Cloning, phylogenetic research, and prokaryotic expression study of the metabolic detoxification gene EoGSTs1 in Empoasca onukii Matsuda

Yujie Zhang 1, 2, Wenlong Chen 1, 2, Ming Li 3, Lin Yang 1, 2, Xiangsheng Chen

1 Institute of Entomology, Guizhou University, Guiyang, China
2 Guizhou Key Laboratory for Plant Pest Management of Mountainous Region, Guizhou University, Guiyang, China
3 The Provincial Special Key Laboratory for Development and Utilization of Insect Resources of Guizhou, Guiyang, China

Corresponding Author: Xiangsheng Chen
Email address: chenxs3218@163.com

Due to the misuse of chemical pesticides, small green leafhoppers (Empoasca onukii Matsuda) have developed resistance to pesticides, thereby posing a serious problem to the tea industry. Glutathione S-transferases (GSTs) are an important family of enzymes that are involved in pesticide resistance in Empoasca onukii Matsuda. Empoasca onukii GST sigma 1 (EoGSTs1, GenBank Accession number MK443501) is a member of the GST family. In this study, the full-length cDNA of EoGSTs1 was cloned by reverse transcription polymerase chain reaction (qPCR), and its taxonomic identity was examined. Furthermore, we performed bioinformatics and phylogenetic analyses of the gene and structural and functional domain prediction of the protein. The results demonstrate that EoGSTS1 belongs to the Sigma family of GSTs; the full-length EoGSTs1 cDNA is 841 bp with a 624-bp coding region that encodes a 23.68932-kDa protein containing 207 amino acids. The theoretical isoelectric point (IEP) was calculated to be 6.00. Phylogenetic analysis indicates that EoGSTS1 is closely related to the Sub psaltriayangi subfamily of the Cicadoidea superfamily in order Hemiptera, whereas it is distantly related to Periplaneta americana of order Blattodea. Amino acid sequence alignment of EoGSTS1 and GSTs from four other insects of order Hemiptera revealed protein sequence conservation. Tertiary structure analysis and structural domain functional predictions of the protein revealed that EoGSTS1 contains nine α helices and two β sheets with one conserved GST domain. The results of enzyme activity assay showed that recombinant EoGSTs1 (rEoGSTs1) protein had catalytic activity for substrate 1-chloro-2,4-dinitrobenzene (CDNB) and exhibited the highest activity at pH 7 and 25°C. The Michaelis constant Km of rEoGSTs1 protein was 0.07782 ± 0.01990 mmol/L, and the maximum reaction rate Vmax was 12.15 ± 1.673 μmol/min·mg. Our study clarified the taxonomic identity of small green leafhopper EoGSTs1 and revealed some properties of the gene and its encoded protein sequence. According to the catalytic
activity of the rEoGSTs1 enzyme on the model substrate CDNB, we infer that it functions in the degradation of exogenous substances.
Cloning, phylogenetic research, and prokaryotic expression study of the metabolic detoxification gene \textit{EoGSTs1} in \textit{Empoasca onukii} Matsuda

Yujie Zhang$^{1,2}$, Wenlong Chen$^{1,2}$, Ming Li$^3$, Lin Yang$^{1,2}$, Xiangsheng Chen$^{1,2*}$

1 Institute of Entomology, Guizhou University, Guiyang, China
2 Guizhou Key Laboratory for Plant Pest Management of Mountainous Region, Guizhou University, Guiyang, China
3 The Provincial Special Key Laboratory for Development and Utilization of Insect Resources of Guizhou, Guizhou University, Guiyang, China

Corresponding Author: Xiangsheng Chen$^{1,2}$
Guizhou University, Guiyang, Guizhou, 550025, China
Email address: chenxs3218@163.cm

ABSTRACT

Due to the misuse of chemical pesticides, small green leafhoppers (\textit{Empoasca onukii} Matsuda) have developed resistance to pesticides, thereby posing a serious problem to the tea industry. Glutathione S-transferases (GSTs) are an important family of enzymes that are involved in pesticide resistance in \textit{Empoasca onukii} Matsuda. \textit{Empoasca onukii} GST sigma 1 (\textit{EoGSTs1}, GenBank Accession number MK443501) is a member of the GST family. In this study, the full-length cDNA of \textit{EoGSTs1} was cloned by reverse transcription polymerase chain reaction (qPCR), and its taxonomic identity was examined. Furthermore, we performed bioinformatics and phylogenetic analyses of the gene and structural and functional domain prediction of the protein. The results demonstrate that \textit{EoGSTS1} belongs to the Sigma family of GSTs; the full-length \textit{EoGSTs1} cDNA is 841 bp with a 624-bp coding region that encodes a 23.68932-kDa
protein containing 207 amino acids. The theoretical isoelectric point (IEP) was calculated to be
6.00. Phylogenetic analysis indicates that EoGSTS1 is closely related to the *Sub psaltriayangi*
subfamily of the *Cicadoidea* superfamily in order *Hemiptera*, whereas it is distantly related to
*Periplaneta americana* of order *Blattodea*. Amino acid sequence alignment of EoGSTS1 and
GSTs from four other insects of order *Hemiptera* revealed protein sequence conservation.

Tertiary structure analysis and structural domain functional predictions of the protein revealed
that EoGSTS1 contains nine α helices and two β sheets with one conserved GST domain. The
results of enzyme activity assay showed that recombinant EoGSTS1 (rEoGSTS1) protein had
catalytic activity for substrate 1-chloro-2,4-dinitrobenzene (CDNB) and exhibited the highest
activity at pH 7 and 25°C. The Michaelis constant Km of rEoGSTS1 protein was 0.07782 ±
0.01990 mmol/L, and the maximum reaction rate Vmax was 12.15 ± 1.673 μmol/min·mg. Our
study clarified the taxonomic identity of small green leafhopper *EoGSTS1* and revealed some
properties of the gene and its encoded protein sequence. According to the catalytic activity of the
rEoGSTS1 enzyme on the model substrate CDNB, we infer that it functions in the degradation of
exogenous substances.

INTRODUCTION

Small green leafhoppers or tea green leafhoppers, *Empoasca onukii* Matsuda, belong to the
*Cicadidae* family in genus *Jacobiasca formosana*. Small green leafhoppers are one of the most
predominant insects that infiltrate tea production fields. Small green leafhoppers are extremely
harmful to tea trees, causing a 15%–50% loss of production of tea each year in mainland China
and Taiwan (*Niu, et al., 2015; Hazarika, Bhuyan & Hazarika, 2009; Fu, Han & Xiao, 2014;
Saha & Mukhopadhyay, 2013; Qin et al., 2015; Wang, 2004; Zhang &Chen, 2015; Mu et al.,
To date, the pest control and prevention of small green leafhoppers still mainly rely on chemicals. Due to long-term misuse and abuse of chemical compounds, insects have developed resistance to most chemical pesticides (Zhan et al., 2012). For instance, Wang et al. (2004) performed drug membrane interaction tests on the resistance of small green leafhoppers to 14 chemical pesticides and found that their LC50 values and resistance gradually increased with the higher pesticide doses used in tea tree facilities. In particular, the resistance to thiamethoxam increased by 2.2- to 13.3-fold, which is the largest increase among all the chemical pesticides.

