RESEARCH ARTICLE

De novo transcriptome analysis and identification of genes associated with immunity, detoxification and energy metabolism from the fat body of the tephritid gall fly, *Procecidochares utilis*

Lifang Li[1], **Xi Gao**[1], **Mingxian Lan**[1], **Yuan Yuan**[1], **Zijun Guo**[1], **Ping Tang**[1], **Mengyue Li**[1], **Xianbin Liao**[1], **Jiaying Zhu**[2], **Zhengyue Li**[1], **Min Ye**[1]*, **Guoxing Wu**[1]*

1 State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, China, **2** Key Laboratory of Forest Disaster Warning and Control of Yunnan Province, Southwest Forestry University, Kunming, China

* wugx1@163.com (GW); yeminpc@163.com (MY)

Abstract

The fat body, a multifunctional organ analogous to the liver and fat tissue of vertebrates, plays an important role in insect life cycles. The fat body is involved in protein storage, energy metabolism, elimination of xenobiotics, and production of immunity regulator-like proteins. However, the molecular mechanism of the fat body's physiological functions in the tephritid stem gall-forming fly, *Procecidochares utilis*, are still unknown. In this study, we performed transcriptome analysis of the fat body of *P. utilis* using Illumina sequencing technology. In total, 3.71 G of clean reads were obtained and assembled into 30,559 unigenes, with an average length of 539 bp. Among those unigenes, 21,439 (70.16%) were annotated based on sequence similarity to proteins in NCBI's non-redundant protein sequence database (Nr). Sequences were also compared to NCBI's non-redundant nucleotide sequence database (Nt), a manually curated and reviewed protein sequence database (SwissProt), and KEGG and gene ontology annotations were applied to better understand the functions of these unigenes. A comparative analysis was performed to identify unigenes related to detoxification, immunity and energy metabolism. Many unigenes involved in detoxification were identified, including 50 unigenes of putative cytochrome P450s (P450s), 18 of glutathione S-transferases (GSTs), 35 of carboxylesterases (CarEs) and 26 of ATP-binding cassette (ABC) transporters. Many unigenes related to immunity were identified, including 17 putative serpin genes, five peptidoglycan recognition proteins (PGRPs) and four lysozyme genes. In addition, unigenes potentially involved in energy metabolism, including 18 lipase genes, five fatty acid synthase (FAS) genes and six elongases of very long chain fatty acid (ELOVL) genes, were identified. This transcriptome improves our genetic understanding of *P. utilis* and the identification of a numerous transcripts in the fat body of *P. utilis* offer a series of valuable molecular resources for future studies on the functions of these genes.

Introduction

Crofton weed, *Eupatorium adenophorum* Spreng, is a perennial, hazardous invading species of the family Asteraceae, which is native to Mexico. It has successfully invaded many regions on the globe in a wide variety of natural and anthropogenic ecosystems that range from forest and grassland to farmland [1] and is now one of the most dangerous exotic invasive plants, causing great economic losses and environmental problems worldwide [2]. The tephritid stem gall-forming fly, *Procecidochares utilis* Stone (Diptera: Trypetidae), is an important natural enemy of the noxious invasive *E. adenophorum*. The larvae of flies can bore into the *E. adenophorum* leading to gall formation, which can inhibit the growth and development of its host plant and can effectively restrict the dispersal of *E. adenophorum* [3, 4]. In 1945, *P. utilis* was introduced from Mexico to Hawaii to combat *E. adenophorum* and it proved an effective biological control agent for *E. adenophorum* [5, 6]. Many countries such as New Zealand, Australia, and China have used this insect to control the damage caused by *E. adenophorum* [7, 8]. Since *P. utilis* is utilized as a biocontrol agent, many of its ecological and biological characteristics have been studied [9], but the molecular research of *P. utilis* is relatively limited, and the entire genome is not yet available. The transcriptome sequencing of its alimentary tract of *P. utilis* has been performed successfully, and has proved to be an effective method to gather genetic information of the alimentary tract [7]. These data have provided comprehensive gene expression information regarding detoxification. However, the fat body transcriptome has not been investigated.

The insect fat body is a multifunctional organ that fills the adipocytes, which play important roles in the life of insects [10]. The fat body controls the synthesis and use of energy reserves, lipid and glycogen, and has multiple biochemical functions in intermediate metabolism, including lipid, carbohydrate, amino acid and nitrogen metabolism, and protein synthesis [11]. It is a major storage depot for nutrients and fat, and these reserves can be utilized to release energy to meet the energy demands of the insect. Furthermore, the insect fat body participates in the process of immune regulation and production of antimicrobial peptides, which provides an immune barrier between the internal and external environment [12, 13]. In recent studies, a number of genes associated with immune response were identified in transcriptome of the insect fat body. For example, 71, 60 and 44 unigenes encoding various putative immune-related enzymes were identified in *Bactrocera dorsalis*, *Glossina morsitans morsitans* and *Aedes aegypti*, respectively [13–15]. In addition, the insect fat body is also involved in metabolism and detoxification of xenobiotics, harboring several detoxification enzymes [13, 16, 17]. In the *B. dorsalis* fat body transcriptome, 37 P450s, 18 GSTs and 29 ABC transporters were identified [13]. With the development of novel next generation high-throughput sequencing technology, the fact that insect fat body serves as a multifunctional organ of great biosynthetic and metabolic importance has also been confirmed by the fat body transcriptomes of *Drosophila melanogaster* [18], the tsetse fly (*G. morsitans morsitans*) [14], the yellow fever mosquito (*A. aegypti*) [15,19], the Oriental Fruit Fly (*B. dorsalis*) [13], the brown planthopper (*Nilaparvata lugens*) [11], the wing polymorphic cricket (*Gryllus firmus*) [20]. Despite the importance of this organ, little is known about its molecular machinery as it relates to its physiology, immunity and energy metabolism. In this study, we performed high-throughput Illumina HiSeqTM ^{2000} to acquire a transcriptome of the fat body for *P. utilis*. We mainly focused on genes associated with immune defense, detoxification and energy metabolism, as these genes are important for insect development. The results obtained from this study will improve our genetic understanding of *P. utilis*. At the same time, a large number of genes identified in the fat body of *P. utilis* will provide a series of valuable resources for further study of the physiological functions of these genes.

Materials and methods

Ethics statement

No specific permits were required for the insects collected in this study. The *P. utilis* used for this study is neither an endangered or protected species. Galls of *E. adenophorum* were collected in Kunming, China and no specific site permission was required for this location as it is not privately-owned or protected.

Insect samples

Galls of *E. adenophorum* were originally collected from Kunming, Yunnan Province, China. All galls were kept in cages and reared under laboratory conditions at 25°C and 75% relative humidity as described previously [8]. Fresh galls were dissected with a scalpel and then larvae were collected for later experiments.

Fat body collection and RNA isolation

Fat body samples of *P. utilis* were obtained from third instar larvae. Firstly, 75% alcohol was used to disinfect the skin of *P. utilis* larvae, then skins of the larvae were dissected from head to tail and the Malpighian tubules, cuticula, salivary glands, midgut and other residues were discarded. The fat body samples were transferred to an Eppendorf tube with Trizol on ice. They were stored at -80°C until RNA extraction.

Total RNA was isolated from the fat bodies using the Trizol reagent (Invitrogen, USA) according to the manufacturer's instructions. The concentration and quality of total RNA were determined by an Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit, Agilent Technologies, USA) and NanoDrop™ spectrophotometer (Thermo Fisher, Waltham, MA, USA), respectively.

cDNA library preparation and Illumina sequencing

The cDNA library of *P. utilis* fat body was constructed from a pool of 100 fat bodies using the mRNA-Seq Sample Preparation Kit (Illumina, San Diego, CA USA) following the manufacturer's instruction. In brief, the Poly (A) mRNA was purified from 20 μg of total RNA using Oligo (dT) magnetic beads. It was then chemically fragmented into short sequences in the presence of fragmentation buffer at 94°C for 5 min. These short sequences were used as templates to synthesize first-strand cDNA using random hexamer-primers. Subsequently, second-strand cDNAs were synthesized using buffer, dNTPs, RNaseH and DNA polymerase IDNA polymerase I (New England BioLabs, Ipswich, MA). After that, these cDNA short fragments were purified using QiaQuick PCR extraction kit (Qiagen) and then resolved with EB buffer for end reparation and single nucleotide A (adenine) addition. Finally, the cDNAs were connected with sequencing adapters. Suitable fragments (200–250 bp), as judged by agarose gel electrophoresis, were collected and used as templates for PCR amplification. The quantification of cDNA library was validated and quantified via Agilent 2100 bioanalyzer and ABI StepOnePlus Real-Time PCR System. The mean read lengths of the cDNA library were 350 bp. Finally, the cDNA libraries (three biological replicates) were pooled and sequenced in one lane on the Illumina HiSeq™ 2000 platform using paired-end technology in a single run at Beijing Genomics Institute (BGI; Shenzhen, China). A total of 45,639,322 raw reads were obtained and the raw transcriptome data was deposited in the NCBI Short Read Archive (SRA) with the accession number: SRR8521423.

Transcriptome *De novo* assembly and bioinformatics analysis

Prior to assembly, reads with adaptors, the number of reads with unknown nucleotides content larger than 5% and low quality reads which the percentage of low quality bases (base quality≤10) was more than 20% were removed from raw data by using filter_fq (BGI internal software). Meanwhile the clean data were calculated based on Q20, Q30, GC-content and sequence duplication level. The subsequent sequencing on Illumina HiSeq[TM] 2000 for paired read of 100 or 105 bases following manufacturer's manual. The clean reads were then *de novo* assembled into unigenes using Trinity v2.0.6 with a Kmer_length of 25 and all other parameters set to default [21]. Next, TIGR Gene Indices clustering tools (TGICL) was used to perform sequence clustering and generate unigenes [22]. Briefly, clean reads with a certain length of overlap (the overlap length between k-mer was equal to k-1 to extend the seed until it could no longer be extended) were combined to form contigs. Then, using paired-end reads, reads were then mapped back to the contigs and contigs from the same transcript, as well as the distances between these contigs were detected. Trinity connected these contigs to produce sequences that could not be extended on either end, were considered unigenes. After clustering, the obtained unigenes were divided into two classes: clusters, with the prefix CL, and singletons, having the prefix unigene.

After assembly, homology searches and annotation of all unigenes were performed using the BLASTn programs against nucleotide sequence databases (Nt, e-value≤ $1.0e^{-5}$) in NCBI [23]. Unigenes were also aligned by BLASTx with the non-redundant protein (Nr, e-value≤ $1.0e^{-5}$), euKaryotic Ortholog Groups (KOG), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Swiss-Prot databases [23]. Functional annotation by Gene Ontology (GO) terms (http://www.geneontology.org) was analyzed by Blast2GO v2.5.0 software with e-value less than $1.0e^{-6}$ [24] and InterPro was performed by InterProScan5 [25].

Identification and analysis of interesting genes

Sequences putatively encoding genes related to detoxification, immunity and energy metabolism, such as detoxification enzymes (GSTs, CarEs, P450s and ABC transporters), immune-related enzymes (serpin, PGRP and lysozyme) and energy metabolism enzymes (lipase, FAS and ELOVL) were identified by the BALSTx alignment against the nr database with a cut-off E-value of 10^{-5}. Contigs from the same cluster represent the same unigene. If a cluster contains several contigs, the longest one was selected for further analysis. The full open reading frames of interesting genes were determined (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and were further verified by protein BLAST results. Amino acid sequences were aligned using the Clustal W. The interesting gene families in *P. utilis* were compared to different model insect species, *D. melanogaster*, *B. dorsalis* and *C. capitate*. The GST, CarE, P450, ABC transporter, serpin, PGRP, lysozyme, lipase, FAS and ELOVL sequences of *D. melanogaster*, *B. dorsalis* and *C. capitate* were downloaded from NCBI. Subsequently, the phylogenetic trees were constructed based on the amino acid sequence alignment using Neighbor-joining (NJ) method in software MEGA 5 [26]. Bootstrap analysis of 1,000 replication trees was performed to evaluate the branch strength of each tree.

Results and discussion

Illumina sequencing and de novo assembly

Illumina sequencing was used to sequence a cDNA library of the *P. utilis* fat body, and generated about 45.64 Mb raw reads. After filtration, a total number of 41,241,514 clean reads were assembled into 40,467 contigs with a total length of 19,806,004 bp and a mean length of 489

bp. The contigs were further assembled into 30,559 unigenes with an average length of 539 bp (Table 1), which is shorter than that obtained from the *N. lugens* (avirulent TN1 population with a mean length of 656 bp and the virulent Mudgo (M) population with a mean length of 676 bp) [11] and *B. dorsalis* [13]. Among all unigenes, 4,304 unigenes (14.08%) were longer than 1,000 bp, 9,325 unigenes (30.5%) were longer between 500 and 1,000 bp, and 20,665 unigenes (67.6%) ranged from 300 to 500 bp (Fig 1).

Functional annotation of the *P. utilis* fat body unigenes

For functional annotation, a total of 30,559 unigenes were searched against the Nr, Nt, SwissProt, KEGG, KOG, InterPro and GO databases. The results showed that 21,439 (70.16%), 13,657 (44.69%), 12,188 (39.88%), 12,405 (40.59%), 12,925 (42.30%), 13,149 (43.03%) and 9,354 (30.61%) unigenes matched to Nr, Nt, SwissProt, KEGG, KOG, InterPro and GO known protein databases, respectively (Table 2). After Nr database annotation, the species distributions demonstrated that 58.58% of the unigenes had best matches against sequences of *Ceratitis capitata*, followed by *Trypanosoma brucei gambiense* DAL972 (18.83%), *Trypanosoma brucei brucei* TREU927 (10.9%) and *Musca domestica* (2.19%) (Fig 2). Even though a large percentage of transcripts matched to trypanosomes, PCR analysis confirmed the lack of contamination in *P. utilis* fat bodies and thus, the transcripts are likely not contaminants and are likely truly derived from the fat bodies of *P. utilis* (S1 Text). The species distribution was consistent with our previous study of transcriptome analysis of the alimentary tract of *P. utilis*, in which more than 50% of genes were most matched to *C. capitate* [27]. However, in the fat body of *B. dorsalis*, the highest percentage of unigene sequences were matched with *Drosophila* (over 79%) [13].

Table 1. Sequencing summary for the fat body transcriptome of *P. utilis*.

Sequencing Summary	Fat body specific transcriptome
Total raw reads(Mb)	45.64
Total clean reads	41,241,514
Total clean bases (Gb)	3.71
Q20 percentage (%)	98.72
Q30 percentage (%)	94.19
Number of contigs	40,467
Total length of contigs (bp)	19,806,004
Mean length of contigs (bp)	489
N50 length of contigs (bp)	623
N70 length of contigs (bp)	362
N90 length of contigs (bp)	232
GC percentage (%)	39.53
Number of unigenes	30,559
Total length of unigenes (bp)	16,485,836
Mean length of unigenes (bp)	539
N50 length of unigenes (bp)	685
N70 length of unigenes (bp)	403
N90 length of unigenes (bp)	255
GC percentage (%)	39.84
Total number of complete ORFs	2,808
Total number of partial ORFs	11,164

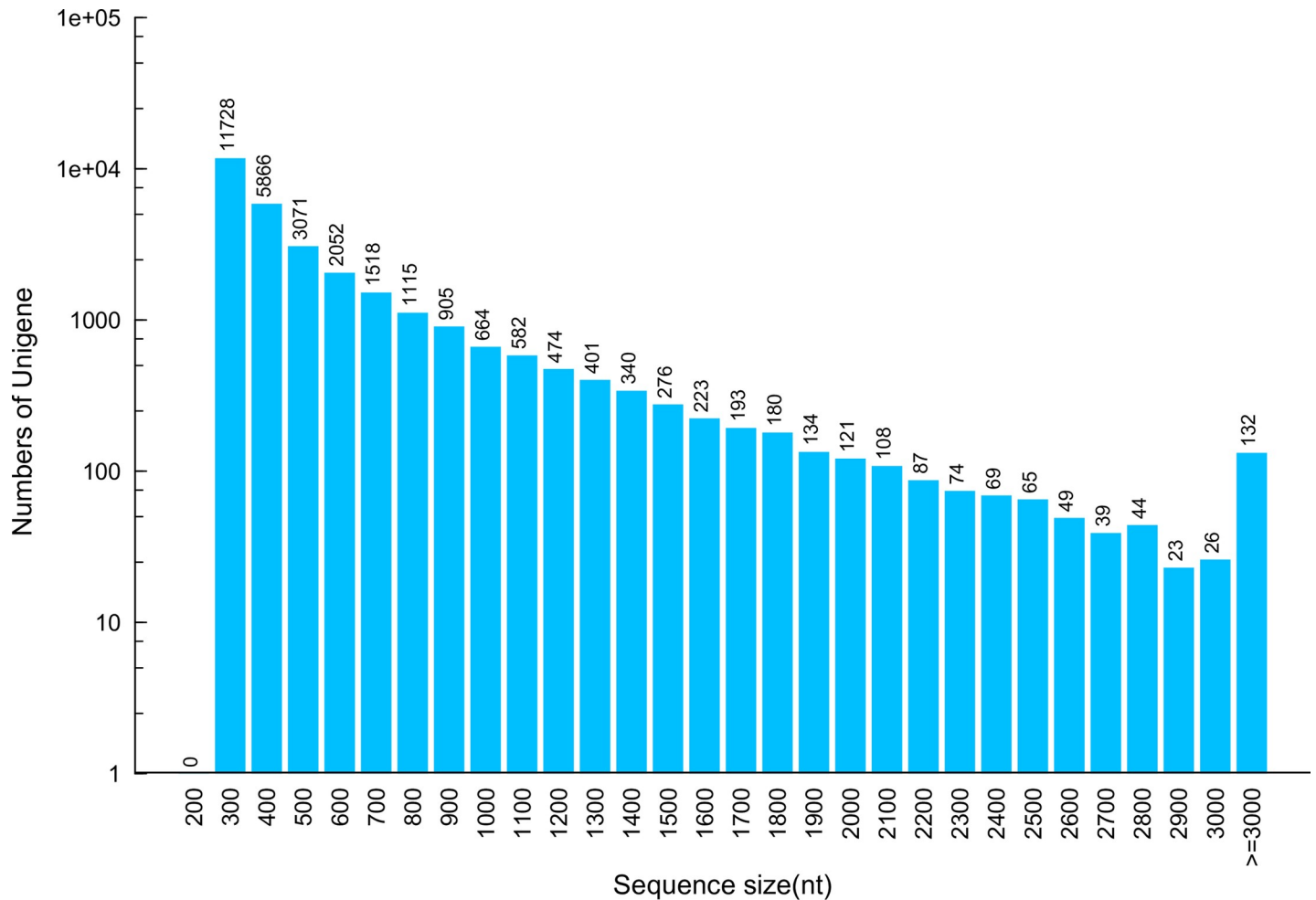https://doi.org/10.1371/journal.pone.0226039.t001

Fig 1. Length distribution of unigenes in transcriptome assembly for *P. utilis* fat body.

