National Center for Biotechnology Information (NCBI) with an e-value cut-off $\leq 1e^{-5}$. 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 were

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). Five replicate samples for each treatment were used for qPCR analysis, and each sample was analyzed 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 analyzed 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 UGT (Ahn *et al.* 2012; Huang *et al.* 2008) 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). 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,

191 *Danaus plexippus* (Zhan *et al.* 2011). Chromosomal assignments were based on published
192 mapping of scaffolds in the *H. melpomene* reference genome (*Heliconius* Genome Consortium
193 2012). The validated GSTs and other detoxification genes of *H. melpomene* were used for
194 tBLASTn (E-value < 0.01) search against the genome assemblies of other *Heliconius* species,
195 including *H. pachinus*, *H. timareta*, *H. wallacei*, *H. hecuba*, *H. doris*, *H. hierax*, *H. xanthocles*, *H.*
196 *telesiphe* and *H. clysonymus*. To improve the quality of the sequences found, the coding sequences
197 were performed by read mapping in CLC Genomics Workbench v. 6.5.1. Conservative parameters
198 (mismatch, insertion and deletion cost of 3; length fraction and similarity fraction of 0.9) were
199 used to prevent mis-mapping of paralogous sequences. All read-mappings were inspected by eye.

200 *Phylogenetic analysis and gene nomenclature*

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

211 *Molecular evolution of detoxifying genes in Heliconius butterflies*

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

226 *Detection of positive selection*. To detect site-specific positive selection, the CODEML program as
227 implemented in PAML4.5 (Yang 2007) was used. Model M0 (one ratio), M1a (nearly neutral),
228 M2a (positive selection), M7 (β distribution), and M8 (β distribution and ω) were run for each set
229 of orthologs (Table S3). To avoid being trapped at local optima, three different initial ω-values of
230 0.5, 1.1, and 2.0 were used in the estimation of the log likelihood for model 7 and model 8 (Low *et*
231 *al.* 2007). Models M2a vs. M1a and M8 vs. M7 were compared using likelihood ratio tests (LRT),
232 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 H. melpomene

The lepidopteran gut plays key roles in nutrient digestion and allocation. The guts of the fifth instar larvae reared with the host plants *P. menispermifolia* and *P. 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 *H. melpomene* reference genome with TopHat (Trapnell *et al.* 2009). The expression level of each gene was estimated by Cufflinks (Trapnell *et al.* 2010). In order 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, 7,993 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). 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 H. 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 non-host 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) from ten guts of the fifth instar caterpillars reared on host and non-host 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 $\geq$ 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), 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 Figure 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 analyze 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). 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 Figure S2. 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).

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 non-redundant 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). 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 non-host plant. We found 30 detoxifying genes that were 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 out 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).

296 Twenty-nine transporter genes were identified as DEGs, in which more than half were upregulated
297 in native host-reared larvae. Functional annotation showed that 12 of the 29 transporters belonged
298 to the major facilitator superfamily (MFS). Previous studies suggested that MFS might be key
299 players facilitating insect adaptation and survival in response to new diets (Dermauw *et al.* 2013).
300 Chitin and peritrophin are the major components of the peritrophic matrix (Lehane 1997). One
301 chitin synthase and eight chitin-degrading enzymes (chitin deacetylase and chitinase) responded to
302 the host plant shift. Six genes encoding peritrophin also showed differential expression. This
303 regulation of genes related to PM biosynthesis might affect the penetration of allelochemicals in
304 the gut. These results suggest that multiple gene categories are involved in the defense network for
305 plant secondary metabolites in *H. melpomene*.

306 *Specific responses of detoxifying enzymes to plant allelochemicals*

307 Allelochemicals are the major defense chemicals in plants, and are also therefore the selective
308 agents on the detoxification systems of insect herbivores (Li *et al.* 2007; Ragland *et al.* 2015).
309 Accordingly, we observed major plastic changes in expression of detoxification-related genes after
310 the host plant shift (Fig.S2, Table S5). Based on functional annotations of the DEGs, we found 9,
311 9, 8 and 4 members of GSTs, UGTs, P450s and COEs respectively that were significantly
312 differentially expressed (Table S5). Generally, GSTs and UGTs families contained fewer genes
313 than P450s and COEs in insect genomes (Ahn *et al.* 2012; Ranson *et al.* 2002). However, GST
314 and UGT gene families have more members differentially regulated by the host-plant shift in *H.*
315 *melpomene*. Thus, it would be interesting to understand whether the two families were expanded
316 in the genome of *Heliconius* butterflies and the role of duplicated genes in host plant adaptation.

317 Using the amino acid sequences of silkworm GSTs (Yu *et al.* 2008) and UGTs (Ahn *et al.*
318 2012; Huang *et al.* 2008) as queries, we manually annotated the detoxification genes through a
319 local tBLASTn search in the *H. melpomene* reference genome (*Heliconius* Genome Consortium
320 2012). In total, 33 GSTs and 52 UGTs were identified in the *H. melpomene* genome (Fig. 2).
321 GSTs had the highest percentage (9 out of 33, 27.3%) of differential regulation, followed by
322 UGTs (9 out of 52, 17.3%). For P450s, one hundred genes have been found in *H. melpomene*
323 (Chauhan *et al.* 2013). One new candidate (*HMEL004608*) was found in our analysis (Table S5).
324 Thus, only 7.9% of P450 genes showed differential regulation after the host plant shift.
325 For comparison, detoxification-related genes have also been characterized in another
326 butterfly, *D. plexippus* (Fig. 2). Phylogenetic analysis was performed for GSTs and UGTs of
327 lepidopteran model insects (Fig. 3; Fig. 4). The GSTs and UGTs of *H. melpomene* and *D.*
328 *plexippus* were named using published nomenclature (Ahn *et al.* 2012; Yu *et al.* 2008).
329 Interestingly, it was not insect-specific delta and epsilon classes, which are known to be related to
330 insecticide resistance and detoxification in insects (Li *et al.* 2007), but rather the omega and sigma
331 classes that are more ubiquitously distributed in all organisms, that were most commonly
332 duplicated in the *H. melpomene* genome (Fig. 2; Fig. 3). *H. melpomene* is the first species in
333 which omega and sigma classes both showed expansion (Fig. 2). The intron positions of the
334 members of the ubiquitous sigma and omega classes are highly conserved and show very strong
335 class-specificity (Fig. S3). Furthermore, omega and sigma classes were mainly clustered on
336 chromosome 7 and scaffold HE669239 respectively (Fig. S4). RNA-Seq data showed that most of
337 the significantly differentially regulated GSTs belonged to the sigma (4) and omega (3) classes

(Table S5).

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). 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). Tandem duplication produced more recent copies (Fig. S5A). These duplicated UGT33s share a common signature motif (FhTQhGLQSTxExxxxxVPhhxxPhxxDQ) (Mackenzie *et al.* 1997) with UGT33 members from other lepidopteran species (Fig. S5B). Transcriptome data indicated that five out of nine differentially expressed UGTs belonged to the UGT33 subfamily (Table S5).

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 analyzed (Table S3). 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

359 that binds the hydrophobic substrates (Ranson & Hemingway 2005). The comparison indicated
360 that the selective constraint on the C-terminus of oGSTs was more relaxed than on the N-terminus
361 (t-test, $p << 0.01$), while the selective pressure on the C-terminus and N-terminus was similar
362 among the differentially expressed GSTs (Fig. 5).
363 To detect the evolutionary rates of different sites, CODEML (model 3 with K =3)
364 implemented in PAML was used to calculated three different ω values and the corresponding
365 proportions of sites. The results indicated that some of the detoxification genes might have
366 positively selected sites ($\omega > 1$) (Table S7). We used two standard PAML comparisons (model
367 M2a vs. M1a and M8 vs.M7) to test whether an adaptive model fits the data better than the neutral
368 model. Only *GSTe3* was found to be significant with Bayes Empirical Bayes (BEB) method and P
369 < 0.01 (Table 1). The amino acid state of the positively selected site in N-terminal region of
370 GSTE3 was shown in Fig. 6A. As shown in Fig. 6A, a histidine/alanine/glycine substitution at site
371 40 has occurred across the six *Heliconius* species. The tertiary protein structure of HmelGSTE3
372 was predicted by homology modeling (Fig. 6B). The positively selected site (H40) was located
373 near the glutathione-binding site (G-site) in N-terminal domain.

374 **Discussion**

375 *Differentially expressed genes*

376 *Digestive enzymes.* We anticipated that digestion-related genes of herbivorous insects might
377 respond to exposure to different hosts. Previous studies suggested that serine proteases are key
378 enzymes allowing larvae to adapt to many different diets (Celorio-Mancera *et al.* 2013; Chikate *et*
379 *al.* 2013). In this study, 16 digestion-related genes showed plastic responses in *H. melpomene*

larvae after the host-plant shift (Table S5). 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) were identified. These 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). 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 *H. melpomene*.

