Title for authors: *X. L. Ma et al.*

Title for subject: *Sulfs and sumfs in arthropod*

Correspondence: Min-Sheng You, State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Institute of Applied Ecology, Fujian Agriculture and Forestry University, Fuzhou 350002, China. Email: msyou@iae.fjau.edu.cn

* These authors contribute equally to this work.

Original article

Structure and expression of sulfatase and sulfatase modifying factor genes in the diamondback moth, *Plutella xylostella* (L.)

Xiao-Li Ma^{1,2,3,4*}, Wei-Yi He^{1,2,3,4*}, Wei Chen^{1,2,3,4}, Xue-Jiao Xu^{1,2,3,4}, Wei-Ping Qi^{1,2,3,4}, Ming-Min Zou^{1,2,3,4}, Yan-Chun You^{1,2,3,4}, Simon W. Baxter^{1,6}, , Ping Wang^{1,5}, and Min-Sheng You^{1,2,3,4}

¹ State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Fujian Agriculture and Forestry University, Fuzhou 350002, China, ² Institute of Applied Ecology, Fujian Agriculture and Forestry University, Fuzhou 350002, China, ³ Fujian-Taiwan Joint Centre for Ecological Control of Crop Pests, Fujian Agriculture and Forestry University,

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/1744-7917.12487](#).

This article is protected by copyright. All rights reserved.

Fuzhou 350002, China, ⁴ Key Laboratory of Integrated Pest Management for Fujian-Taiwan Crops, Ministry of Agriculture, Fuzhou 350002, China, ⁵ Department of Entomology, Cornell University, New York State Agricultural Experiment Station, Geneva, NY 14456, USA, and ⁶ School of Biological Sciences, the University of Adelaide, South Australia, Australia.

Abstract The diamondback moth, *Plutella xylostella* (L.), uses sulfatases (SULF) to counteract the glucosinolate–myrosinase defensive system that cruciferous plants have evolved to deter insect feeding. Sulfatase activity is regulated by post-translational modification of a cysteine residue by sulfatase modifying factor 1 (SUMF1). We identified 12 SULF genes (*PxylSulfs*) and two SUMF1 genes (*PxylSumf1s*) in the *P. xylostella* genome. Phylogenetic analysis of SULFs and SUMFs from *P. xylostella*, *Bombyx mori*, *Manduca sexta*, *Heliconius melpomene*, *Danaus plexippus*, *Drosophila melanogaster*, *Tetranychus urticae* and *Homo sapiens* showed that the SULFs were clustered into five groups, and the SUMFs could be divided into two groups. Profiling of the expression of *PxylSulfs* and *PxylSumfs* by RNA-seq and by qRT-PCR showed that two glucosinolate sulfatase genes (*GSSs*), *PxylSulf2* and *PxylSulf3*, were primarily expressed in the midgut of 3rd- and 4th-instar larvae. Moreover, expression of sulfatases *PxylSulf2*, *PxylSulf3* and *PxylSulf4* were correlated with expression of the sulfatases modifying factor *PxylSumf1a*. The findings from this study provide new insights into the structure and expression of *SUMF1* and *PxylSulf* genes that are considered to be key

factors for the evolutionary success of *P. xylostella* as a specialist herbivore of cruciferous plants.

Keywords: evolutionary arms race; glucosinolate detoxification; glucosinolate sulfatase; phylogenetic analysis; transcriptome analysis

Introduction

Plants produce phytochemicals in a constitutive or inducible manner to deter herbivory by insects (Ahuja *et al.*, 2011; Gloss *et al.*, 2013). Successful herbivory of insects relies on their ability to counteract or adapt to the plants' defensive chemicals (Furstenberg-Hagg *et al.*, 2013; Gloss *et al.*, 2013) and this evolutionary arms race is thought to shape the coevolution of plants and herbivores (Futuyma & Agrawal, 2009; Textor & Gershenzon, 2009). Glucosinolates are known as feeding deterrents to most insects (Newman *et al.*, 1992; Hopkins *et al.*, 2009; Müller, 2009; Winde & Wittstock, 2011). More than 120 different types of glucosinolates have been identified from plants so far, and this diversity is believed to be of ecological importance (Rask *et al.*, 2000; Hopkins *et al.*, 2009). Cruciferous plants (family Cruciferae) have the glucosinolate - myrosinase defense system, a well-characterized plant chemical defense mechanism known as the "mustard oil bomb" (Müller, 2009; Winde & Wittstock, 2011). However, numerous herbivorous insects that specialize on crucifers have evolved mechanisms to cope with the defensive glucosinolate-myrosinase system of their host plants (Newman *et*

al., 1992; Ratzka *et al.*, 2002). In host plant cells, the glucosinolates and myrosinases are stored in separate compartments (Hopkins *et al.*, 2009; Müller, 2009; Winde & Wittstock, 2011; Furstenberg-Hagg *et al.*, 2013). Rupture of the plant cells by insect feeding allows interaction between these compounds whereby myrosinase cleaves β -glucose to destabilize glucosinolates and generate toxic products (Hopkins *et al.*, 2009; Müller, 2009; Winde & Wittstock, 2011; Furstenberg-Hagg *et al.*, 2013). A crucifer specialist insect, the diamondback moth (DBM), *Plutella xylostella* (L.) (Lepidoptera: Plutellidae), circumvents this sophisticated plant defense system by modifying glucosinolates using glucosinolate sulfatase (Ratzka *et al.*, 2002). The *P. xylostella* GSSs (*PxylSulf2* and *PxylSulf3*) may rapidly desulfate glucosinolates from plants to limit the production of toxic breakdown products (Ratzka *et al.*, 2002; You *et al.*, 2013).

