

Full-length transcriptome analysis of *Spodoptera frugiperda* larval brain reveals detoxification genes

Lei Yang¹, Binglin Xing¹, Fen Li¹, Li Kui Wang¹, Linlin Yuan¹, Amosi Leonard Mbuji^{1,2}, Zhengqiang Peng³, Farag Malhat⁴ and Shaoying Wu¹

¹ Hainan University, Haikou, Hainan, China

² Department of Resources Utilization and Plant Protection, College of Resources and Environmental Science, China Agricultural University, Beijing, Beijing, China

³ The Ministry of Agriculture and Rural Affairs Key Laboratory of Integrated Pest Management of Tropical Crops, Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China

⁴ Pesticide Residues and Environmental Pollution Department, Agricultural Research Center, Dokki, Giza, Egypt

ABSTRACT

Background: *Spodoptera frugiperda* (J. E. Smith), commonly known as fall armyworm (FAW), is one of the most destructive agricultural pests in the world and has posed a great threat to crops. The improper use of insecticides has led to rapid development of resistance. However, the genetic data available for uncovering the insecticide resistance mechanisms are scarce.

Methods: In this study, we used PacBio single-molecule real-time (SMRT) sequencing aimed at revealing the full-length transcriptome profiling of the FAW larval brain to obtain detoxification genes.

Results: A total of 18,642 high-quality transcripts were obtained with an average length of 2,371 bp, and 11,230 of which were successfully annotated in six public databases. Among these, 5,692 alternative splicing events were identified.

Subjects Biochemistry, Entomology, Molecular Biology, Toxicology

Keywords *Spodoptera frugiperda*, Resistance, Full-length transcriptome, Detoxification gene

INTRODUCTION

Spodoptera frugiperda (J. E. Smith), known as fall armyworm (FAW), belongs to the Noctuidae of Lepidoptera. It is one of the most dangerous invasive pests originated from the tropical and sub-tropical regions of the Americas (*Capinera, 2002*). FAW possesses a wide range of hosts, including at least 353 plants belonging to 76 families, such as corn, wheat, rice, sorghum and others (*Wan et al., 2020*), it has caused great damage to crops, thus leading to huge economic losses to agriculture. FAW is an alien invasive pest in China and quickly invaded the Yunnan Province, southern China in December 2018 (*Sun et al., 2021; Wu, Jiang & Wu, 2019*). At the end of 2020, it had colonized corn fields of 1,338 counties of 27 provinces in China, and the damaged area caused by FAW has reached as high as 11,125–12,780 hectares (*Zhou et al., 2021*). FAW adult has a strong capability of long-distance migration, and the average flight distance was more than

Submitted 24 May 2021
Accepted 5 August 2021
Published 24 August 2021

Corresponding author
Shaoying Wu,
wsywsy6000@hainanu.edu.cn

Academic editor
Joseph Gillespie

Additional Information and
Declarations can be found on
page 12

DOI 10.7717/peerj.12069

© Copyright
2021 Yang et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

100 km per night under proper environmental conditions (*Tendeng et al., 2019*), which aggravating its rapid spreading (*Wan et al., 2020*). Additionally, the fecundity of FAW is extremely high, the average number of laying eggs per female is 1,052–1,323 (*Wan et al., 2020*). More exaggeratedly, the maximum number of egg production per female reached 2,000 in Africa (*Prasanna et al., 2018*). To sum up, the gluttony, strong migration ability and high fertility of FAW have attracted extensive attention. Therefore, it was classified as the top first destructive pest in China (*Zhou et al., 2021*).

Transcriptome especially the short-reading sequencing technique has long been developed and used as one of the most efficient methods to obtain reference gene sequences, quantify gene expression levels and characterize gene functions not only in vertebrates, but also in invertebrates such as insects (*Djebali et al., 2012; Ekblom & Galindo, 2011; Nagalakshmi et al., 2008; Yin et al., 2014, 2016*). Although booming advances in technology have been made recently, there were many limitations (*Zhou et al., 2010*). For instance, this sequencing technology failed to identify genes longer than 5,000 bp upon most occasions, and the excavation of alternative splicing spanning entire transcripts was rather difficult (*Koren et al., 2012*). Moreover, incorrect annotations were often conducted due to low-quality transcripts. To overcome these limitations, third generation full-length transcriptome based on PacBio single-molecule real-time (SMRT) sequencing technology was developed (*Picelli et al., 2013*). Compared to the short-reading sequencing technique, SMRT sequencing greatly contributes to our understanding of the physiological processes of objects at a molecular level, such as the availability of accurate gene full-length (maximum > 20,000 bp), identifiable alternative splicing isoforms, and easy to screen novel transcripts.

Insects are naturally under the selection pressure of various pesticides in local ecological context. As a coping strategy, several defensive mechanisms have been formed to enhance the metabolism of chemicals, decrease their toxic effect and ensure their survival during long-term evolution (*Despres, David & Gallet, 2007*). Among these, metabolic resistance is of vital importance, which involves many detoxification enzymes acting in the biotransformation of hazardous chemicals to non-toxic metabolites. To date, many detoxification gene families have been characterized typically classified into five categories according to their different metabolic phases, among which the ATP-binding cassettes (ABCs) are involved in the transport of a broad variety of substrates across the cell membrane (*Atsumi et al., 2012*), and carboxylesterases (CEs) hydrolyze an ester *via* the addition of water to form the corresponding alcohol and acid (*Wheelock, Shan & Ottea, 2005*). In contrast, cytochrome P450s (CYPs) are terminal oxidases involved in drug and steroid hydroxylation reactions (*Scott, 1999*), and glutathione S-transferases (GSTs) can catalyze the conjugation of glutathione to insecticides or directly metabolize the insecticides to less toxic soluble conjugate or non-catalytically bind to the insecticides (*Kostaropoulos et al., 2001*), while UDP-glucuronosyl transferases (UGTs) conjugate the uridine diphosphate glucuronic acid to metabolites to decrease hydrophobicity (*Li et al., 2018*). The roles of these detoxification genes had been well documented in numerous phytophagous pests such as lepidopteran, coleopteran, hemipteran, dipteran, and other

species (*Enayati, Ranson & Hemingway, 2005; Stevens et al., 2000; Yang et al., 2005*). However, the full-length data of detoxification genes are scarce.

Here, we used PacBio SMRT sequencing technology to profile the full-length transcriptome of the FAW larval brain. Next, five main gene families associated with the detoxification function were classified, and the predicted lncRNA targeting these transcripts were further analyzed. These findings undoubtedly provide important information for exploring the scenario of these genes and revealing their functions.

MATERIALS & METHODS

Insect rearing and sampling

FAW larvae were collected from the corn field in Baoting, Hainan, Southern China (19°14'45.12"N, 109°89'40.40"E) on May 5th, 2019. We established a laboratory colony by feeding the larvae on artificial diet under conditions of 25 ± 1 °C, 14 h:10 h (L:D) photoperiod, and $55 \pm 5\%$ relative humidity. The third instar larvae were individually transferred into a small box and reared until pupation. Adults were transferred into a home-made insect raise cage (Patent no. 201921652702.2) after inclusion and fed with 10% hydromel (v/v). Once adequate fourth instar larvae were obtained, we rinsed them in 75% ethanol (v/v) and sterile phosphate-buffered saline (PBS, pH 7.2) twice, respectively. Next, the brain was dissected in PBS containing 1 unit/ μ L Murine RNase inhibitor (Vazyme, Nanjing, China) under an Olympus SZ61 stereomicroscope (Olympus, Tokyo, Japan). We collected the sample into a 1.5 ml Eppendorf tube for subsequent RNA extraction.