Although we might expect detoxification genes to be upregulated in order to deal with the allelochemicals from a non-host plant, in this case *P. 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

401 expressed P450s (7 out of 8) and COEs (3 out of 4) were upregulated in larvae reared on the
402 native host plant (Table S5). For some species including *H. melpomene*, *Passiflora* natural product
403 diversity does not cause major deleterious effects on *Heliconius* growth suggesting that larval
404 feeding specialization is not primarily driven by biochemical coevolution (Smiley 1978; Smiley &
405 Wisdom 1985). *Heliconius-Passiflora* interactions are therefore mediated primarily by adult
406 female butterflies correctly identifying suitable host plants for oviposition (Briscoe *et al.* 2013).
407 So, these differentially expressed phase I detoxifying genes showed higher constitutive
408 expressions and might be responsible for routine detoxification of allelochemicals in *H.*
409 *melpomene* on its natural host in Central America.
410 Cytosolic GSTs are important enzymes involved in detoxification of various plant
411 xenobiotics, which can be induced by xanthotoxin, indoles and flavonoids (Li *et al.* 2007).
412 Previous studies indicated that insect specific delta and epsilon classes primarily mediated
413 allelochemical tolerance and insecticide resistance (Li *et al.* 2007). In our study, nine GST genes
414 showed plastic responses to the host plant shift in *H. melpomene*, two of which (*HmelGSTd2* and
415 *HmelGSTe3*) belonged to insect specific classes (Table S5). In addition, only *HmelGSTe3* showed
416 a similar response to phase I detoxifying genes and was upregulated on the native host plant. Thus,
417 it is surprising that a large number of ubiquitous omega and sigma GSTs were upregulated in
418 non-host feeding larvae. The GSTs offer passive protection against pyrethroid insecticides in
419 *Tenebrio molitor* (Kostaropoulos *et al.* 2001), and in *Petunia* AN9 is a flavonoid-binding GST
420 protein that acts as a cytoplasmic flavonoid carrier protein (Mueller *et al.* 2000). It may be that
421 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 up-regulation 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). In addition, other transporters were also found, such as Zinc transporter, proton-coupled amino acid transporter, organic anion transporter, and etc. 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

443 proteins, glycoproteins, and chitin microfibrils, which may protect insects from insult by
444 pathogens and toxins (Lehane 1997). Peritrophins are a major fraction of the PM proteins and
445 noncovalently bonded to chitin. Chitin, a homopolymer of β-(1-4)-linked *N*-acetyl-D-glucosamine,
446 can be hydrolyzed by chitinase and chitin deacetylase which breaks down glycosidic bonds and
447 acetamido groups, respectively (Arakane *et al.* 2009; Lehane 1997). After host plant shift, six
448 genes encoding peritrophin proteins and one chitin synthase were differentially expressed, all of
449 which were upregulated in non-host-feeding larvae. In addition, 3 out of 8 differentially expressed
450 chitin-degrading enzymes were upregulated in non-host treatment. Taken together, more
451 peritrophins and chitins might be synthesized in the non-host feeding larval gut. The strengthened
452 cuticular component of PM might be another important mechanism to adapt to host plants in *H.*
453 *melpomene*.

454 *Detoxification of cyanogenic glucosides*. Cyanogenic glycosides (CNglc, cyanogen) are important
455 defensive compounds in *Passiflora* plants (Olafsdottir *et al.* 1988; Spencer 1988). After long-term
456 coevolution between *Heliconius* butterflies and *Passiflora* hosts, some *Heliconius* not only
457 specifically sequester simple monoglycoside cyclopentenyl (SMC) cyanogens, but also most
458 species synthesize CNglc linamarin and lotaustralin *de novo* for their own defense against
459 predators (Engler-Chaouat & Gilbert 2007; Nahrstedt & Davis 1981). In this study, *H. melpomene*
460 was exposed to two non-SMC plants, *P. menispermifolia* and *P. biflora*, with the former having a
461 higher cyanogen content (Engler-Chaouat & Gilbert 2007). The ingested non-SMC cyanogens
462 might be degraded by β-glucosidases and α-hydroxynitrile lyases (Ketudat Cairns & Esen 2010;
463 Zagrobelny *et al.* 2004). Although no specific β-glucosidase involved in CNglc catabolism has

464 been characterised in insects yet, a β-glucosidase (ZfBGD1) has been identified in
465 CNglc-containing droplets of *Zygaena filipendulae* larvae (Pentzold *et al.* 2016). Among the 326
466 DEGs in the *H. melpomene* gut after host plant shift, three homologs of β-glucosidase enzymes
467 (HMEL013290, HMEL014474 and HMEL014483) were identified. All of these were significantly
468 upregulated in larvae reared on the native host, *P. menispermifolia* (Table S5). Except for
469 sequestration and degradation of cyanogens, insects have evolved to cope with the toxic effects of
470 HCN, through conversion of HCN into β-cyanoalanine by β-cyanoalanine synthase (CAS)
471 (Despres *et al.* 2007; Wybouw *et al.* 2014). Three candidates of CAS (HMEL013489,
472 HMEL016300 and HMEL002400) were found in *H. melpomene* (Wybouw *et al.* 2014). Although
473 none of them showed significantly differential expression after host plant shift, HMEL002400 was
474 expressed higher in larvae reared on the host (FPKM=46.62) as compared to non-host larvae
475 (FPKM=20.12), with an FDR value of 0.052. Our results suggested that the differentially
476 expressed β-glucosidases and elevated CAS might play important roles in detoxifying cyanogens
477 and its toxic product, HCN.

478 *Comparison of DEGs detected by Cuffdiff2 and DESeq2.* In this study, we compared the DEGs
479 detected with Cuffdiff2 and DESeq2 methods. Generally, 217 DEGs were detected by DESeq2
480 (Table S8), in which 147 were found in the differentially expressed gene lists detected by
481 Cuffdiff2 (Table S5). Although the total number of DEGs detected by DESeq2 were decreased,
482 more than half the genes related to detoxification, peritrophic matrix biosynthesis, transporter and
483 nutrient digestion were also identified (Table S8), which would support the results based on
484 Cuffdiff2. For detoxification gene group, 7 out of 9 differentially expressed GST genes (Table S5)

were characterized as DEGs using DESeq2, meanwhile, *HmelGSTs10* and *HmelGSTs11* were identified as new DEGs. Although *HmelGSTe3* was identified as non-differentially expressed gene using DESeq2, it was a differentially expressed gene validated by qPCR (Fig. 1). UGT genes showed an exception that only 2 out of 9 UGTs were detected as DEGs using DESeq2. In typical high-throughput sequencing experiments, as few as two or three replicates per condition were used *(*Love *et al.* 2014*)*. To obtain robust results, five replicates were performed in our study. While more replicates might result in higher variance of read counts for a certain gene within group *(*Love *et al.* 2014*)*. DESeq2 is a method for differential analysis of count data, which enables a more quantitative analysis focused on the strength rather than the mere presence of differential expression (Love et al. 2014). In this study, Cuffdiff2 method was mainly used to identify DEGs, and it might find some evidences for the mere presence of differential expression.

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 (Briscoe *et al.* 2013; Brown 1981). It seems likely that the detoxification system has played an important role in host adaptation among these butterflies. In order 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), while omega and sigma classes expanded before the radiation of *Heliconius* (Fig. S6).

After the host plant shift in *H. melpomene*, 30 detoxification genes showed significant

differential expression (Table S5). 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 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 coevolution (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 analyzed 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 defense 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). Most of the differentially expressed UGTs/GSTs belonged to lineage-specific duplications (Fig. 3, Fig.4), and were upregulated in larvae reared on non-host 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. 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.

Beltran M, Jiggins CD, Brower AVZ *et al.* (2007) Do pollen feeding, pupal-mating and larval gregariousness have a single origin in *Heliconius* butterflies? Inferences from multilocus DNA

554 sequence data. *Biological Journal of the Linnean Society of London,* **92**: 221-239.
555 Benson WW, Brown KS, Gilbert LE (1976) Coevolution of plants and herbivores: passion flower
556 butterflies. *Evolution,* **29**: 659-680.
557 Briscoe AD, Macias-Munoz A, Kozak KM *et al.* (2013) Female behaviour drives expression and
558 evolution of gustatory receptors in butterflies. *PLoS Genetics,* **9**: e1003620.
559 Brower AVZ (1997) The evolution of ecologically important characters in *Heliconius* butterflies
560 (Lepidoptera: Nymphalidae): a cladistic review. *Zoological Journal of the Linnean Society,* **119**:
561 457-472.
562 Brown KS (1981) The biology of *Heliconius* and related genera. *Annual Review of Entomology,*
563 **26**: 427-456.
564 Celorio-Mancera MDLP, Wheat CW, Vogel H *et al.* (2013) Mechanisms of macroevolution:
565 polyphagous plasticity in butterfly larvae revealed by RNA-Seq. *Molecular Ecology,* **22**:
566 4884-4895.
567 Chauhan R, Jones R, Wilkinson P *et al.* (2013) Cytochrome P450-encoding genes from the
568 *Heliconius* genome as candidates for cyanogenesis. *Insect Molecular Biology,* **22**: 532-540.
569 Chikate YR, Tamhane VA, Joshi RS *et al.* (2013) Differential protease activity augments
570 polyphagy in *Helicoverpa armigera*. *Insect Molecular Biology,* **22**: 258-272.
571 Claudianos C, Ranson H, Johnson RM *et al.* (2006) A deficit of detoxification enzymes: pesticide
572 sensitivity and environmental response in the honeybee. *Insect Molecular Biology,* **15**: 615-636.
573 Conesa A, Gotz S, Garcia-Gomez JM *et al.* (2005) Blast2GO: a universal tool for annotation,
574 visualization and analysis in functional genomics research. *Bioinformatics,* **21**: 3674-3676.
575 Dermauw W, Wybouw N, Rombauts S *et al.* (2013) A link between host plant adaptation and
576 pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *Proceedings of the*
577 *National Academy of Sciences of the United States of America,* **110**: E113-122.
578 Despres L, David JP, Gallet C (2007) The evolutionary ecology of insect resistance to plant
579 chemicals. *Trends in Ecology & Evolution,* **22**: 298-307.
580 Dhawan K, Dhawan S, Sharma A (2004) *Passiflora*: a review update. *Journal of*
581 *Ethnopharmacology,* **94**: 1-23.
582 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput.
583 *Nucleic Acids Research,* **32**: 1792-1797.
584 Ehrlich PR, Raven PH (1964) Butterflies and plants: a study in coevoution. *Evolution,* **18**:
585 586-608.
586 Engler-Chaouat HS, Gilbert LE (2007) De novo synthesis vs. sequestration: negatively correlated
587 metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. *Journal*
588 *of Chemical Ecology* **33**: 25-42.
589 Glavinas H, Krajcsi P, Cserepes J *et al.* (2004) The role of ABC transporters in drug resistance,
590 metabolism and toxicity. *Current drug delivery,* **1**: 27-42.
591 Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for
592 comparative protein modeling. *Electrophoresis,* **18**: 2714-2723.
593 Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of
594 mimicry adaptations among species. *Nature,* **487**: 94-98.
595 Hoang K, Matzkin LM, Bono JM (2015) Transcriptional variation associated with cactus host
596 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, 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.