Sulfatases (SULFs) are evolutionarily conserved enzymes, found in both prokaryotes and eukaryotes (Sardiello *et al.*, 2005; Schlotawa *et al.*, 2007). They catalyze the cleavage of sulfate ester bonds in a variety of substrates, ranging from complex molecules, such as glycosaminoglycans and sulfolipids, to simple steroid sulfates (Cosma *et al.*, 2003; Cosma *et al.*, 2004; Sardiello *et al.*, 2005; Schlotawa *et al.*, 2007). The activity of SULFs is regulated by various mechanisms. Sulfatases activated by modification of a cysteine residue at the strategy activation center into C α -formylglycine (FGly) by a formylglycine-generating enzyme sulfatase-modifying factor 1 (SUMF1) (Cosma *et al.*, 2003; Fraldi *et al.*, 2007). In mammals, expression of SUMF1 results in an increase of transcripts of SULFs, so it has been suggested that SUMF1 is essential to govern SULF activities (Cosma *et al.*, 2003; Cosma *et al.*, 2004;

Zito *et al.*, 2007). In addition to oxidation of the cysteine residue at the activation site, in bacteria SULF activity is regulated by oxidation of a serine instead of a cysteine, which is located at the same position of the activation site (Fang *et al.*, 2004; Grove *et al.*, 2013). However, in nematodes and fungi there are no SUMF orthologs, so they may have alternative mechanisms for cysteine-to-FGly conversion, or they may even not be able to modify the cysteine residue to FGly (Fraldi *et al.*, 2007).

At present, SULF and SUMF genes are not yet well understood in arthropods. The increasing number of insect genome sequences becoming available provides opportunities for genome-wide analysis of genes involved in detoxification of plant glucosinolates (Claudianos *et al.*, 2006). In this study, we identified SULFs and SUMFs genes from *P. xylostella*, *Bombyx mori*, *Manduca sexta*, *Heliconius melpomene*, *Danaus plexippus*, *Drosophila melanogaster*, and *Tetranychus urticae* genomes (Celniker & Rubin, 2003; Duan *et al.*, 2010; Grbic *et al.*, 2011, Zhan *et al.*, 2011; Heliconius Genome, 2012; You *et al.*, 2013; www.agripestbase.org), and performed phylogenetic analysis to further understand the evolutionary relationships among different groups. We further characterized the gene structure and motifs and the expression of the *PxylSulfs* and *PxylSumfs* to provide new molecular information for understanding the genes for detoxification of glucosinolates in a crucifer specialist insect.

Materials and methods

Experimental DBM strain

The DBM strain used in this study was derived from an insecticide susceptible strain, Fuzhou-S (You *et al.*, 2013). DBM was reared on potted radish seedlings (*Raphanus sativus* L.) at $25\pm1^{\circ}\text{C}$, $65\%\pm5\%\text{RH}$ and L:D = 16:8h in greenhouse without exposure to insecticides since its collection from the field in 2004.

Identification of SULFs and SUMFs

To identify the SULF and SUMF protein sequences in the *P. xylostella* genome, a systematic tBLASTn search of the DBM genome database (You *et al.*, 2013; Tang *et al.*, 2014) was performed using arthropod SULF and SUMF protein sequences available from NCBI as queries (<http://www.ncbi.nlm.nih.gov/>) (Figs. S1 and S2), with the e-value cutoff set at $<10^{-6}$. Candidate *PxylSulf* and *PxylSumf* sequences were further examined to identify SULF and SUMF domains by searching the conserved protein domain database at NCBI (<http://www.ncbi.nlm.nih.gov/>). The same BLAST strategy was used to identify *SULF* and *SUMF* genes from the genomes of *B. mori* (*Bm*) (Duan *et al.*, 2010), *D. plexippus* (*Dp*) (Zhan *et al.*, 2011), *M. sexta* (*Ms*) (www.agripestbase.org), *H. melpomene* (*Hm*) (Heliconius Genome, 2012), *D. melanogaster* (*Dm*) (Diptera) (Celniker & Rubin, 2003), *T. urticae* (*Tu*) (Acarina) (Grbic *et al.*, 2011). The *SULF* sequences that were previously documented, including 17 *SULFs* from *H. sapiens*, three from *Caenorhabditis elegans*, and six from five

species of fungi (*Debaryomyces hansenii*, *Kluyveromyces lactis*, *Yarrowia lipolytica*, *Schizosaccharomyces pombe*, *Neurospora crassa*) (Sardiello *et al.*, 2005), were similarly identified and used for phylogenetic analysis of the gene families.

The SULF and SUMF genes identified from these seven species were used as queries to BLAST against their genomes for a thorough search for genes of the two families. We used the online FGENESH and FGENESH+ programs (<http://linux1.softberry.com/berry.html>) to predict the gene structures of the *PxylSulf* and *PxylSumf* genes on the genomic DNA sequences from the DBM genome.