RNA isolation and SMRT sequencing

We extracted the total RNA of FAW larval brain using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the description of manufacturer and checked the RNA purity (OD 260/280) using Nanodrop 2000 spectrophotometer (NanoDrop Products, Wilmington, DE, USA). Then, RNA degradation and DNA contamination were detected by 1% agarose gel electrophoresis. After a quality inspection, 12 μ g RNA was used to gather mRNA molecules by oligo (dT) magnetic beads. We constructed a full length cDNA library using the Clontech SMARTer PCR cDNA synthesis kit (Clontech, Mountain View, CA, USA). The synthesized cDNA was purified by AMPure XP beads (Beckman Coulter, Brea, CA, USA) and sequenced on the PacBio sequel II platform (Pacific Biosciences, Menlo Park, CA, USA) based on the manufacturer's instructions. Raw data of FAW transcriptome were available from the NCBI SRA database (Project number: [PRJNA729608](https://www.ncbi.nlm.nih.gov/sra/PRJNA729608)).

Data processing of full-length transcriptome

We used SMRTlink 6.0 software (Pacific Biosciences, Menlo Park, CA, USA) to process the sequence data. Then, reads containing adaptors, Poly-N, and low-quality from raw data were removed, and the filtered reads were converted into cyclic consensus sequence (CCS) according to the criteria of full passes ≥ 1 and sequence accuracy >0.9 . Full length and non-full length reads were respectively predicted followed by arrow polishing using

Quiver. Subsequently, we removed redundant sequences using the CD-hit program and aligned these terms against the FAW reference genome (SRA project number: [PRJNA647344](https://www.ncbi.nlm.nih.gov/sra/PRJNA647344)) using GMAP (<http://research-pub.gene.com/gmap/>) (*Li & Godzik, 2006*). The obtained transcripts were used for the next analysis.

Transcript annotation and alternative splicing analysis

All transcripts were functionally annotated based on the six databases, NCBI non-redundant protein sequences (NR) database using blastx with e-value $<1e^{-5}$ (*Pruitt, Tatusova & Maglott, 2007*), Protein family (Pfam) database using Hmmscan, a HMMER-based approach for detecting matches to Pfam families (*Finn, Clements & Eddy, 2011; Finn et al., 2009*), a manually annotated and reviewed protein sequence (Swiss-Prot) database (*Bairoch & Apweiler, 2000*), Clusters of Orthologous Groups of proteins (KOG) database (*Tatusov et al., 2000*), Kyoto Encyclopedia of Genes and Genomes (KEGG) database using Diamond v0.8.36 software with e-value $<1e^{-5}$ (*Kanehisa & Goto, 2000*), Gene Ontology (GO) database using Blast2GO v2.5 (*Conesa et al., 2005; Consortium, 2004*), individually. To identify alternative splicing events, transcripts were further processed using Astalavista software with default parameters (*Foissac & Sammeth, 2007*), and we defined alternative splicing events based on the criteria of previous description (*Shen et al., 2021*).

Identification and phylogeny analysis of detoxification genes

Five detoxification gene families, CYP, ABC, CES, GST, and UGT, were screened based on the annotation results and further manually verified. A simple modular architecture research tool (SMART) was used for predicting the conserved domains (*Letunic & Bork, 2017*). We conducted multiple sequence alignments using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) followed by constructing phylogenetic trees based on the maximum likelihood method using Mega-X software with 1,000 bootstrap values (*Kumar et al., 2018*), which was further edited and visualized using FigTree v1.4.3 software (Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK).

LncRNA identification, target prediction and motif analysis

We used four analysis tools, Pfam (*Finn et al., 2014*), Coding-Non-Coding-Index (CNCI) (*Sun et al., 2013*), Coding Potential Calculator (CPC) (*Kong et al., 2007*), and Coding Potential Assessment Tool (CPAT) to screen potential lncRNAs (*Wang et al., 2013*). To ensure the accuracy of these results, transcripts simultaneously recognized by these tools were defined as lncRNAs. In the following, LncTar tool was used to predict the lncRNA-mRNA interactions (*Li et al., 2015*), and motif analysis was conducted using the online tool MEME (*Bailey et al., 2015*).

RESULTS

Overview of full-length transcriptome data

The 61.50 GB clean data of fall armyworm (FAW) larval brain full-length transcriptome were obtained by SMRT sequencing. A total of 297,682 circular consensus sequences

Table 1 Overview of FAW larval brain full-length transcriptome.

Category	Number
Circular consensus sequences (CCS)	297,682
Full-length sequences	248,554
Non-chimeric full-length sequences with Poly-A Tail	234,863
Average read length	2,371
High quality non-redundant transcripts	18,642
Alternative splicing events	5,692
Predicted lncRNA	1,319
Transcripts with functional annotation	11,230
Transcripts with ORF	11,013

(CCSs) were generated, and 248,554 sequences containing poly-A tail were identified as full length sequences (Table 1). Among these, 234,863 reads (94.49%) was classified as non-chimeric sequences with a mean read length of 2371 bp. Next, redundant sequences within different transcript clusters were removed using CD-hit program, and 18,642 high quality non-redundant transcripts were identified and used for subsequent analysis.

Transcript annotation

To reveal the function of 18,642 transcripts, we annotated these terms on six databases. Notably, 9,834, 7,805, 6,821, 6,715, 5,513 and 219 transcripts were annotated in NR, Pfam, Swiss-Prot, GO, KEGG, and KOG databases, respectively. In total, we obtained 11,230 transcripts with valid functional prediction in at least one database. What stood out in Fig. 1A was that 7,737 homologous transcripts (68.90%) best matched the *S. litura*, and the top seven hits belonged to the Noctuidae of Lepidoptera based on the annotation results in NR database, showing high similarity. Next, all transcripts were annotated in GO database and 6,715 terms were classified into three GO categories, biological process, cellular component, and molecular function (Fig. 1B). In biological processes, we found that most of the transcripts were associated with cellular processes followed by metabolic processes and biological regulation. In contrast, the transcripts involved in cell and cell parts were the most abundant terms in the cellular component category. Additionally, in molecular function, the top three annotated transcripts having binding, catalytic and transporter activity were respectively identified (Fig. 1B). We also sorted these transcripts into different groups according to their participation in KEGG metabolic pathways. Among the top 20 annotated terms, the most enriched categories pertained to adrenergic signaling in the cardiomyocyte pathway (268) followed by the cardiac muscle contraction pathway (254) (Fig. 1C). Finally, we annotated 219 transcripts in KOG databases, transcripts participating in “Cytoskeleton”, “Energy production and conversion” and “Post-translational modification, protein turnover, chaperones” were the top three abundant terms (Fig. 1D).