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*

640 *Evolutionary Biology,* **26**: 1959-1967.
641 Mueller LA, Goodman CD, Silady RA *et al.* (2000) AN9, a petunia glutathione S-transferase
642 required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiology,* **123**:
643 1561-1570.
644 Nahrstedt A, Davis RH (1981) Cyanogenic glycosides in butterflies: detection and synthesis of
645 linamarin and lotaustralin in the heliconiinae. *Planta Medica,* **42**: 124-125.
646 Nakamura C, Yajima S, Miyamoto T *et al.* (2013) Structural analysis of an epsilon-class
647 glutathione transferase from housefly, *Musca domestica*. *Biochemical and Biophysical Research*
648 *Communications,* **430**: 1206-1211.
649 Nam J, Nei M (2005) Evolutionary change of the numbers of homeobox genes in bilateral animals.
650 *Molecular Biology and Evolution,* **22**: 2386-2394.
651 Oakeshott JG, Johnson RM, Berenbaum MR *et al.* (2010) Metabolic enzymes associated with
652 xenobiotic and chemosensory responses in *Nasonia vitripennis*. *Insect Molecular Biology,* **19**:
653 147-163.
654 Olafsdottir ES, Andersen JV, Jaroszewski JW (1988) Cyanohydrin glycosides of Passifloraceae.
655 *Phytochemistry,* **28**: 127-132.
656 Pentzold S, Zagrobelny M, Khakimov B, *et al.* (2016) Lepidopteran defence droplets - a
657 composite physical and chemical weapon against potential predators. *Scientific Reports,* **6**:
658 22407.
659 Pringle EG, Baxter SW, Webster CL *et al.* (2007) Synteny and chromosome evolution in the
660 lepidoptera: evidence from mapping in Heliconius melpomene. *Genetics,* **177**: 417-426.
661 Ragland GJ, Almskaar K, Vertacnik KL *et al.* (2015) Differences in performance and
662 transcriptome-wide gene expression associated with *Rhagoletis* (Diptera: Tephritidae) larvae
663 feeding in alternate host fruit environments. *Molecular Ecology,* **24**: 2759-2776.
664 Ranson H, Claudianos C, Ortelli F *et al.* (2002) Evolution of supergene families associated with
665 insecticide resistance. *Science,* **298**: 179-181.
666 Ranson H, Hemingway J (2005) Glutathione transferases. In: *Comprehensive molecular insect*
667 *science-pharmacology Volume 5*: *Glutathione transferases.* (eds Gilbert LI, Iatrou K, Gill SS),
668 pp. 383-398.Oxford: Elsevier.
669 Reddy VS, Shlykov MA, Castillo R *et al.* (2012) The major facilitator superfamily (MFS)
670 revisited. *FEBS Journal,* **279**: 2022-2035.
671 Saier MH Jr, Yen MR, Noto K *et al.* (2009) The transporter classification database: recent
672 advances. *Nucleic Acids Research,* **37**: D274-278.
673 Smiley J (1978) Plant chemistry and the evolution of host specificity: new evidence from
674 *Heliconius* and *Passiflora*. *Science,* **201**: 745-747.
675 Smiley JT, Wisdom CS (1985) Determinants of growth rate on chemically heterogeneous host
676 plants by specialist insects. *Biochemical Systematics and Ecology,* **13**: 305-312.
677 Spencer KC (1988) Chemical mediation of coevolution in the *Passiflora-Heliconius* interaction. In:
678 *Chemical mediation of coevolution* (eds Spencer KC), pp. 167-240. London: Academic Press.
679 Strode C, Wondji CS, David JP *et al.* (2008) Genomic analysis of detoxification genes in the
680 mosquito *Aedes aegypti*. *Insect Biochemistry and Molecular Biology,* **38**: 113-123.
681 Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence
682 alignments into the corresponding codon alignments. *Nucleic Acids Research,* **34**: W609-612.

683 Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis
684 using maximum likelihood, evolutionary distance, and maximum parsimony methods.
685 *Molecular Biology and Evolution*, **28**: 2731-2739.
686 Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq.
687 *Bioinformatics*, **25**: 1105-1111.
688 Trapnell C, Roberts A, Goff L *et al.* (2012) Differential gene and transcript expression analysis of
689 RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, **7**: 562-578.
690 Trapnell C, Williams BA, Pertea G *et al.* (2010) Transcript assembly and quantification by
691 RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.
692 *Nature Biotechnology*, **28**: 511-515.
693 Tribolium Genome Sequencing Consortium (2008) The genome of the model beetle and pest
694 *Tribolium castaneum*. *Nature*, **452**: 949-955.
695 Turlings TC, Loughrin JH, McCall PJ *et al.* (1995) How caterpillar-damaged plants protect
696 themselves by attracting parasitic wasps. *Proceedings of the National Academy of Sciences of*
697 *the United States of America*, **92**:4169-74.
698 Wang Y, Tang H, Debarry JD *et al.* (2012) MCScanX: a toolkit for detection and evolutionary
699 analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**: e49.
700 Wybouw N, Dermauw W, Tirry L, *et al.* (2014) A gene horizontally transferred from bacteria
701 protects arthropods from host plant cyanide poisoning. *Elife*, **3**: e02365.
702 Wybouw N, Zhurov V, Martel C *et al.* (2015) Adaptation of a polyphagous herbivore to a novel
703 host plant extensively shapes the transcriptome of herbivore and host. *Molecular Ecology*, **24**:
704 4647-63.
705 Xie C, Mao X, Huang J *et al.* (2011) KOBAS 2.0: a web server for annotation and identification
706 of enriched pathways and diseases. *Nucleic Acids Research*, **39**: W316-322.
707 Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and*
708 *Evolution*, **24**: 1586-1591.
709 Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under
710 positive selection. *Molecular Biology and Evolution*, **22**: 1107-1118.
711 Yu Q, Lu C, Li B *et al.* (2008) Identification, genomic organization and expression pattern of
712 glutathione S-transferase in the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular*
713 *Biology*, **38**: 1158-1164.
714 Zagrobelny M, Bak S, Rasmussen AV, *et al.* (2004) Cyanogenic glucosides and plant-insect
715 interactions. *Phytochemistry*, **65**: 293-306.
716 Zhan S, Merlin C, Boore JL *et al.* (2011) The monarch butterfly genome yields insights into
717 long-distance migration. *Cell*, **147**: 1171-1185.

718 **Data accessibility**

719 Illumina reads for each of the libraries were deposited in Sequence Read Archive (SRA) database
720 under the accession number: SRP074338. Supplementary data is deposited in the Dryad repository:
721 doi:10.5061/dryad.65k4n.
722

Author Contributions

Y.Q.Y. and C.D.J. conceived and designed the study. Y.Q.Y. performed the experiments, analyzed 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.

Figure and Table captions

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.

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 *B. mori*, *D. plexippus* and *H. melpomene* was from Pringle *et al*. (2007). Gene numbers of the other detoxification enzymes were obtained from previous literature (Ahn *et al.* 2012; Claudianos *et al.* 2006; Oakeshott *et al.* 2010; Strode *et al.* 2008; Tribolium Genome Sequencing Consortium 2008; Yu *et al.* 2008). De: delta; Ep: epsilon; Om: omega; Si: sigma; Th: theta; Ze: zeta; Un: unclassified class.

Fig. 3 Phylogenetic tree of GSTs in *H. melpomene*, *B. mori* and *D. plexippus*. The tree was reconstructed with MEGA5.0 using neighbor-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 *B. mori* were retrived from Yu *et al.* (2008). GST sequences of *H. melpomene* and *D. plexippus* are in Table S2.

Fig. 4 Phylogenetic tree of UGTs in *H. melpomene*, *D. plexippus*, *B. 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 retrived from Huang *et al.* (2008) and Ahn *et al.* (2012). UGT sequences of *H. melpomene* and *D. plexippus* were from Table S2.