Glutathione S-transferases (GSTs) are a family of enzymes that widely exist in aerobic organisms and play an important role in detoxifying insects from endogenous or exogenous toxins (Enayati, Ranson & Hemingway, 2005). For example, Wu et al. (2016) used cycloxaprid as a toxin to induce the activities of detoxification enzymes in Alfalfa aphid and found that the activities of GSTs in this species that survived treatment were significantly higher than those in control insects. Thus, they suggested that GSTs play a critical role in detoxifying cycloxaprid in Alfalfa aphid. Zhang et al. (2014) studied the levels of pest resistance to the neonicotinoid class of insecticides and the synergistic effect of enzyme inhibitors in Aphis gossypii in different regions of Shandong Province, China. They discovered a noticeable synergistic effect on this class of insecticides by the enzyme inhibitor diethyl maleate (DEM). Cytoplasmic GSTs in insects are divided into six subfamilies, which include Epsilon, Delta, Zeta, Omega, Sigma, and Theta (Chelvanayagam, Parker & Board, 2001). To date, studies on resistance to pesticides related to GSTs in insects have focused on the Epsilon and Delta subfamilies, whereas investigations on the involvement of other subfamilies of GSTs are limited (Yu et al., 2010). In recent years, with the rapid development of techniques in molecular biotechnology, studies on...
other GST subfamilies related to pesticide resistance have grown in number, although studies on
GST-related pesticide resistance in *E. onukii* Matsuda remain scarce.

Pesticide resistance in *E. onukii* can further increase in the foreseeable future because of
current continuous misuse of chemical pesticides. Reducing or slowing down this process of
increasing pesticide resistance has thus become a major goal. Research on genes involved in the
resistance of small green leafhoppers can provide critical and fundamental knowledge for the
development of new pesticides, recipes for pesticide cocktails, or protocols for alternating the
usage of different chemicals (*Zhou & Han, 2017*). GSTs are the major enzymes involved in
metabolic resistance and have been proven essential in the detoxification of small green
leafhoppers that are resistant to chlorpyrifos and thiamethoxam. In this study, we screened
transcriptome data and identified genes that are differentially expressed in *E. onukii* upon
chemical treatment and discovered that the expression level of the GST cluster-166.0 was
significantly upregulated. Therefore, we cloned GST cluster-166.0 using reverse transcription
polymerase chain reaction (qPCR) with proofreading. Bioinformatics analysis, phylogenetic
reconstruction, and homology modeling of the full-length cDNA and its encoded protein
sequence were conducted to obtain biological information on the *E. onukii* GST gene (*EoGSTs1,
cluster-166.0*), including its physical and chemical properties, taxonomic classification, and the
spatial structure of the EoGSTS1 protein. In addition, the expression vector of the EoGSTS1
protein was constructed for the expression in prokaryotic cells, and the expression of fusion
protein was detected by western blotting. Finally, the enzyme activity of purified EoGSTS1
protein was determined, and its enzymatic characteristics were verified. This information may
serve as the foundation for future applications of *EoGSTs1*, may be used as a theoretical
reference for the detection and comprehensive management of chemical resistance in small green leafhoppers, as well as employed in the development and utilization of new pesticides.

**MATERIALS AND METHODS**

**Preparation of insect samples**

*E. onukii* Matsuda used in this study is a sensitive strain that has been cultured in the laboratory for many generations. Insects were maintained in a light incubator with a constant temperature of 25°C ± 1°C, a humidity of 70% ± 5%, and a photoperiod of 16:8 (L:D). Small green leafhoppers of approximately the same size were starved for 1 h and then treated with 2.5 μg/mL of thiamethoxam for 48 h. Individuals who survived the treatment were collected at 20 insects/tube, flash frozen in liquid nitrogen, and stored in a -80°C freezer.

**Extraction and examination of total RNA**

Total RNA was extracted using the TRIzol® PlusRNA Purification Kit (Cat. No. 12183-555, Invitrogen, Carlsbad, CA, USA) following the manufacturer’s instructions. The extracted RNAs were examined by NanoDrop quantification and electrophoresis. Three microliters of each extracted RNA sample were used in agarose gel (1%) electrophoresis.

**Primer design and template synthesis**

Primers were designed using Primer Premier 6.0 (Premier Biosoft International, Palo Alto, CA, USA). Specific primer sequences are listed in Table 1. The 5′ rapid amplification of cDNA ends (RACE) and 3′ RACE templates were synthesized using a GeneRacer™ Kit (Cat. No. 12183-555, Invitrogen, Carlsbad, CA, USA). The cDNA template used for intermediate fragment PCR was synthesized using the SuperScript™ III First-Strand Synthesis SuperMix (Cat. No. 18080-400, Invitrogen, Carlsbad, CA, USA) for qPCR Kit.
PCR: The reaction system and conditions for 5’ RACE PCR, intermediate fragment PCR, and 3’ RACE PCR are presented in Tables S1–S10.

**Prokaryotic expression and purification**

Full-length splicing primers were designed based on PCR-based accurate synthesis (PAS).

Protective bases were designed at both ends of the primers to synthesize the EoGSTs1-TARGET gene, which connected the site between the NdeI and XbaI sites of vector PCZN1. The obtained recombinant plasmid PCZN1-EoGSTs1-TARGET was transferred into *Escherichia coli* strain Top10, and positive clones were selected for sequencing and restriction enzyme digestion to ensure the correct expression plasmid was obtained through subcloning.

Approximately 1 μL of the constructed expression plasmid was added to 100 μL of Arctic Express *E. coli*, placed on the ice for 20 min, heat-shocked for 90 s at 42°C, placed on ice for 5 min, added 600 mL of LB culture medium, and placed on a shaker for 1 h at the frequency of 220 rpm and 37°C. After centrifugation, all the samples were coated onto LB plates that contained 50 μg/mL Amp (ampicillin), and incubated overnight at 37°C. Monoclones were selected and inoculated into 3 mL of LB medium containing 50 μL/mL, placed on a shaker overnight at the frequency of 220 rpm and 37°C, inoculated into 30 mL of LB medium (containing 50 μL/mL AMP) at a ratio of 1:100 the next day, and placed on a shaker at a frequency of 220 rpm and 37°C when the OD600 value of the culture reached 0.6–0.8. Then, 1 mL of the culture was collected and centrifuged for 2 min at 10,000 rpm at room temperature. The supernatant discarded, and the pellet was resuspended in 100 μL 1× sample loading buffer. IPTG was added to the remaining culture until the final concentration was 0.5 mM and placed on a shaker overnight at a frequency of 220 rpm and 11°C to induce the expression of the fusion protein. Then, 1 mL of the culture was collected and centrifuged for 2 min at 10,000 rpm at room
temperature, the supernatant was discarded, and the pellet was resuspended in 100 μL of 1× sample loading buffer. The remaining culture was centrifuged for 10 min at 4000 rpm, the supernatant was discarded, and the pellet was resuspended in phosphate-buffered saline (PBS). The cells in suspension were then lysed by sonication, the supernatant and pellet were resuspended in sample loading buffer, detected by 12% SDS-PAGE, and stained with Coomassie brilliant blue.