The identified SULF and SUMF genes were then named according to their sequence homology with the silkworm SULFs which were named by their positions in chromosomes (Table S1). The protein sequences of the *PxylSulf* and *PxylSumf* were examined using EXPASY PROTPARAM (<http://www.expasy.org/tools/protparam.html>) to obtain their predicted molecular weights and isoelectric points (pI). Full length sequences of *PxylSulf2*, *PxylSulf3* and *PxylSumf1a* were confirmed by PCR (Figs. S3 & S4).

Sequence alignment and phylogenetic tree construction

Multiple alignments of the SULF and SUMF protein sequences were performed using Clustal X2.0 (Larkin *et al.*, 2007). Phylogenetic trees based on the protein sequences were

constructed using the Neighbor-joining (NJ) method with p-distance and the pairwise deletion in MEGA 5.10 software (Tamura *et al.*, 2011). Maximum-likelihood (ML) analyses were performed using RAXML (<https://github.com/stamatak/ExaML>). Images of the phylogenetic trees were produced using iTOL software (Letunic & Bork, 2007) (<http://itol.embl.de/index.html/>).

Examination of gene structure and motifs

The *PxylSulfs* and *PxylSumfs* were mapped onto the corresponding scaffolds of the DBM genome sequence assembly v. 2 (You *et al.*, 2013; Tang *et al.*, 2014). The distributions of *PxylSulfs* and *PxylSumfs* on scaffolds were determined using the Softberry program (<http://www.softberry.com/>). Conserved motifs in the candidate sequences were identified using the MEME software (<http://meme-suite.org/>) with the following parameters: number of repetitions = any, maximum number of motifs = 20, and optimum motif width = 6 to 20 residues.

Expression profiling of PxylSulfs and PxylSumfs transcripts

RNA-seq data of the *PxylSulfs* and *PxylSumfs* were downloaded from the DBM genome database (He *et al.*, 2012; You *et al.*, 2013; Tang *et al.*, 2014) (<http://iae.fafu.edu.cn/DBM/>).

Expression profiles of the DBM genes in different developmental stages and tissues were analyzed, including eggs (24 h), 1st to 4th instar larvae, pupae, male and female adults, heads of the 4th-instar larvae and male/female adults, fat body-, hemolymph-, Malpighian tubules-, silk glands-, salivary glands-, and midguts of the 4th-instar larvae. The Newick output format of the *PxylSulfs* and *PxylSumfs* phylogenetic trees with the logarithmic values of RPKMs was used to visualize the expression patterns using the online tool iTOL (Letunic & Bork, 2007).

To confirm the stage- and tissue-favored expression of the *PxylSulfs* and *PxylSumfs*, qRT-PCR analysis was performed using SYBR-green fluorescence with gene-specific primers (Giulietti *et al.*, 2001; Bustin *et al.*, 2009) (Table S2). The PCR products were examined by dissociation curve analysis after the PCR reaction to confirm the specific detection of target transcripts by the qRT-PCR analysis. Three independent biological replicates were included for qRT-PCR, each of which had three technical replicates. The first-strand cDNA was synthesized from total RNA using the reverse transcriptase kit from Promega (Madison, WI, USA). The qRT-PCR reactions were prepared using the SYBR SELECT MASTER MIX FOR CFX from Invitrogen (Carlsbad, CA, USA) following manufacturer's instructions and run on a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA), following the program: 95°C for 3 min; 45 cycles of 95°C for 15 s, and 57°C for 35 s, and a final melt curve at 60°C for 5 s to 95°C at 0.5°C increments. The DBM ribosomal protein L32 (RPL32) gene (GenBank acc. no. AB180441) was used as an internal

reference. Standard curves were generated by five dilutions of the cDNA templates to determine the best range of template concentration for all genes analyzed.

Statistical analysis

The $\Delta\Delta C_t$ method was used to analyze the qRT-PCR-based expression patterns (Livak & Schmittgen, 2001; Yuan *et al.*, 2006; Yuan *et al.*, 2008). One-way ANOVA, using SPSS statistics 22 (Armonk, NY: IBM Corp), followed by a Tukey's test to evaluate the significance of differences among expression levels in different samples. The results were presented by means $\pm$ standard deviations of the mean of the relative mRNA expressions.

Results and Discussion

Identification and structure of SULFs

A total of 12 putative *PxylSulf* genes were identified in the DBM genome, which were predicted to code for proteins of 328 to 1131 amino acid (aa) residues in length (Table 1). All *PxylSulf* proteins contain an alkaline phosphatase superfamily domain (cl21476). Most *PxylSulf* genes coded for approximately 500 amino acids, but *PxylSulf4* and *PxylSulf7* were around double sized in length to the other genes. This may indicate a functional or regulatory difference among *PxylSulf* genes. Variable numbers of *SULF* gene were also identified from

the genomes of six other arthropod species including: nine SULF genes from *B. mori* (*Bm*), eight from *D. plexippus* (*Dp*), eight from *M. sexta* (*Ms*), six from *H. melpomene* (*Hm*), seven from *D. melanogaster* (*Dm*) (Diptera) and three from *T. urticae* (*Tu*) (Acarina) (Table 2).