CYPs

Based on the functional annotation, 84 transcripts were characterized as CYPs with a mean length of 2,146 bp, further being manually confirmed in NR database. Among these,

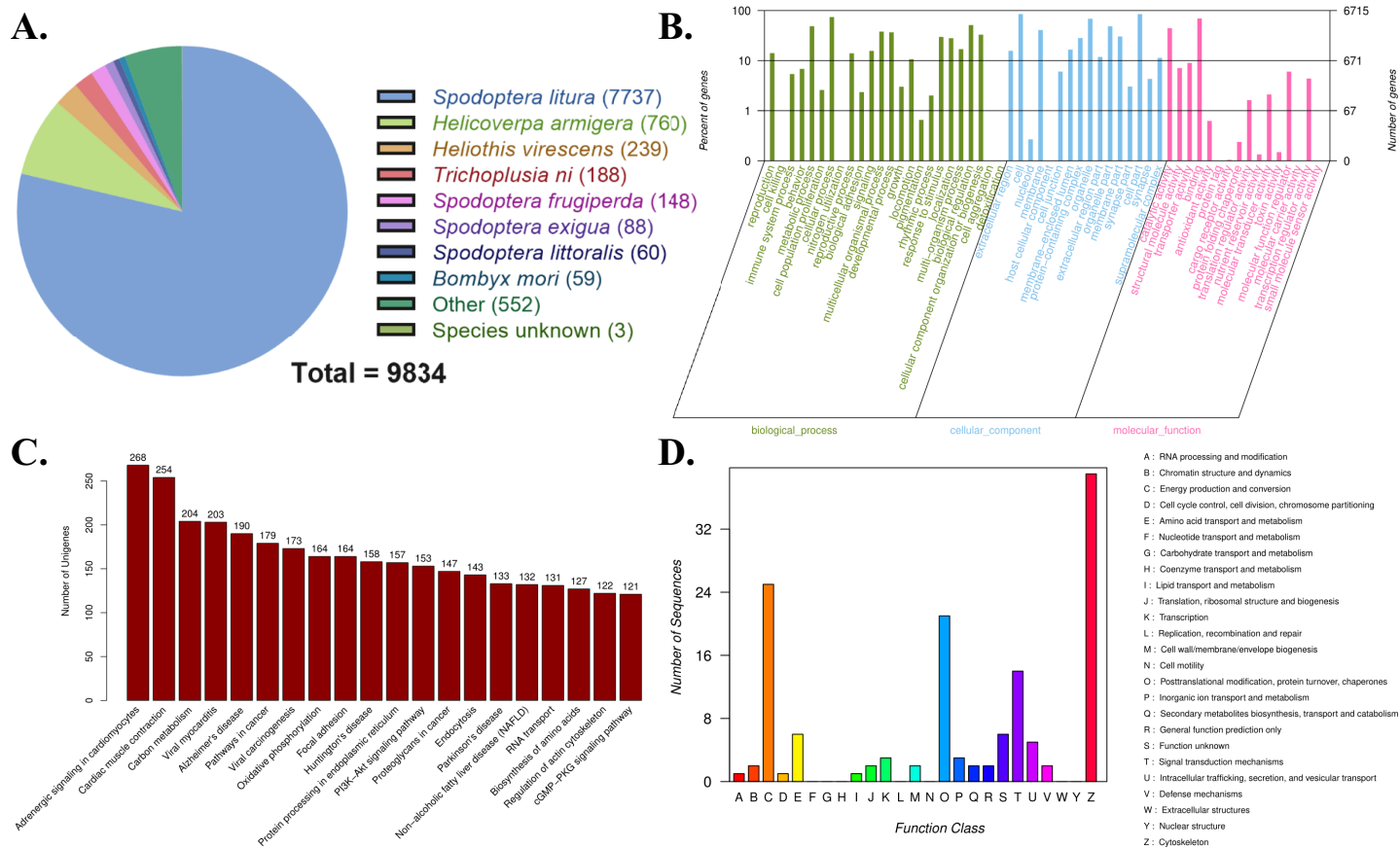


Figure 1 Full-length transcript annotation of FAW larval brain. (A) Homologous species distribution of transcripts in NR database. (B) GO functional categorization of transcripts. (C) Top 20 KEGG functional classifications of transcripts. (D) COG functional classification of transcripts. Full-size [DOI: 10.7717/peerj.12069/fig-1](https://doi.org/10.7717/peerj.12069/fig-1)

64 transcripts with full length were identified. A phylogenetic tree based on the conserved “CYP domain” was constructed, and results showed that these transcripts were distributed in four clades, CYP4, CYP6, CYP9 and CYP12. Among these, CYP9 concluding 28 transcripts made up the largest clade (Fig. 2A), indicating their functional redundancy. We further identified a lncRNA targeting the transcript “PB.1240.2”, and the predicted binding site was located in the CDS region (Fig. 2B). Besides, the conserved motif analysis presented an overview of the conserved amino acid residues among 84 transcripts (Fig. 2C).

ABCs

Fifty-four transcripts encoding ABCs were obtained from the transcriptome of the FAW larval brain, and nine of which did not contain full length. The mean length of 49 ABCs was 4,074 bp. In typical cases, ABCs were embedded with one or more “ABC transmembrane domain”, we thus constructed a phylogenetic tree based on the conserved domains. From the data in Fig. 3A, it was apparent that 54 transcripts clustered into five clades, namely ABC-A, ABC-B, ABC-E, ABC-F, and ABC-G. Among these, ABC-A occupied the majority, which contained 18 members. None of the lncRNAs were predicted

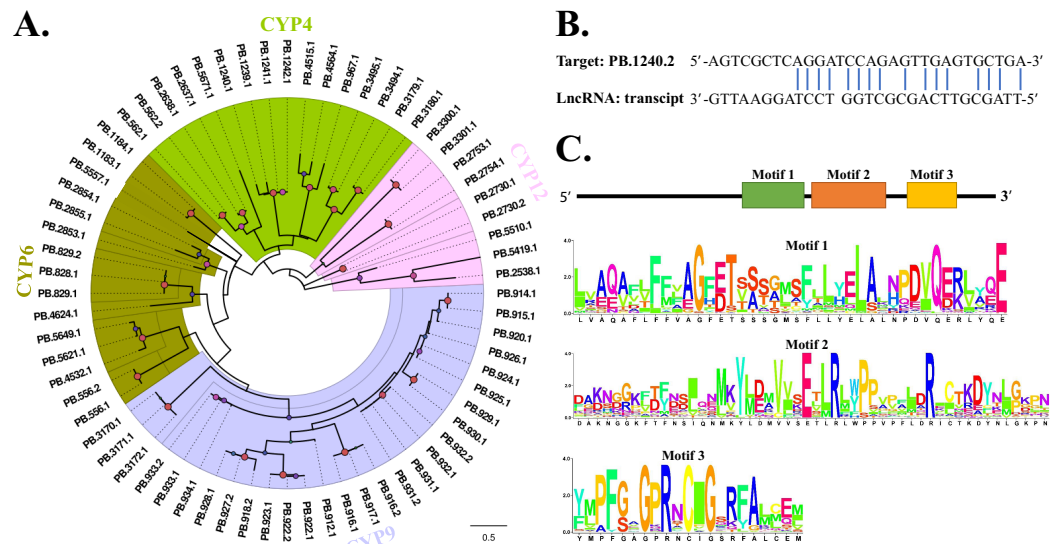


Figure 2 Phylogenetic construction, lncRNA target prediction and motif analysis of CYPs. (A) Phylogenetic tree of 84 FAW CYP transcripts. The evolutionary tree was constructed based on the maximum likelihood method using the program Mega X with 1,000 bootstrap values. Forty-eight different amino acid substitution models were tested and “LG+G” was the best model. (B) The binding sites of lncRNA targeting transcript “PB.1240.2”. LncTar was used to predict RNA targets of lncRNAs with a cutoff of -0.1 normalized deltaG. (C) Motif analysis of 84 CYPs. We used MEME tool to identify the conserved motifs. [Full-size !\[\]\(fcc3264021d438d9732560e78099f674_img.jpg\) DOI: 10.7717/peerj.12069/fig-2](https://doi.org/10.7717/peerj.12069/fig-2)