The results of the preliminary experiment showed that the target protein was soluble. Therefore, the supernatant was purified by Ni affinity chromatography. The supernatant solution (flow rate: 0.5 mL/min) was loaded onto the Ni-IDA-Sepharose Cl-6B affinity chromatography column (pre-equilibrated with Ni-IDA binding buffer) with a low-pressure chromatography system and washed with Ni-IDA binding buffer at the flow rate of 0.5 mL/min until the OD280 value of the effluent reached the baseline. The column was washed with Ni-IDA washing buffer (20 mM Tris-HCl, 20 mM imidazole, 0.15 M NaCl, pH 8.0) at the flow rate of 1 mL/min until the OD280 value of the effluent reached the baseline. Then, the target protein was eluted with Ni-IDA elution buffer (20 mM Tris-HCl, 250 mM imidazole, 0.15 M NaCl, pH 8.0) at a flow rate of 1 mL/min, and effluent was collected. The above-mentioned protein solution was poured into a dialysis bag, and PBS (3.58 g Na₂HPO₄·12H₂O, 8.8 g NaCl, 0.2 g KCl, 0.27 g KH₂PO₄) was used for overnight dialysis, then detected by 12% SDS-PAGE.

**Western blot analysis**

A 3-μL sample was loaded into a well of a polyacrylamide gel. After running the stacking gel at 90 V and then increased to 200 V until the end of electrophoresis, conducted PVDF transmembrane (constant pressure 100V, 1.5 h, constant current 250 mA), the membrane was washed four times (5 min each time) with phosphate-buffered saline-Tween 20 (PBST) and then
kept in the blocking buffer containing 5% non-fat milk for 1 h at 37°C. The membrane was incubated overnight with a term primary antibody that was diluted in the blocking buffer (dilution rate, 1:1000). The next day, the membrane was washed four times (5 min each time) with PBST. The membrane was then incubated with a secondary HRP antibody that was diluted with the blocking buffer with 5% milk (dilution rate, 1:5000) for 1 h at 37°C. Then, the membrane was washed four times (5 min each time) in a clean box, developed, and exposed by the ECL (electrochemiluminescence) method.

**Enzyme activity determination**

Parameters of enzymatic kinetic: The determination method was based on Li et al. (2018) with minor modifications. The reaction system consisted of phosphate buffer (100 mmol/L), rEoGSTs1 protein (0.12 mg), and GSH (0.01, 0.02, 0.04, 0.06, 0.08, and 0.1 mmol/L). After mixing the above solution, the mixture was incubated with 1-chloro-2,4-dinitrobenzene (CDNB) solution (10 mmol/L) at 35°C for 15 min, then 20 μL of the CDNB solution was added into the mixture, and the volume was set to 200 μL. Subsequently, the absorbance was measured at a wavelength of 340 nm in the Multiskan GO 1510 spectrophotometer (Thermo Fisher Scientific, Vantaa, Finland), 30 s for each time and for a total of 2 min. Enzyme activity was calculated using the recombinant-inactivating protein as control. The specific activity was calculated according to the following formula, and the obtained data were submitted to GraphPad Prism 8.0.1 to calculate for enzyme kinetics using the following equation:

$$\text{Specific activity (μmol/min·mg)} = \frac{\Delta OD_{340} \times V}{ε \times T \times L \times E}.$$  

Note: OD340: Change in absorbance per unit time; V: enzymatic reaction volume (200 mL); ε: extinction coefficient of product (9.6 mmol/cm); T: reaction time (2 min); L: optical path (1 cm); and E: amount of enzyme added (0.12 mg).
To determine the effect of temperature on the activity of the purified rEoGSTs1 protein, Glutathione S-transferase (GSH-ST) assay kit (Colorimetric method) was used (Cat. No. A004, Nanjing Jiancheng Bioengineering Institute, Nanjing, Jiangsu, China), following the manufacturer’s protocol. Approximately 0.12 mg of the rEoGSTs1 protein was added into the reaction system and then was incubated at 5°C, 15°C, 25°C, 35°C, 45°C, and 60°C for 15 min.

To determine the effect of pH on the activity of the purified rEoGSTs1 protein, Glutathione S-transferase (GSH-ST) assay kit (Colorimetric method) (Cat. No. A004, Nanjing Jiancheng Bioengineering Institute, Nanjing, Jiangsu, China) was used, following the manufacturer’s protocol. The pH of the reaction system was adjusted with PBS to 5.0, 6.0, 7.0, 8.0, and 9.0.

**Data processing**

The *EoGSTs1* full-length cDNA was assembled from the DNA fragments using Seqman 7.1.0 software (DNASTar, Madison, WI, USA). Homology analysis of EoGSTS1 protein sequences in the NCBI database was performed using BLAST (*Johnson et al., 2008*). The physical and chemical properties of the EoGSTS1 protein were analyzed using software tools provided in the ExPASy-ProParam website (http://www.expasy.org), and cluster analysis was performed using MEGA 6.0 (*Tamura et al., 2013*). Clustal Omega (*Sievers et al., 2011*) was used in amino acid sequence alignment of EoGSTS1 and the GSTs of four other insects of order Hemiptera. The secondary structure of EoGSTS1 was predicted by CLC Genomics Workbench 11.0.1 (CLC Bio, Aarhus, Denmark), whereas the three-dimensional (3D) structure was predicted using I-TASSER software (*Roy, Kucukural & Zhang, 2010*). The functional domains were predicted based on the amino acid sequence deduced from the nucleotide sequence of the *EoGSTs1* gene by searching the conserved domain database (CDD) in the NCBI database (*Marchler-Bauer et al., 2006*). Phylogenetic reconstruction of the GSTs was performed using the maximum likelihood (ML)
method and Bayesian inference (BI) with the assistance of IQtree 1.6.1 (Nguyen et al., 2015) and MrBayes 3.2 (Ronquist et al., 2012).