Gene ontology (GO) assignments were used to functionally classify the predicted unigenes from the fat body of *P. utilis*. According to the sequence similarity, 9354 unigenes (30.61%) were annotated and classified into 61 functional groups of three main ontologies: biological

Table 2. Number of unigenes annotated in seven databases.

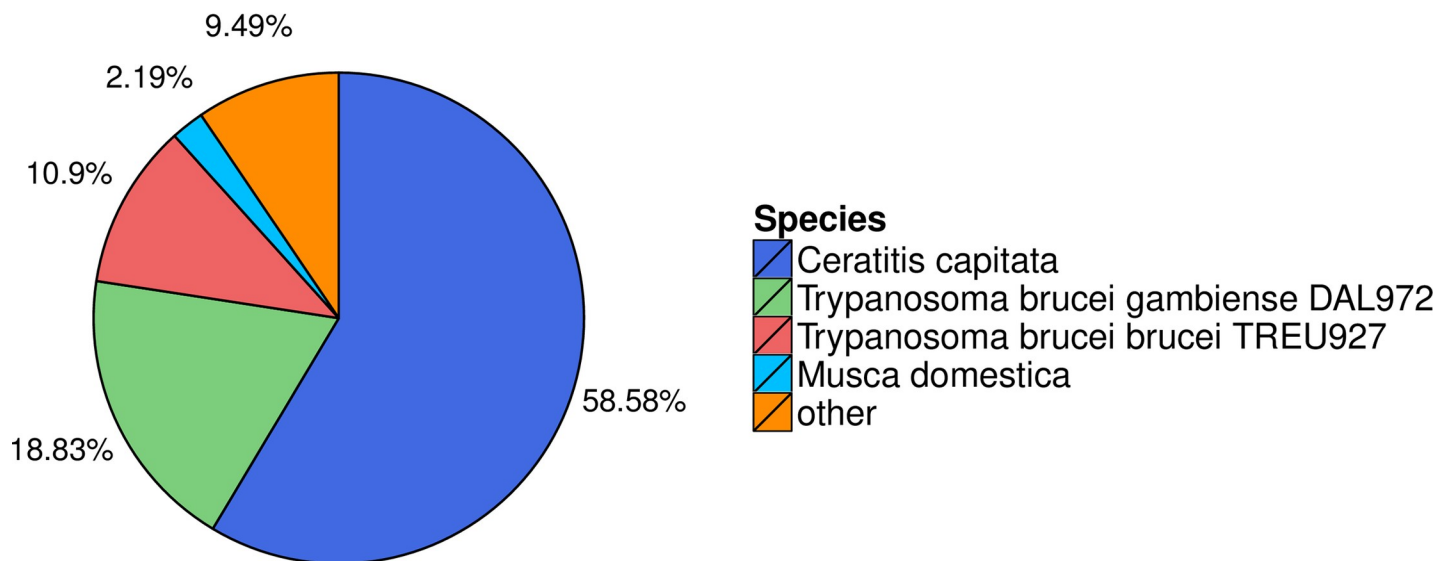| Database	The number of unigenes	Percentage (%)
Nr	21,439	70.16
Nt	13,657	44.69
Swissprot	12,188	39.88
KEGG	12,405	40.59
KOG	12,925	42.30
Interpro	13,149	43.03
GO	9,354	30.61
Total number of unigenes	30,559	100
All databases	2,692	8.81
At least one database	24,395	79.83

Species Distribution

Fig 2. Species distribution of the top BLASTX matches to the *P. utilis* fat body unigenes.

processes (46.48%) was the largest category, followed by cellular components (16.88%) and molecular function (36.64%) (Fig 3). In the biological processes category, "cellular process" and "metabolic process" terms were the two largest groups. In the cellular components category, "cell" and "cell parts" terms were the most abundant. In the molecular function, "binding" and "catalytic activity" terms were the two largest groups. The results were similar to the transcriptome study of the fat body from *N. lugens* [13]. Meanwhile, the "cell killing" term in the biological processes category was the smallest group, containing only one unigene. Only one unigene each was assigned in the functional groups "cell killing", "collagen trimer" and "translation regulator activity" (Fig 3).

In addition, to obtain more detailed annotation information of the transcriptome, all unigenes were compared against the KOG for functional prediction and classification. In all, 12,925 (42.30%) unigenes were annotated and classified into 25 KOG categories (Fig 4). Among all these categories, the "General function prediction only" contained 2,345 unigenes, which was the largest category, followed by "Signal transduction mechanisms" (1660), "Post-translational modification, protein turnover, chaperones" (1,540) (Fig 4). Among these unigenes, 1,140 were classified as "function unknown" categories, which revealed that these may be unique or novel genes in the fat body transcriptome of *P. utilis*. In contrast, "Defense mechanisms" and "Cell motility" represented relatively smaller KOG groups, containing 94 and 36 unigenes, respectively (Fig 4).

Subsequently, the KEGG pathway assignment was also performed on unigenes to identify the biological pathways including metabolic and regulatory pathways that will be actively involved in the fat body of *P. utilis*. In total, 12,405 unigenes (40.59%) were annotated to the KEGG database and mapped to the cellular processes, environmental information processing, genetic information processing, human diseases, metabolism, and organismal systems pathways (Fig 5). Among these pathways, the most abundant pathway is the global and overview maps (2,126, 17.14%), followed by signal transduction (1,436, 11.58%), translation (1,126,

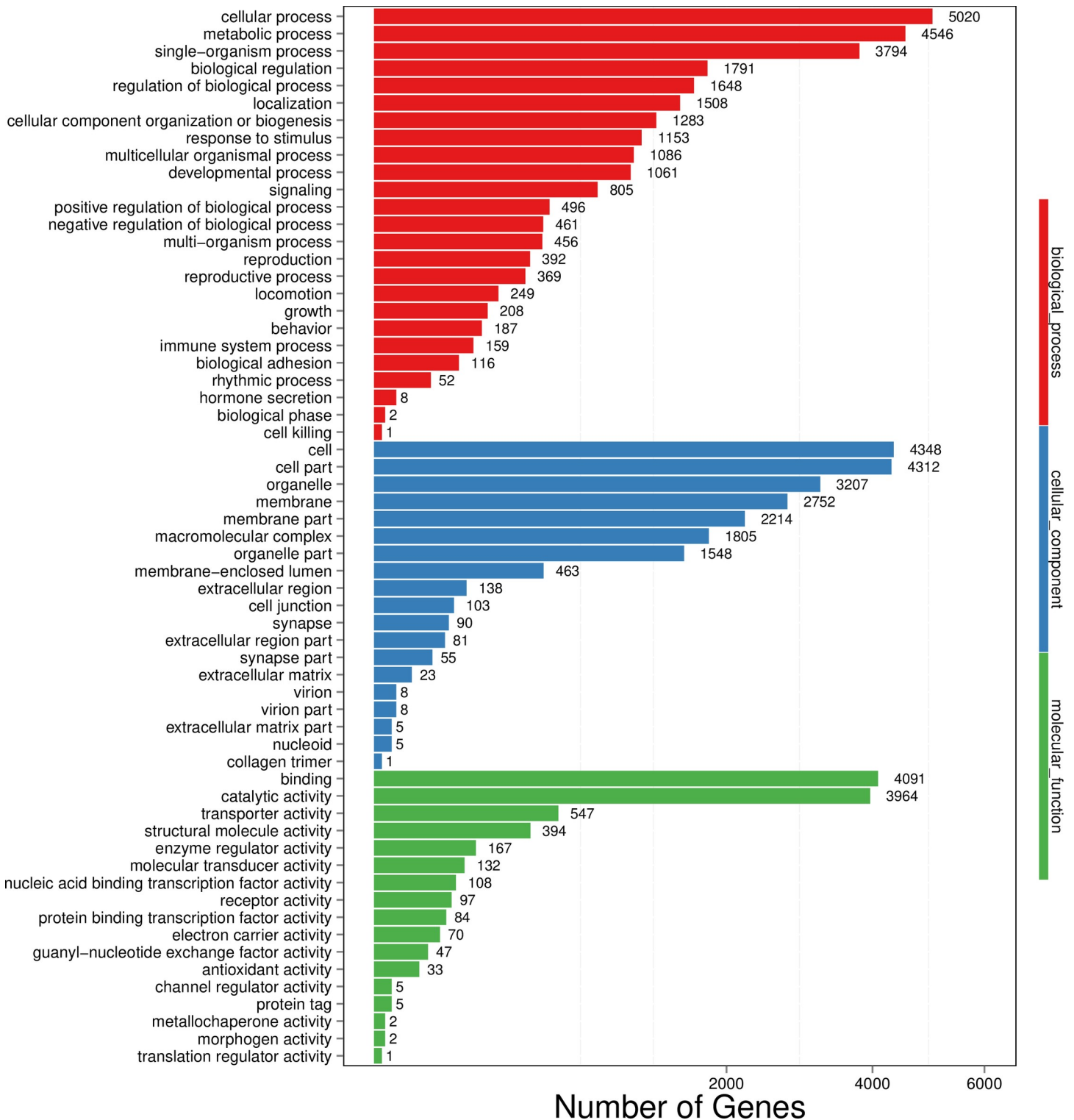**Fig 3. Classification of the gene ontology (GO) for the transcriptome of the fat body from *P. utilis*.**
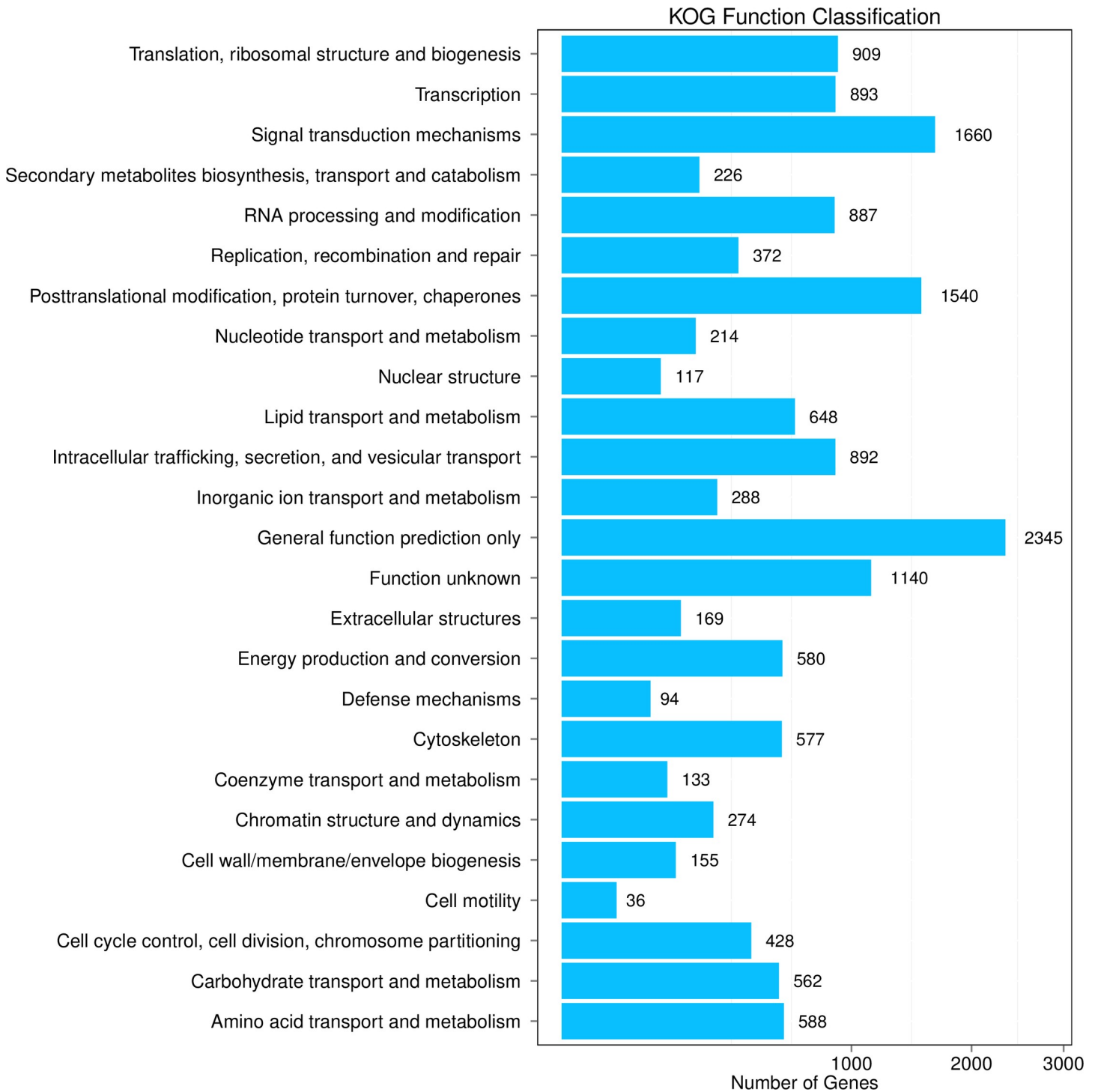
Fig 4. KOG classification of the fat body transcriptome of *P. utilis*.

9.08%), and transport and catabolism (1,030, 8.3%) (Fig 5). The KEGG pathway assignment will be helpful to further research specific biological processes, functions and pathways that exist in the fat body of *P. utilis*.
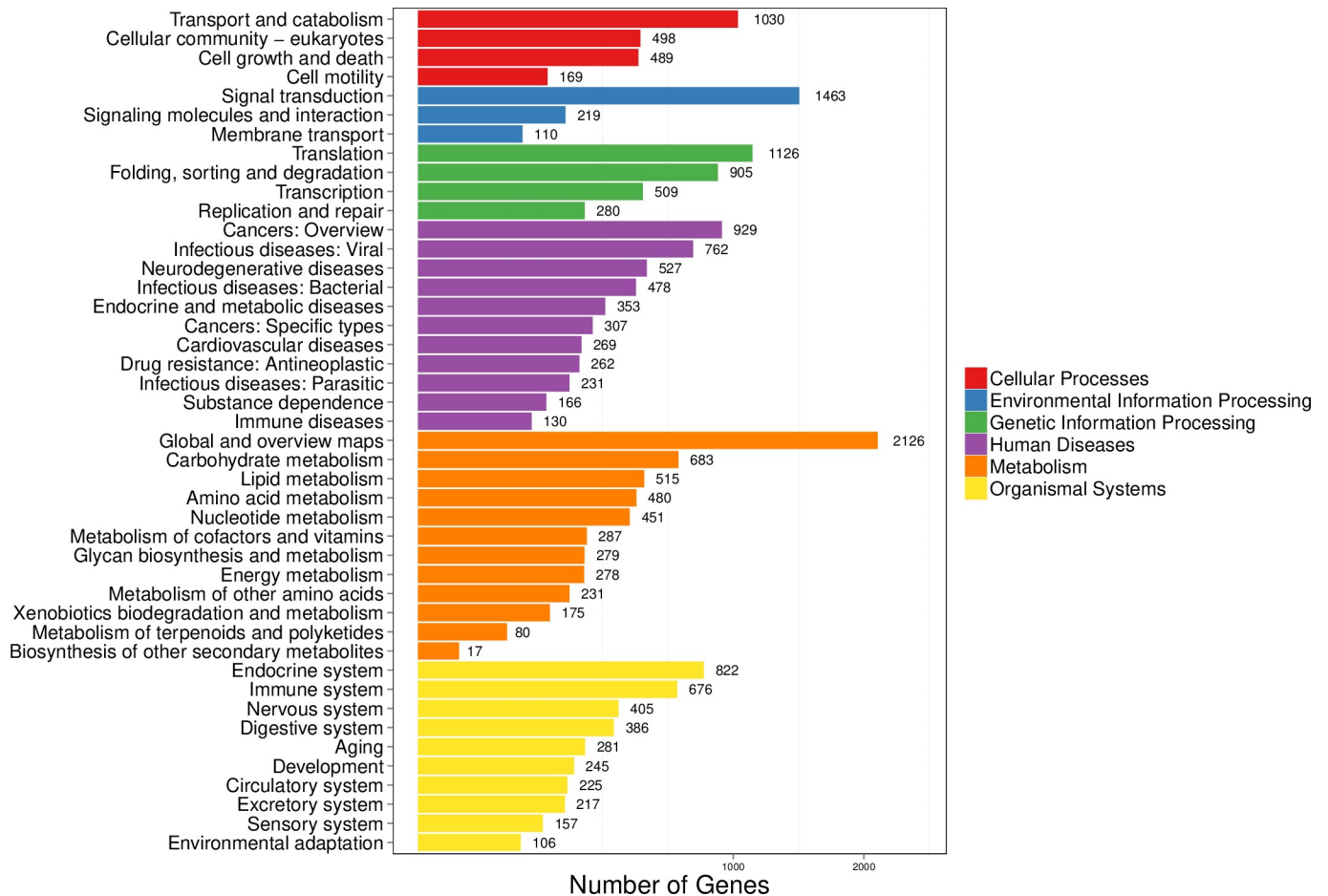
Fig 5. KEGG pathway classification of unigenes in the fat body of *P. utilis*.

Simple sequence repeats discovery

Based on the unigene sequences, a total of 2,839 sequences containing 3,585 SSRs were identified from 30,559 unigenes. Among all SSRs, 544 sequences contained more than one SSR, and trinucleotide repeats (43.65%) represented the largest proportion of microsatellite repeat units, followed by mononucleotide (24.16%), dinucleotide (23.7%), quadranucleotide (4.88%), hexanucleotide (2.00%) and pentanucleotide (1.59%) repeats (Fig 6). Among all repeat types, A/T (24.1%), AAC/GTT (23.88%) and AC/GT (10.79%) are the most abundant ones (S1 Table). In total, 3,585 SSRs were identified. The SSRs identified in this research will be helpful for further studies related to population genetic structure in *P. utilis*, such as genetic variation and gene flow.