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 non-differentially expressed detoxifying genes, 21 non-differentially expressed GST (oGST) genes were characterized in a further nine *Heliconius* species (Table S3). GSTs were divided into N-terminus (GST-N) and C-terminus (GST-C) regions to compare evolutionary rates in different functional domains.

Fig. 6 Homology modeling 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 modeling 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 modeling (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 gray, respectively.

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.

Fig. S1 Gene ontology classification of the postman butterfly genes expressed in the gut. GEG: gene expressed in the gut. GHG: gene highly-rexpressed in the gut.

Fig. S2 Scatterplot of enriched pathways for DEGs in larval gut after host plant shift. The size and color of the dots represent the gene number and the range of the FDR value, respectively. Rich factor is the ratio of the differentially expressed gene number to the total gene number in a certain pathway. Four pathway databases were used in our analysis, including Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, Reactome, BioCyc and Protein Analysis Through Evolutionary Relationships (PANTHER).

Fig. S3 NJ tree and intron positions of HmelGST genes. Phase 0, 1, and 2 introns are shown by black, blue, and red solid lines, respectively.

Fig. S4 Distributions of *H. melpomene* GST and UGT genes on chromosmes. Scaffold arrangement is based on the published linkage map (*Heliconius* Genome Consortium 2012).

Fig. S5 Duplication mechanisms and conserved motif of UGT33 family in *H. melpomene*. MCScanX (max gaps = 50) was used to identify the segmental and tandem duplications of detoxification genes in *H. melpomene*. This annotated tree was predicted by the collinear and tandem relationships of the output from 'family tree plotter of MCScanX'. The characters T and S on the nodes mean tandem and segmental duplications, respectively. All the UGT33 family members in Fig. 4 were used to create sequence logo of signature region (http://weblogo.berkeley.edu/logo.cgi).

Fig. S6 Gene gain and loss of GSTs superfamily in *Heliconius* butterflies. The species tree was derived from a phylogeny based on independent nuclear and mitochondrial DNA sequences (Beltran *et al.* 2007). Each class of GSTs from 10 *Heliconius* species was used to reconstruct the NJ phylogenetic tree. Gene gain and loss was estimated with the method of Nam & Nei (2005).

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*. "host" and "non" represent the samples from individuals reared with host and non-host plant, respectively.

The number after the sample name means replicates.

Table S5 Annotations and expression signals of 326 DEGs in larval gut after host plant shift. The protein sequences of all the genes were used as queries to BLASTp search against nr database in NCBI (http://www.ncbi.nlm.nih.gov/). Transporter Classification DataBase (TCDB) was searched for the best hits of transporter genes (Saier *et al.* 2009). Peritrophin and genes related to nutrient digestion were classified by search against PFAM database (http://pfam.xfam.org/).

Table S6 GO enrichment ananlysis of the differentially expressed genes.

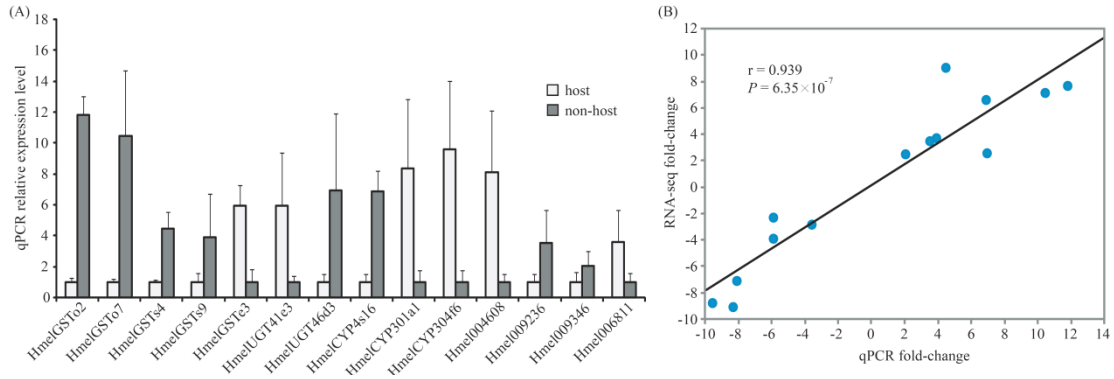
Table S7 Orthologous divergence of detoxification genes among *Heliconius* butterflies.

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

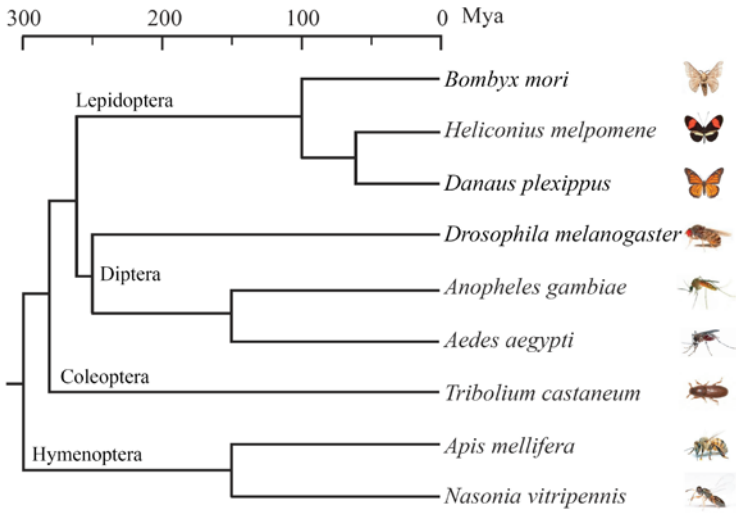
Model	No. of parameters	Estimates of parameters	ln L	P-value	BEB*
M0: one ratio	1	$\omega_0 = 0.29855$	-446.40		
M1a: neutral	1	$p_0 = 0.82404$, $p_1 = 0.17596$	-443.02		
		$\omega_0 = 0.05580$, $\omega_1 = 1.00000$			
M2a: selection	3	$p_0 = 0.98698$, $p_1 = 0.00000$, $p_2 = 0.01302$	-437.75	0.0051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

*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.

813 Fig. 1

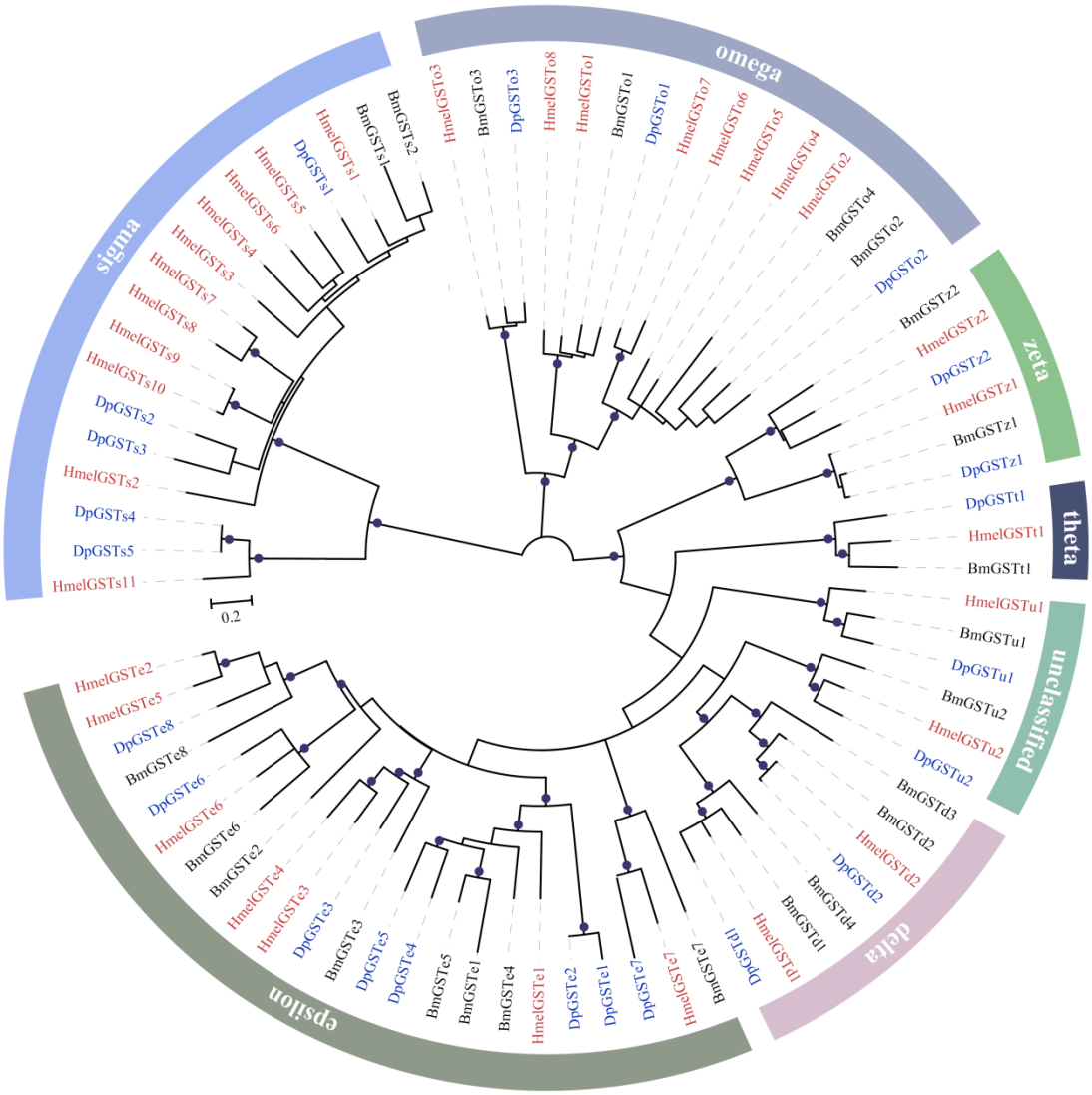

(A)
qPCR relative expression level
host
non-host
HmelGSTo2
HmelGSTo7
HmelGSTs4
HmelGSTs9
HmelGSTe3
HmelUGT41e3
HmelUGT46d3
HmelCYP4s16
HmelCYP301a1
HmelCYP304f6
Hmel004608
Hmel009236
Hmel009346
Hmel006811
(B)
r = 0.939
P = 6.35×10-7
RNA-seq fold-change
qPCR fold-change