RESULTS

Detection results of total RNA

Both the 28S RNA and 18S RNA bands were observed as distinct bands (Fig. 1), suggesting that the extracted RNAs were of good quality. The OD260/280 ratio of the RNA was 1.92, indicating that the collected RNA was of high purity with no degradation.

Amplification of the full-length cDNA

The 5’ end of the EoGSTs1 cDNA fragment obtained by 5’ RACE was 223 bp in length. The middle fragment of the EoGSTs1 amplified by intermediate fragment PCR was 480 bp in size. The 3’ end of the EoGSTs1 cDNA obtained by 3’ RACE was 431 bp in length (Fig. 2A–C). The full-length cDNA of EoGSTs1 was an 841-bp sequence obtained through sequence assembly of the three DNA fragments; its coding region is 624-bp long, encoding a total of 207 amino acids, including the start codon ATG and the stop codon TGA (Dataset S1). The noncoding region at the 5’ end is 99 bp in length, and the 3’ end noncoding region is 118 bp in length followed by a 16-bp poly(A) tail (Fig. 3).

Taxonomic identification of EoGST

The amino acid sequence of EoGST is highly homologous to GSTs in Subpsaltriayangi (AVC68800.1), Locusta migratoria (AHC08043.1), Daktulosphaira vitifoliae (AUN35388.1), and Leptinotarsa decemlineata (APX61045.1), with a similarity of 53%, 52%, 52%, and 50%, respectively.
The amino acid sequence of *EoGST* obtained in small green leafhoppers is shown in Table 2 along with 54 other GST sequences belonging to six GST families. Cluster analysis using the adjacency method demonstrated that the *EoGST* gene is clustered with the Sigma subfamily of GSTs (a high bootstrap value of 99%), indicating that *EoGST* is closely related to this particular subfamily (Fig. 4). However, it did not cluster with GSTs in the Delta or Epsilon subfamilies, indicating that it is distantly related to these two subfamilies. Based on these findings and mammalian GST gene nomenclature, the nucleotide sequence of the *E. onukii* GST (*EoGST, cluster-166.0*) was initially named as *EoGSTs1*, and the encoded protein sequence was designated as EoGSTS1.

**Bioinformatics analysis**

**Physical and chemical properties of the EoGSTS1 protein**

The molecular weight of EoGSTS1 is 23.68932 kDa, with a calculated isoelectric point (IEP) of 6.00. The number of negatively and positively charged residues is 30 and 29, respectively. Its N-terminus begins with methionine, and the half-life of the protein is 30 h. The instability index of the protein in solution is 31.87, the fat coefficient is 78.26, and the total average hydrophobicity index is -0.376. The protein contains a total of 3331 atoms, including 1078 carbon atoms, 1662 hydrogen atoms, 276 nitrogen atoms, 305 oxygen atoms, and 10 sulfur atoms. It consists of 20 different types of amino acids, among which the top three are alanine (Ala, A), lysine (Lys, K), and glutamic acid (Glu, E), with respective counts of 19 (9.2%), 19 (9.2%), and 16 (7.7%). The rarely occurring amino acids included Cys (C) and His (H), with counts of 2 (1%) and 1 (0.5%), respectively (Table 3).

**EoGSTS1 protein sequence alignment**
The catalytic domain of the protein contains 69 amino acid residues, as revealed in the SMART software analysis (Letunic et al., 2014). Comparison of the protein sequences of EoGSTS1 and the GSTs of four other insects of order Hemiptera using Clustal Omega (Fig. 5) revealed the existence of multiple conserved domains.

Analysis of the EoGSTS1 protein structure

Based on its protein sequence, neither signal peptide nor transmembrane helix domains are present in the protein, but four potentially linear antigenic epitopes may exist in the protein sequence (amino acids 31–51, 81–88, 112–123, and 169–180).

The EoGSTS1 protein contains nine α helices, two β sheets with a protein kinase C phosphorylation site, and one N-glycosylation site (Fig. 6). The protein belongs to the superfamily of GSTs and shares sequence similarities with GSTs in other insects (Fig. 7).

The protein sequence of EoGSTS1 contains two Asn-X-Ser/Thr motifs, one of which is an N-glycosylation site. The protein sequence also contains eight other potential phosphorylation sites: four serine phosphorylation sites located at positions 95, 118, 152, and 166 in the peptide chain; two threonine phosphorylation sites located at positions 47 and 125 in the peptide chain; and two tyrosine phosphorylation sites located at positions 148 and 188 in the peptide chain (Table 4).

The 3D structure of the EoGSTS1 protein was predicted using the Blattella germanica GST (BgGST) protein (PDB ID: 4Q5R) as reference (as shown in Fig. 8, the C-terminus is denoted in red and the N-terminus in blue). The C-terminus is composed of 207 amino acids, and the N-terminus contains 204 amino acids. The C-score of the protein structural comparison is 1.21, indicating that it is highly similar to 4Q5R in terms of protein folding and secondary structure. The TM-score, which represents the structural similarity between the target sequence (EoGSTS1) and the template protein sequence (BgGST), is 0.88 ± 0.07.
Phylogenetic analysis

The phylogenetic tree of the full-length EoGSTS1 protein sequence and GST sequences of insects from five other orders, including Hemiptera, Orthoptera, Coleoptera, Odonata, and Hymenoptera (Table 5), were constructed using BI phylogenetic analysis and an ML-based method. The topological structures of the phylogenetic trees generated using either ML or BI displayed substantial similarities (Fig. 9). Cluster analysis indicated that the EoGSTS1 is clustered with known insect GSTs of order Hemiptera but not with GSTs from insects of order Odonata or Hymenoptera. These results suggest that E. onukii Matsuda is most closely related to insects of order Hemiptera and relatively closely related to those in order Orthoptera but distantly related to insects of orders Hymenoptera, Coleoptera, and Odonata.

Expression of the rEoGSTs1 protein

The recombinant expression plasmid pCzn1-EoGSTs1-target was transformed into E. coli. Arctic Express and the pCzn1/EoGSTs1 fusion protein was successfully expressed with the induction of IPTG. SDS-PAGE analysis showed that rEoGSTs1 was expressed in the supernatant and mainly occurred in a soluble form (Fig. 10A). Western blotting revealed a specific band in the corresponding position, which indicated that this band is the target band (Fig. 10B). The concentration of the target protein that was determined by the BCA method was 1.2 mg/mL.

Enzymatic analysis of EoGSTS1

The kinetic data of EoGSTS1 showed that the rEoGSTs1 protein could catalyze CDNB with Km of $0.21 \pm 0.06 \text{ mmol/L}$ and $V_{\text{max}}$ of $14.02 \pm 1.40 \mu\text{mol/min\cdot mg}$ (Fig. 11A). At different temperatures, the specific activity of EoGSTS1 initially gradually increased, and then decreased after reaching the highest value at $25^\circ\text{C}$, and reached the lowest value at $60^\circ\text{C}$. The lowest enzyme activity was about 4.13% of the highest one (Fig. 11B). At different pH levels, the
enzyme activity of EoGSTS1 initially increased and then decreased with increasing pH, peaked at pH 7, and at pH 5, the enzyme activity was 73.67% of that at pH 7 (Fig. 11C).