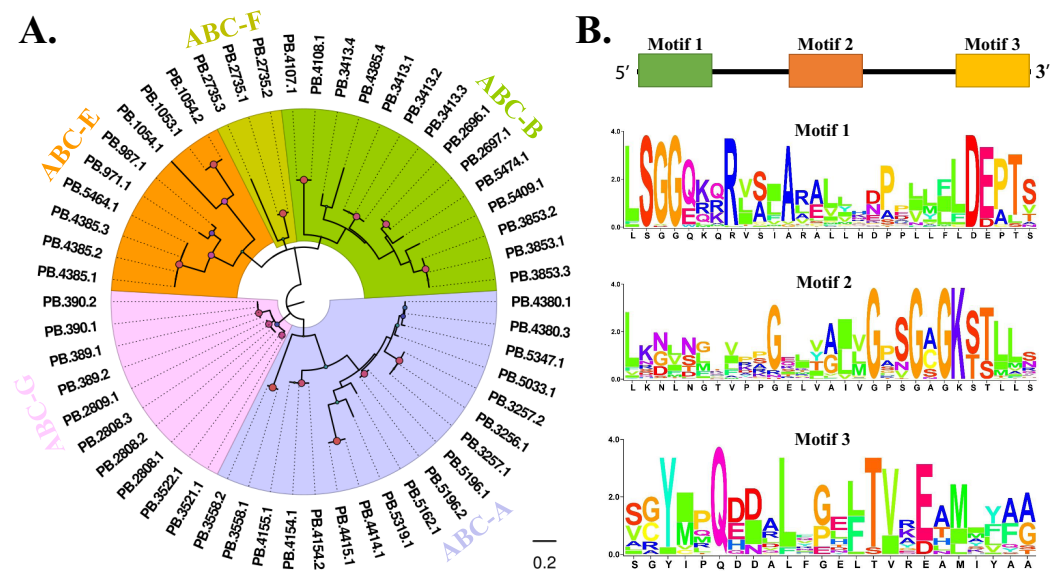


Figure 3 Phylogenetic and motif analysis of ABCs. (A) Phylogenetic tree of 54 FAW ABC transcripts. The evolutionary tree was constructed based on the maximum likelihood method using the program Mega X with 1,000 bootstrap values. Forty-eight different amino acid substitution models were tested and “LG+G” was the best model. (B) Motif analysis of 54 ABCs. We used MEME tool to identify the conserved motifs. [Full-size !\[\]\(9d188a796ceef961be962a3cd4b57b68_img.jpg\) DOI: 10.7717/peerj.12069/fig-3](https://doi.org/10.7717/peerj.12069/fig-3)

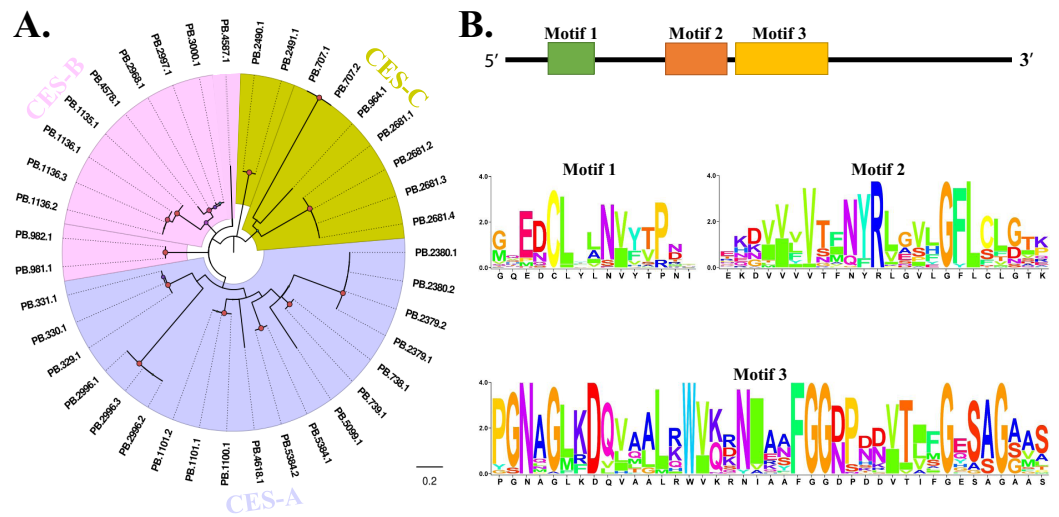


Figure 4 Phylogenetic and motif analysis of 46 CESs. (A) Phylogenetic tree of 46 FAW CES transcripts. The evolutionary tree was constructed based on the maximum likelihood method using the program Mega X with 1,000 bootstrap values. Forty-eight different amino acid substitution models were tested and “LG+G+I” was the best model. (B) Motif analysis of 46 CESs. We used MEME tool to identify the conserved motifs. [Full-size !\[\]\(1663bb69f307a960345edb0e712f8c02_img.jpg\) DOI: 10.7717/peerj.12069/fig-4](https://doi.org/10.7717/peerj.12069/fig-4)

to bind to the 54 ABC members. Subsequently, several conserved amino acid sites were identified based on the motif analysis (Fig. 3B).

CESs

After removal of redundant sequences, 46 transcripts were annotated as CESs and half of these were equipped with full length with a mean length of 2,493 bp. To classify these transcripts into different categories, a phylogenetic analysis based on the “CES domain” was constructed. What stood out in Fig. 4A was that 46 CESs clustered into three branches, namely CES-A, CES-B and CES-C. Again, no predicted lncRNAs targeting these transcripts were disclosed. According to the results of motif analysis, we identified several amino acid residues highly conserved among the CES transcripts (Fig. 4B).

GSTs

According to the functional annotation of transcriptome data in six databases, 33 transcripts were identified as GSTs, in which 18 terms were available of full length and the mean length was 1,781 bp. GST was embedded with at least one “GST domain” in the N-terminal and one “GST domain” in the C-terminal, respectively, we thus assigned these transcripts into different groups based on the evolutionary relationship of GST domains. From the graph below, we could see that three clusters, including “Delta and Epsilon”, “Sigma” and “Omega” gathered together, indicating their functional diversity (Fig. 5A). The lncRNA target prediction revealed a regulatory lncRNA could bind to the CDS of “PB.3383.1” transcript (Fig. 5B). Additionally, the results obtained from motif analysis of 33 transcripts were shown in Fig. 5C, and many conserved sites were also revealed.