814 Fig. 2

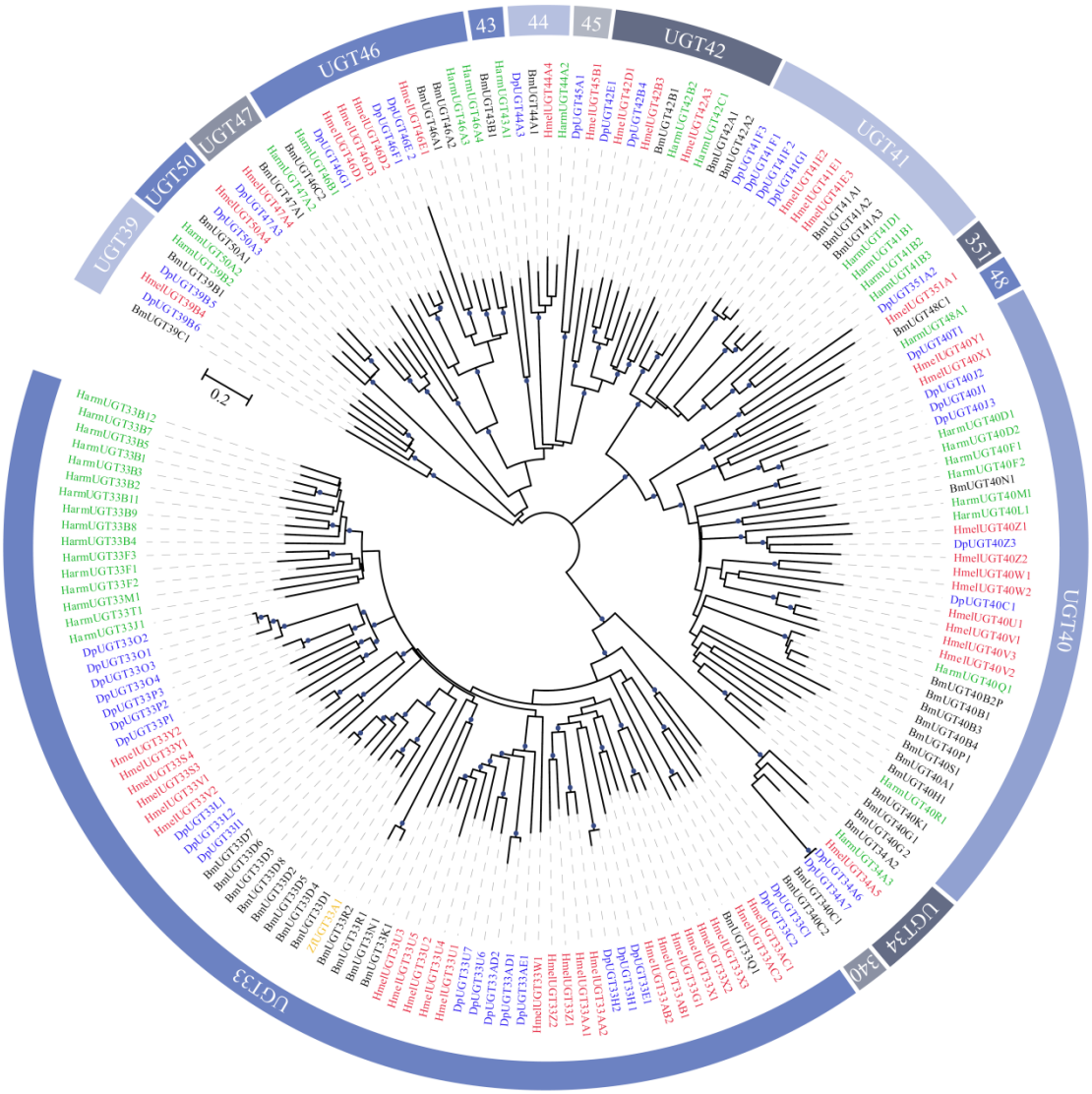


	UGT	GST							
		Total	De	Ep	Om	Si	Th	Ze	Un
Bombyx mori	45	23	4	8	4	2	1	2	2
Heliconius melpomene	52	33	2	7	8	11	1	2	2
Danaus plexippus	49	23	2	8	3	5	1	2	2
Drosophila melanogaster	34	37	11	14	5	1	4	2	0
Anopheles gambiae	26	28	12	8	1	1	2	1	3
Aedes aegypti	34	27	8	8	1	1	4	1	3
Tribolium castaneum	43	35	26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

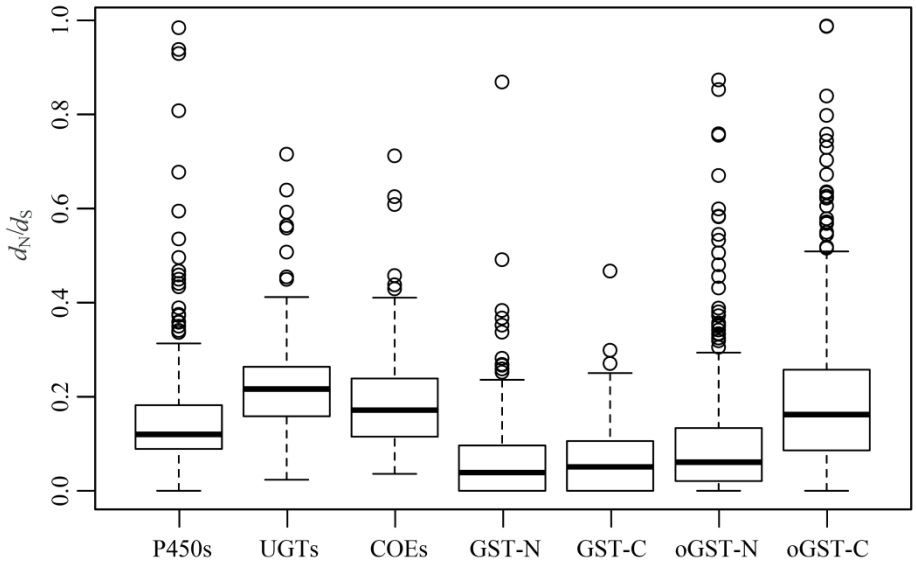
815 Fig. 3


omega
zeta
theta
unclassified
delta
epsilon
sigma
0.2
HmelGSTo3
BmGSTo3
DpGSTo3
HmelGSTo8
HmelGSTo1
BmGSTo1
DpGSTo1
HmelGSTo7
HmelGSTo6
HmelGSTo5
HmelGSTo4
HmelGSTo2
BmGSTo4
BmGSTo2
DpGSTo2
BmGSTz2
HmelGSTz2
DpGSTz2
HmelGSTz1
BmGSTz1
DpGSTz1
DpGSTt1
HmelGSTt1
BmGSTt1
HmelGSTu1
BmGSTu1
DpGSTu1
BmGSTu2
HmelGSTu2
DpGSTu2
BmGSTd3
BmGSTd2
HmelGSTd2
DpGSTd2
BmGSTd4
BmGSTd1
HmelGSTd1
DpGSTd1
BmGSTe7
HmelGSTe7
DpGSTe7
DpGSTe1
DpGSTe2
HmelGSTe1
BmGSTe4
BmGSTe1
BmGSTe5
DpGSTe4
DpGSTe5
BmGSTe3
DpGSTe3
HmelGSTe3
HmelGSTe4
BmGSTe2
BmGSTe6
HmelGSTe6
DpGSTe6
BmGSTe8
DpGSTe8
HmelGSTe5
HmelGSTe2
HmelGSTs11
DpGSTs5
DpGSTs4
HmelGSTs2
DpGSTs3
DpGSTs2
HmelGSTs10
HmelGSTs9
HmelGSTs8
HmelGSTs7
HmelGSTs3
HmelGSTs4
HmelGSTs6
HmelGSTs5
DpGSTs1
HmelGSTs1
BmGSTs1
BmGSTs2