**DISCUSSION**

GSTs represent a superfamily of genes that are widely present in many organisms and play critical roles in the resistance to foreign substances in insects. In this study, we used transcriptome analysis to screen overexpressed genes induced by thiamethoxam at the early stage of induction. Based on the results, the cDNA of *EoGSTs1* cluster-166.0 was selected, and its full-length cDNA was amplified by qPCR. The end products showed that the full-length cDNA sequence of *EoGSTs1* was 841 bp, with a 624-bp coding region that is predicted to generate a protein consisting of 207 amino acids.

The results from the homology search and cluster analysis suggest that the EoGSTS1 protein is highly similar to *Sub psaltriayangi* (AVC68800.1) GST and *Locusta migratoria* GST (AHC08043.1) and clustered with the GSTs in the Sigma subfamily with stronger support (bootstrap value>90%). Thus, it is believed to belong to the Sigma GST gene subfamily, which coincides with previous findings that the Sigma subfamily of GSTs is involved in chemical resistance in other insects. For example, Zhou *et al.* (2012) reported that the upregulation of *LsGSTs2* gene expression (Sigma subfamily) can be induced by cypermethrin, chlorpyrifos, and fipronil in *Laodelphax striatellus*. The results from the gene expression study of GST by You *et al.* showed that *PxGSTs1* (another gene in the Sigma subfamily) was significantly upregulated in the resistant strain of *Plutella xylostella* (You *et al.*, 2015). In the present study, the expression of *EoGSTs1* in the chemically induced group of small green leafhoppers was 6.6538 times that of
the untreated group; therefore, we believe that EoGSTs1 plays an important role in resistance to thiamethoxam in small green leafhoppers.

The instability index of the EoGSTS1 protein in solution is below the threshold of 40, and the average hydrophobicity is a negative value, indicating that the protein is a stable hydrophilic protein (Nie, Wu & Zhang, 2006; Wanyonyi et al., 2011). Using SignalP 4.1 and subcellular localization analysis, we found that the protein lacks sequences that direct subcellular localization such as signal peptides, lysosomes, peroxidase, or mitochondria. There are no transmembrane domains in EoGSTS1. Therefore, we conclude that EoGSTS1 is a nontransmembrane GST protein located in the cytoplasm with a single functional structure (Li et al., 2018).

Functional domain prediction based on the protein structure of EoGSTS1 showed that it contains a GST conserved domain and belongs to the protein superfamily. It consists of an N-terminal domain starting with a methionine and a C-terminal domain with multiple α helices and β sheets dispersing throughout the structure. These motifs play an important role in linking polypeptides together and forming the protein structure (Sun, 1994). Multiple sites for potential phosphorylation also exist in the protein sequence, which can be divided into three types, with serine phosphorylation as the major one. Phosphorylation site predictions can provide insights into the important biological processes occurring in E. onukii such as cell growth and development, signal transduction, and gene expression. Proteins participate in various essential biological activities in organisms, including signal transduction, enzyme catalysis, and cellular transportation (Wu, 2010). Therefore, the 3D structural information of a protein can be very valuable for its related biological or medical studies such as configuring the structural and functional composition of the protein and designing protein-targeted drug molecules (Kuhlman et al., 2018).
In the present study, we used the threading method to predict the 3D structure of EoGSTS1. The results demonstrated that EoGSTS1 in 3D is highly similar to that of the 4Q5R protein from the German cockroach *Blattella germanica*. The C-score of the compared proteins is $>-5$ and $<2$, and the TM-score is $>0.5$ (Yang & Zhang, 2015; Xu & Zhang, 2010), indicating that the prediction model is sound with correct topology and high reliability in terms of quality assessment.

Using BLAST alignment, we established that the amino acid sequence encoded by *EoGSTs1* is highly similar to that of *Sub psaltriayangi*, but its similarities with insects belonging to orders *Orthoptera, Coleoptera, Odonata,* and *Hymenoptera* did not show notable differences, suggesting that GSTs are not highly conserved across different orders. The EoGSTS1 clusters with a known insect GST in order *Hemiptera* in ML-based and BI phylogenetic trees; however, within that gene cluster, EoGSTS1 forms another branch with *Cicadidae* GST, indicating a close genetic association between the two. However, its genetic relationships with GSTs in insects that are not in order *Hemiptera* appear to be distantly related, as it is completely separated in the phylogenetic tree. This result further verifies the close genetic association between EoGSTS1 and the known GST gene in order *Hemiptera* and the conservation of the GST genes within order *Hemiptera*.

The target protein was purified by Ni affinity chromatography, the results of SDS-PAGE detection and western blot detection coincided with prediction results of the molecular weight of protein (23.68932 kDa), indicating that the expression product had not been extensively processed and modified. The determination of enzyme activity showed that the specific activity of rEoGSTs1 protein reached the highest value at 25°C and a pH of 7, which is similar to the results of glutathione transferase in silkworm (Yamamoto et al., 2006). CNDB is a kind of
insoluble and poisonous chemical compound that has high reactivity with many GST isoforms; therefore, it is often used as a model substrate to detect whether GSTs have detoxification activity (Tan et al., 2014; Deponte, 2013). In this study, the kinetic parameters of EoGSTS1 showed that it has catalytic activity on the model substrate CDNB, and compared to the results of the activity of BmGSTD in silkworm (Yamamoto et al., 2012), Vmax of the EoGSTS1 protein is higher than that of BmGSTD, which indicates that rEoGSTs1 protein has good detoxification function. Together with the differential gene profiling results, we propose that EoGSTS1 plays an important role in the detoxification of *E. onukii* to thiamethoxam.

**CONCLUSIONS**

In this study, the full-length cDNA of the *E. onukii* GST gene *EoGSTs1* was successfully amplified by qPCR. EoGSTS1 belongs to the Sigma subfamily of GSTs and is a stable hydrophilic protein located in the cytoplasm and a nontransmembrane GST protein with a single functional structure. The EoGSTS1 protein has multiple motifs of \(\alpha\) helices and \(\beta\) sheets distributed throughout its N- and C-termini. The protein sequence indicates the existence of multiple potential phosphorylation sites, with serine phosphorylation sites as the major type. The 3D structure of the EoGSTS1 protein is highly similar to that of 4Q5R, a GST protein in the German cockroach *B. germanica*. EoGSTS1 is closely related to insect GSTs of order *Hemiptera* but is distantly related to GSTs of insects of order *Blattodea*. The results of recombinant expression and purification of EoGSTS1 *in vitro* and the enzyme activity assay of purified protein show that EoGSTS1 possesses detoxification activity.