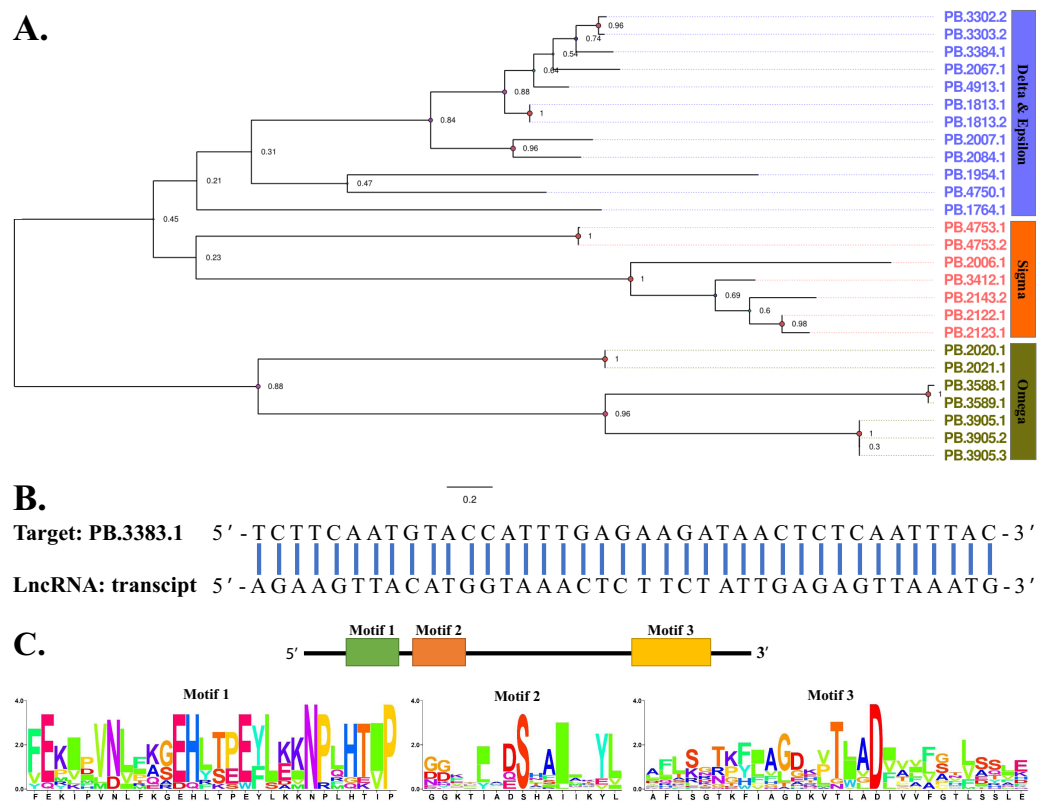


Figure 5 Phylogenetic construction, lncRNA target prediction and motif analysis of GSTs. (A) Phylogenetic tree of 33 GSTs. The evolutionary tree was constructed based on maximum likelihood method using the program Mega X with 1,000 bootstrap values. Forty-eight different amino acid substitution models were tested and “LG+G” was the best model. (B) The binding sites of lncRNA targeting transcript “PB.3383.1”. LncTar was used to predict RNA targets of lncRNAs with a cutoff of -0.1 normalized deltaG. (C) Motif analysis of 33 GSTs. We used MEME tool to identify the conserved motifs. Full-size [DOI: 10.7717/peerj.12069/fig-5](https://doi.org/10.7717/peerj.12069/fig-5)

UGTs

Thirty-one transcripts containing “UGT domain” were annotated as UGTs and among these 19 full-length sequences were obtained. The mean length of the fully sequenced transcripts was 2,104 bp. Phylogenetic analysis revealed that these transcripts evolved into two branches, “UGT1” and “UGT2” (Fig. 6A). Besides, no lncRNAs were identified as the regulator of UGTs. As shown in Fig. 6B, there were several conserved amino acid residues identifiable from 31 UGTs.

DISCUSSION

The full transcriptome analysis of FAW larval brain revealed that 18,642 high-quality transcripts, 11,230 of which were categorized into different functional groups based on homologous blast in six public databases. From the results, we could see that most of the transcripts showed a higher homology with those of *S. litura*, a closely related species of FAW, which can provide a valuable reference for subsequent gene function characterization of FAW. Moreover, GO enrichment analysis, KEGG metabolic pathway, and KOG annotations also provided us a reference for the function of these transcripts.

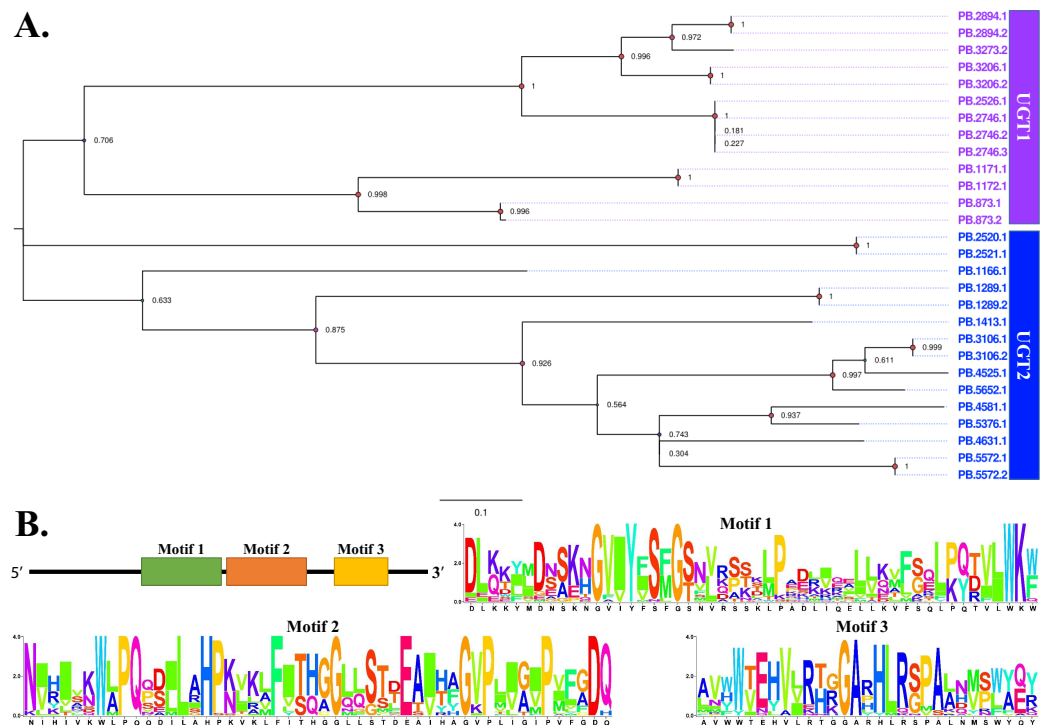


Figure 6 Phylogenetic and motif analysis of UGTs. (A) Phylogenetic tree of 31 UGTs. The evolutionary tree was constructed based on maximum likelihood method using the program Mega X with 1,000 bootstrap values. Forty-eight different amino acid substitution models were tested and “LG+G+I” was the best model. (B) Motif analysis of 31 UGTs. We used MEME tool to identify the conserved motifs.

Full-size DOI: [10.7717/peerj.12069/fig-6](https://doi.org/10.7717/peerj.12069/fig-6)

Alternative splicing events form diverse gene variants, and 5,692 alternative splicing events were identified in this assay, possibly accounting for the diversification of gene functions of these transcripts.