The 12 *P. xylostella* *Sulf* genes were located on seven genomic scaffolds and they showed considerable variation in exon number (Fig. 1A). Three *PxylSulf* classes (*PxylGSSs*, *PxylSulf5*, and *PxylSulf6*) all contained two or more paralogs. Scaffold 272 contained three *PxylGSS* paralogous genes (*PxylSulf2*, 3, 4) and scaffold 104 contained all three *PxylSulf5* genes (Table 1, Fig. 1A). Genes clustered on the same scaffold share similar intron/exon structures and length. For example, *PxylSulf2* and 3 all have six introns, and are about 1800 bp in length (Fig. 1A). *PxylSulf2* and *PxylSulf3* are highly similar in ORF sequence (96% of sequence identity) (Figs. S4 and S5). Ratzka (2002) cloned the first GSS gene in *P. xylostella*, which was most similar to *PxylSulf2* (99%), which are likely to be allelic variants from different insect strains (Fig. S5). The exon/intron organization and gene structure in of the *Sulf* genes appear highly conserved to reflect their evolutionary relationship. However, how similar of their biochemical and physiological functions in insects remains to be understood.

Phylogenetic analysis of SULF proteins from *B. mori*, *M. sexta*, *H. melpomene*, *D. plexippus*, *D. melanogaster*, *T. urticae*, fungi, *C. elegans* and *H. sapiens* showed that the SULFs clustered into five different groups (bootstrap value > 55) (Fig. 2). The distribution of

SULF genes in the five groups varied among species, but group A is the largest group (Table 2, Fig. 2). The arthropod genes in Group A form two distinct branches (bootstrap value > 80), with one only containing genes from Lepidoptera and the other containing *Sulf5* genes from both insect and non-insect arthropods (Fig. 2). The *GSS* genes (*PxylSulf2*, 3) are among the members that belong to the Lepidoptera-specific branch. The *SULF5* genes are all in the same branch regardless of the species from *M. sexta*, *P. xylostella* and *D. melanogaster* to *Homo sapiens*, exhibiting the most conserved *Sulf* genes across different taxa.

Identification and structure of SUMFs

SUMFs were also identified from the genomes by blast searches. Two *SUMF* genes, SUMF1 and SUMF2, were found in most of the species examined in this study, except for fungi and *C. elegans*. All SUMF1 genes from arthropods contain a sulfatase-modifying factor enzyme domain (pfam03781). *P. xylostella* is the only species which has two SUMF1 genes, but lacks a SUMF2 gene. The *PxylSumf* genes were located on different genomic scaffolds, suggesting they were not a recent tandem duplication and coded for proteins with 348 and 267 amino acid (aa) residues in length (Table 1).

Phylogenetic analysis of SUMFs from *B. mori*, *M. sexta*, *H. melpomene*, *D. plexippus*, *D. melanogaster*, *T. urticae*, and *H. sapiens* indicated that the SUMFs could be divided into two main groups (bootstrap value > 89), SUMF1 and SUMF2 (Fig. 3). The two SUMF genes in

P. xylostella, *PxylSumf1a* and *PxylSumf1b*, are in the same SUMF1 group with a higher amino acid sequence similarity (73%) than that between SUMF1 and SUMF2 sequences from other species (8%-13% in arthropods, 36% in human). *P. xylostella* is a species diverged early from the other lepidopteran species analyzed (Kawahara & Breinholt, 2014). Therefore, SUMF2 appears to have been lost from *P. xylostella* and an additional SUMF1 gained through a gene duplication event after divergence of the species from the other lepidopterans. Genome analysis of species closely related to *P. xylostella* may enable approximate timing of this event.

Conserved domain organization of the PxylSulf

Twenty distinct motifs were identified in the SULFs from each of the analyzed species. These genes showed a common motif composition (Fig. 1B), especially for those within the three paralogous gene groups (*PxylGSSs*, *PxylSulf5*, and *PxylSulf6*) of *P. xylostella*. The catalytic core motif, C/SXPXR, in SULFs has been documented to be the minimum sequence required for activity of SULF modification (Sardiello *et al.*, 2005; Annunziata *et al.*, 2007; Schlotawa *et al.*, 2007). This motif was found in at least one sulfatase in all species examined (Fig. 4), but not all the sulfatases contain this core motif (Fig. 1B). Therefore, activation of SULFs by oxidation of the core motif may not be universally true for all sulfatases (Fang *et al.*, 2004; Grove *et al.*, 2013).

Based on sequence analysis of *GSS* orthologs available from the NCBI sequence database, we found that only the *PxylGSSs* contained a serine at position two instead of a threonine in its orthologs from 38 species (Fig. 5). Serine is structurally similar to threonine, so a threonine to serine substitution may not lead to a significant structural or biochemical change of the protein. The alanine at position four of *GSSs*, instead of threonine/serine found in other analyzed species, appears to be conserved in Lepidoptera (Fig. 5). *BmorSulf3* has a serine at this position, which is the only exception among the lepidopteran *GSSs* in this study. Whether this difference leads to a unique functional change is unknown.