**REFERENCES**
Chelvanayagam G, Parker MW, Board P. 2001. Fly fishing for GSTs: a unified nomenclature for mammalian and insect glutathione transferases. *Chemico-biological interactions* 133: 256-260

Deponte M. 2013. Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1830(5): 3217-3266

Enayati AA, Ranson H, Hemingway J. 2005. Insect glutathione transferases and insecticide resistance. *Insect Molecular Biology* 14(1): 3-8

Fu JY, Han BY, Xiao Q. 2014. Mitochondrial COI and 16sRNA evidence for a single species hypothesis of *E. vitis*, *J. formosana* and *E. onukii* in East Asia. *PloS One* 9(12): e115259

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic acids research* 36(suppl_2): W5-W9

Hazarika LK, Bhuyan M, Hazarika BN. 2009. Insect pests of tea and their management. *Annual Review of Entomology* 54: 267-284

Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science* 302(5649):1364-1368

Letunic I, Doerks T, Bork P. 2015. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Research* 43(D1): D257-D260

Li LD, Wang DF, Li HL, Zhang H, Zeng MS, Wu GY. 2018. Cloning and bioinformatics of CYP6A14 gene from *Empoasca flavescens* (Hemiptera: cicadas). *Acta Tea Sinica* 59(1): 1-5

Li S, Su L, Li BL, Li YP, Li GW, Wu JX. 2018. cDNA cloning, prokaryotic expression and enzymatic characteristics of the glutathione S-transferase GmolGST6 in *Grapholita molesta* (Lepidoptera: Tortricidae). *Acta Entomologica Sinica* 61(4): 398-409
Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, Ke Z, Krylov D, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Thanki N, Yamashita RA, Yin JJ, Zhang D, Bryant SH. 2006. CDD: a conserved domain database for interactive domain family analysis. *Nucleic acids research* **35**(suppl_1): D237-D240

Mu D, Cui L, Wang MX, Liu LF, Yu XP, Zhang QH, Han BY. 2012. Behavioral responses for evaluating the attractiveness of specific tea shoot volatiles to the tea green leafhopper, *Empoaca vitis*. *Insect Science* **19**(2): 229-238

Nie L, Wu G, Zhang W. 2006. Correlation between mRNA and protein abundance in Desulfovibrio vulgaris: a multiple regression to identify sources of variations. *Biochemical and Biophysical Research Communications* **339**(2): 603-610

Niu YQ, Wang MX, Cui L, Ye HX, Pan C, Han BY. 2015. Control of the tea green leafhopper via volatile compounds of rosemary: the potential for further development of a Push-Pull strategy. *Acta Ecologica Sinica* **35**(7): 2380-2387

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* **32**(1): 268-274

Qin D, Zhang L, Xiao Q, Dietrich C, Matsumura M. 2015. Clarification of the identity of the tea green leafhopper based on morphological comparison between Chinese and Japanese specimens. *PloS One* **10**(9): e0139202

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**(3): 539-542
Roy A, Kucukural A, Zhang Y. 2010. I-tasser: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**(4): 725-38

Saha D, Mukhopadhyay A. 2013. Insecticide resistance mechanisms in three sucking insect pests of tea with reference to North-East India: an appraisal. *International Journal of Tropical Insect Science* **33**(1): 46-70

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**(1): 539

Sun ZR. 1994. The classification of peptide connecting α–helices and β–strands in proteins. *Journal of Tsinghua University* **34**: 58-62

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**(12): 2725-2729

Tan X, Hu XM, Zhong XW, Chen QM, Xia QY, Zhao P. 2014. Antenna-specific glutathione S-transferase in male silkmoth *Bombyx mori*. *International journal of molecular sciences* **15**(5): 7429-7443

Wang NW. 2004. Studies on the threshold and pesticide resistance comparison of green leafhopper *Empoasca vitis* Göthe. *Master Thesis. Fuzhou: Fujian Agriculture and Forestry University*

Wanyonyi SS, Sharp JA, Khalil E, Lefevre C, Nicholas KR. 2011. Tammar wallaby mammary cathelicidins are differentially expressed during lactation and exhibit
antimicrobial and cell proliferative activity. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 160(3): 431-439

Wu JZ. 2010. The study on applying parallel ant colony optimization to protein structure prediction. Suzhou: Suzhou University 1

Wu YC, Xu ZP, Shao XS, Cheng JG, Li Z. 2016. Toxicity of cycloxaprid to Aphis craccivora (Koch) and its effects on detoxification enzymes. Chinese Journal of Pesticide Science 18(6): 710-716

Xu J, Zhang Y. 2010. How significant is a protein structure similarity with TM-score= 0.5?. Bioinformatics 26(7): 889-895

Yamamoto K, Usuda K, Kakuta Y, Kimura M, Higashiura A, Nakagawa A, Aso Y, Suzuki M. 2012. Structural basis for catalytic activity of a silkworm Delta-class glutathione transferase. Biochimica et Biophysica Acta (BBA)-General Subjects 1820(10): 1469-1474

Yamamoto K, Zhang PB, Banno Y, Fujii H. 2006. Identification of a sigma-class glutathione-S-transferase from the silkworm, Bombyx mori. Journal of Applied Entomology 130(9-10): 515-522

Yang J, Zhang Y. 2015. Protein structure and function prediction using I-TASSER. Current Protocols in Bioinformatics 52(1): 5.8.1-5.8.15

Yu QY, Fang SM, Zuo WD, Zhang Z, Lu C. 2010. Molecular cloning, sequence analysis and tissue expression characterization of glutathione s-transferase gene bmgstz1 in bombyx mori. Acta Entomologica Sinica 53(10): 1061-1068

You Y, Xie M, Ren N, Cheng X, Li J, Ma X, Zou M, Vasseur L, Gurr GM, You M. 2015. Characterization and expression profiling of glutathione S-transferases in the diamondback moth, Plutella xylostella (L.). BMC genomics 16(1): 152-159
Zhan MK, Fang SM, Chen M, Wang B, Li ZZ. 2012. Screening for a highly virulent entomogenous fungal strain to *Empoasca vitis* (Homoptera: Cicadellidae) and its field efficacy. *Chinese Journal of Biological Control* **28**(1): 41-46

Zhang GF, Li LL, Li BJ, Wang KY, Xia XM. 2014. Resistance detection and synergism of enzyme inhibitors on neonicotinoids to *Aphis gossypii* in Shandong province. *Chinese Journal of Pesticide Science* **16**(6): 673-680

Zhang Z, Chen Z. 2015. Non-host plant essential oil volatiles with potential for a ‘push-pull’ strategy to control the tea green leafhopper, *Empoasca vitis*. *Entomologia Experimentalis et Applicata* **156**(1): 77-87