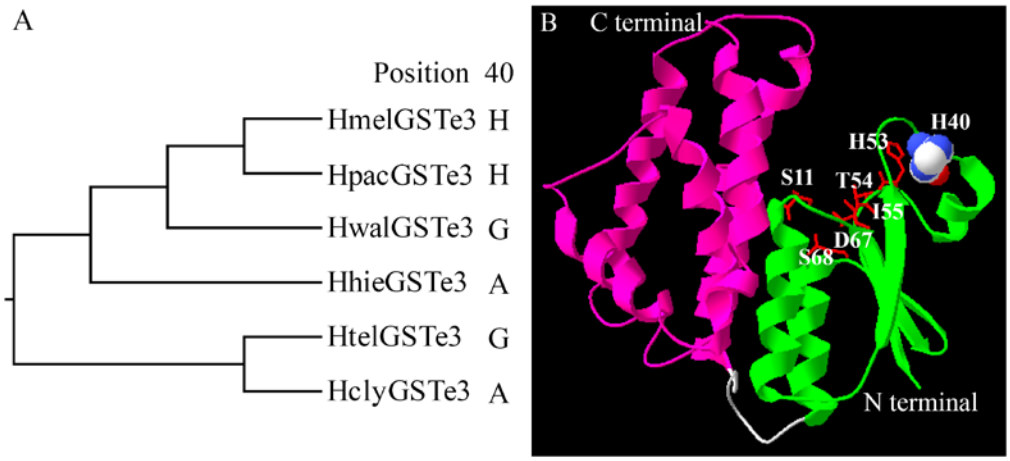
816 Fig. 4


UGT46
43
44
45
UGT42
UGT41
351
48
UGT40
UGT34
340
UGT33
UGT39
UGT50
UGT47
0.2

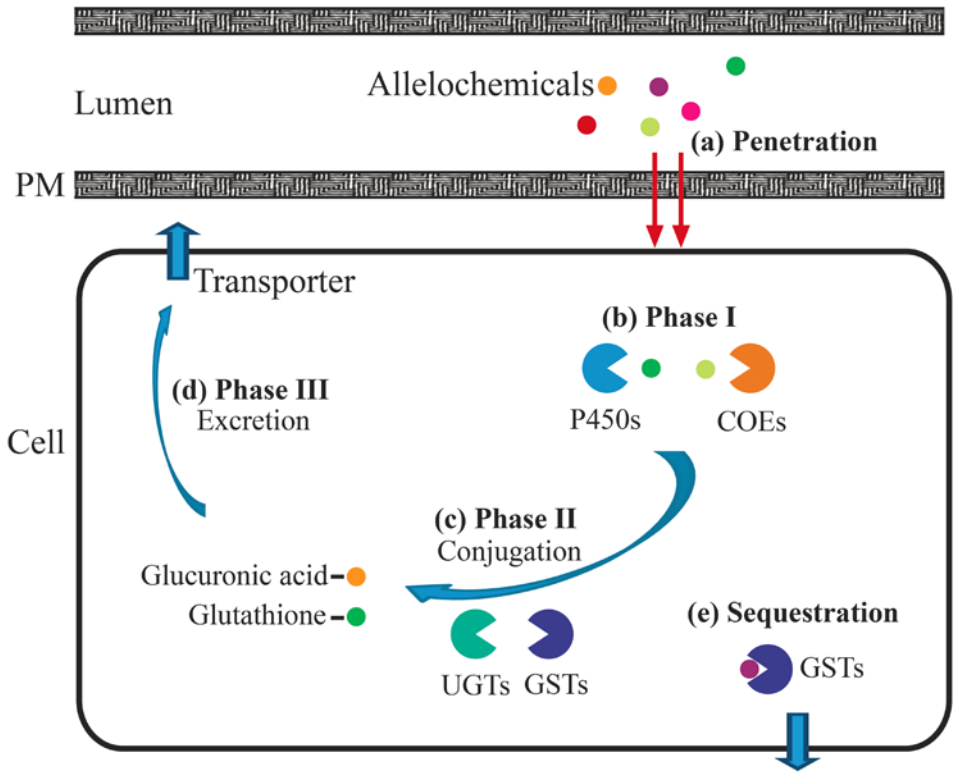
817 Fig. 5



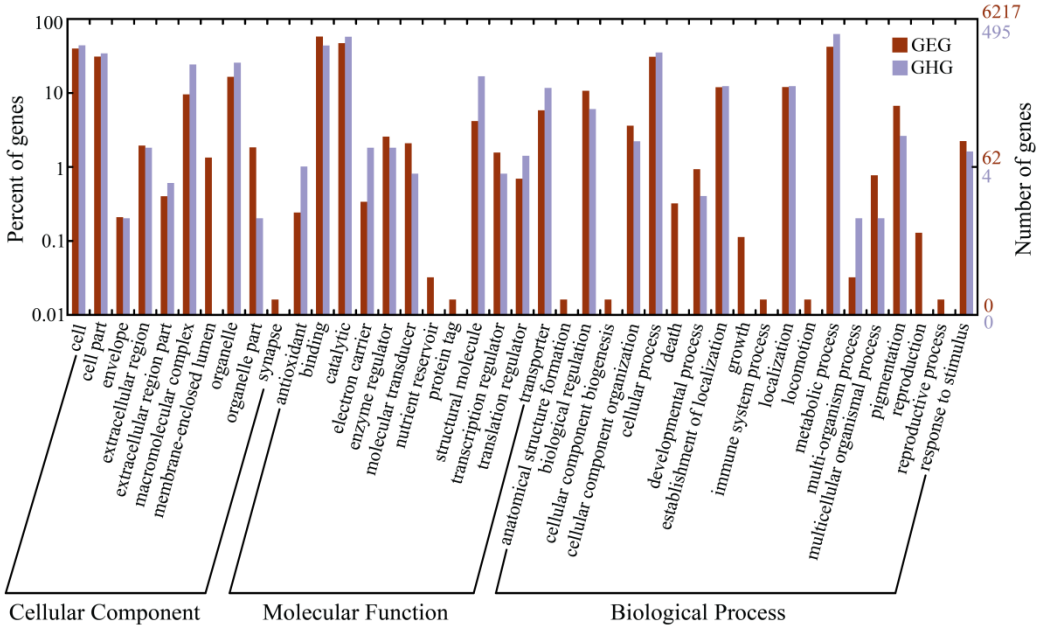
818 Fig. 6


A
Position 40
HmelGSTe3 H
HpacGSTe3 H
HwalGSTe3 G
HhieGSTe3 A
HtelGSTe3 G
HclyGSTe3 A
B
C terminal
H40
H53
S11
T54
I55
D67
S68
N terminal

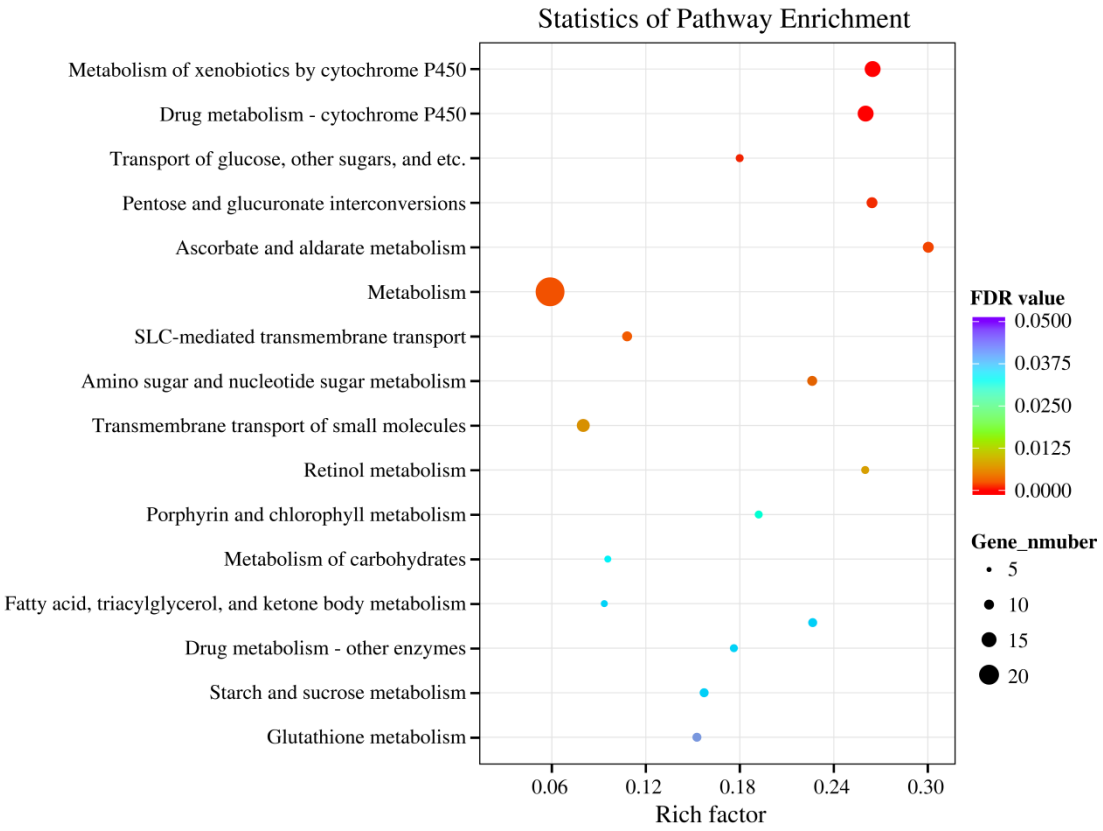
819 Fig. 7


Lumen
Allelochemicals
(a) Penetration
PM
Transporter
(b) Phase I
P450s
COEs
(d) Phase III
Excretion
Cell
(c) Phase II
Conjugation
Glucuronic acid
Glutathione
UGTs
GSTs
(e) Sequestration
GSTs