Identification of the major detoxification-related genes

The insect fat body is an endocrine organ, which involved in metabolism and detoxification of xenobiotics, harboring several detoxification enzymes [13, 16, 17]. In this study, we obtained abundant fat body transcriptome data of *P. utilis*, and identified many unigenes that are highly similar to those related to xenobiotic metabolism in other insects, including 50 cytochrome P450 (P450s), 18 glutathione S-transferases (GSTs), 35 carboxylesterases (CarEs), 26 ATP-binding cassette (ABC) transporters.
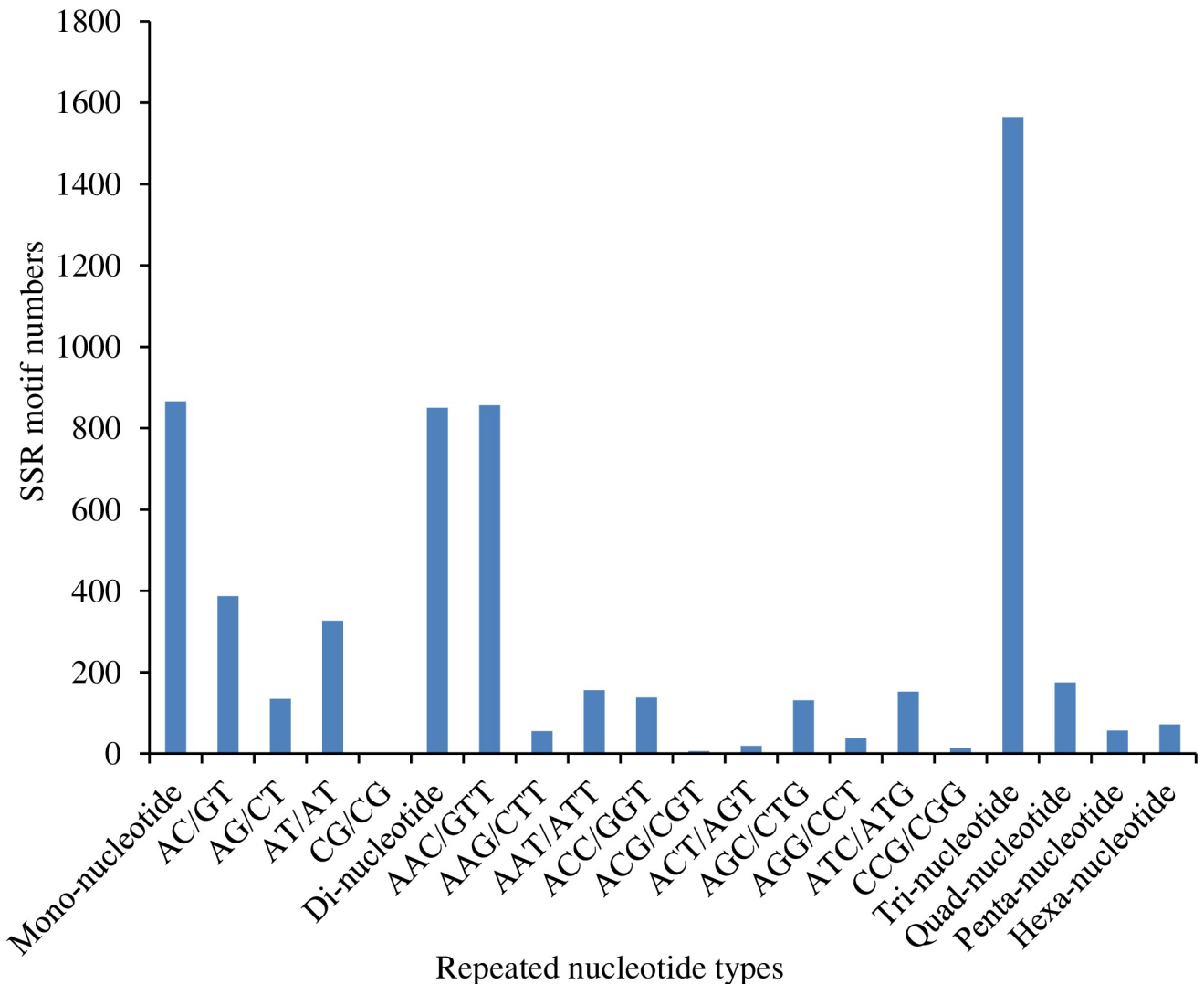
Fig 6. Statistics of SSR nucleotide classes found in the transcriptome of *P. utilis* fat body.

Cytochrome P450s. Cytochrome P450 (P450s) phase I enzymes constitute an old and widely distributed protein superfamily that participated in the metabolism and detoxification of a wide variety of plant secondary metabolites and pesticides [28, 13]. The functional and

Table 3. Number and class distribution of P450s genes in other insects.

Classification	Pu	Dm	Ss	Gm	Bm	Tc	Ag	As
CYP2 clan	1	7	6	14	7	8	10	8
CYP3clan	30	36	44	30	28	79	40	44
CYP4 clan	11	32	19	18	33	47	46	34
Mitochondrial clan	5	12	10	15	10	9	9	7
Total	50	87	79	77	78	143	105	93

Pu, *P. utilis*; Dm, *D. melanogaster*; Ss, *Shirakiacris shirakii*; Gm, *Grapholita Molesta*; Bm, *Bombyx mori*; Tc, *T. castaneum*; Ag, *A. gambiae*; As, *Anopheles sinensis*. Data of Dm, Gm, Bm, Tc, Ag from Guo *et al* (2017). Data of Ss from Qiu *et al* (2016). Data of As from Zhou *et al* (2015).

evolutionary diversity of P450s play key roles in insect adaptation to various ecological environments [29]. They are divided into four clades, CYP2, CYP3, CYP4 and mitochondrial. To date, P450s have been identified from many insects (Table 3) [29–31]. In this study, a total of 50 unigenes corresponding to P450s were identified in the *P. utilis* fat body transcriptome based on the Nr annotation results. The number was less than the number of CYP450s identified in the fat body transcriptome of *B. dorsalis*. According to the closest blast match in the NCBI nr database, all the identified P450s were divided into four clades: CYP2 (1), CYP3 (30), CYP4 (11) and the mitochondrial CYP clade (5) (S2 Table). After removing short sequences, 26 unigenes were used for phylogenetic analysis. The results showed that these unigenes were divided into seven families: Cyp9 (1), Cyp6 (13), Cyp4 (5), Cyp12 (2), Cyp309 (2), Cyp314 (2) and Cyp18 (1) (Fig 7). The majority of the P450s in *P. utilis* fat body belong to the Cyp6 family and Cyp4 family. In our previous studies, 22 unigenes were assigned to eight families according to phylogenetic analysis. Five, five, four, three and two genes belonged to the Cyp 4, Cyp 9, Cyp 6, Cyp 307 and Cyp 12 families, respectively. Lastly, single unigene expression in the alimentary canal were identified as members of Cyp 315, Cyp 314, and Cyp 302 families and most of *P450* unigenes in the alimentary tract of *P. utilis* belong to Cyp4, Cyp6 and Cyp9 families. By comparing these unigenes identified in the fat body and alimentary tract of *P. utilis*, 17 and 12 unigenes were unique to the fat body and alimentary tract, respectively, and nine unigenes were found in both tissues [27].

In previous studies, most of the P450s belong to the CYP3 and CYP4 clades. These clades in insects are involved in xenobiotic metabolism, such as pesticides and plant secondary metabolites [32–37]. For instance, three of the four CYP6F subfamily genes and *CYP9AQ2* in *Locusta migratoria* were related to the detoxification of deltamethrin or carbaryl [38, 39]. *CcCYP6A51* in *C. capitata* was associated with the detoxification of lambda-cyhalothrin [40]. *CYP4g1*, *CYP6g1* and *CYP12d1* confer DDT resistance in *D. melanogaster* [41]. In *Plutella xylostella*, *CYP6BG1* conferred resistance to pyrethroid [42], recently, overexpression of *contigs00326* and *02103*, both members of the CYP6 P450 family were known to be associated with pyrethroid detoxification in *Bactrocera oleae* [43]. Moreover, the expression levels of P450s were differentially induced by feeding on the different host plants of herbivorous insects. *Oedaleus asiaticus* grasshoppers feeding on *Artemisia frigida* (a species with low nutrient content and a high level of secondary compounds), had higher P450s enzyme activity and the gene expression level of cytochrome *P450 6K1* was highest compared to feeding on the grasses *Cleistogenes squarrosa*, *Leymus chinensis*, and *Stipa krylovii*, that level showed a significant positive correlation with the secondary compounds of host plant [33]. *Rhynchophorus ferrugineus* larvae fed on diet containing methyl eugenol, methyl isoeugenol, and rosmarinic acid tremendously enhanced the expression of P450 gene [32, 44]. Several CYP6AS subfamily genes and *CYP9Q1* in *Apis mellifera* were related to the detoxification of quercetin, a flavonol widely present in the honey bee's distinctive diet of honey and bee bread [45, 46]. Moreover, in *A. mellifera* *CYP9Q2*, and *CYP9Q3* detoxify both tau-fluvalinate and coumaphos; these twogenes can also be up-regulated by quercetin [47]. Furthermore, the expression of *CYP6AB14*, *CYP321A7* and *CYP321A9* genes of *Spodoptera litura* were induced by the toxic allochemicals xanthotoxin, coumarin, and flavones [48]. Likewise, the expression of *CYP6A8*, *CYP6D5*, *CYP6W1*, *CYP9B2*, and *CYP12D1* of *D. melanogaster* were induced by *Piper nigrum* extracts, and *CYP6A2* and *CYP6A8* expression was induced by caffeine [49, 50]. It is possible that during the evolution, *P. utilis* has produced P450s to adapt to the toxins of its host plant *E. adenophorum*.

Glutathione S-transferases. Glutathione S-transferases (GSTs) are multifunctional enzymes that play a crucial role in phase II detoxification of endogenous and xenobiotic compounds including plant secondary metabolites and pesticides [51]. In insects, GSTs are
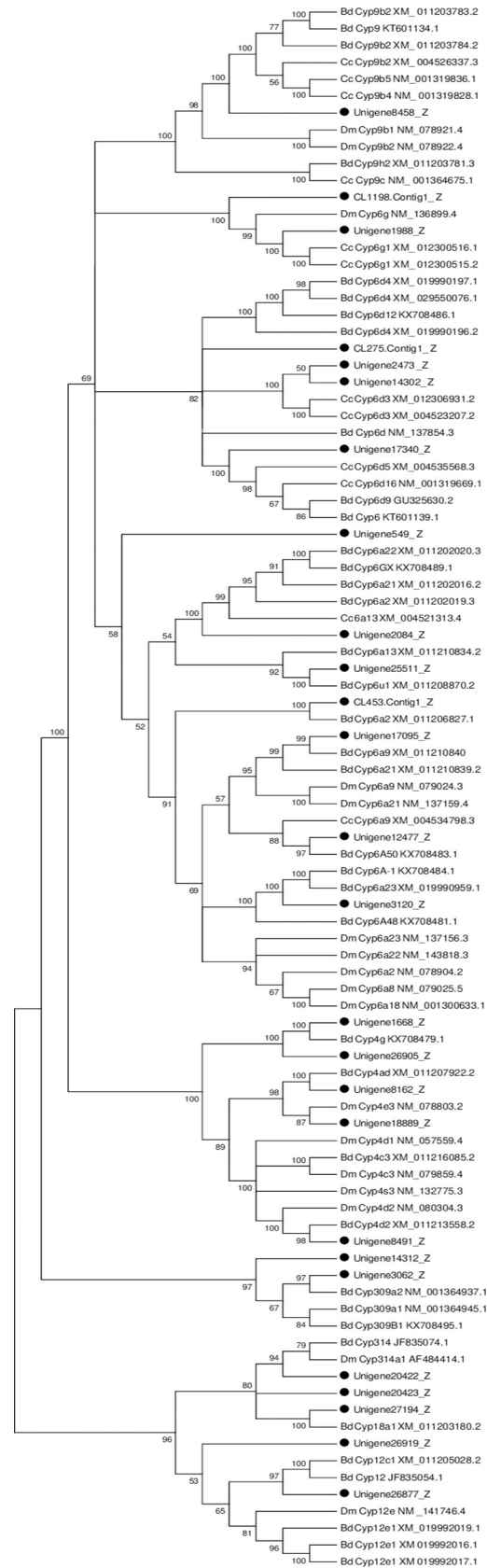
Fig 7. Neighbor-joining phylogenetic tree of *P450* genes from the *P. utilis* fat body (●) and other insects. Numbers at each branch node represent bootstrap values.

ordinarily divided into several major subclasses: Delta, Epsilon, Sigma, Omega, Theta, Zeta, Microsomal and others [52, 53]. In particular, Delta and Epsilon were two insect-specific classes and have been widely associated with responses of environmental stress, especially during xenobiotic detoxification [54, 55]. In *D. melanogaster*, *B. dorsalis*, *G. molesta* and *Bactrocera minax* 37, 14, 28, 27 GSTs genes have been identified [30, 56–58]. In this study, a total of 18 sequences encoding GSTs were identified in the fat body of *P. utilis* (S3 Table). After removing short sequences, 13 GSTs unigenes were manually selected for phylogenetic analysis. The results showed that these unigenes were assigned to six classes: seven unigenes belonged to the Epsilon class, one unigene belonged to the Delta class, and two unigenes belonged to the Theta class, one unigene was assigned to each of the Omega, Sigma and Microsomal GST classes (Fig 8). No genes belonged to the Zeta class. Most of the genes belonged to Epsilon class, which is in agreement with the previous studies from the fat body of *B. dorsalis* [13]. In our previous analysis of the alimentary canal transcriptome, 21 unigenes were used for phylogenetic analysis after removing overly short sequences. The result indicated that they belonged to seven families: Epsilon (8), Delta (4), Theta (2), Zeta (1), Omega (1), Sigma (1), and Microsomal (4) families, and like the fat body, most of the unigenes were assigned to the Epsilon family. Interestingly, by comparing these unigenes identified in the fat body and alimentary canal of *P. utilis*, eight unigenes were unique to fat body, 13 unigenes were unique to alimentary canal, and five unigenes were found in both tissue.

It has been demonstrated that Delta and Epsilon two classes of GSTs were unique in insect species and responsible for detoxification and adaptation to environmental selection pressures [57, 59]. For example, in *B. dorsalis*, three Epsilon GSTs have been implicated in resistance to malathion and overexpression of the genes increased the detoxification of malathion in *B. dorsalis* [30, 60]. The *LmGSTd1* gene might be related to the resistance of chlorpyrifos in *L. migratoria*, and the expression of this gene was induced by chlorpyrifos [61]. The expression level of *Drosophila GSTD2* and *GSTD5* genes were upregulated after exposure to the heavy metals, cadmium and zinc [62]. In addition, some studies suggest that *Metaseiulus occidentalis* and *Tetranychus urticae* Microsomal GSTs may be involved in eliminating toxic xenobiotics, protection against oxidation and pesticide resistance [63–65]. In *M. occidentalis* and *B. mori*, the Zeta class GSTs not only participated in the oxidative stress response and catalyzed the degradation of tyrosine and phenylalanine, but they also probably played a role in pesticide resistance [63, 66, 67]. The Sigma GSTs of *B. dorsalis*, *D. melanogaster* and *S. litura* play a role in detoxification of toxic compounds and oxidative stress resistance and may also be involved in muscle structure [13, 58, 68]. Moreover, the GSTs have remarkable connections with the secondary compounds in the host plant. The GSTs of *Hyphantria cunea* play important roles in degrading plant secondary metabolites [69]. *Rhynchophorus ferrugineus* larvae fed on diet containing α-asarone, eugenol, coniferyl aldehyde, rosmarinic acid, methyl isoeugenol and methyl eugenol enhanced the expression of GSTs [32, 70]. The *P. utilis* GSTs unigenes in these classes may play a similar role as their counterparts in other insect species.