Zhou C, Han ZJ. 2017. Aquaporin gene cloning and its possible contribution to insecticide metabolism in *laodelphax striatellus*. *Journal of Nanjing Agricultural University* **40**(6): 1014-1022

Zhou WW, Li XW, Quan YH, Cheng J, Zhang CX, Gurr G, Zhu ZR. 2012. Identification and expression profiles of nine glutathione S-transferase genes from the important rice phloem sap-sucker and virus vector *Laodelphax striatellus* (Fallén) (Hemiptera: Delphacidae). *Pest Management Science* **68**(9): 1296-1305
Table 1 (on next page)

Primers and sequences
| Primer   | Sequence (5′ to 3′)            |
|----------|--------------------------------|
| rGST-R1  | CTCCTGCCTTACCCTCGATCTCAAGAAT   |
| rGST-R2  | CAGGCCATTCAACCCCATTTGACTCTG    |
| mGST-F   | GGGCTTGGAGAACCATAAGATT         |
| mGST-R   | GATATCTGACCCGTTGTTGC           |
| rGST-F1  | GAGAGGGGAGTTGGCGAAATACTACTACGA |
| rGST-F2  | GGCGAAATACTACTACGAGCTGATGAAGAG |
Table 2 (on next page)

GST genes of different insect species and GenBank accession numbers
## Table 2 GST genes of different insect species and GenBank accession numbers

| Types   | Species                  | Gene code | Accession number |
|---------|--------------------------|-----------|------------------|
| Delta   | *Plutella xylostella*    | PxGSTD    | BAJ10978.1       |
|         | *Lasioderma serricorne*  | LsGSTD    | AUO28661.1       |
|         | *Chilo suppressalis*     | CsGSTD    | AKS40338.1       |
|         | *Cnaphalocrocis medinalis* | CmGSTD | AIL29308.1       |
|         | *Spodoptera litura*      | SiGSTD    | AIH07596.1       |
|         | *Drosophila willistoni*  | DwGSTE    | XP_002068808.1   |
|         | *Spodoptera exigua*      | SeGSTE    | AHB18378.1       |
|         | *Bombyx mori*            | BmGSTE    | NP_001108460.1   |
|         | *Anopheles funestus*     | AfGSTE    | AHC31033.1       |
| Epsilon | *Cnaphalocrocis medinalis* | CmGSTE | AIL29313.1       |
|         | *Spodoptera litura*      | SiGSTE    | AIH07589.1       |
|         | *Aedes aegypti*          | AaGSTE    | NP_001345908.1   |
|         | *Bicyclus anynana*       | BaGSTE    | AON96571.1       |
|         | *Anopheles cracens*      | AeGSTE    | ACY95463.1       |
|         | *Alitta succinea*        | AsGSTT    | ABQ82132.1       |
|         | *Tenebrio molitor*       | TmGSTT    | AIL23552.1       |
|         | *Mus musculus*           | MmGSTT    | CAA66666.1       |
|         | *Carassius auratus*      | CaGSTT    | ABW96271.1       |
| Theta   | *Cyprinus carpio*        | CcGSTT    | BAS29977.1       |
|         | *Channa punctata*        | CpGSTT    | ABY83769.1       |
|         | *Procambarus clarkii*    | PeGSTT    | AXR98486.1       |
|         | *Andrias davidianus*     | AdGSTT    | AYG85510.1       |
|         | *Medicago truncatula*    | MtGSTT    | AET05000.1       |
|         | *Sus scrofa*             | SsGSTO    | AAF71994.2       |
|         | *Halocynthia roretzi*    | HrGSTO    | BAD77935.1       |
|         | *Schistosoma mansoni*    | SmGSTO    | AAO49385.1       |
|         | *Tenebrio molitor*       | TmGSTO    | AIL23546.1       |
| Omega   | *Perna viridis*          | PvGSTO    | AGN03944.1       |
|         | *Tigriopus japonicus*    | TjGSTO    | ACE81246.1       |
|         | *Cyprinus carpio*        | CcGSTO    | BAS29980.1       |
|         | *Kryptolebias marmoratus* | KmGSTO | AEM65182.1       |
|         | *Spodoptera exigua*      | SeGSTO    | AHB18379.1       |
| Zeta    | *Brassica napus*         | BrGSTZ    | AAO60042.1       |
| Species                        | Short Name | Accession    |
|-------------------------------|------------|--------------|
| Arabidopsis thaliana          | AtGSTZ     | AAO60039.1   |
| Tigriopus japonicus           | TjGSTZ     | ACE81250.1   |
| Tenebrio molitor              | TmGSTZ     | AIL23553.1   |
| Panonychus citri              | PcGSTZ     | AFD36889.1   |
| Azumapecten farreri           | AfGSTZ     | ADD82544.1   |
| Vibrio sinaliensis            | VsGSTZ     | EGA70514.1   |
| Tetrahymena thermophila       | TtGSTZ     | EAR89084.1   |
| Cyprinus carpio               | CcGSTZ     | BAS29981.1   |
| Antheraea pernyi              | ApGSTS     | ADC32118.1   |
| Bombyx mori                   | BmGSTS     | BAD91107.1   |
| Chilo suppressalis            | CsGSTS     | ADD14027.1   |
| Tetrahymena molitor           | TmGSTS     | AIL23551.1   |
| Daphnia magna                 | DmGSTS     | AOQ25845.1   |
| Lasioderma serricorne         | LsGSTS     | AUO28662.1   |
| Operophtera brumata           | ObGSTS     | KOB58098.1   |
| Mytilus galloprovincialis     | MgGSTS     | AFQ35985.1   |
| Cnaphalocrocis medinalis      | CmGSTS     | AI46905.1    |
Table 3 (on next page)

The chemical composition of the cloned fragment of EoGSTs1
Table 3 The chemical composition of the cloned fragment of EoGSTs1

| Amino acid | Number | Proportion of quantity | Mass ratio | Amino acid | Number | Proportion of quantity | Mass ratio |
|------------|--------|------------------------|------------|------------|--------|------------------------|------------|
| Ala (A)    | 19     | 9.20%                  | 6.18%      | Lys (K)    | 19     | 9.20%                  | 10.14%     |
| Arg (R)    | 10     | 4.80%                  | 6.36%      | Met (M)    | 8      | 3.90%                  | 4.36%      |
| Asn (N)    | 6      | 2.90%                  | 2.89%      | Phe (F)    | 9      | 4.30%                  | 5.43%      |
| Asp (D)    | 14     | 6.80%                  | 6.80%      | Pro (P)    | 12     | 5.80%                  | 5.04%      |
| Cys (C)    | 2      | 1.00%                  | 0.88%      | Ser (S)    | 7      | 3.40%                  | 2.68%      |
| Gln (Q)    | 6      | 2.90%                  | 3.20%      | Thr (T)    | 8      | 3.90%                  | 3.48%      |
| Glu (E)    | 16     | 7.70%                  | 8.59%      | Trp (W)    | 6      | 2.90%                  | 4.47%      |
| Gly (G)    | 14     | 6.80%                  | 3.84%      | Tyr (Y)    | 10     | 4.80%                  | 6.61%      |
| His (H)    | 1      | 0.50%                  | 0.57%      | Val (V)    | 13     | 6.30%                  | 5.56%      |
| Ile (I)    | 13     | 6.30%                  | 6.22%      | Pyl (O)    | 0      | 0.00%                  | 0.00%      |
| Leu (L)    | 14     | 6.80%                  | 6.70%      | Sec (U)    | 0      | 0.00%                  | 0.00%      |
Table 4 (on next page)