Stage- and tissue-favored expression of PxylSulfs and PxylSumfs

RNA-seq analyses of *PxylSulf* and *PxylSumf* expression showed that these genes were differentially expressed in eggs, larvae, pupae, and adults (Fig. 6). The expression of *PxylGSSs* and *PxylSumfla* was abundant in the midgut and Malpighian tubules at larval stage, consistent with the assumption that these genes play important roles in metabolizing plant glucosinolates produced by host plants as the main defense components (Dow, 2009; Hopkins *et al.*, 2009; Textor & Gershenzon, 2009; Winde & Wittstock, 2011; Badenes-Perez *et al.*, 2013). The expression pattern of *PxylSulf1* is similar to that of its ortholog *BmorSulf1* in *B. mori* (sw19788), which is abundantly expressed in the head and has been suggested having the function of odorant processing and/or xenobiotic metabolism (Yu *et al.*, 2009;

Badenes-Perez *et al.*, 2013; You *et al.*, 2015; Yu *et al.*, 2015; Qi *et al.*, 2016). The expression of the orthologs of *GSSs* in *B. mori* is reported to be very low level in midguts and malpighian tubules and at all developmental stages (Table S2) (Dow, 2009; Duan *et al.*, 2010). Therefore, the *BmorGSSs* expression patterns may implicate that the *GSS* in *B. mori* does not play a major role in detoxification of glucosinolates in *B. mori*, resulting from the co-evolution between *B. mori* and its non-cruciferous host plants (Dow, 2009). However, the *BmorSumf1* and *BmorSumf2* exhibited a relatively higher level than most *BmorSulfs*, which was consistent with the function of *Sumfs* serve as a modifying factor for all *BmorSulfs*.

Determination of gene expression using qRT-PCR analysis confirmed the *PxylSulf* and *PxylSumf* expression profiles obtained from the RNAseq data (Fig. 7). Notably, most *PxylSulfs* in group A showed larval-favored expression. Especially the *PxylGSSs* exhibited a high expression level in the midgut and Malpighian tubules of 4th-instar (Fig. 8), the major feeding stage of *P. xylostella*. The remaining *PxylSulf* genes showed various expression levels, but overall were lower than the expression of *GSSs*. *PxylSumf1a* exhibited similar expression profiles to *PxylSulf2* but showed a lower level of expression at all developmental stages. Analysis of the expression of 12 *PxylSulfs* and two *PxylSumfs* from different developmental stages showed that the expression of *PxylSulf2*, *PxylSulf3* and *PxylSulf4* were correlated with the expression of *PxylSumf1a* (Table 3). The presence of two *Sumf1* genes was only found in *P. xylostella* and both *Sumf1* genes may also be candidates for the sulfatase FGly-generating system in *P. xylostella*. No significantly correlation was observed between the 12 *PxylSulfs* and the *PxylSumf1b*. However, the role of *PxylSumf1s* needs to be further understood.

Conclusions

Novel evolutionary correlations between functionally associated gene families can be made through combining comparative genomic analysis across species with gene expression data within a species. Here we show the most comprehensive study on identification, gene structure and expression profiles of the SULF and SUMF genes from *P. xylostella*, in

comparison with which from six other arthropod species. Variable features and different expression patterns of these gene families revealed that the SULFs and SUMFs have diversified in *P. xylostella*, and this is likely to have been an important factor which lead to the adaptive success of this Brassica pest. Thus, further functional research on the *PxylSulfs* and *PxylSumf1s* is essential to identify the molecular mechanism underlying *P. xylostella* in adaptation of its host plant defense system.

Acknowledgements

We are grateful to Jianlin Bai for his advice on insect dissection experiments, and Yanting Chen, Yaqiong Chen, Renfu Dong, Xiujing Shen, Liying yu, Nanjing chen and Fengqin Song for *P. xylostella* rearing and sampling. We appreciate Liette Vasseur, Mark Goettel and Guillaume Tetreau for their comments and suggestions on the manuscript revision. The work is supported by National Natural Science Foundation of China (Nos. 31320103922, 31230061 and 31301677), National Key Project of Fundamental Scientific Research (“973” Programs, No. 2011CB100404) in China, Natural Science Foundation of Fujian Province, China (2014 J01086) and Outstanding Youth Fellowships for WYH (xjq201403) at FAFU. SMB and PW were both supported by the Jinshan Scholar Fellowships at FAFU.

Authors’ contributions

XLM, WYH and MSY planned and designed the study. XLM performed the computational analysis, executed the experiments, and generated the figures. XLM, PW, SWB, WYH and MSY contributed to development and/or revision of the manuscript. WC and XJX contributed to sample preparation. YCY, WPW, MMZ and XJX performed the qRT-PCR experiments and data analyses. YCY, MMZ, PW, SWB, WYH and MSY contributed to the discussion. All authors read and approved the final manuscript.

Disclosure

The authors declare that they have no competing interests.