Carboxylesterases. Carboxylesterases (CarEs) is a multigene family and ubiquitous in animals, plants, insects, and microbes, which play an important role in insecticide resistance, allelochemical metabolism and tolerance, defense, and regulating development [71]. For example, CarEs are involved in metabolism of phenolic glycosides in *Papilio canadensis* [72]. The CarEs of *Anoplophora glabripennis* play important roles in promoting *A. glabripennis* colonization and survival in trees that produce phenolic glycosides in the family Salicaceae [73, 74].
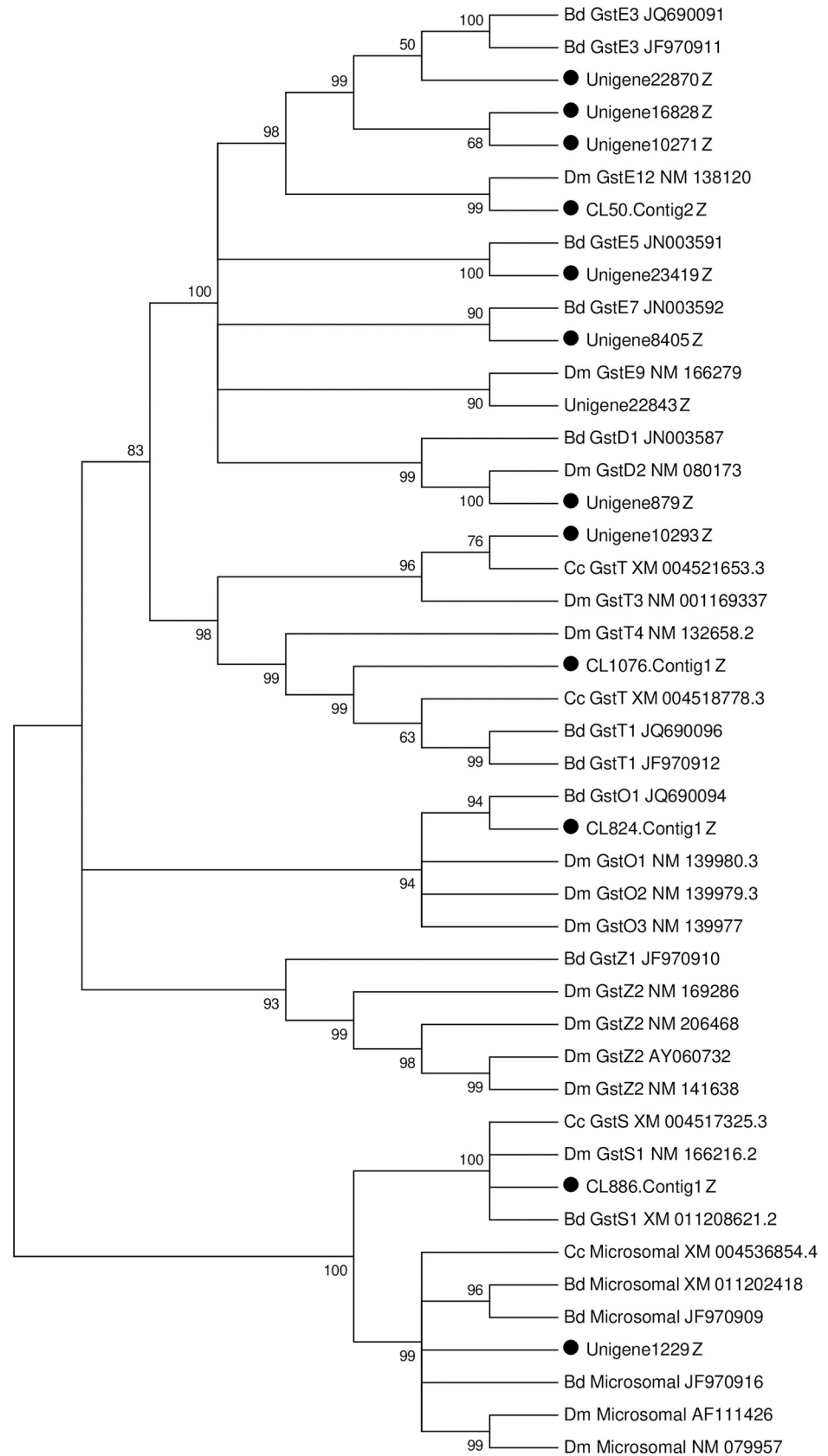
Fig 8. Neighbor-joining phylogenetic tree of glutathione S-transferase (GST) genes from the *P. utilis* fat body (•) and other insects. Numbers at each branch node represent bootstrap values.

https://doi.org/10.1371/journal.pone.0226039.g008

Furthermore, CarEs are also implicated in neurogenesis, hormones and pheromones degradation [75]. Insects CarEs can be divided into 13 clades. These clades fall into three main groups, A-C clades belong to the dietary/detoxification group, D-G clades belong to the hormone/semiochemical processing group, and I-M clades belong to the neuro/developmental group [76]. 35, 24, 44 CarEs have been identified in the genome of *D. melanogaster*, *A. mellifera* and *Metaseiulus occidentalis* respectively [63, 77], while 15 and 38 putative CarEs have been identified in the transcriptome of *B. oleae* and *B. dorsalis* [76, 78]. In this study, 35 putative CarEs have been identified in the fat body transcriptome of *P. utilis* (S4 Table).

Based on phylogenetic analysis with other known insect CarEs, 17 CarEs identified in the fat body transcriptome of *P. utilis* were divided into seven clades: seven in clade C (α-esterases), three in clade H (glutactins and glutactin-like enzymes), two in each M (neurotactins) and I (uncharacterized) clade, one in each K (gliotactin), L (neuroligins), G (lepidopteran JhE). No CarEs in the fat body transcriptome of *P. utilis* belonged to A, B, D, E, F and J clades (Fig 9). Interestingly, 12 of these unigenes coding for CarE were unique to the fat body and were not detected in our previous analysis of the alimentary canal [27]. In this study, most of the CarEs (seven α-esterases) belong to clade C of the dietary detoxification group. This phenomenon was consistent with the CarEs clade in *L. migratoria* and *S. shirakii* [79, 31]. This finding is consistent with that described in *L. migratoria* and *S. shirakii*, a large number of detoxification genes can be used to detoxify many different plant secondary metabolites and to develop insecticide resistance [79].

The ATP-binding cassette (ABC) transporters are membrane-bound proteins belonging to the ABC superfamily located in the cellular membrane in all living organisms from bacteria to humans [80, 81]. ABC transporters have been subdivided into eight subfamilies, from ABC-A to ABC-H [82, 83]. In insects, ABC transporters are not only involved in the molecule transport of molecules, but also play a major role in the metabolism of xenobiotics [76]. From the fat body transcriptome of *P. utilis*, 26 unigenes sequences encoding putative ABC transporters were identified (S5 Table). After removing short sequences, 22 unigenes were subjected to phylogenetic analysis with *B. dorsalis* and the results showed that these unigenes were divided into six subfamilies: five belong to the ABCB subfamily (22.73%), four belong to the ABCC (18.19%), one belong to the ABCD (4.55%), three belong to the ABCF (13.63%), nine belong to the ABCG (40.9%). No ABC transporters belong to the ABCA and ABCH subfamily (Fig 10). Genome-wide analysis of ABC transporters in the *B. dorsalis*, *D. melanogaster*, *A. gambiae*, *T. castaneum*, *A. mellifera*, *B. mori*, *Daphnia pulex*, *Tetranychus urticae*, *Bemisia tabaci* and *Laodelphax striatellus* led to the identification of 47, 56, 52, 73, 41, 55, 65, 103, 55 and 40 ABC transporters, respectively (Table 4) [80, 84, 85]. Furthermore, in the *B. dorsalis* fat body transcriptome, 29 ABC transporters were identified and assigned to 7 subfamilies, including ABCA (4), ABCB (4), ABCC (1), ABCD (2), ABCE (1), ABCF (3), ABCG (4) [13]. Compared with number of ABC transporters in other insect species, there were fewer ABCs in *P. utilis*, especially in the ABCA, ABCD and ABCH subfamilies (Table 4). The missing ABC genes in the current transcriptome database may await further discovery.

ABC transporters have been reported to be involved in the translocation of xenobiotics and phytochemicals. For example, in larvae of *Anopheles stephensi*, the ABC transporters were involved in the detoxification of permethrin [86]. In *B. dorsalis*, *BdABCF1*, *BdABCF3*, *BdABCG1*, *BdABCG11*, *BdABCH1*, and *BdABCH2* genes were upregulated after treatment with three different pesticides; however the expression level of *BdABCB7* was significantly

Fig 9. Neighbor-joining phylogenetic analysis of CarE genes from the *P. utilis* fat body (●) and other insects.
Numbers at each branch node represent bootstrap values.

higher after treatment with malathion than that after treatment with beta-cypermethrin and avermectin, and the toxicity of malathion to *B. dorsalis* increased after *BdABCB7* was silenced [85]. In *B. tabaci*, the ABC transporters responded to imidacloprid, a neonicotinoid insecticide, and most of the significantly up-regulated ABC transporters of *B. tabaci* (80%) involved in the
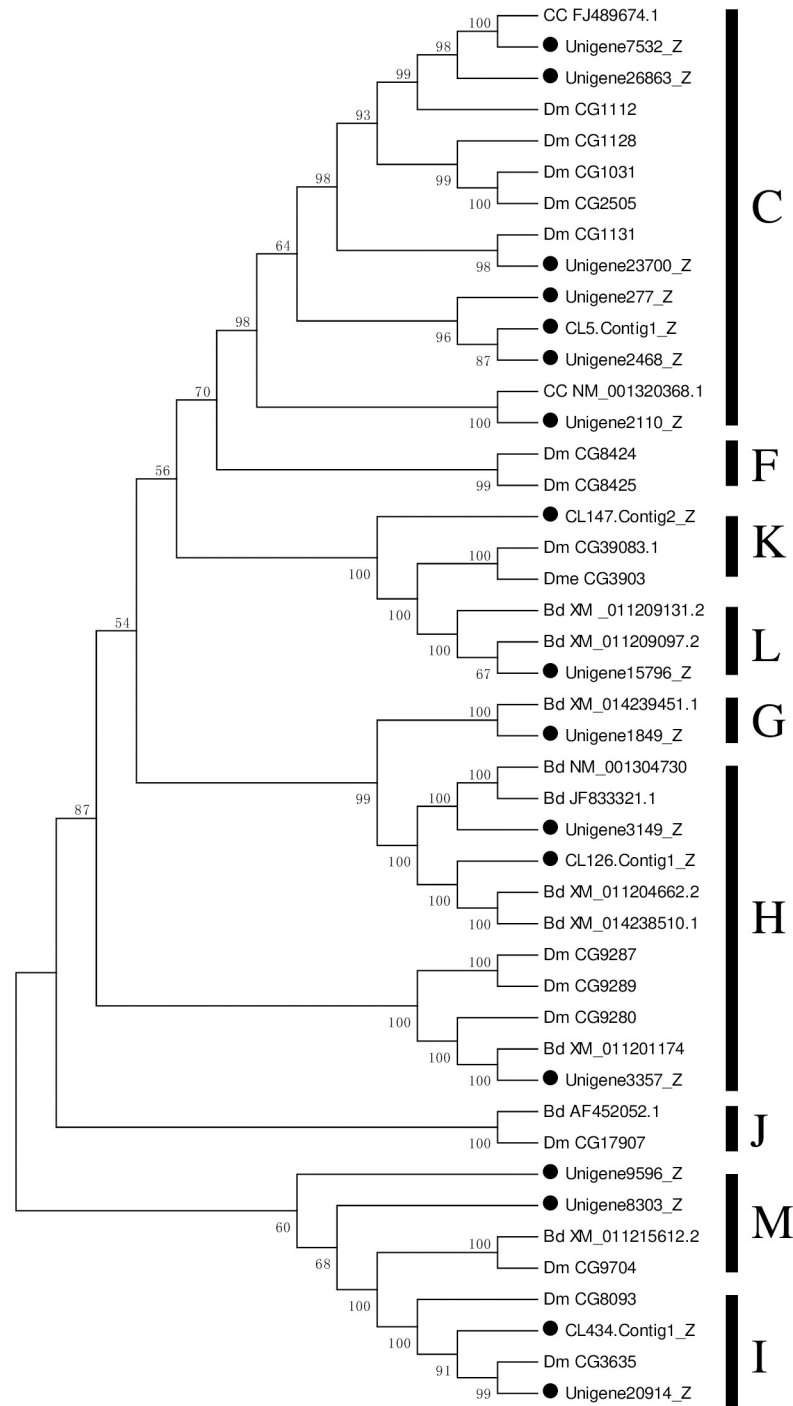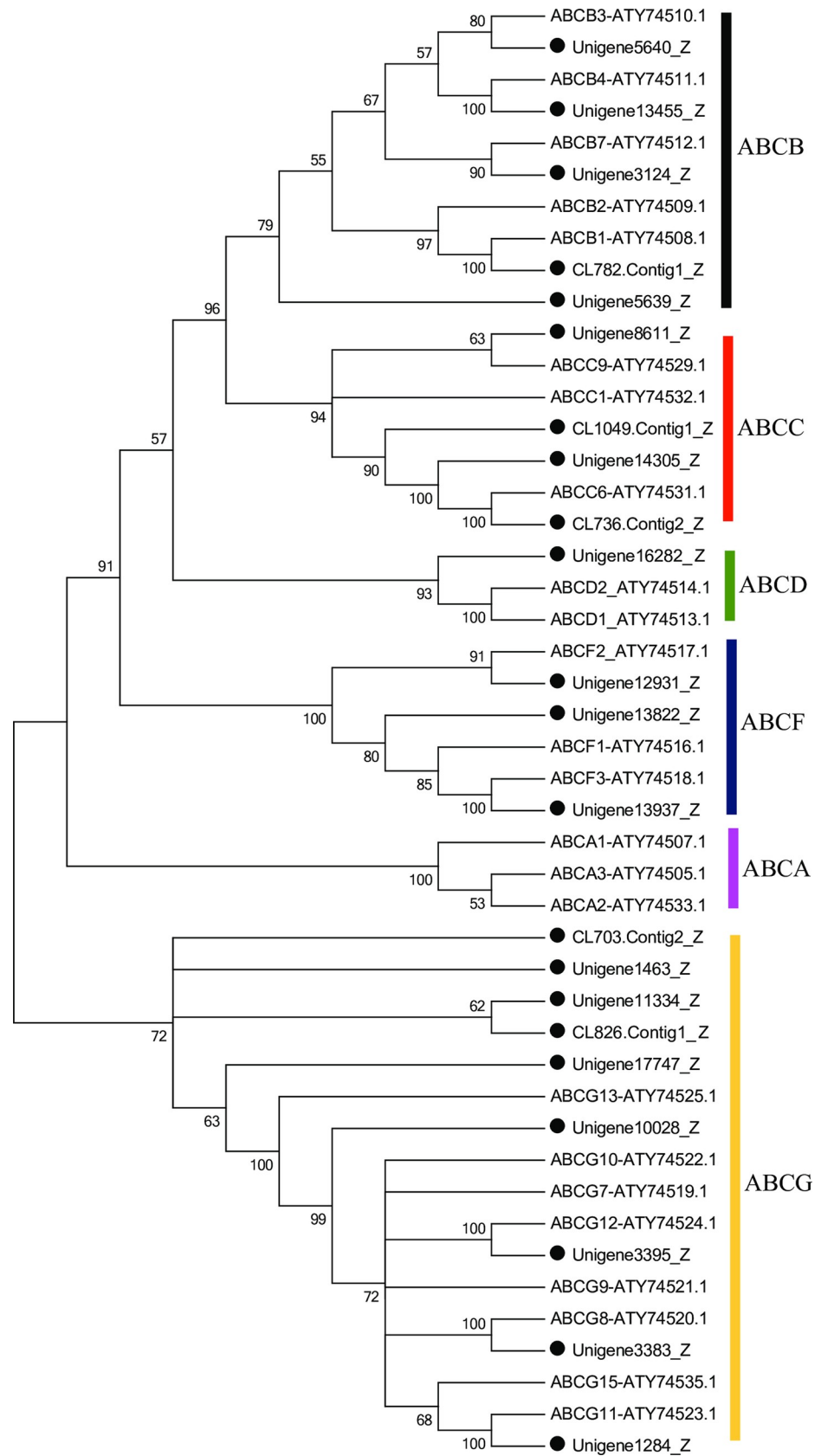
Fig 10. Neighbor-joining phylogenetic analysis of ABC genes from the *P. utilis* fat body (●) and *B. dorsalis*. Numbers at each branch node represent bootstrap values.

detoxification of imidacloprid belong to the ABCG family [87]. In addition, the ABC transporters were regulated by heavy metals and implicated in the biochemical detoxification of zinc and copper in *D. melanogaster* [62]. Likewise, in the cotton bollworm, *Helicoverpa armigera*, the ABC transporters were associated with the degradation of plant secondary metabolites [88]. Previous studies have shown that most of the ABC transporter genes of other insects involved in detoxification belonged to the three families: A, B and C [13, 86]. Here we identified 22 ABC transporters in the fat body of *P. utilis* for the first time and more than half of them also belong to A, B and C families. This was the first analysis of the ABC transporters. It remains to be further studied whether these ABC transporters play roles in the adaptation of *P. utilis* to the toxic host *E. adenophorum*.

Identification of the putative immunity-related genes

The insect fat body acts as an immune barrier between the internal and external environment, which plays a major role in preventing infection by microbial, protozoan pathogens and other foreign objects [12, 13, 89]. In recent studies, a number of genes associated with immune response were identified from the fat body transcriptome of *B. dorsalis*, *G. morsitans morsitans* and *A. aegypti*, including lysozyme, serpin, antimicrobial peptide and other genes [13–15, 90]. In this study, 17 putative serpin, five peptidoglycan recognition proteins (PGRP) and four lysozyme unigenes were identified, which played important roles in insect immune defense.

Serpins are the largest family of serine proteinase inhibitors with members spread over eukaryotes and prokaryotes. At present, many serpin genes have been identified in a number of insect genomes and transcriptomes (S6 Table) [91–96]. The fat body is a main site for the serpin synthesis, indicating that this organ participates in the immune response of insects. For example, in *Manduca sexta*, serpin-9 and -13 genes were mainly produced in fat body [97]. In this study, a total of 17 putative serpin unigenes were identified (S7 Table). This number was less than in the fat body transcriptome of *B. dorsalis* (25) [13]. After removing short sequences, 11 unigenes were used for phylogenetic analysis. The results showed that these unigenes were divided into different clades (Fig 11). Previous studies have illustrated that serpins play vital roles in regulation of innate immune responses via inhibition of serine proteinase cascade

Table 4. Gene numbers in ABC subfamilies of twelve arthropod species.

ABC subfamily	Pu	Bd	Dm	Bo	Ag	Tc	Am	Bm	Dp	Tu	Bt	Ls
A	0	7	10	3	9	10	3	9	4	9	8	2
B	5	7	8	4	5	6	5	5	7	4	3	6
C	4	9	14	2	13	35	9	9	7	39	6	5
D	1	2	2	1	2	2	2	2	2	2	2	2
E	0	1	1	1	1	1	1	1	1	1	1	1
F	3	3	3	3	3	3	3	3	3	3	3	2
G	9	15	15	4	16	13	15	13	24	23	23	14
H	0	3	3	0	3	3	3	3	15	22	9	8
Total	22	47	56	18	52	73	41	55	65	103	55	40

Pu, *P. utilis*; Bd, *B. dorsalis*; Dm, *D. melanogaster*; Ag, *A. gambiae*; Tc, *T. castaneum*; Am, *A. mellifera*; Bm, *B. mori*; Dp, *D. pulex*; Tu, *T. urticae*; Bt, *B. tabaci*; L. *Striatellus*. Numbers of other nine species were derived from Dermauw and Van Leeuwen (2014), Xiao *et al* (2018).

pathways. Serpin-5 and serpin-9 two genes in *H. armigera* controlled melanization by directly inhibiting their target proteases clip-domain serine protease 4 and clip-domain serine protease 6, respectively [98].