Predicted phosphorylation sites
| Position | Sequence   | Score | Classification | Position | Sequence   | Score | Classification |
|----------|------------|-------|----------------|----------|------------|-------|----------------|
| 95       | DEIVSVVDE  | 0.986 | *S*            | 47       | IKPNTPWGK  | 0.801 | *T*            |
| 118      | ERKASLKEP  | 0.998 | *S*            | 125      | EPVLQTVP   | 0.597 | *T*            |
| 152      | NGKFSWADV  | 0.985 | *S*            | 145      | ENKGYLANG  | 0.932 | *Y*            |
| 166      | SDHMSMNNG  | 0.546 | *S*            | 188      | RERVYAIKP  | 0.890 | *Y*            |
### Table 5 (on next page)

GST genes of different insect species and GenBank accession numbers
## Table 5 GST genes of different insect species and GenBank accession numbers

| Species                        | Gene code | Accession number  |
|-------------------------------|-----------|-------------------|
| *Sogatella furcifera*         | SyGST     | AFJ75815.1        |
| *Halyomorpha halys*           | HhGST     | XP_014283901.1    |
| *Laodelphax striatella*       | LsGST     | AEY80032.1        |
| *Nilaparvata lugens*          | NlGST     | XP_022201420.1    |
| *Riptortus pedestris*         | RpGST     | BAN21228.1        |
| *Cimex lectularius*           | CIGST     | XP_024082892.1    |
| *Sogatella furcifera*         | SyGST     | AFJ75815.1        |
| *Locusta migratoria*          | LmGST     | AHC08043.1        |
| *Schistocerca gregaria*       | SgGST     | AEV89756.1        |
| *Leptinotarsa decemlineata*   | LdGST     | APX61045.1        |
| *Tribolium castaneum*         | TcGST     | XP_967475.1       |
| *Anoplophora glabripennis*    | AgGST     | XP_018560985.1    |
| *Tenebrio molitor*            | TmGST     | AIL23548.1        |
| *Aethina tumida*              | AtGST     | XP_019878993.1    |
| *Agrius planipennis*          | ApGST     | XP_018334284.1    |
| *Blattella germanica*         | BgGST     | PSN56155.1        |
| *Zootermopsis nevadensis*     | ZnGST     | XP_021913534.1    |
| *Periplaneta americana*       | PaGST     | AVA17428.1        |
| *Apis cerana cerana*          | AcGST     | PBC25817.1        |
| *Athalia rosae*               | ArGST     | XP_012268124.1    |
| *Copidosoma floridanum*       | CfGST     | XP_014210579.1    |
| *Apis florea*                 | AfGST     | XP_003694329.2    |
| *Apis mellifera*              | AmGST     | XP_026295805.1    |
| *Megachile rotundata*         | MrGST     | XP_003703954.1    |
| *Microplitis demolitor*       | MdGST     | XP_008550325.1    |
Figure 1

1% gel electrophoresis results of total RNA.
Manuscript to be reviewed
Figure 2

1.5% gel electrophoresis results of amplification products.

A, 5’ RACE PCR; B, middle fragment PCR; C, 3’ RACE PCR.
Figure 3

Complete nucleotide sequence encoding *E. onukii* Matsuda and deduced amino acids of the cloned EoGSTs1.

Start codon (ATG), black bold; Putative polyadenylation signal, underline; Stop codon, black bold with an asterisk; Putative catalytic domain, rectangle; N-linked glycosylation sites, italic bold.
Manuscript to be reviewed
Figure 4

Clustering analysis of GST genes of different insect species based on the neighbouring method.

Numbers on the branch represent the bootstrap value (bootstrap >70%).
Figure 5

Comparison of GST amino acid sequence of some insects (Hemiptera).

'*' represents identical amino acid residues, ' - ' represents sequence deletion of RNA at this position, '"' and ':' represent conservatism of RNA sequence at different degrees, no identifier represents the sequence does not have conservation.
Figure 6

Predicted secondary structure and annotated functional sites of the EoGSTS1 protein.

The protein kinase C phosphorylation site is shown in purple and the N-glycosylation site is shown in yellow.
Figure 7

Prediction of the functional domains of the EoGSTS1 protein.

Small triangle: conservative character or amino acid of the conserved sites, specific hits: specific hits domains, non-specific hits: non-specific hits domains, superfamilies: superfamilies that corresponds to the above-mentioned matching domains.
Figure 8

Figure 8 Predicted 3D structure and main chain interface structure of the EoGSTS1 protein.

A: Predicted 3D structure of the EoGSTs1 protein of *E. onukii* Matsuda. B: Superimposed prediction model and native cartoon structures of *E. onukii* Matsuda EoGSTs1 and *Blattella germanica* GST. Rainbow structure represents *E. onukii* Matsuda EoGSTs1. Purple line represents the alpha carbon backbone trace of *B. germanica* GST.
The phylogenetic tree was constructed with ML (best model: LG+I+G4) and BI (best model: LG+I+G, -lnL=-7163.4499) methods based on the GST genes of different insects.

Numbers on the branch indicate the ML bootstrap values (>70%), whereas Bayesian posterior possibilities (>90%) are shown under the branch.
Figure 10

SDS-PAGE analysis of expression (A) and Western Blot analysis of fusion protein purification (B).

A, M1: Protein molecular weight marker; 1: Expressed product of recombinant EoGSRs1 protein without induction with IPTG at 11 °C; 2: Expressed product of recombinant EoGSRs1 protein with induction with IPTG at 11 °C; 3: The supernatant of recombinant EoGSRs1 protein with induction with IPTG at 11 °C; 4: The precipitant of recombinant EoGSRs1 protein with induction with IPTG at 11 °C. B, M2: Protein molecular weight marker; 1: Purified sample.
Figure 11

Enzymatic kinetics of the purified recombinant EoGSRs1 protein (A) and influence of temperature (B) and pH (C) on the enzyme activity of the purified recombinant EoGSRs1 protein.

Data in the figure are mean ± SE.