Table 1
Overview on 12 *SULF* and 2 *SUMF* genes identified in the *P. xylostella* genome

S.	Gene	Gene ID	Position (Strand)	ORF(b)	Protein ^b			Gene	No.
					Size(aa)	p	MW(KD)		
1	<i>PxylSulf1</i>	PX01055	scaffold39:	1803	600	7.2	67,29	5224	8
2	<i>PxylSulf2</i>	PX01810	scaffold272:	1644	547	5.0	60,79	3231	7
3	<i>PxylSulf3^c</i>	PX01810	scaffold272:	1644	547	5.1	60,79	4445	7
4	<i>PxylSulf4</i>	PX00768	scaffold272:	2850	949		103,99	7077	7
5	<i>PxylSulf5</i>	PX00051	scaffold104:	1605	534	6.2	58,63	9422	1
6	<i>PxylSulf5</i>	PX00050	scaffold104: 241,	1596	531	5.8	57,73	13909	1
7	<i>PxylSulf5</i>	PX00051	scaffold104:	1509	502	6.1	54,87	4386	1
8	<i>PxylSulf6</i>	PX00809	scaffold289:	1772	573	7.9	59,29	5097	6
9	<i>PxylSulf6</i>	PX01242	scaffold49:	987	328	5.6	37,31	3813	4
10	<i>PxylSulf7</i>	PX01247	scaffold49:	3396	1131	8.9	163.135	13020	1
11	<i>PxylSulf8</i>	PX01548	scaffold72:	1362	453	5.7	54.97	16154	1

12	<i>PxylSulf9</i>	PX00363	scaffold159:	1263	420	5.9	58.67	1993	8
13	<i>PxylSumf</i>	PX00412	scaffold166:	1047	348	4.6	35,17	2871	4
14	<i>PxylSumf</i>	PX01646	scaffold8:	804	267	5.4	30.92	1563	4

^a Systematic designation given to *PxylSulfs* and *PxylSumfs* in this study;

^b Protein characterization of *PxylSulfs* and *PxylSumfs* by Expasy (<http://www.expasy.org/>);

^c Represents complete coding sequences by PCR;

S.N., serial number; ORF, open reading frame; bp, base pair; aa, amino acids; MW, molecular weight; pI, isoelectric point.

Table 2

Comparison of the *SULF* genes in analyzed genomes and their distribution across clades

Species	group A	group B	group C	group D	group E	Total	reference
<i>P. xylostella</i>	7	0	1	2	2	12	The present study
<i>B. mori</i>	5	0	1	1	2	9	The present study
<i>D. plexippus</i>	2	0	1	1	2	8	The present study
<i>H. melpomene</i>	2	0	1	1	2	6	The present study
<i>M. sexta</i>	4	0	1	1	2	8	The present study
<i>D. melanogaster</i>	4	0	1	1	1	7	The present study
<i>T. urticae</i>	1	2	0	0	0	3	The present study
<i>C. elegans</i>	1	1	0	0	1	3	(Sardiello <i>et al.</i> , 2005)
<i>Fungi</i>	4	0	0	1	1	6	(Sardiello <i>et al.</i> , 2005)
<i>H. sapiens</i>	3	8	1	1	4	17	(Sardiello <i>et al.</i> , 2005)

Table 3

Correlation analysis between the *PxylSumfs* and *PxylSulfs* using their gene expression values of different developmental stages in the susceptible strain. (Tukey's test, ** $P < 0.01$)

Correlation coefficient	<i>PxylSumf1b</i>	<i>PxylSulf1</i>	<i>PxylSulf2</i>	<i>PxylSulf3</i>	<i>PxylSulf4</i>	<i>PxylSulf5a</i>	<i>PxylSulf5b1</i>
<i>PxylSumf1a</i>	.424	.197	.794**	.733**	.695**	-.692**	-.300
<i>PxylSumf1b</i>	1.000	-.082	.224	.222	.119	-.304	-.123
	<i>PxylSulf5b2</i>	<i>PxylSulf6a1</i>	<i>PxylSulf6a2</i>	<i>PxylSulf7</i>	<i>PxylSulf8</i>	<i>PxylSulf9</i>	-
<i>PxylSumf1a</i>	-.370	-.414	-.600**	-.736**	-.234	-.426	-
<i>PxylSumf1b</i>	-.350	-.181	-.703**	-.443	-.290	-.329	-

Figure legends

Figure 1. Illustration for gene structure (A) and conserved motifs (B) of the *P. xylostella* *PxylSulfs* and *PxylSumfs*. Acronyms in (A) are detailed as CDSf: first coding segment (starting with start codon); CDSi: internal coding segment (internal exon); CDSl: last coding segment (ending with stop codon); CDSo: gene containing one coding exon only; PolA: terminal polyA signal; TSS: transcription start site. Boxes with different colors in (B) represent different motifs.

Figure 2. Maximum-likelihood tree estimated in RAXML for the *SULF* genes. The number at the branch point of the node represents the value >80 resulting from 1000 replications and gaps. The branches are colored to represent different taxa. Taxon abbreviations were used in the legend with Lep for Lepidoptera (including *Plutella xylostella*; *Bombyx mori*; *Danaus plexippus*; *Manduca sexta* and *Heliconius Melpomene*); *Dm* for *Drosophila melanogaster*; *Tu* for *Tetranychus urticae*; *Hs* for *Homo sapiens*; *Ce* for *Caenorhabditis elegans*. Fungi include five species (*D. hansenii*, *K. lactis*, *Y. lipolytica*, *S. pombe* and *N. crassa*).

Figure 3. Phylogenetic tree of the *SUMF* genes in selected species. The phylogenetic tree was constructed using a neighbor-joining technique with MEGA5.10. The purple dot at the branch point of the node represents the value resulting from 1000 replications. The tree is drawn in scale and the evolutionary distances, indicated by the scale bar and the number next to the branches.

Figure 4. Logo representations of the catalytic cores of sulfatases in selected species. The overall height of each column is proportional to the information content at that position, and within each of the columns the conservation of each residue is visualized as the relative height of symbols representing amino acids.