Peptidoglycan recognition proteins (PGRPs) form a large family of proteins that recognize peptidoglycan (PGN) in both gram-negative and gram-positive bacteria cell walls and induce the activation of immune responses [99, 100]. The PGRPs are widely distributed and highly conserved from invertebrates to mammals [100]. According to the length of corresponding mRNA, insects PGRPs are divided into two categories: long-type PGRP (PGRP-L) that encodes secreted proteins and short-type PGRP (PGRP-S) that encodes transmembrane or intracellular products, which play distinct roles in innate immune response [101]. In *Drosophila*, PGRP-SA and PGRP-SD can recognize lysine-type peptidoglycan and activate the Toll pathway, while PGRP-LE is essential for activation of the prophenoloxidase (proPO) cascade [102]; DAP-type peptidoglycan is recognized by PGRP-LC and PGRP-LC can interact with PGRP-LE to trigger Imd pathway signaling activation [101]. In *Anopheles*, PGRPLC activates the immune-deficiency (Imd) pathway to resist microbiota load and *Plasmodium* infection [103]. At present, many PGRP genes have been identified from *Drosophila*, *T. castaneum*, *A. gambiae*, *B. dorsalis* and other insects (S6 Table) [13, 104–106]. In this study, five PGRP genes were identified from the fat body transcriptome of *P. utilis*, of which one belongs to the short (S) subfamily (PGRPS1, S2, and S3), while four belong to the long (L) subfamily (PGRPLA, LB, LC, and LD) based on the phylogenetic analysis with genes from other insects (S8 Table and Fig 12). The number of PGRP genes in *P. utilis* was similar to the number identified in the fat body of *B. dorsalis* [13].

The innate insect immune system includes humoral and cellular immunity that can be activated by invasion of pathogens [107]. An array of potent antimicrobial peptides and proteins (AMPs), lysozymes and lectins make up the humoral response. Lysozyme is well-known as an immune effector that is found in both vertebrates and invertebrates, which play protective roles in defense against infections [107, 108]. To date, with genomic and transcriptomic analysis many lysozyme genes have been identified from several insects such as *Galleria mellonella*, *T. castaneum*, *Reticulitermes speratus*, *M. sexta*, *Acyrthosiphon pisum* and *Harmonia axyridis* (S6 Table) [108–113]. Previous reports have shown that insect lysozyme is synthesized in hemocytes and much more so in the fat body, and the lysozyme is released into the hemolymph [107, 114]. Generally insect lysozyme genes are significantly expressed but upregulated in response to microbial challenge in specific tissues, mainly the fat body and hemocytes [107, 115]. Six hours after injection with *E. coli*, *SgLys* transcripts were upregulated 2-fold in the fat body of *Schistocerca gregaria*, indicating that *SgLys* plays a role in *S. gregaria* innate immunity [107]. Low levels of transcripts were observed in the fat body of non-infected larvae of *M. sexta*, while after treatment with peptidoglycan the levels of transcripts increased rapidly and remained elevated for several days [107, 116].

We identified four gene sequences encoding putative lysozymes in the fat body transcriptome of *P. utilis*. Among them, three belong to lysozyme genes and one encodes a lysozyme-like proteins (LLPs) gene according to the phylogenetic analysis with other insects (S9 Table and Fig 13), which is similar to or less than the number found in other insects (S6 Table) [91, 106, 108, 110, 117]. LLPs exhibit antimicrobial activity against both gram-positive and gram-negative bacteria, which probably inhibit bacterial membrane, but do not dissolve the cell wall [118]. Lysozyme not only participates in immunity, but also has a digestive function. In *D. melanogaster* two lysozyme genes were recruited for the digestion of symbiotic bacteria in the stomach [119]. Therefore, whether lysozymes in *P. utilis* are involved in immunization needs further study. This is the first analysis of the lysozymes in the fat body transcriptome of *P. utilis*.
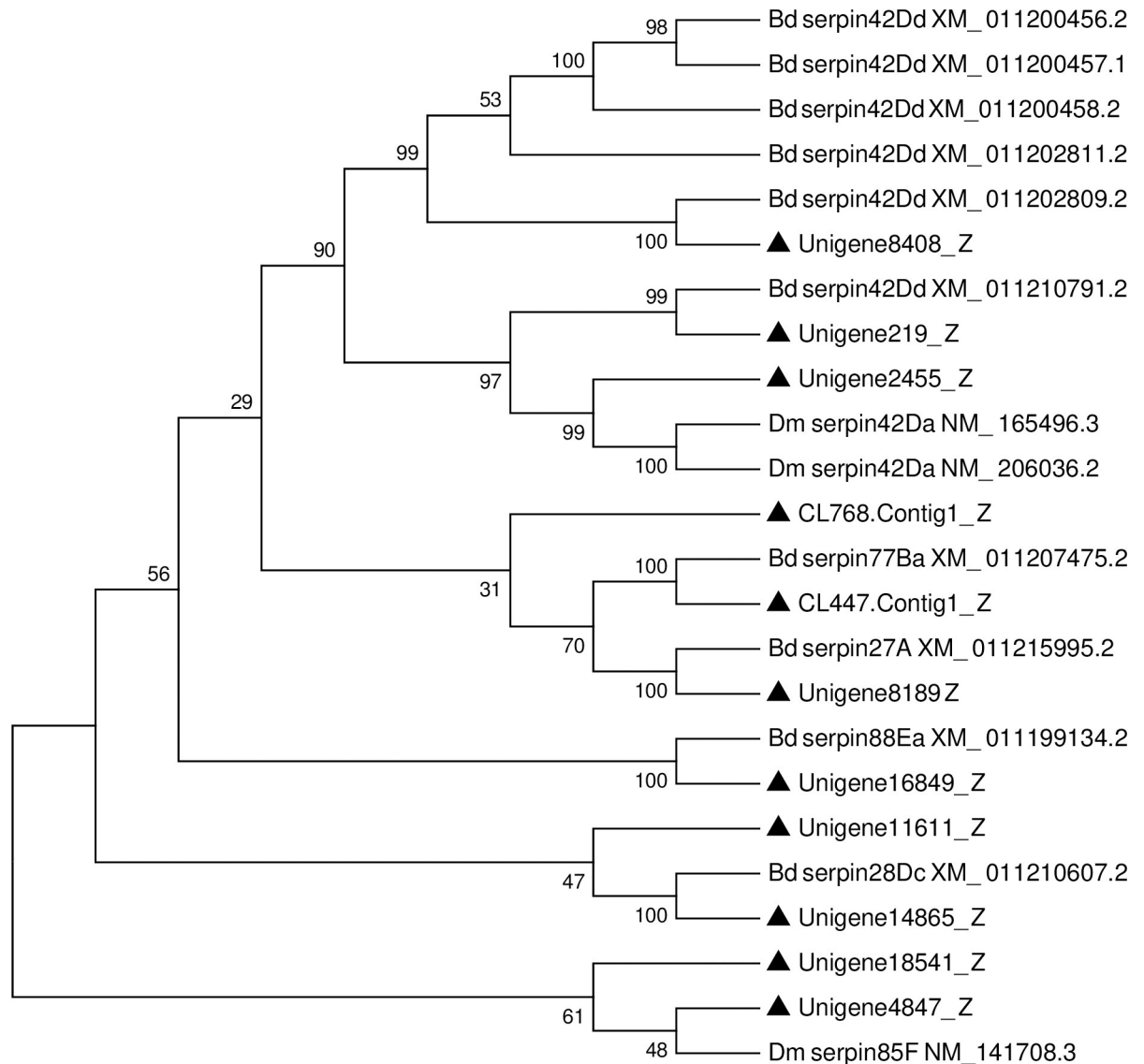
Fig 11. Neighbor-joining phylogenetic analysis of serpin genes from the *P. utilis* fat body (▲) and other insects. Numbers at each branch node represent bootstrap values.

Energy metabolism functions of fat body

The fat body plays major roles in insect metabolism and nutrient storage, which is similar to the liver and fat tissue of vertebrates. The fat body is distributed all over the insect body, but mainly in the abdomen and under the epidermis [10]. Insects store energy in the form of glycogen, triglycerides and amino acids, which form storage proteins in fat body. It is a major storage depot for nutrients stores, and releases energy in response to the energy demands of the insect [10, 13]. In this study, we found many unigenes that are highly similar to those related to energy metabolism in other insects, including 18 lipases, five fatty acid synthase and six elongases of very long chain fatty acid (ELOVL).

Lipids are the main component of the fat body and the major source of metabolic energy [13]. Lipases are a special type of serine hydrolase, which are widely found in vertebrates and
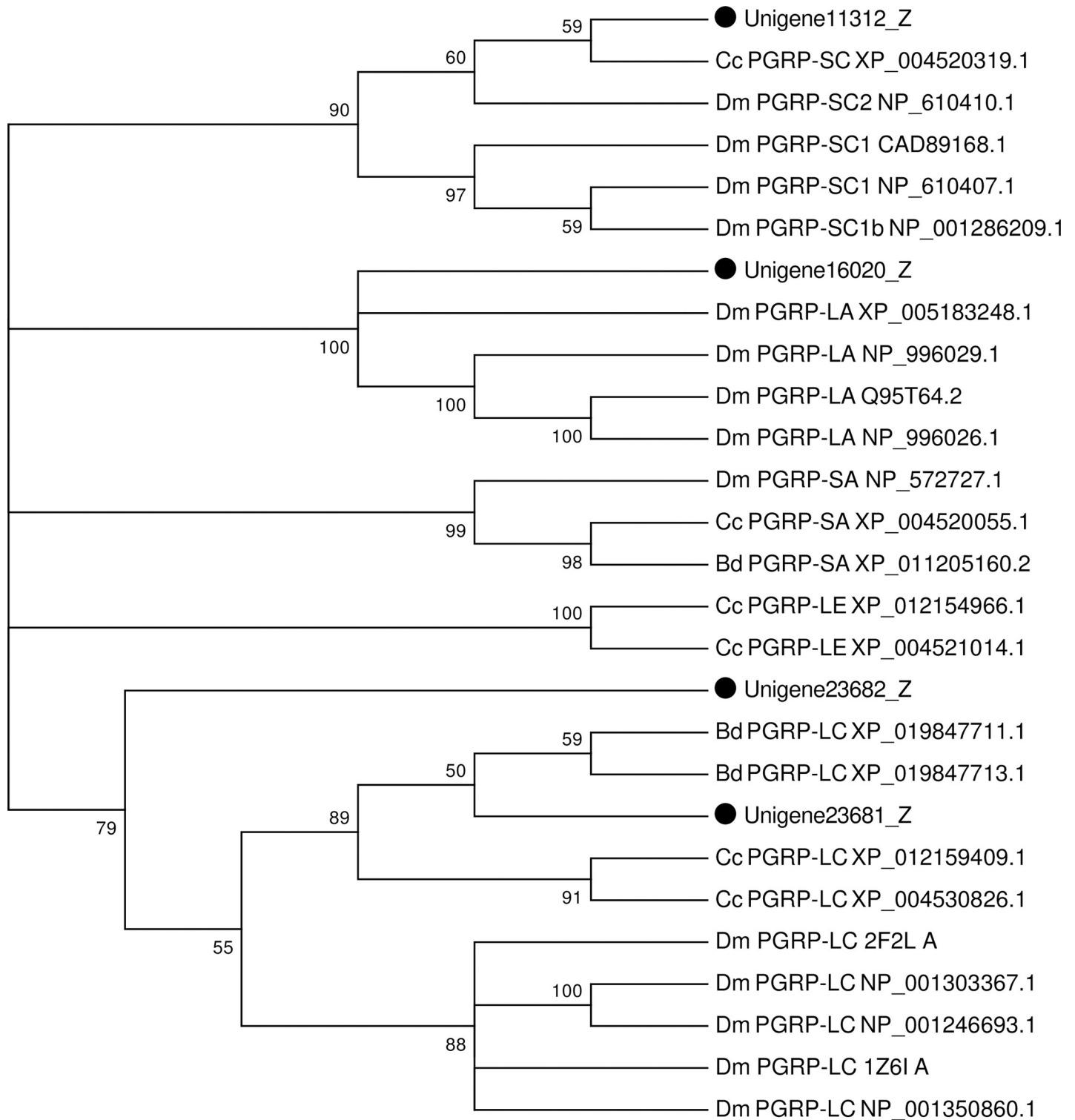
Fig 12. Neighbor-joining phylogenetic analysis of PGRP genes from the *P. utilis* fat body (●) and other insects. Numbers at each branch node represent bootstrap values.
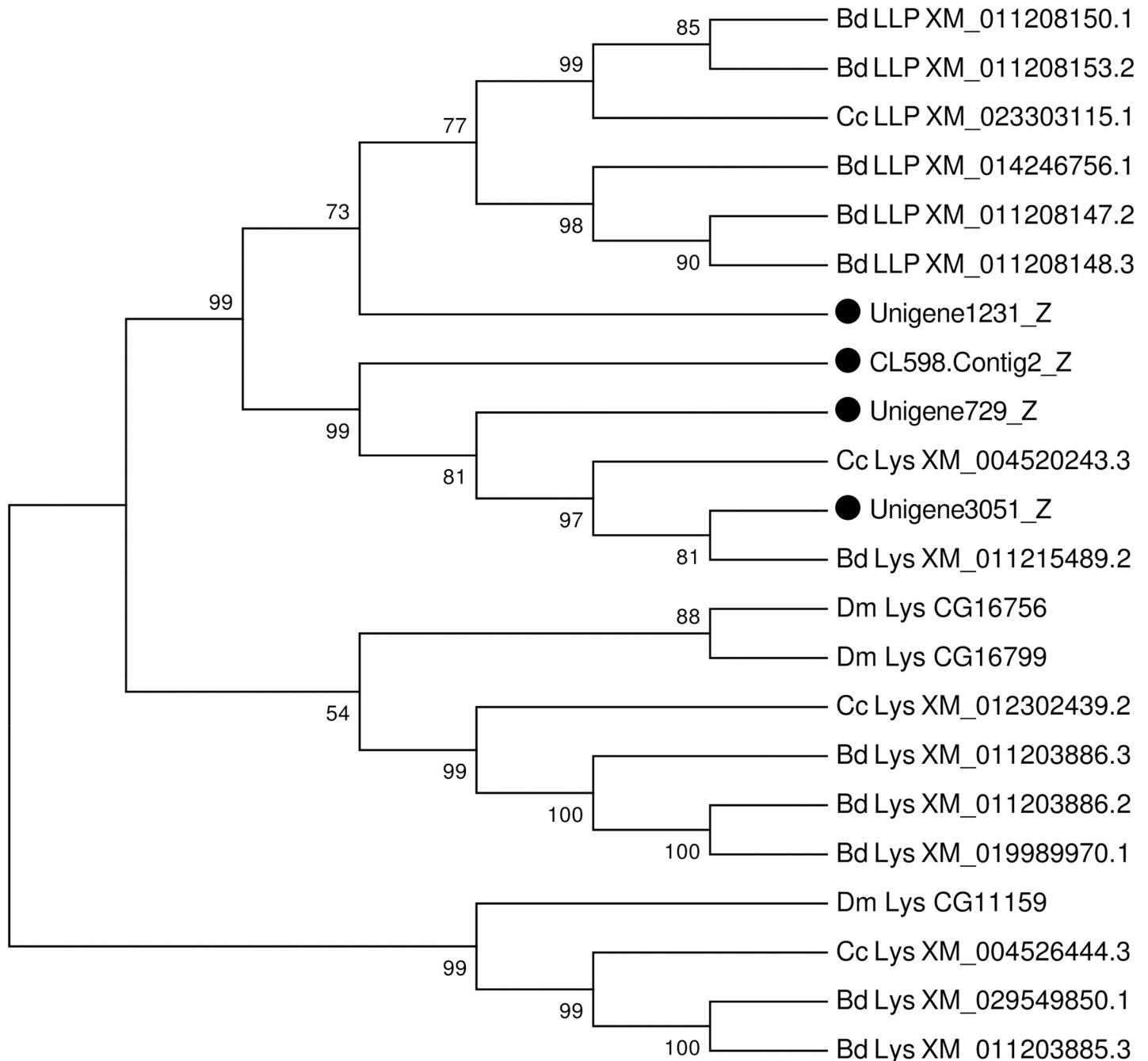
Fig 13. Neighbor-joining phylogenetic analysis of lysozymes and lysozyme-like proteins from the *P. utilis* fat body (●) and other insects. Numbers at each branch node represent bootstrap values.

invertebrates and play a crucial role in fat metabolism [14]. According to sequence relationships within the α/β hydrolase fold superfamily of proteins, lipases are divided into six families: neutral (Pfam: PF00151), acid (Pfam: PF04083), lipase 2 (Pfam: PF01674), lipase 3 (Pfam: PF01764), GDSL (Pfam: PF00657) and hormone sensitive lipases (HSL) [120, 121]. The genes encoding lipases were identified in the genomes of many insects (S6 Table) [121, 122]. According to our research on the *P. utilis* fat body transcriptome, a total of 18 unigenes putatively

encoding lipase were identified (S10 Table), which is less than the number found in other insects [121, 122]. After manually removing short sequences, 12 unigenes were used for the phylogenetic analysis and the result indicated that they were divided into four families, five belonged to the neutral lipase, four belonged to the acid lipase, two belonged to the Brummer and one belonged to the lipase 3. No unigene belonged to lipase 2 and hormone sensitive lipases (Fig 14). The features and functions of lipase vary with families. In insects, neutral lipases have been characterized to play important roles in regulating lipolysis in the fat body. For example, a triacylglycerol lipase with phospholipase A1 activity in the fat body of *M. sexta* plays a crucial role in regulating lipolysis [10, 123]. Brummer lipase, or insect adipose triglyceride lipase (ATGL) has been identified in some insects such as *D. melanogaster*, *B. mori* and *G. morsitans* [121, 123–125]. During starvation, Brummer is crucial for lipolysis in the fat body of *Drosophila* [124]. Furthermore, a mutation of Brummer in *Drosophila* reduced fat body lipid accumulation, nevertheless overexpression of Brummer results in lean flies [123, 124]. The acid lipase family includes gastric lipases, pregastric or lingual lipase, pregastric esterase and lysosomal acid lipases (LAL) [126], which function in the intracellular hydrolysis of cholesteryl esters and triglycerides that are internalized through receptor mediated endocytosis of lipoprotein particles.