The next section of the survey was concerned with detoxification genes. An abundance of evidence has accumulated over the last few decades, indicating that insect herbivores have evolved a set of detoxification genes to detoxify chemicals (Rane *et al.*, 2019). In typical cases, the detoxification enzymes include CYPs, ABCs, CEs, GSTs, UGTs and so on (Georghiou, 2012). As one of the largest superfamily, CYPs commonly exist in insects while the number varies a lot. For instance, there was a total of 143 genes in *Tribolium castaneum* (Zhu *et al.*, 2013). Comparatively, only 38 CYPs were identified in *Pediculus humanus humanus* (Lee *et al.*, 2010). A previous study has examined CYPs of FAW by whole genome-wide sequencing and finally identified 152 CYP gene clusters (Gui *et al.*, 2020). These results also corroborated the ideas of Xiao *et al.* (2020), who discovered 169 FAW CYPs with widespread expression at different developmental stages (Xiao *et al.*, 2020). In this study, only 84 transcripts encoding CYPs were annotated in full-length transcriptome data of the FAW larval brain, showing that numerous CYPs were not specifically expressed in the brain. Phylogenetic analysis clustered these transcripts into four clans, among these 15 unique terms were classified into the CYP6 clade. It has been widely known that CYP6 family members act in detoxifying pesticides (Ibrahim *et al.*,

2016; Shi et al., 2018; Yang et al., 2020). Therefore, our findings will provide an important reference for exploring the molecular mechanism of FAW CYPs in detoxification.

The motif analysis displayed several conserved amino acid sites among 84 CYPs possibly crucial for its function.

Relative to CYPs, ABCs are a set of newly identified detoxification enzymes, among which the ABC-C2 genes are largely responsible for transport of activated Cry toxins into the cell, and the insecticide resistance is derived from loss of function in ABC-C2 (Merzendorfer, 2014). Recently, the whole genome sequencing of various insect species provides a great convenience to identify the ABC gene family. For example, 104, 73, 56 and 53 ABCs were characterized in *Tetranychus urticae* (Dermauw et al., 2013), *T. castaneum* (Broehan et al., 2013), *Drosophila melanogaster* (Liu et al., 2011), and *Bombyx mori* (Liu et al., 2011), respectively. A recent study revealed that 57 ABCs were expressed in FAW combining genome and transcriptome analyses (Gui et al., 2020). In this assay, 54 transcripts were identified, which was in accordance with the previous results. These ABCs were categorized into five functional groups, ABC-A, ABC-B, ABC-E, ABC-F and ABC-G. Several evidences have shown that ABCs are closely related to insecticide resistance in insects. For instance, in *Trichoplusia ni*, a significant increase in transcript levels in ABC-B genes was recorded after exposure to deltamethrin (Simmons et al., 2013). Besides, five ABC-B and 13 ABC-C genes of *Anopheles gambiae* were thought to be responsible for detoxification (Roth et al., 2003). Many conserved amino acid residues were also identified among 54 ABCs by motif analysis, implying the potential targets for developing novel pesticides.

As a member of the large carboxylesterase family, CESs are also implicated in the detoxification of various xenobiotic compounds, especially the organophosphate insecticides (Wheelock, Shan & Ottea, 2005). CESs have been well characterized in many flies and aphids, such as *Lucilia cuprina* (Jackson et al., 2013), *D. melanogaster* (Cui et al., 2015), *Sitobion avenae* (Xu et al., 2014), *Aphis gossypii* (Cao et al., 2008), *Myzus persicae* (Devonshire, Moores & Ffrench-Constant, 1986). Many studies supported that CESs act in alleviating insecticidal efficacy through gene amplification or up-regulated their transcriptional level (Hemingway, 2000). Here, we annotated 46 FAW CESs embedded with many conserved amino acid residues, more than the 30 terms previously reported (Gui et al., 2020), and then assigned these terms into three clades from an evolutionary perspective. We thus speculated that these CESs may account for the rapid development of FAW resistance to pesticides.

GSTs can produce resistance to all main classes of insecticides. Advances in genomic and transcriptomic sequencing technologies triggered an exhaustive identification of GSTs in many insects, including 11 crop pests and four species of mosquito (Pavliidi, Vontas & Van Leeuwen, 2018). Our present investigation identified 33 GSTs distributed in three clades equipped with many conserved amino acid residues. In contrast, a recent study reported 29 GSTs of FAW combining genomic and transcriptomic analyses (Gui et al., 2020). Among these, 15 genes were remarkably down-regulated after pesticide treatment.

Also, in *S. litura*, a closely related species of FAW, 37 transcripts were annotated and their expression levels after chlorpyrifos treatment were analyzed (Zhang et al., 2016). These findings undoubtedly provided important references for further functional study of GSTs in detoxification.

UGTs also belong to the relatively newly identified classes. As a kind of detoxification enzyme, UGTs widely exist in lepidopteran insects, including *B. mori* (Huang et al., 2008), *S. litura* and many other species (Shen et al., 2021; Shi et al., 2019). In typical cases, UGTs are classified into two distinct subfamilies, UGT1 and UGT2. Compared with the 49 UGTs previously detected in FAW genome and transcriptome (Gui et al., 2020), a small number of UGTs representing 31 transcripts grouped into two families were annotated in this study. It was noted that 17 FAW UGTs showed significant differential expression after pesticide treatment (Gui et al., 2020). Similarly, 10 UGTs were remarkably up-regulated in both indoor and field indoxacarb-resistant strains of *S. litura* (Shi et al., 2019). Taken all these findings together, we speculated that the 31 UGT transcripts aid in the detoxification process. We also identified several conserved amino acid residues in these UGTs, providing promising targets for developing novel pesticides.

lncRNAs are implicated in various biological functions such as protein synthesis, RNA transport and *etc.*, one of the most important roles in which is the gene expression regulation through gene silencing (Wang & Chang, 2011). The full transcriptome sequencing provides an alternative method to predict lncRNAs with the potential binding ability to transcripts. Thus, we analyzed the lncRNAs targeting the transcripts of five main detoxification gene families, and the results, while preliminary, suggested that only two lncRNAs potentially bond to the transcript PB.1240.2, a member of CYPs, and the transcript PB.3383.1, a member of GSTs, respectively. These findings may help us understand the underlying molecular mechanism that lncRNAs regulate detoxification genes.

CONCLUSIONS

Full-length transcriptome of the FAW larval brain was obtained based on the PacBio SMRT sequencing and analyses involving gene annotation, lncRNA prediction and alternative splicing events were conducted. A further in-depth study examined the five detoxification gene families, which provided an abundant genetic resource and laid a solid foundation for revealing the resistance mechanisms of FAW to pesticides.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Hainan Major Science and Technology Project (No. ZDKJ202002), the Hainan Provincial Natural Science Foundation of China (No. 320RC473, No. 320RC489), Monitoring and control of major invasive organisms in Hainan Province (No. 1630042020020), Hainan University (No. KYQD (ZR) 1963, No. KYQD (ZR) 1951, No. RZ2100003220). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Hainan Major Science and Technology Project: ZDKJ202002.

Hainan Provincial Natural Science Foundation of China: 320RC473 and 320RC489.

Monitoring and Control of Major Invasive Organisms in Hainan Province:
1630042020020.

Hainan University: KYQD (ZR) 1963, KYQD (ZR) 1951, and RZ2100003220.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Lei Yang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Binglin Xing conceived and designed the experiments, performed the experiments, prepared figures and/or tables, and approved the final draft.
- Fen Li conceived and designed the experiments, prepared figures and/or tables, and approved the final draft.
- Li Kui Wang analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Linlin Yuan analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Amosi Leonard Mbuji analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Zhengqiang Peng conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Farag Malhat conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Shaoying Wu conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The raw data and assembled transcriptome of full-length transcriptome of *Spodoptera frugiperda* larval brain described here are available at NCBI SRA: [PRJNA729608](https://www.ncbi.nlm.nih.gov/sra/PRJNA729608).