Figure 5. Amino acid sequence alignment of GSS genes across 39 different species. The alignment was performed with CLUSTAL2.0 and MUSCLE. Identical amino acids are highlighted in red. The catalytic cores of Lepidoptera insects are delimited within a black

box. The red arrow indicates that only *PxylGSS* has a serine on this position rather than threonine in other species.

Figure 6. Stage- and tissue-favored expression patterns of the *PxylSulfs* and *PxylSumfs* based on RPKM values. Relative expressions levels are illustrated with seven graded colors. Tissue-favored expression was performed using the 4th instar larvae.

Figure 7. Stage-favored expression of *PxylSulfs* and *PxylSumfs* based on qRT-PCR.

Figure 8. Tissue-favored expression of *PxylSulfs* and *PxylSumfs* in the 4th instar larvae based on qRT-PCR.

References

- Ahuja, I., Rohloff, J. and Bones, A.M. (2011) Defence mechanisms of Brassicaceae: implications for plant–insect interactions and potential for integrated pest management. *Sustainable Agriculture Volume 2* (eds. E. Lichtfouse, M. Hamelin, M. Navarrete & P. Debaeke), pp. 623–670. Springer Netherlands.
- Annunziata, I., Bouche, V., Lombardi, A., Settembre, C. and Ballabio, A. (2007) Multiple sulfatase deficiency is due to hypomorphic mutations of the *SUMF1* gene. *Human Mutation*, 28, 928.
- Badenes-Perez, F.R., Reichelt, M., Gershenzon, J. and Heckel, D.G. (2013) Interaction of glucosinolate content of *Arabidopsis thaliana* mutant lines and feeding and oviposition by generalist and specialist lepidopterans. *Phytochemistry*, 86, 36–43.
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Biochemistry*, 55, 611–622.
- Celniker, S.E. and Rubin, G.M. (2003) The *Drosophila melanogaster* genome. *Annual Review of Genomics and Human Genetics*, 4, 89–117.
- Claudianos, C., Ranson, H., Johnson, R., Biswas, S., Schuler, M., Berenbaum, M., *et al.* (2006) A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Molecular Biology*, 15, 615–636.

-
- Cosma, M.P., Pepe, S., Annunziata, I., Newbold, R.F., Grompe, M., Parenti, G., *et al.* (2003) The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. *Cell*, 113, 445–456.
- Cosma, M.P., Pepe, S., Parenti, G., Settembre, C., Annunziata, I., Wade-Martins, R., *et al.* (2004) Molecular and functional analysis of SUMF1 mutations in multiple sulfatase deficiency. *Human Mutation*, 23, 576–581.
- Dow, J.A. (2009) Insights into the Malpighian tubule from functional genomics. *Journal of Experimental Biology*, 212, 435–445.
- Duan, J., Li, R., Cheng, D., Fan, W., Zha, X., Cheng, T., *et al.* (2010) SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Research*, 38, D453–D456.
- Fang, Q.H., Peng, J.H. and Dierks, T. (2004) Post-translational formylglycine modification of bacterial sulfatases by the radical S-adenosylmethionine protein AtsB. *Journal of Biological Chemistry*, 279, 14570–14578.
- Fraldi, A., Biffi, A., Lombardi, A., Visigalli, I., Pepe, S., Settembre, C., *et al.* (2007) SUMF1 enhances sulfatase activities *in vivo* in five sulfatase deficiencies. *Biochemical Journal*, 403, 305–312.
- Furstenberg-Hagg, J., Zagrobelny, M. and Bak, S. (2013) Plant defense against insect herbivores. *International Journal of Molecular Sciences*, 14, 10242–10297.
- Futuyma, D.J. and Agrawal, A.A. (2009) Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences USA*, 106, 18054–18061.
- Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R. and Mathieu, C. (2001) An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. *Methods*, 25, 386–401.

-
- Gloss, A.D., Nelson Dittrich, A.C., Goldman-Huertas, B. and Whiteman, N.K. (2013) Maintenance of genetic diversity through plant–herbivore interactions. *Current Opinion in Plant Biology*, 16, 443–450.
- Grbic, M., Van Leeuwen, T., Clark, R.M., Rombauts, S., Rouze, P., Grbic, V., *et al.* (2011) The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature*, 479, 487–492.
- Grove, T.L., Ahlum, J.H., Qin, R.M., Lanz, N.D., Radle, M.I., Krebs, C., *et al.* (2013) Further characterization of Cys-type and Ser-type anaerobic sulfatase maturing enzymes suggests a commonality in the mechanism of catalysis. *Biochemistry*, 52, 2874–2887.
- He, W.Y., You, M.S., Vasseur, L., Yang, G., Xie, M., Cui, K., *et al.* (2012) Developmental and insecticide-resistant insights from the de novo assembled transcriptome of the diamondback moth, *Plutella xylostella*. *Genomics*, 99, 169–177.
- Heliconius Genome, C. (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487, 94–98.
- Hopkins, R.J., van Dam, N.M. and van Loon, J.J. (2009) Role of glucosinolates in insect–plant relationships and multitrophic interactions. *Annual Review of Entomology*, 54, 57–83.
- Kawahara, A.Y. and Breinholt, J.W. (2014) Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proceedings of the Royal Society B-Biological Sciences*, 281, 1788.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., Mcwilliam, H., *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947–2948.
- Letunic, I. and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23, 127–128.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)). *Methods*, 25, 402–408.