Lipids accumulated as fat reserves in the fat body provide energy for insect growth and development. Fatty acid synthase (FAS) plays an important role in lipid synthesis, which is a key enzyme for biosynthesis of endogenous fatty acid in both vertebrates and invertebrates [127]. In *D. melanogaster* genome, three distinct type I FAS genes (*CG3523*, *CG3524* and *CG17374*) were identified [122]. RNA *in situ* hybridization of these three genesin *D. melanogaster* adult males showed that *CG3523* was expressed only in the adult fat body and the other two were expressed in oenocytes [122, 128]. Three FAS I genes (*RPRC00269*, *RPRC002909* and *RPRC000123*) were identified from the *R. prolixus* genome [122]. In this study, a total of five unigenes putatively encoding FAS were identified (S11 Table). Phylogenetic analysis with genes from other insects was carried out, and two unigenes (Unigene17317_Z, Unigene17318_Z) showed homology with FAS1 from *B. dorsalis*, *D. melanogaster* and *A. aegypti*. Unigene12878_Z and Unigene20112_Z were exclusively detected in *Drosophila* oenocytes (*CG17374)* and *B. dorsalis* FAS gene, respectively (Fig 15). It has been reported that biochemical deficiencies in FAS1 inhibit *de novo* lipid biosynthesis in *A. aegypti* blood fed mosquitoes, as well as impact eggshell formation and blood digestion [10]. Moreover, when FAS1 expression was knocked down at mRNA levels inhibition of fatty acid biosynthesis in *A. aegypti* blood fed mosquitoes by RNAi [10]. In *Colaphellus bowringi*, FAS plays a central role in lipid metabolism during diapause, and knockdown of FAS2 reduced lipid accumulation as well as affected stress tolerance genes expression and water content of this species [127].

Both FAS products and fatty acids taken up from the diet are further elongated into very long chain fatty acids (VLCFA) [122]. Elongases of very long chain fatty acid (ELOVL) are enzymes that are essential for the biosynthesis of VLCFAs and elongation of long-chain fatty acids [129]. ELOVLs, which are conserved from microorganisms to mammals, play key roles in various biological processes [130]. In insects, a large number of ELOVL genes have been identified from their genomes although most of their functions remain mostly unknown, including *D. melanogaster*, *B. mori* and *R. prolixus* [122, 129, 131]. In the present study, we obtained a total of six unigenes encoding ELOVL proteins from the *P. utilis* fat body transcriptome (S12 Table). All of these unigenes were used for phylogenetic analysis with other insects. Phylogenetic analysis revealed that they divided into three clades: ELOVL 3/6, ELOVL 1/7 and "ELOVL X clades" (Fig 16). No genes belong to ELOVL 2/5 and ELOVL 4 clade. The "ELOVL 1/7 clade" includes the *P. utilis* genes Unigene3318, this clade possibly plays an important role in elongating saturated/monounsaturated fatty acids. Within these clades, functional

Fig 14. Neighbor-joining phylogenetic analysis of lipases genes from the *P. utilis* fat body (●) and other insects. Numbers at each branch node represent bootstrap values.
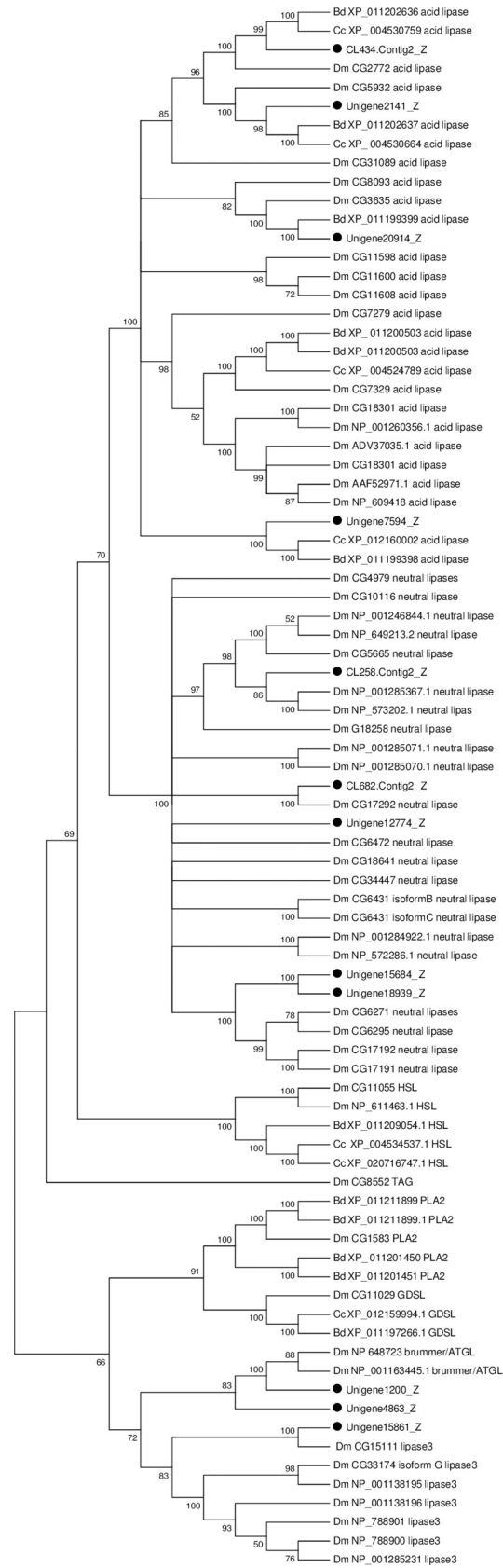
characterization is available for the mosquito *Aedes albopictus* (*ACS37245*), which play a critical role to control the dehydration resistance of diapause eggs by regulating hydrocarbon formation from VLCFA precursors [132]. Four ELOVL unigenes of *P. utilis* (Unigene15322, Unigene25406, Unigene10414 and Unigene1062) belong to the ELOVL 3/6, this clade includes an ELOVL gene of the fruit fly gene *baldspot* (*CG3971*) (Fig 16), which is involved in the sex pheromone production, viability, and sperm development [131]. The remaining unigene (Unigene19540) in the insect-specific ELOVL clusters "ELOVL X clades," which related to both the ELOVL 1/7 and ELOVL 4 clades (Fig 16). This clade includes two ELOVL genes of the fruit fly: elongase F (*CG16905*) and james bond (*CG6921*), which are also involved in sex pheromone production, viability, and sperm development [131].

Conclusions

Here we report the first comprehensive analysis of the fat body transcriptome of *P. utilis* using Illumina sequencing technology. We assembled the raw reads into 30,559 unigenes with an average length of 539 bp. Among these unigenes, 21,439, 13,657, 12,188, 12,405, 12,925, 13,149 and 9,354 unigenes matched to Nr, Nt, SwissProt, KEGG, KOG, InterPro and GO databases, respectively. A number of unigenes that are associated with detoxification (P450s, GSTs, CarEs and ABC transporters), immunity (serpin, PGRPs and lysozymes) and energy metabolism (lipase, FAS and ELOVL) were identified from the annotated unigenes. Specific information of these unigenes identified in this study is shown in S2–S12 Tables. In addition, the fat body-specific transcriptome analysis generated a great deal of unigenes newly identified in *P. utilis*, which provided a major genomic resource for investigating the fat body of *P. utilis* and laid the foundation for future functional genomics studies.
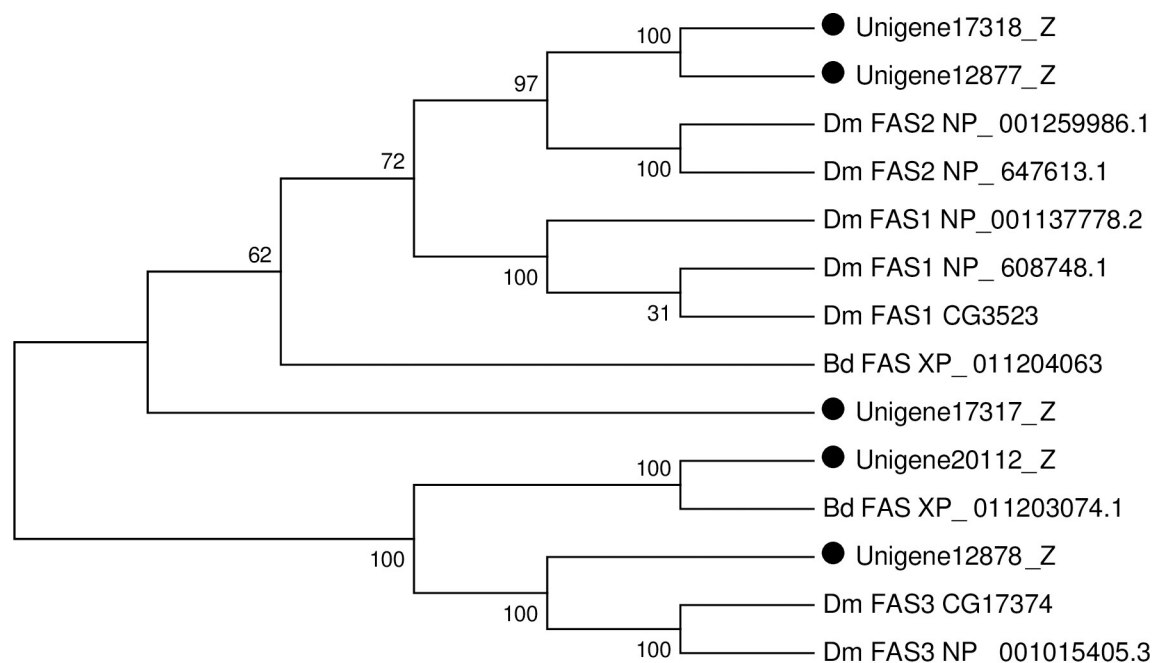
Fig 15. Gene identification of *P.utilis* FASs. Numbers at each branch node represent bootstrap values.
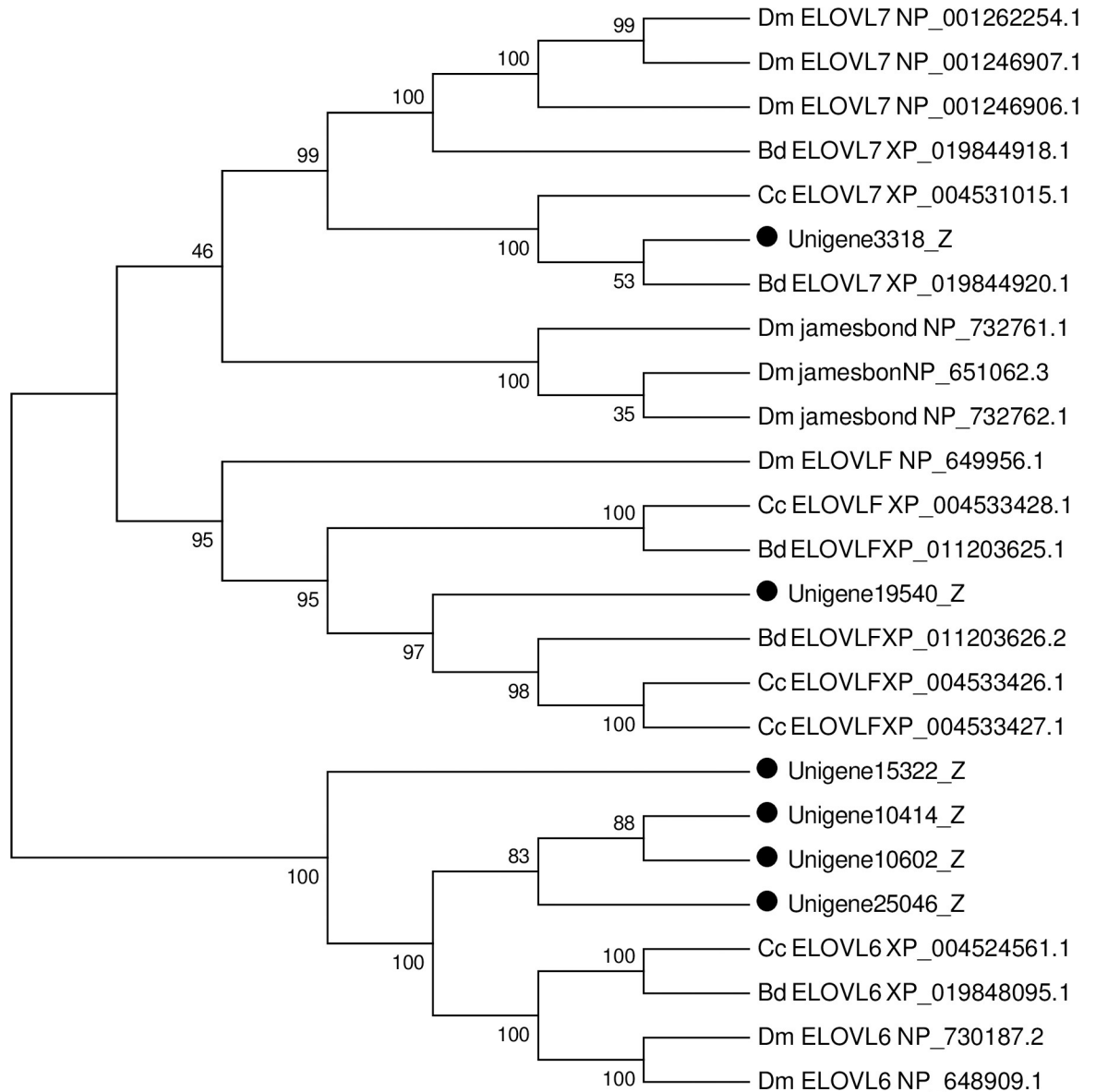
Fig 16. Neighbor-joining phylogenetic analysis of ELOVL genes from the *P. utilis* fat body(●) and other insects. Numbers at each branch node represent bootstrap values.

Supporting information

S1 Table. Distribution of simple sequence repeat (SSR) types found in the *Procecidochares utilis* transcriptome unigenes.
(XLSX)

S2 Table. Unigene sequences for cytochrome oxidase P450s identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S3 Table. Unigene sequences for the glutathione s-transferases identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S4 Table. Unigene sequences for carboxylesterases identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S5 Table. Unigene sequences for ATP-binding cassette (ABC) transporters identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S6 Table. Number of serpins, PGRPs, lysozymes and lipases in other insects.
(XLSX)

S7 Table. Unigene sequences for serpins identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S8 Table. Unigene sequences for peptidoglycan recognition proteins identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S9 Table. Unigene sequences for lysozymes identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S10 Table. Unigene sequences for lipases identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S11 Table. Unigene sequences for fatty acid synthases identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S12 Table. Unigene sequences for elongases of very long chain fatty acids identified in the *Procecidochares utilis* transcriptome.
(XLSX)

S1 Text.
(TXT)

Acknowledgments

Author Contributions

Conceptualization: Lifang Li, Xi Gao, Jiaying Zhu, Guoxing Wu.

Data curation: Lifang Li, Xi Gao, Mingxian Lan, Zijun Guo, Xianbin Liao, Min Ye, Guoxing Wu.

Formal analysis: Lifang Li, Min Ye, Guoxing Wu.

Investigation: Lifang Li, Yuan Yuan, Zijun Guo.

Methodology: Mingxian Lan, Yuan Yuan, Ping Tang.

Project administration: Xi Gao.

Resources: Ping Tang, Mengyue Li, Zhengyue Li.

Software: Lifang Li.

Writing – original draft: Lifang Li.

Writing – review & editing: Lifang Li, Min Ye, Guoxing Wu.

References

1. Sang W, Zhu L, Axmacher JC. Invasion pattern of *Eupatorium adenophorum* Spreng in southern China. Biolo Invasions, 2010; 12(6):1721–1730.

2. Zhou ZX, Jiang H, Yang C, Yang MZ, Zhang HB. Microbial community on healthy and diseased leaves of an invasive plant *Eupatorium adenophorum* in Southwest China. J Microbiol. 2010; 48(2):139–145. https://doi.org/10.1007/s12275-010-9185-y PMID: 20437143.

3. Erasmus DJ, Bennett PH. The effect of galls induced by the gall fly *Procecidochares utilis* on vegetative growth and reproductive potential of crofton weed, *Ageratina andenophora*. Ann Appl Biolo. 2010; 120(1):173–181. https://doi.org/10.1111/j.1744-7348.1992.tb03414.x

4. Wang Y. Predicting the potential geographic distribution of crofton weed (*Ageratina adenophora*) around the world using maxent modeling. Int J Plant Res. 2012; 25(2):324–335.

5. Bess HA and Frank HH. Biological control of Pamakani, *Eupatorium adenophorum*, in Hawaii by a tephritid gall fly, *Procecidochares utilis*. 2. Population studies of the weed, the fly and the parasites of the fly. Ecolo. 1959; 40(2): 244–249. https://doi.org/10.2307/1930034

6. Rahman O, Agarwal ML. Biological control of crofton weed (*Eupatorium adenophorum* Sprengel) by a fruit fly *Procecidochares utilis* Stone in eastern Himalayas. Indian J Weed Sci. 1990; 22:98–101.

7. Buccellato L, Byrne MJ, Witkowski ETF. Interactions between a stem gall fly and a leaf-spot pathogen in the biological control of *Ageratina adenophora*. Biolo Control. 2012; 61(3):222–229. https://doi.org/10.1016/j.biocontrol.2012.02.004

8. Gao X, Zhu JY, Ma S, Zhang Z, Xiao C, Li Q, et al. Transcriptome profiling of the crofton weed gall fly *Procecidochares utilis*. Genet Mol Res. 2014; 13(2):2857–2864. https://doi.org/10.4238/2014.March.19.1 PMID: 24682983.

9. Li AF, Gao XM, Dang WG, Huang RX, Deng ZP, Tang HC. Parasitism of *Procecidochares utilis* and its effect on growth and reproduction of *Eupatorium adenophorum*. J Plant Ecol. 2006; 30(3): 496–503.

10. Arrese EL, Soulages JL. Insect fat body: energy, metabolism, and regulation. Annu Rev Entomol. 2010; 55(55):207–225. https://doi.org/10.1146/annurev-ento-112408-085356 PMID: 19725772.

11. Yu H, Ji R, Ye W, Chen H, Lai W, Fu Q, Lou Y. Transcriptome analysis of fat bodies from two brown planthopper (*Nilaparvata lugens*) populations with different virulence levels in rice. PLoS ONE. 2014; 9(2):e88528. https://doi.org/10.1371/journal.pone.0088528 PMID: 24533099.