Data Availability

The following information was supplied regarding data availability:

The raw data are available in the [Supplemental Table](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12069#supplemental-information>.

REFERENCES

- Atsumi S, Miyamoto K, Yamamoto K, Narukawa J, Kawai S, Sezutsu H, Kobayashi I, Uchino K, Tamura T, Mita K. 2012. Single amino acid mutation in an ATP-binding cassette transporter gene causes resistance to Bt toxin Cry1Ab in the silkworm, *Bombyx mori*. *Proceedings of the National Academy of Sciences of the United States of America* **109**:E1591–E1598.
- Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. *Nucleic Acids Research* **43**:W39–W49.
- Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* **28**:45–48.
- Broehan G, Kroeger T, Lorenzen M, Merzendorfer H. 2013. Functional analysis of the ATP-binding cassette (ABC) transporter gene family of *Tribolium castaneum*. *BMC Genomics* **14**:1–19.
- Cao CW, Zhang J, Gao XW, Liang P, Guo HL. 2008. Overexpression of carboxylesterase gene associated with organophosphorous insecticide resistance in cotton aphids, *Aphis gossypii* (Glover). *Pesticide Biochemistry and Physiology* **90**:175–180.
- Capinera JL. 2002. Fall Armyworm, *Spodoptera frugiperda* (J.E. Smith) (Insecta: Lepidoptera: Noctuidae). *University of Florida* **2002**:1–7.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**:3674–3676.
- Consortium GO. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* **32**:D258–D261.
- Cui F, Li MX, Chang HJ, Mao Y, Zhang HY, Lu L-X, Yan SG, Lang ML, Liu L, Qiao CL. 2015. Carboxylesterase-mediated insecticide resistance: quantitative increase induces broader metabolic resistance than qualitative change. *Pesticide Biochemistry and Physiology* **121**:88–96.
- Dermauw W, Osborne EJ, Clark RM, Grbić M, Tirry L, Van Leeuwen T. 2013. A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*. *BMC Genomics* **14**:1–22.
- Despres L, David JP, Gallet C. 2007. The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution* **22**(6):298–307 DOI [10.1016/j.tree.2007.02.010](https://doi.org/10.1016/j.tree.2007.02.010).
- Devonshire A, Moores G, Ffrench-Constant R. 1986. Detection of insecticide resistance by immunological estimation of carboxylesterase activity in *Myzus persicae* (Sulzer) and cross reaction of the antiserum with *Phorodon humuli* (Schrank)(Hemiptera: Aphididae). *Bulletin of Entomological Research* **76**:97–107.
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F. 2012. Landscape of transcription in human cells. *Nature* **489**:101–108.
- Eklblom R, Galindo J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**:1–15.
- Enayati AA, Ranson H, Hemingway J. 2005. Insect glutathione transferases and insecticide resistance. *Insect Molecular Biology* **14**:3–8.
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J. 2014. Pfam: the protein families database. *Nucleic Acids Research* **42**:D222–D230.
- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**:W29–W37.

- Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A. 2009. The Pfam protein families database. *Nucleic Acids Research* 38(Suppl. 1):D211–D222 DOI 10.1093/nar/gkp985.
- Foissac S, Sasmeth M. 2007. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Research* 35:W297–W299.
- Georghiou GP. 2012. *Pest resistance to pesticides*. Berlin: Springer Science & Business Media.
- Gui F, Lan T, Zhao Y, Guo W, Dong Y, Fang D, Liu H, Li H, Wang H, Hao R, Cheng X, Li Y, Yang P, Sahu SK, Chen Y, Cheng L, He S, Liu P, Fan G, Lu H, Hu G, Dong W, Chen B, Jiang Y, Zhang Y, Xu H, Lin F, Slipper B, Postma A, Jackson M, Abate BA, Tesfaye K, Demie AL, Bayeleygne MD, Degefu DT, Chen F, Kuria PK, Kinyua ZM, Liu TX, Yang H, Huang F, Liu X, Sheng J, Kang L. 2020. Genomic and transcriptomic analysis unveils population evolution and development of pesticide resistance in fall armyworm *Spodoptera frugiperda*. *Protein & Cell* 1(9):21 DOI 10.1007/s13238-020-00795-7.
- Hemingway J. 2000. The molecular basis of two contrasting metabolic mechanisms of insecticide resistance. *Insect Biochemistry and Molecular Biology* 30:1009–1015.
- Huang FF, Chai CL, Zhang Z, Liu ZH, Dai FY, Lu C, Xiang ZH. 2008. The UDP-glucosyltransferase multigene family in *Bombyx mori*. *BMC Genomics* 9(1):563 DOI 10.1186/1471-2164-9-563.
- Ibrahim SS, Riveron JM, Stott R, Irving H, Wondji CS. 2016. The cytochrome P450 CYP6P4 is responsible for the high pyrethroid resistance in knockdown resistance-free *Anopheles arabiensis*. *Insect Biochemistry and Molecular Biology* 68:23–32.
- Jackson CJ, Liu JW, Carr PD, Younus F, Coppin C, Meirelles T, Lethier M, Pandey G, Ollis DL, Russell RJ. 2013. Structure and function of an insect α -carboxylesterase (α Esterase7) associated with insecticide resistance. *Proceedings of the National Academy of Sciences of the United States of America* 110:10177–10182.
- Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28:27–30.
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research* 35:W345–W349.
- Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology* 30(7):693–700 DOI 10.1038/nbt.2280.
- Kostaropoulos I, Papadopoulos AI, Metaxakis A, Boukouvala E, Papadopoulou-Mourkidou E. 2001. Glutathione S-transferase in the defence against pyrethroids in insects. *Insect Biochemistry and Molecular Biology* 31(4–5):313–319 DOI 10.1016/S0965-1748(00)00123-5.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35(6):1547–1549 DOI 10.1093/molbev/msy096.
- Lee SH, Kang JS, Min JS, Yoon KS, Strycharz JP, Johnson R, Mittapalli O, Margam VM, Sun W, Li HM. 2010. Decreased detoxification genes and genome size make the human body louse an efficient model to study xenobiotic metabolism. *Insect Molecular Biology* 19(5):599–615 DOI 10.1111/j.1365-2583.2010.01024.x.
- Letunic I, Bork P. 2017. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research* 46(D1):D493–D496 DOI 10.1093/nar/gkx922.