-
- Müller, C. (2009) Interactions between glucosinolate-and myrosinase-containing plants and the sawfly *Athalia rosae*. *Phytochemistry Reviews*, 8, 121–134.
- Newman, R.M., Hanscom, Z. and Kerfoot, W.C. (1992) The watercress glucosinolate-myrosinase system: a feeding deterrent to caddisflies, snails and amphipods. *Oecologia*, 92, 1–7.
- Qi, W.P., Ma, X.L., He, W.Y., Chen, W., Zou, M., Gurr, G.M., *et al.* (2016) Characterization and expression profiling of ATP-binding cassette transporter genes in the diamondback moth, *Plutella xylostella* (L.). *BMC Genomics*, 17, 760.
- Rask, L., Andréasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B. and Meijer, J. (2000) Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Molecular Biology*, 42, 93–114.
- Ratzka, A., Vogel, H., Kliebenstein, D.J., Mitchell-Olds, T. and Kroymann, J. (2002) Disarming the mustard oil bomb. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 11223–11228.
- Sardiello, M., Annunziata, I., Roma, G. and Ballabio, A. (2005) Sulfatases and sulfatase modifying factors: an exclusive and promiscuous relationship. *Human Molecular Genetics*, 14, 3203–3217.
- Schlotawa, L., Steinfeld, R., Figura, K., Dierks, T. and Gärtner, J. (2007) Molecular analysis of SUMF1 mutations: stability and residual activity of mutant formylglycine-generating enzyme determine disease severity in multiple sulfatase deficiency. *Human mutation*, 29, 205–205.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28, 2731–2739.
- Tang, W.Q., Yu, L.Y., He, W.Y., Yang, G., Ke, F.S., Baxter, S.W., *et al.* (2014) DBM-DB: the diamondback moth genome database. *Database*, 2014, bat087.

-
- Textor, S. and Gershenzon, J. (2009) Herbivore induction of the glucosinolate–myrosinase defense system: major trends, biochemical bases and ecological significance. *Phytochemistry Reviews*, 8, 149–170.
- Winde, I. and Wittstock, U. (2011) Insect herbivore counteradaptations to the plant glucosinolate–myrosinase system. *Phytochemistry*, 72, 1566–1575.
- You, M.S., Yue, Z., He, W.Y., Yang, X.H., Yang, G., Xie, M., *et al.* (2013) A heterozygous moth genome provides insights into herbivory and detoxification. *Nature Genetics*, 45, 220–225.
- You, Y.C., Xie, M., Ren, N.N., Cheng, X.M., Li, J.Y., Ma, X.L., *et al.* (2015) Characterization and expression profiling of glutathione S-transferases in the diamondback moth, *Plutella xylostella* (L.). *BMC Genomics*, 16, 152.
- Yu, L.Y., Fang, W.Q., He, W.Y., Ma, X.L., Vasseur, L., Baxter, S.W., *et al.* (2015) Characterization and expression of the cytochrome P450 gene family in diamondback moth, *Plutella xylostella* (L.). *Scientific Reports*, 5, 8952.
- Yu, Q.Y., Lu, C., Li, W.L., Xiang, Z.H. and Zhang, Z. (2009) Annotation and expression of carboxylesterases in the silkworm, *Bombyx mori*. *BMC Genomics*, 10, 553.
- Yuan, J.S., Reed, A., Chen, F. and Stewart, C.N., Jr. (2006) Statistical analysis of real-time PCR data. *BMC Bioinformatics*, 7, 85.
- Yuan, J.S., Wang, D. and Stewart, C.N., Jr. (2008) Statistical methods for efficiency adjusted real-time PCR quantification. *Biotenology Journal*, 3, 112–123.
- Zhan, S., Merlin, C., Boore, J.L. and Reppert, S.M. (2011) The monarch butterfly genome yields insights into long-distance migration. *Cell*, 147, 1171–1185.
- Zito, E., Buono, M., Pepe, S., Settembre, C., Annunziata, I., Surace, E.M., *et al.* (2007) Sulfatase modifying factor 1 trafficking through the cells: from endoplasmic reticulum to the endoplasmic reticulum. *The EMBO Journal*, 26, 2443–2453.

Manuscript received November 09, 2016

Final version received March 29, 2017

Accepted April 25, 2017

Author Manuscript

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Ma, X-L;He, W-Y;Chen, W;Xu, X-J;Qi, W-P;Zou, M-M;You, Y-C;Baxter, SW;Wang, P;You, M-S

Title:

Structure and expression of sulfatase and sulfatase modifying factor genes in the diamondback moth, *Plutella xylostella*

Date:

2018-12-01

Citation:

Ma, X. -L., He, W. -Y., Chen, W., Xu, X. -J., Qi, W. -P., Zou, M. -M., You, Y. -C., Baxter, S. W., Wang, P. & You, M. -S. (2018). Structure and expression of sulfatase and sulfatase modifying factor genes in the diamondback moth, *Plutella xylostella*. *INSECT SCIENCE*, 25 (6), pp.946-958. <https://doi.org/10.1111/1744-7917.12487>.

Persistent Link:

<http://hdl.handle.net/11343/293306>