12. Dunn PE. Biochemical aspects of insect immunology. Annu Rev Entomol. 1986; 31(31):321–339. https://doi.org/10.1146/annurev.en.31.010186.001541

13. Yang WJ, Yuan GR, Cong L, Xie YF, Wang JJ. De novo cloning and annotation of genes associated with immunity, detoxification and energy metabolism from the fat body of the oriental fruit fly, *Bactrocera dorsalis*. PLoS ONE. 2014; 9(4):e94470. https://doi.org/10.1371/journal.pone.0094470 PMID: 24710118.

14. Attardo GM, Strickler-Dinglasan P, Perkin SA, Caler E, Bonaldo MF, Soares MB, et al. Analysis of fat body transcriptome from the adult tsetse fly, *Glossina morsitans morsitans*. Insect Mol Biol. 2006; 15(4):411–24. https://doi.org/10.1111/j.1365-2583.2006.00649.x PMID: 16907828.

15. Price DP, Nagarajan V, Churbanov A, Houde P, Milligan B, Drake LL, et al. The fat body transcriptomes of the yellow fever mosquito *Aedes aegypti*, pre- and post- blood meal. PLoS ONE. 2011; 6(7): e22573. https://doi.org/10.1371/journal.pone.0022573 PMID: 21818341.

16. Snyder MJ, Stevens JL, Andersen JF, Feyereisen R. Expression of cytochrome P450 genes of the CYP4 family in midgut and fat body of the tobacco hornworm, *Manduca sexta*. Arch Biochem Biophys. 1995; 321(1):13–20. https://doi.org/10.1006/abbi.1995.1362 PMID: 7639512.

17. Yang J, McCart C, Woods DJ, Terhzaz S, Greenwood KG, ffrench-Constant RH, et al. A *Drosophila* systems approach to xenobiotic metabolism. Physiol Genomics. 2007; 30(3):223–231. https://doi.org/10.1152/physiolgenomics.00018.2007 PMID: 17488889.

18. Jiang Z, Wu XL, Michal JJ, McNamara JP. Pattern profiling and mapping of the fat body transcriptome in *Drosophila melanogaster*. Obes Res. 2005; 13(11):1898–904. https://doi.org/10.1038/oby.2005.233 PMID: 16339120.

19. Feitosa FM, Calvo E, Merino EF, Durham AM, James AA, de Bianchi AG, et al. A transcriptome analysis of the *Aedes aegypti* vitellogenic fat body. J Insect Sci. 2006; 6(6):1–26. https://doi.org/10.1673/1536-2442(2006)6[1:ATAOTA]2.0.CO;2 PMID: 19537968.

20. Nanoth Vellichirammal N, Zera AJ, Schilder RJ, Wehrkamp C, Riethoven JJ, Brisson JA. De novo transcriptome assembly from fat body and flight muscles transcripts to identify morph-specific gene expression profiles in *Gryllus firmus*. PLoS ONE. 2014; 9(1):e82129. https://doi.org/10.1371/journal.pone.0082129 PMID: 24416137.

21. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full–length transcriptome assembly from RNA–Seq data without a reference genome. Nat Biotechnol. 2011; 29(7):644–652. https://doi.org/10.1038/nbt.1883 PMID: 21572440.

22. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, et al. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics. 2003; 19(5):651–652. https://doi.org/10.1093/bioinformatics/btg034 PMID: 12651724.

23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool (BLAST). J Mol Biol. 1990; 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712.

24. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21(18):3674–3676. https://doi.org/10.1093/bioinformatics/bti610 PMID: 16081474.

25. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifer. Nucleic Acids Res. 2005; 33:116–20. https://doi.org/10.1093/nar/gki442 PMID: 15980438.

26. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 2011; 28(10):2731–2739. https://doi.org/10.1093/molbev/msr121 PMID: 21546353.

27. Li LF, Lan MX, Lu WF, Li ZY, Xia T, Zhu JY, et al. De novo transcriptomic analysis of the alimentary tract of the tephritid gall fly, *Procecidochares utilis*. PLoS ONE. 2018; 13(8):e0201679. https://doi.org/10.1371/journal.pone.0201679 PMID: 30138350.

28. Feyereisen R. Insect P450 enzymes. Annu Rev Entomol. 1999; 44(44):507–533. https://doi.org/10.1146/annurev.ento.44.1.507 PMID: 9990722.

29. Zhou D, Liu X, Sun Y, Ma L, Shen B, Zhu C. Genomic analysis of detoxification supergene families in the mosquito *Anopheles sinensis*. PLoS ONE. 2015; 10(11):e0143387. https://doi.org/10.1371/journal.pone.0143387 PMID: 26588704.

30. Guo Y, Chai Y, Zhang L, Zhao Z, Gao LL, Ma R. Transcriptome analysis and identification of major detoxification gene families and insecticide targets in *Grapholita Molesta* (Busck) (Lepidoptera: Tortricidae). J Insect Sci. 2017; 17(2):43. https://doi.org/10.1093/jisesa/iex014 PMID: 28365764.

31. Qiu Z, Liu F, Lu H, Yuan H, Zhang Q, Huang Y. De Novo assembly and characterization of the transcriptome of *Grasshopper Shirakiacris shirakii*. Int J Mol Sci. 2016; 17(7):1110. https://doi.org/10.3390/ijms17071110 PMID: 27455245.

32. Aljabr AM, Hussain A, Rizwan-Ul-Haq M, Al-Ayedh H. Toxicity of plant secondary metabolites modulating detoxification genes expression for natural red palm weevil pesticide development. Molecules. 2017; 22(1):169–181. https://doi.org/10.3390/molecules22010169 PMID: 28117698.

33. Huang X, Ma J, Qin X, Tu X, Cao G, Wang G, et al. Biology, physiology and gene expression of grasshopper *Oedaleus asiaticus* exposed to diet stress from plant secondary compounds. Sci Rep. 2017; 7(1):8655–8664. https://doi.org/10.1038/s41598-017-09277-z PMID: 28819233.

34. Qin Q, Li Y, Zhong D, Zhou N, Chang X, Li C, et al. Insecticide resistance of *Anopheles sinensis* and *An. vagus* in Hainan Island, a malaria-endemic area of China. Parasit Vectors. 2014; 7:92. https://doi.org/10.1186/1756-3305-7-92 PMID: 24589247.

35. Wang RL, Staehelin C, Xia QQ, Su YJ, Zeng RS. Identification and characterization of *CYP9A40* from the tobacco cutworm moth (*Spodoptera litura*), a cytochrome P450 gene induced by plant allelochemicals and insecticides. Int J Mol Sci. 2015; 16(9):22606–20. https://doi.org/10.3390/ijms160922606 PMID: 26393579.

36. Rupasinghe SG, Wen Z, Chiu TL, Schuler MA. *Helicoverpa zea CYP6B8* and *CYP321A1*: different molecular solutions to the problem of metabolizing plant toxins and insecticides. Protein Eng Des Sel. 2007; 20(12):615–624. https://doi.org/10.1093/protein/gzm063 PMID: 18065401.

37. Yu L, Tang W, He W, Ma X, Vasseur L, Baxter SW, et al. Characterization and expression of the cytochrome P450 gene family in diamondback moth, *Plutella xylostella* (L.). Sci Rep. 2015; 5:8952. https://doi.org/10.1038/srep08952 PMID: 25752830.

38. Guo Y, Zhang X, Wu H, Yu R, Zhang J, Zhu KY, et al. Identification and functional analysis of a cytochrome P450 gene *CYP9AQ2* involved in deltamethrin detoxification from *Locusta migratoria*. Pestic Biochem Physiol. 2015; 122:1–7. https://doi.org/10.1016/j.pestbp.2015.01.003 PMID: 26071800.

39. Guo Y, Wu H, Zhang X, Ma E, Guo Y, Zhu KY, Zhang J. RNA interference of cytochrome P450 CYP6F subfamily genes affects susceptibility to different insecticides in *Locusta migratoria*. Pest Manag Sci. 2016; 72(11):2154–2165. https://doi.org/10.1002/ps.4248 PMID: 26853074.

40. Arouri R, Le Goff G, Hemden H, Navarro-Llopis V, M'saad M, Castañera P, et al. Resistance to lambda-cyhalothrin in Spanish field populations of *Ceratitis capitata* and metabolic resistance mediated by P450 in a resistant strain. Pest Manag Sci. 2015; 71(9):1281–1291. https://doi.org/10.1002/ps.3924 PMID: 25296621.

41. Gellatly KJ, Yoon KS, Doherty JJ, Sun W, Pittendrigh BR, Clark JM. RNAi validation of resistance genes and their interactions in the highly DDT-resistant 91-R strain of *Drosophila melanogaster*. Pestic Biochem Physiol. 2015; 121:107–115. https://doi.org/10.1016/j.pestbp.2015.01.001 PMID: 26047118.

42. Bautista MA, Miyata T, Miura K, Tanaka T. RNA interference-mediated knockdown of a cytochrome P450, *CYP6BG1*, from the diamondback moth, *Plutella xylostella*, reduces larval resistance to permethrin. 2009; 39(1):38–46. https://doi.org/10.1016/j.ibmb.2008.09.005 PMID: 18957322.

43. Pavlidi N, Kampouraki A, Tseliou V, Wybouw N, Dermauw W, Roditakis E, et al. Molecular characterization of pyrethroid resistance in the olive fruit fly *Bactrocera oleae*. Pestic Biochem Physiol. 2018; 148:1–7. https://doi.org/10.1016/j.pestbp.2018.03.011 PMID: 29891359.

44. Hussain A, Rizwan-Ul-Haq M, Al-Ayedh H, Aljabr AM. Toxicity and Detoxification Mechanism of Black Pepper and Its Major Constituent in Controlling *Rhynchophorus ferrugineus* Olivier (Curculionidae: Coleoptera). Neotrop Entomol. 2017; 46(6):685–693. https://doi.org/10.1007/s13744-017-0501-7 PMID: 28326461.

45. Mao W, Rupasinghe SG, Johnson RM, Zangerl AR, Schuler MA, Berenbaum MR. Quercetin-metabolizing CYP6AS enzymes of the pollinator *Apis mellifera* (Hymenoptera: Apidae). Comp Biochem Physiol B Biochem Mol Biol. 2009; 154(4):427–434. https://doi.org/10.1016/j.cbpb.2009.08.008 PMID: 19737624.

46. Mao W, Schuler MA, Berenbaum MR. Disruption of quercetin metabolism by fungicide affects energy production in honey bees (*Apis mellifera*). Proc Natl Acad Sci USA. 2017; 114(10):2538–2543. https://doi.org/10.1073/pnas.1614864114 PMID: 28193870.

47. Mao W, Schuler MA, Berenbaum MR. CYP9Q-mediated detoxification of acaricides in the honey bee (*Apis mellifera*). Proc Natl Acad Sci USA. 2011; 108(31):12657–62. https://doi.org/10.1073/pnas.1109535108 PMID: 21775671.

48. Wang RL, He YN, Staehelin C, Liu SW, Su YJ, Zhang JE. Identification of Two Cytochrome Monooxygenas P450 Genes, *CYP321A7* and *CYP321A9*, from the Tobacco Cutworm Moth (*Spodoptera Litura*) and Their Expression in Response to Plant Allelochemicals. Int J Mol Sci. 2017; 18(11): pii: E2278; https://doi.org/10.3390/ijms18112278 PMID: 29084173.

49. Jensen HR, Scott IM, Sims S, Trudeau VL, Arnason JT. Gene expression profiles of *Drosophila melanogaster* exposed to an insecticidal extract of Piper nigrum. J Agric Food Chem. 2006; 54(4):1289–1295. https://doi.org/10.1021/jf052046n PMID: 16478250.

50. Bhaskara S, Dean ED, Lam V, Ganguly R. Induction of two cytochrome P450 genes, *Cyp6a2* and *Cyp6a8*, of *Drosophila melanogaster* by caffeine in adult flies and in cell culture. Gene. 2006; 377(1):56–64. https://doi.org/10.1016/j.gene.2006.02.032 PMID: 16713132.

51. Bass C, Hebsgaard M B, Hughes J. Genomic resources for the brown planthopper, *Nilaparvata lugens*: Transcriptome pyrosequencing and microarray design. Insect Sci. 2012; 19(1):1–12. https://doi.org/10.1111/j.1744-7917.2011.01440.x

52. Sparks ME, Rhoades JH, Nelson DR, Kuhar D, Lancaster J, Lehner B, et al. A transcriptome survey spanning life stages and sexes of the harlequin bug, *Murgantia histrionica*. Insects. 2017; 8(2): 55. https://doi.org/10.3390/insects8020055 PMID: 28587099.

53. Friedman R. Genomic organization of the glutathione S-transferase family in insects. Mol Phylogenet Evol. 2011; 61(3):924–32. https://doi.org/10.1016/j.ympev.2011.08.027 PMID: 21930223.

54. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, et al. Evolution of supergene families associated with insecticide resistance. Sci. 2002; 298(5591):179–81. https://doi.org/10.1126/science.1076781 PMID: 12364796.

55. Zhang J, Zhang Y, Li J, Liu M, Liu Z. Midgut transcriptome of the cockroach *Periplaneta americana* and its microbiota: digestion, detoxification and oxidative stress response. PLoS ONE. 2016; 11(5): e0155254. https://doi.org/10.1371/journal.pone.0155254 PMID: 27153200.

56. Shen GM, Dou W, Niu JZ, Jiang HB, Yang WJ, Jia FX, et al. Transcriptome analysis of the oriental fruit fly (*Bactrocera dorsalis*). PLoS ONE. 2011; 6(12):e29127. https://doi.org/10.1371/journal.pone.0029127 PMID: 22195006.

57. Wang J, Xiong KC, Liu YH. *De novo* Transcriptome analysis of chinese citrus fly, *Bactrocera minax* (Diptera: Tephritidae), by high-throughput illumina sequencing. PLoS ONE. 2016; 11(6):e0157656. https://doi.org/10.1371/journal.pone.0157656 PMID: 27331903.

58. Singh SP, Coronella JA, Benes H, Cochrane BJ, Zimniak P. Catalytic function of *Drosophila melanogaster* glutathione S-transferase *DmGSTS1-1* (GST-2) in conjugation of lipid peroxidation end products. Eur J Biochem. 2001; 268(10):2912–23. https://doi.org/10.1046/j.1432-1327.2001.02179.x PMID: 11358508.

59. Lumjuan N, Rajatileka S, Changsom D, Wicheer J, Leelapat P, Prapanthadara LA, et al. The role of the *Aedes aegypti* Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. Insect Biochem Mol Biol. 2011; 41(3):203–9. https://doi.org/10.1016/j.ibmb.2010.12.005 PMID: 21195177.

60. Lu XP, Wang LL, Huang Y, Dou W, Chen CT, Wei D, et al. The epsilon glutathione S -transferases contribute to the malathion resistance in the oriental fruit fly, *Bactrocera dorsalis* (Hendel). Comp Biochem Physiol C Toxicol Pharmacol. 2016; 180:40–48. https://doi.org/10.1016/j.cbpc.2015.11.001 PMID: 26610787.

61. Qin G, Liu T, Guo Y, Zhang X, Ma E, Zhang J. Effects of chlorpyrifos on glutathione S -transferase in migratory locust, *Locusta migratoria*. Pestic Biochem Physiol. 2014; 109(1):1–5. https://doi.org/10.1016/j.pestbp.2013.12.008 PMID: 24581378.

62. Yepiskoposyan H, Egli D, Fergestad T, Selvaraj A, Treiber C, Multhaup G, et al. Transcriptome response to heavy metal stress in *Drosophila* reveals a new zinc transporter that confers resistance to zinc. Nucleic Acids Res. 2006; 34(17):4866–77. https://doi.org/10.1093/nar/gkl606 PMID: 16973896.

63. Wu K, Hoy MA. The Glutathione-S-Transferase, Cytochrome P450 and carboxyl/cholinesterase gene superfamilies in predatory mite *Metaseiulus occidentalis*. PLoS One. 2016; 11(7):e0160009. https://doi.org/10.1371/journal.pone.0160009 PMID: 27467523.

64. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. Annu Rev Pharmacol Toxicol. 2005; 45:51–88. https://doi.org/10.1146/annurev.pharmtox.45.120403.095857 PMID: 15822171.

65. Pavlidi N, Tseliou V, Riga M, Nauen R, Van Leeuwen T, Labrou NE, et al. Functional characterization of glutathione S-transferases associated with insecticide resistance in *Tetranychus urticae*. Pestic Biochem Physiol. 2015; 121:53–60. https://doi.org/10.1016/j.pestbp.2015.01.009 PMID: 26047112.

66. Board PG, Baker RT, Chelvanayagam G, Jermiin LS. Zeta, a novel class of glutathione transferases in a range of species from plants to humans. Biochem J. 1997; 328 (Pt 3):929–35. https://doi.org/10.1042/bj3280929 PMID: 9396740.

67. Yamamoto K, Shigeoka Y, Aso Y, Banno Y, Kimura M, Nakashima T. Molecular and biochemical characterization of a Zeta-class glutathione S-transferase of the silkmoth. Pestic Biochem Physiol. 2009; 94(1):30–35.

68. Huang Y, Xu Z, Lin X, Feng Q, Zheng S. Structure and expression of glutathione S-transferase genes from the midgut of the common cutworm, *Spodoptera litura* (Noctuidae) and their response to xenobiotic compounds and bacteria. J Insect Physiol. 2011; 57(7):1033–44. https://doi.org/10.1016/j.jinsphys.2011.05.001 PMID: 21605564.

69. Pan L, Ren L, Chen F, Feng Y, Luo Y. Antifeedant Activity of Ginkgo biloba Secondary Metabolites against *Hyphantria cunea* Larvae: Mechanisms and Applications. PLoS ONE. 2016; 11(5):e0155682. https://doi.org/10.1371/journal.pone.0155682 PMID: 27214257.

70. Al-Ayedh H, Hussain A, Rizwan-Ul-Haq M, Al-Jabr AM. Status of insecticide resistance in field-collected populations of *Rhynchophorus ferrugineus* (Olivier) (Coleoptera: Curculionidae). Int J Agric Biol. 2016; 18(1):103–110. https://doi.org/10.17957/IJAB/15.0070

71. Yu QY, Lu C, Li WL, Xiang ZH, Zhang Z. Annotation and expression of carboxylesterases in the silkworm, *Bombyx mori*. BMC Genomics. 2009; 10(1):553. https://doi.org/10.1186/1471-2164-10-553 PMID: 19930670.