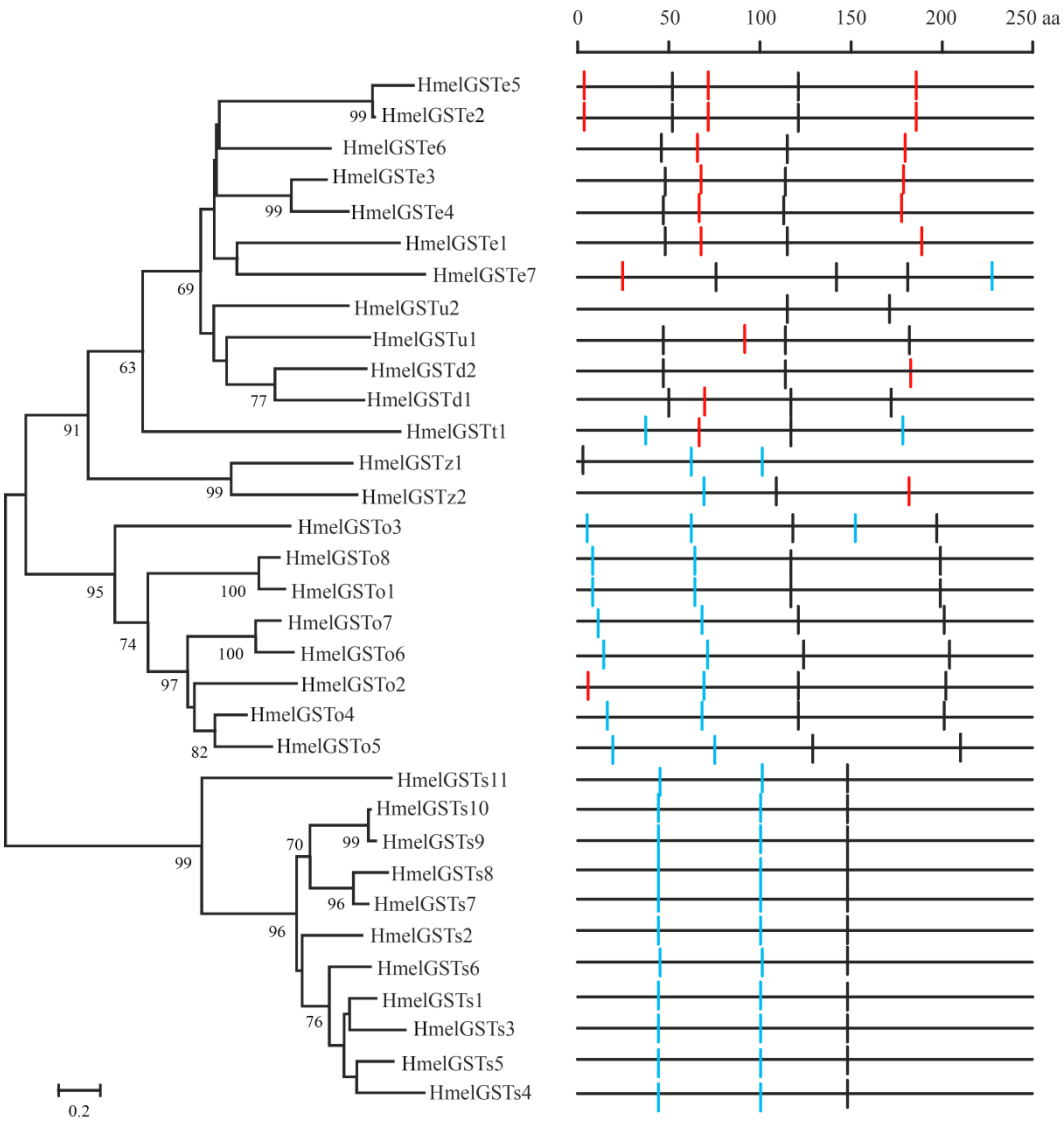
820 Fig. S1


Percent of genes
100
10
1
0.1
0.01
Number of genes
6217
495
62
4
0
0
GEG
GHG
cell
cell part
envelope
extracellular region
extracellular region part
macromolecular complex
membrane-enclosed lumen
organelle
organelle part
synapse
antioxidant
binding
catalytic
electron carrier
enzyme regulator
molecular transducer
nutrient reservoir
protein tag
structural molecule
transcription regulator
translation regulator
transporter
anatomical structure formation
biological regulation
cellular component biogenesis
cellular component organization
cellular process
death
developmental process
establishment of localization
growth
immune system process
localization
locomotion
metabolic process
multi-organism process
multicellular organismal process
pigmentation
reproduction
reproductive process
response to stimulus
Cellular Component
Molecular Function
Biological Process

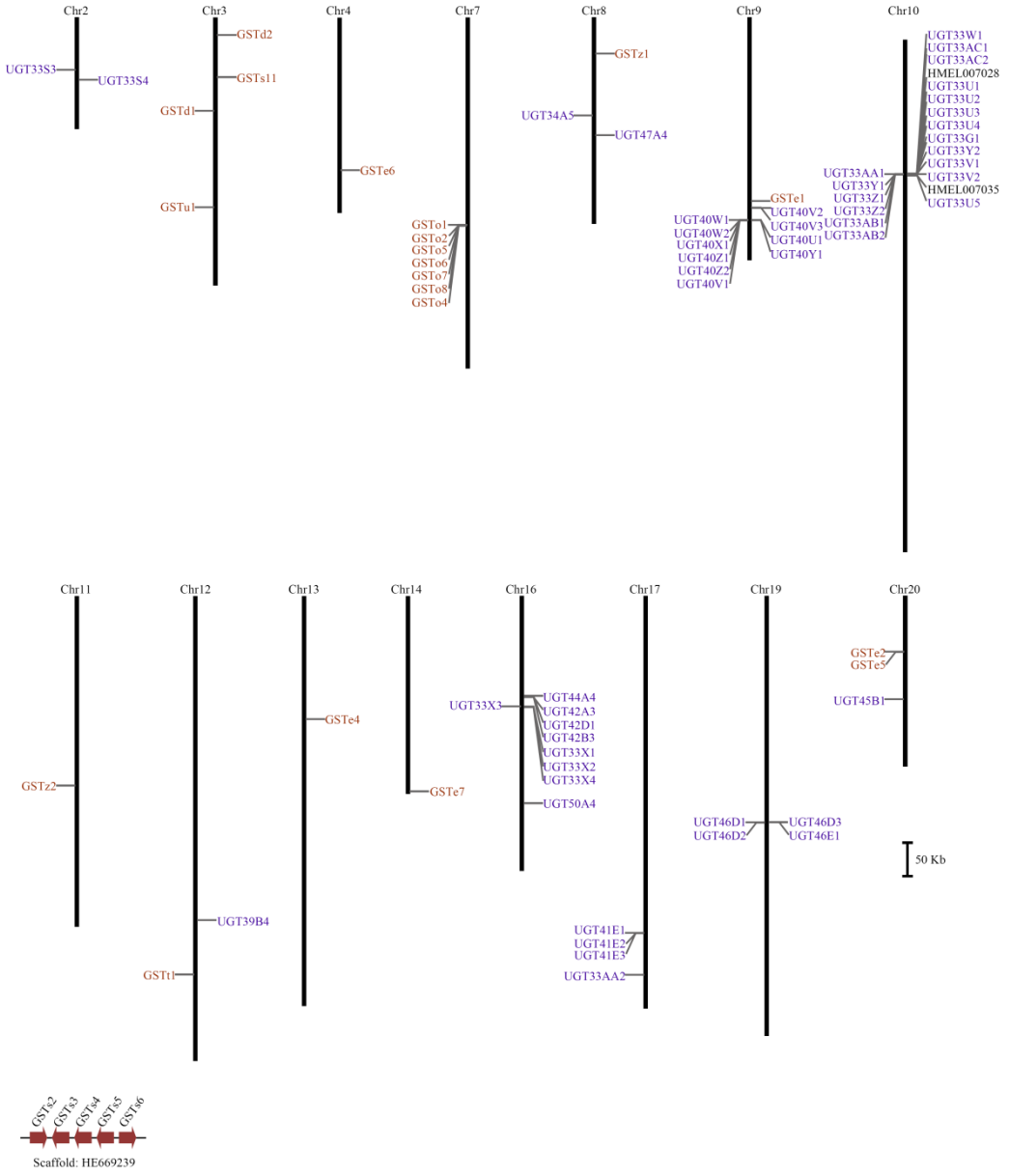
821 Fig. S2


Statistics of Pathway Enrichment
Metabolism of xenobiotics by cytochrome P450
Drug metabolism - cytochrome P450
Transport of glucose, other sugars, and etc.
Pentose and glucuronate interconversions
Ascorbate and aldarate metabolism
Metabolism
SLC-mediated transmembrane transport
Amino sugar and nucleotide sugar metabolism
Transmembrane transport of small molecules
Retinol metabolism
Porphyrin and chlorophyll metabolism
Metabolism of carbohydrates
Fatty acid, triacylglycerol, and ketone body metabolism
Drug metabolism - other enzymes
Starch and sucrose metabolism
Glutathione metabolism
0.06
0.12
0.18
0.24
0.30
Rich factor
FDR value
0.0500
0.0375
0.0250
0.0125
0.0000
Gene_nmuber
5
10
15
20