- Li J, Ma W, Zeng P, Wang J, Geng B, Yang J, Cui Q. 2015. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Briefings in Bioinformatics* **16**:806–812.
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**:1658–1659.
- Li X, Shi H, Gao X, Liang P. 2018. Characterization of UDP-glucuronosyltransferase genes and their possible roles in multi-insecticide resistance in *Plutella xylostella* (L.). *Pest Management Science* **74**:695–704.
- Liu S, Zhou S, Tian L, Guo E, Luan Y, Zhang J, Li S. 2011. Genome-wide identification and characterization of ATP-binding cassette transporters in the silkworm, *Bombyx mori*. *BMC Genomics* **12**:1–15.
- Merzendorfer H. 2014. ABC transporters and their role in protecting insects from pesticides and their metabolites. *Advances in Insect Physiology* **46**:1–72.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**(5881):1344–1349 DOI [10.1126/science.1158441](https://doi.org/10.1126/science.1158441).
- Pavlidis N, Vontas J, Van Leeuwen T. 2018. The role of glutathione S-transferases (GSTs) in insecticide resistance in crop pests and disease vectors. *Current Opinion in Insect Science* **27**(Pt. 3):97–102 DOI [10.1016/j.cois.2018.04.007](https://doi.org/10.1016/j.cois.2018.04.007).
- Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods* **10**(11):1096–1098 DOI [10.1038/nmeth.2639](https://doi.org/10.1038/nmeth.2639).
- Prasanna B, Huesing J, Eddy R, Peschke V. 2018. *Fall armyworm in Africa: a guide for integrated pest management*. Wallingford: CAB International, 1–109.
- Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **35**:D61–D65 DOI [10.1093/nar/gkl842](https://doi.org/10.1093/nar/gkl842).
- Rane RV, Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. 2019. Detoxifying enzyme complements and host use phenotypes in 160 insect species. *Current Opinion in Insect Science* **31**:131–138 DOI [10.1016/j.cois.2018.12.008](https://doi.org/10.1016/j.cois.2018.12.008).
- Roth CW, Holm I, Graille M, Dehoux P, Rzhetsky A, Wincker P, Weissenbach J, Brey PT. 2003. Identification of the *Anopheles gambiae* ATP-binding cassette transporter superfamily genes. *Molecules and Cells* **15**:150–158.
- Scott JG. 1999. Cytochromes P450 and insecticide resistance. *Insect Biochemistry and Molecular Biology* **29**:757–777.
- Shen GM, Ou SY, He C, Liu J, He L. 2021. Full length sequencing reveals novel transcripts of detoxification genes along with related alternative splicing events and lncRNAs in *Phyllotreta striolata*. *PLOS ONE* **16**(3):e0248749 DOI [10.1371/journal.pone.0248749](https://doi.org/10.1371/journal.pone.0248749).
- Shi L, Shi Y, Zhang Y, Liao X. 2019. A systemic study of indoxacarb resistance in *Spodoptera litura* revealed complex expression profiles and regulatory mechanism. *Scientific Reports* **9**(1):14997 DOI [10.1038/s41598-019-51234-5](https://doi.org/10.1038/s41598-019-51234-5).
- Shi Y, Wang H, Liu Z, Wu S, Yang Y, Feyereisen R, Heckel DG, Wu Y. 2018. Phylogenetic and functional characterization of ten P450 genes from the CYP6AE subfamily of *Helicoverpa armigera* involved in xenobiotic metabolism. *Insect Biochemistry and Molecular Biology* **93**:79–91.
- Simmons J, D'Souza O, Rheault M, Donly C. 2013. Multidrug resistance protein gene expression in *Trichoplusia ni* caterpillars. *Insect Molecular Biology* **22**:62–71.

- Stevens JL, Snyder MJ, Koener JF, Feyereisen R. 2000. Inducible P450s of the CYP9 family from larval *Manduca sexta* midgut. *Insect Biochemistry and Molecular Biology* **30**:559–568.
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research* **41**:e166.
- Sun XX, Hu CX, Jia HR, Wu QL, Shen XJ, Zhao SY, Jiang YY, Wu KM. 2021. Case study on the first immigration of fall armyworm, *Spodoptera frugiperda* invading into China. *Journal of Integrative Agriculture* **20**:664–672.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28**:33–36.
- Tendeng E, Labou B, Diatte M, Djiba S, Diarra K. 2019. The fall armyworm *Spodoptera frugiperda* (J.E. Smith), a new pest of maize in Africa: biology and first native natural enemies detected. *International Journal of Biological and Chemical Sciences* **13**(2) DOI 10.4314/ijbcs.v13i2.35.
- Wan J, Huang C, Li C-Y, Zhou H-X, Ren Y-L, Li Z-Y, Xing L-S, Z B, Qiao X, Liu B, Liu C-H, Xi Y, Liu W-X, Wang W-K, Qian W-Q, McKirdy S, Wan F-H. 2020. Biology, invasion and management of the agricultural invader: Fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Journal of Integrative Agriculture* **19**(3):2–19 DOI 10.1016/S2095-3119(20)63367-6.
- Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Molecular Cell* **43**(6):904–914 DOI 10.1016/j.molcel.2011.08.018.
- Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. 2013. CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Research* **41**:e74.
- Wheelock CE, Shan G, Ottea J. 2005. Overview of carboxylesterases and their role in the metabolism of insecticides. *Journal of Pesticide Science* **30**:75–83.
- Wu Q, Jiang Y, Wu K. 2019. Analysis of migration routes of the fall armyworm *Spodoptera frugiperda* (J.E. Smith) from Myanmar to China. *Plant Protection* **45**:1–6.
- Xiao H, Ye X, Xu H, Mei Y, Yang Y, Chen X, Yang Y, Liu T, Yu Y, Yang W. 2020. The genetic adaptations of fall armyworm *Spodoptera frugiperda* facilitated its rapid global dispersal and invasion. *Molecular Ecology Resources* **20**:1050–1068.
- Xu L, Duan X, Lv Y, Zhang X, Nie Z, Xie C, Ni Z, Liang R. 2014. Silencing of an aphid carboxylesterase gene by use of plant-mediated RNAi impairs *Sitobion avenae* tolerance of Phoxim insecticides. *Transgenic Research* **23**:389–396.
- Yang X, Deng S, Wei X, Yang J, Zhao Q, Yin C, Du T, Guo Z, Xia J, Yang Z. 2020. MAPK-directed activation of the whitefly transcription factor CREB leads to P450-mediated imidacloprid resistance. *Proceedings of the National Academy of Sciences of the United States of America* **117**:10246–10253.
- Yang Z, Zhang F, He Q, He G. 2005. Molecular dynamics of detoxification and toxin-tolerance genes in brown planthopper (*Nilaparvata lugens* Stål., Homoptera: Delphacidae) feeding on resistant rice plants. *Archives of Insect Biochemistry and Physiology* **59**:59–66.
- Yin C, Liu Y, Liu J, Xiao H, Huang S, Lin Y, Han Z, Li F. 2014. ChiloDB: a genomic and transcriptome database for an important rice insect pest *Chilo suppressalis*. *Database* **2014**:1–7.
- Yin C, Shen G, Guo D, Wang S, Ma X, Xiao H, Liu J, Zhang Z, Liu Y, Zhang Y. 2016. InsectBase: a resource for insect genomes and transcriptomes. *Nucleic Acids Research* **44**:D801–D807.
- Zhang N, Liu J, Chen S-N, Huang L-H, Feng Q-L, Zheng S-C. 2016. Expression profiles of glutathione S-transferase superfamily in *Spodoptera litura* tolerated to sublethal doses of chlorpyrifos. *Insect Science* **23**(5):675–687 DOI 10.1111/1744-7917.12202.

- Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J. 2010.** The next-generation sequencing technology and application. *Protein & Cell* **1**:520–536.
- Zhou Y, Wu Q-L, Zhang H-W, Wu K-M. 2021.** Spread of invasive migratory pest *Spodoptera frugiperda* and management practices throughout China. *Journal of Integrative Agriculture* **20**(3):2–10 DOI [10.1016/S2095-3119\(21\)63621-3](https://doi.org/10.1016/S2095-3119(21)63621-3).
- Zhu F, Moural TW, Shah K, Palli SR. 2013.** Integrated analysis of cytochrome P450 gene superfamily in the red flour beetle, *Tribolium castaneum*. *BMC Genomics* **14**:1–12.