72. Lindroth RL. Biochemical detoxication: mechanism of differential tiger swallowtail tolerance to phenolic glycosides. Oecologia. 1989; 81(2):219–224. https://doi.org/10.1007/BF00379809 PMID: 28312541.

73. Scully ED, Hoover K, Carlson JE, Tien M, Geib SM. Midgut transcriptome profiling of *Anoplophora glabripennis*, a lignocellulose degrading cerambycid beetle. BMC Genomics. 2013; 14:850. https://doi.org/10.1186/1471-2164-14-850 PMID: 24304644.

74. Boeckler GA, Gershenzon J, Unsicker SB. Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. Phytochemistry. 2011; 72(13):1497–509. https://doi.org/10.1016/j.phytochem.2011.01.038 PMID: 21376356.

75. Li X, Schuler MA, Berenbaum MR. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. Annu Rev Entomol. 2007; 52:231–53. https://doi.org/10.1146/annurev.ento.51.110104.151104 PMID: 16925478.

76. Pavlidi N, Dermauw W, Rombauts S, Chrysargyris A, Chrisargiris A, Van Leeuwen T, et al. Analysis of the olive fruit fly *Bactrocera oleae* transcriptome and phylogenetic classification of the major detoxification gene families. PLoS ONE. 2015; 10(5):e0128056. https://doi.org/10.1371/journal.pone.0128056 PMID: 25955294.

77. Oakeshott JG, Claudianos C, Campbell PM, Newcomb RD, Russel RJ. Biochemical genetics and genomics of insect esterases. Compreh Mol Insect Science-Pharmacology Vol. 5 (editors Gilbert LI, Iatrou K and Gill SS) 309–381 Elsevier Oxford.

78. Hsu JC, Chien TY, Hu CC, Chen MJ, Wu WJ, Feng HT, et al. Discovery of genes related to insecticide resistance in *Bactrocera dorsalis* by functional genomic analysis of a de novo assembled transcriptome. PLoS ONE. 2012; 7(8):e40950. https://doi.org/10.1371/journal.pone.0040950 PMID: 22879883.

79. Zhang J, Li D, Ge P, Guo Y, Zhu KY, Ma E, et al. Molecular and functional characterization of cDNAs putatively encoding carboxylesterases from the migratory locust, *Locusta migratoria*. PLoS ONE 2014; 9(4):e94809. https://doi.org/10.1371/journal.pone.0094809 PMID: 24722667.

80. Sun H, Pu J, Chen F, Wang J, Han Z. Multiple ATP-binding cassette transporters are involved in insecticide resistance in the small brown planthopper, *Laodelphax striatellus*. Insect Mol Biol. 2017; 26 (3):343–355. https://doi.org/10.1111/imb.12299 PMID: 28299835.

81. Higgins CF, Linton KJ. The ATP switch model for ABC transporters. Nat Struct Mol Biol. 2004; 11 (10):918–26. https://doi.org/10.1038/nsmb836 PMID: 15452563.

82. Hollenstein K, Dawson RJ, Locher KP. Structure and mechanism of ABC transporter proteins. Curr Opin Struct Biol. 2007; 17(4):412–418. https://doi.org/10.1016/j.sbi.2007.07.003 PMID: 17723295.

83. Dean M, Annilo T. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet. 2005; 6(1):123–142. https://doi.org/10.1146/annurev.genom.6.080604.162122 PMID: 16124856.

84. Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol. 2014; 45:89–110. https://doi.org/10.1016/j.ibmb.2013.11.001 PMID: 24291285.

85. Xiao LF, Zhang W, Jing TX, Zhang MY, Miao ZQ, Wei DD, et al. Genome-wide identification, phylogenetic analysis, and expression profiles of ATP-binding cassette transporter genes in the oriental fruit fly, *Bactrocera dorsalis* (Hendel) (Diptera: Tephritidae). Comp Biochem Physiol Part D Genomics Proteomics. 2018; 25:1–8. https://doi.org/10.1016/j.cbd.2017.10.001 PMID: 29121518.

86. Epis S, Porretta D, Mastrantonio V, Comandatore F, Sassera D, Rossi P, et al. ABC transporters are involved in defense against permethrin insecticide in the malaria vector *Anopheles stephensi*. Parasit Vectors. 2014; 7(1):349. https://doi.org/10.1186/1756-3305-7-349 PMID: 25073980.

87. Tian L, Song T, He R, Zeng Y, Xie W, Wu Q, et al. Genome-wide analysis of ATP-binding cassette (ABC) transporters in the sweetpotato whitefly, *Bemisia tabaci*. BMC Genomics. 2017; 18(1):330. https://doi.org/10.1186/s12864-017-3706-6 PMID: 28446145.

88. Bretschneider A, Heckel DG, Vogel H. Know your ABCs: Characterization and gene expression dynamics of ABC transporters in the polyphagous herbivore *Helicoverpa armigera*. Insect Biochem Mol Biol. 2016; 72:1–9. https://doi.org/10.1016/j.ibmb.2016.03.001 PMID: 26951878.

89. Aittomäki S, Valanne S, Lehtinen T, Matikainen S, Nyman TA, Rämet M, et al. Proprotein convertase Furin1 expression in the *Drosophila* fat body is essential for a normal antimicrobial peptide response and bacterial host defense. Faseb J. 2017; 31(11):4770–4782. https://doi.org/10.1096/fj.201700296R PMID: 28705811.

90. Hu C, Aksoy S. Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans*. Mol Microbiol, 2010; 60(5):1194–1204. https://doi.org/10.1111/j.1365-2958.2006.05180.x PMID: 16689795.

91. Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, et al. Immune pathways and defence mechanisms in honey bees *Apis mellifera*. Insect Mol Biol. 2006; 15(5):645–656. https://doi.org/10.1111/j.1365-2583.2006.00682.x PMID: 17069638.

92. Mwangi S, Murungi E, Jonas M, Christoffels A. Evolutionary genomics of *Glossina morsitans* immune-related CLIP domain serine proteases and serine protease inhibitors. Infect Genet Evol. 2011; 11 (4):740–745. https://doi.org/10.1016/j.meegid.2010.10.006 PMID: 21055483.

93. Garrett M, Fullaondo A, Troxler L, Micklem G, Gubb D. Identification and analysis of serpin-family genes by homology and synteny across the 12 sequenced *Drosophilia* genomes. BMC Genomics. 2009; 10(1):489–490. https://doi.org/10.1186/1471-2164-10-489 PMID: 19849829.

94. Zou Z, Evans JD, Lu Z, Zhao P, Williams M, Sumathipala N, et al. Comparative genomic analysis of the *Tribolium* immune system. Genome Biol. 2007; 8(8):R177. https://doi.org/10.1186/gb-2007-8-8-r177 PMID: 17727709.

95. Zou Z, Picheng Z, Weng H, Mita K, Jiang H. A comparative analysis of serpin genes in the silkworm genome. Genomics. 2009; 93(4):367–75. https://doi.org/10.1016/j.ygeno.2008.12.010 PMID: 19150649.

96. Lin H, Lin X, Zhu J, Yu XQ, Xia X, Yao F, et al. Characterization and expression profiling of serine protease inhibitors in the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). BMC Genomics, 2017; 18(1):162. https://doi.org/10.1186/s12864-017-3583-z PMID: 28196471.

97. He Y, Wang Y, Zhao P, Rayaprolu S, Wang X, Cao X, et al. Serpin-9 and -13 regulate hemolymph proteases during immune responses of *Manduca sexta*. Insect Biochem Mol Biol. 2017; 90:71–81. https://doi.org/10.1016/j.ibmb.2017.09.015 PMID: 28987647.

98. Yuan C, Xing L, Wang M, Wang X, Yin M, Wang Q, et al. Inhibition of melanization by serpin-5 and serpin-9 promotes baculovirus infection in cotton bollworm *Helicoverpa armigera*. PLoS Pathog. 2017; 13(9):e1006645. https://doi.org/10.1371/journal.ppat.1006645 PMID: 28953952.

99. Zhao S, Wang X, Cai S, Zhang S, Luo H, Wu C, et al. A novel peptidoglycan recognition protein involved in the prophenoloxidase activation system and antimicrobial peptide production in *Antheraea pernyi*. Dev Comp Immunol. 2018; 86:78–85. https://doi.org/10.1016/j.dci.2018.04.009 PMID: 29734021.

100. Meister S, Agianian B, Turlure F, Relógio A, Morlais I, Kafatos FC, et al. *Anopheles gambiae* PGRPLC-mediated defense against bacteria modulates infections with malaria parasites. PLoS Pathog. 2009; 5(8):e1000542. https://doi.org/10.1371/journal.ppat.1000542 PMID: 19662170.

101. Ping Z, Fei X, Liang J, Guo H, Xu G, Sun Q, et al. Enhanced antiviral immunity against, *Bombyx mori*, cytoplasmic polyhedrosis virus via overexpression of peptidoglycan recognition protein S2 in transgenic silkworms. Dev Comp Immunol. 2018; 87:84–89. https://doi.org/10.1016/j.dci.2018.05.021 PMID: 29902708.

102. Takehana A, Katsuyama T, Yano T, Oshima Y, Takada H, Aigaki T, et al. Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and the prophenoloxidase cascade in *Drosophila* larvae. Proc Natl Acad Sci USA. 2002; 99 (21):13705–13710. https://doi.org/10.1073/pnas.212301199 PMID: 12359879.

103. Gendrin M, Turlure F, Rodgers FH, Cohuet A, Morlais I, Christophides GK. The peptidoglycan recognition proteins *PGRPLA* and *PGRPLB* regulate *Anopheles* immunity to bacteria and affect infection by *Plasmodium*. J Innate Immun. 2017; 9(4):333–342. https://doi.org/10.1159/000452797 PMID: 28494453.

104. Werner T, Liu G, Kang D, Ekengren S, Steiner H, Hultmark D. A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. Proc Natl Acad Sci USA. 2000; 97(25):13772–13777. https://doi.org/10.1073/pnas.97.25.13772 PMID: 11106397.

105. Koyama H, Kato D, Minakuchi C, Tanaka T, Yokoi K, Miura K. Peptidoglycan recognition protein genes and their roles in the innate immune pathways of the red flour beetle, *Tribolium castaneum*. J Invertebr Pathol. 2015; 132:86–100. https://doi.org/10.1016/j.jip.2015.09.003 PMID: 26385528.

106. Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, et al. Immunity-related genes and gene families in *Anopheles gambiae*. Sci. 2002; 298(5591):159–165. https://doi.org/10.1126/science.1077136 PMID: 12364793.

107. Mohamed AA, Zhang L, Dorrah MA, Elmogy M, Yousef HA, Bassal TT, et al. Molecular characterization of a c-type lysozyme from the desert locust, *Schistocerca gregaria* (Orthoptera: Acrididae). Dev Comp Immunol. 2016; 61:60–69. https://doi.org/10.1016/j.dci.2016.03.018 PMID: 26997372.

108. Vogel H, Altincicek B, Glöckner G, Vilcinskas A. A comprehensive transcriptome and immune-gene repertoire of the lepidopteran model host *Galleria mellonella*. BMC Genomics. 2011; 12(1):308. https://doi.org/10.1186/1471-2164-12-308 PMID: 21663692.

109. Altincicek B, Knorr E, Vilcinskas A. Beetle immunity: Identification of immune-inducible genes from the model insect *Tribolium castaneum*. Dev Comp Immunol. 2008; 32(5):585–95. https://doi.org/10.1016/j.dci.2007.09.005 PMID: 17981328.

110. Mitaka Y, Kobayashi K, Matsuura K. Caste-, sex-, and age-dependent expression of immune-related genes in a Japanese subterranean termite, *Reticulitermes speratus*. PLoS ONE. 2017; 12(4): e0175417. https://doi.org/10.1371/journal.pone.0175417 PMID: 28410430.

111. He Y, Cao X, Li K, Hu Y, Chen YR, Blissard G, et al. A genome-wide analysis of antimicrobial effector genes and their transcription patterns in *Manduca sexta*. Insect Biochem Mol Biol. 2015; 62:23–37. https://doi.org/10.1016/j.ibmb.2015.01.015 PMID: 25662101.

112. Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, et al. Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. Genome Biol. 2010; 11(2):R21. https://doi.org/10.1186/gb-2010-11-2-r21 PMID: 20178569.

113. Vilcinskas A, Mukherjee K, Vogel H. Expansion of the antimicrobial peptide repertoire in the invasive ladybird *Harmonia axyridis*. Proc Biol Sci. 2013; 280(1750):20122113. https://doi.org/10.1098/rspb.2012.2113 PMID: 23173204.

114. Chapelle M, Girard PA, Cousserans F, Volkoff NA, Duvic B. Lysozymes and lysozyme-like proteins from the fall armyworm, *Spodoptera frugiperda*. Mol Immunol. 2009; 47(2–3):261–269. https://doi.org/10.1016/j.molimm.2009.09.028 PMID: 19828200.

115. Bulet P, Stöcklin R. Insect antimicrobial peptides: structures, properties and gene regulation. Protein Pept Lett. 2005; 12(1):3–11. https://doi.org/10.2174/0929866053406011 PMID: 15638797.

116. Mulnix AB, Dunn PE. Structure and induction of a lysozyme gene from the tobacco hornworm, *Manduca sext*a. Insect Biochem Mol Biol. 1994; 24(3):271–281. https://doi.org/10.1016/0965-1748(94)90007-8 PMID: 7517269.

117. Tanaka H, Ishibashi J, Fujita K, Nakajima Y, Sagisaka A, Tomimoto K, et al. A genome-wide analysis of genes and gene families involved in innate immunity of *Bombyx mori*. Insect Biochem Mol Biol. 2008; 38(12):1087–1110. https://doi.org/10.1016/j.ibmb.2008.09.001 PMID: 18835443.

118. Gandhe AS, Janardhan G, Nagaraju J. Immune upregulation of novel antibacterial proteins from silk-moths (Lepidoptera) that resemble lysozymes but lack muramidase activity. Insect Biochem Mol Biol. 2007; 37(7):655–666. https://doi.org/10.1016/j.ibmb.2007.03.013 PMID: 17550822.

119. Daffre S, Kylsten P, Samakovlis C, Hultmark D. The lysozyme locus in *Drosophila melanogaster*: an expanded gene family adapted for expression in the digestive tract. Mol Gen Genet. 1994; 242 (2):152–62. https://doi.org/10.1007/bf00391008 PMID: 8159165.

120. Rivera-Perez C. Marine invertebrate lipases: Comparative and functional genomic analysis. Comp Biochem Physiol Part D Genomics Proteomics. 2015; 15:39–48. https://doi.org/10.1016/j.cbd.2015.06.001 PMID: 26114431.

121. Horne I, Haritos VS, Oakeshott JG. Comparative and functional genomics of lipases in holometabolous insects. Insect Biochem Mol Biol. 2009; 39(8):547–567. https://doi.org/10.1016/j.ibmb.2009.06.002 PMID: 19540341.

122. Majerowicz D, Calderón-Fernández GM, Alves-Bezerra M, De Paula IF, Cardoso LS, Juárez MP, et al. Lipid metabolism in *Rhodnius prolixus*: Lessons from the genome. Gene. 2017; 596:27–44. https://doi.org/10.1016/j.gene.2016.09.045 PMID: 27697616.

123. Hossain MS, Liu Y, Zhou S, Li K, Tian L, Li S. 20-Hydroxyecdysone-induced transcriptional activity of FoxO upregulates brummer and acid lipase-1 and promotes lipolysis in *Bombyx* fat body. Insect Biochem Mol Biol. 2013; 43(9):829–838. https://doi.org/10.1016/j.ibmb.2013.06.007 PMID: 23811219.

124. Grönke S, Mildner A, Fellert S, Tennagels N, Petry S, Müller G, et al. Brummer lipase is an evolutionary conserved fat storage regulator in *Drosophila*. Cell Metab. 2005; 1(5):323–30. https://doi.org/10.1016/j.cmet.2005.04.003 PMID: 16054079.

125. Attardo GM, Benoit JB, Michalkova V, Yang G, Roller L, Bohova J, et al. Analysis of lipolysis underlying lactation in the tsetse fly, *Glossina morsitans*. Insect Biochem Mol Biol. 2012; 42(5):360–370. https://doi.org/10.1016/j.ibmb.2012.01.007 PMID: 22509523.

126. Pistillo D, Manzi A, Tino A, Boyl PP, Graziani F, Malva C. The *Drosophila melanogaster* lipase homologs: a gene family with tissue and developmental specific expression. J Mol Biol. 1998; 276(5):877–885. https://doi.org/10.1006/jmbi.1997.1536 PMID: 9566193.

127. Tan QQ, Liu W, Zhu F, Lei CL, Wang XP. Fatty acid synthase 2 contributes to diapause preparation in a beetle by regulating lipid accumulation and stress tolerance genes expression. Sci Rep. 2017; 7:40509. https://doi.org/10.1038/srep40509 PMID: 28071706.

128. Chung H, Loehlin DW, Dufour HD, Vaccaro K, Millar JG, Carroll SB. A single gene affects both ecological divergence and mate choice in *Drosophila*. Sci. 2014; 343(6175):1148–1151. https://doi.org/10.1126/science.1249998 PMID: 24526311.

129. Zuo W, Li C, Luan Y, Zhang H, Tong X, Han M, et al. Genome-wide identification and analysis of elongase of very long chain fatty acid genes in the silkworm, *Bombyx mori*. Genome. 2018; 61(3):167–176. https://doi.org/10.1139/gen-2017-0224 PMID: 29505281.

130. Jakobsson A, Westerberg R, Jacobsson A. Fatty acid elongases in mammals: their regulation and roles in metabolism. Prog Lipid Res. 2006; 45(3):237–249. https://doi.org/10.1016/j.plipres.2006.01.004 PMID: 16564093.

131. Ng WC, Chin JS, Tan KJ, Yew JY. The fatty acid elongase Bond is essential for *Drosophila* sex pheromone synthesis and male fertility. Nat Commun. 2015; 6:8263. https://doi.org/10.1038/ncomms9263 PMID: 26369287.

132. Urbanski JM, Benoit JB, Michaud MR, Denlinger DL, Armbruster P. The molecular physiology of increased egg desiccation resistance during diapause in the invasive mosquito, *Aedes albopictus*. Proc Biol Sci. 2010; 277(1694):2683–2692. https://doi.org/10.1098/rspb.2010.0362 PMID: 20410035.