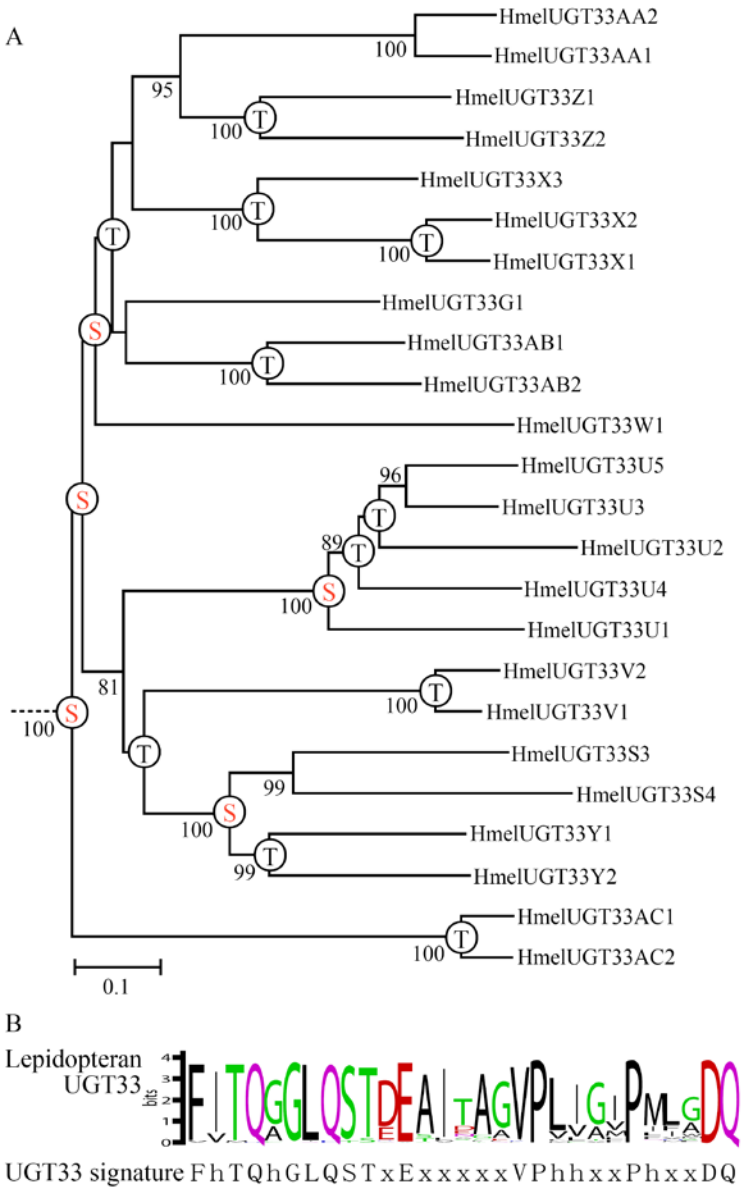
822 Fig. S3


0
50
100
150
200
250 aa
HmelGSTe5
HmelGSTe2
HmelGSTe6
HmelGSTe3
HmelGSTe4
HmelGSTe1
HmelGSTe7
HmelGSTu2
HmelGSTu1
HmelGSTd2
HmelGSTd1
HmelGSTt1
HmelGSTz1
HmelGSTz2
HmelGSTo3
HmelGSTo8
HmelGSTo1
HmelGSTo7
HmelGSTo6
HmelGSTo2
HmelGSTo4
HmelGSTo5
HmelGSTs11
HmelGSTs10
HmelGSTs9
HmelGSTs8
HmelGSTs7
HmelGSTs2
HmelGSTs6
HmelGSTs1
HmelGSTs3
HmelGSTs5
HmelGSTs4
99
69
63
77
91
99
95
100
74
100
97
82
99
70
99
96
96
76
0.2

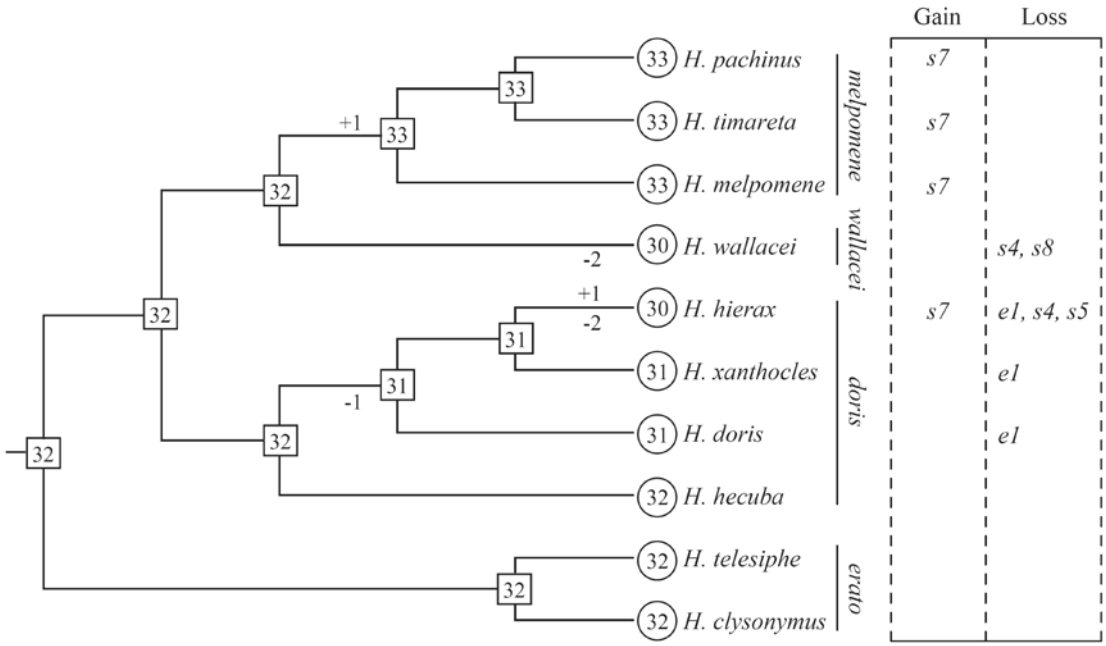
823 Fig. S4


Chr2
UGT33S3
UGT33S4
Chr3
GSTd2
GSTs11
GSTd1
GSTu1
Chr4
GSTe6
Chr7
GSTo1
GSTo2
GSTo5
GSTo6
GSTo7
GSTo8
GSTo4
Chr8
GSTz1
UGT34A5
UGT47A4
Chr9
GSTe1
UGT40V2
UGT40V3
UGT40U1
UGT40Y1
UGT40W1
UGT40W2
UGT40X1
UGT40Z1
UGT40Z2
UGT40V1
Chr10
UGT33W1
UGT33AC1
UGT33AC2
HMEL007028
UGT33U1
UGT33U2
UGT33U3
UGT33U4
UGT33G1
UGT33Y2
UGT33V1
UGT33V2
HMEL007035
UGT33U5
UGT33AA1
UGT33Y1
UGT33Z1
UGT33Z2
UGT33AB1
UGT33AB2
Chr11
GSTz2
Chr12
UGT39B4
GSTt1
Chr13
GSTe4
Chr14
GSTe7
Chr16
UGT33X3
UGT44A4
UGT42A3
UGT42D1
UGT42B3
UGT33X1
UGT33X2
UGT33X4
UGT50A4
Chr17
UGT41E1
UGT41E2
UGT41E3
UGT33AA2
Chr19
UGT46D1
UGT46D2
UGT46D3
UGT46E1
Chr20
GSTe2
GSTe5
UGT45B1
50 Kb
GSTs2
GSTs3
GSTs4
GSTs5
GST46
Scaffold: HE669239

824 Fig. S5


A
HmelUGT33AA2
HmelUGT33AA1
HmelUGT33Z1
HmelUGT33Z2
HmelUGT33X3
HmelUGT33X2
HmelUGT33X1
HmelUGT33G1
HmelUGT33AB1
HmelUGT33AB2
HmelUGT33W1
HmelUGT33U5
HmelUGT33U3
HmelUGT33U2
HmelUGT33U4
HmelUGT33U1
HmelUGT33V2
HmelUGT33V1
HmelUGT33S3
HmelUGT33S4
HmelUGT33Y1
HmelUGT33Y2
HmelUGT33AC1
HmelUGT33AC2
100
95
100
100
100
100
100
96
89
100
100
81
100
99
100
99
100
0.1
B
Lepidopteran UGT33
bits
UGT33 signature FhTQhGLQSTxExxxxxxVPhhxxPhxxDQ

825 Fig. S6


Gain
Loss
33 H. pachinus
33 H. timareta
33 H. melpomene
30 H. wallacei
30 H. hierax
31 H. xanthocles
31 H. doris
32 H. hecuba
32 H. telesiphe
32 H. clysonymus
melpomene
wallacei
doris
erato
+1
-2
+1
-2
-1
s7
s7
s7
s4, s8
s7
e1, s4, s5
e1
e1