IMMUNE SIGNALING PATHWAYS AND DIGESTION-RELATED PROTEINS IN THE *MANDUCA SEXTA* LARVAE

By

ZELONG MIAO

Bachelor of Biological Science
Henan Normal University
Xinxiang, Henan
2016

Master of Biology
Henan Normal University
Xinxiang, Henan
2018

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2022

IMMUNE SIGNALING PATHWAYS AND

DIGESTION-RELATED PROTEINS IN THE *MANDUCA*

SEXTA LARVAE

Dissertation Approved:

Dr. Haobo Jiang

Dissertation Adviser

Dr. Stephen M. Marek

Dr. Bruce H. Noden

Dr. Junpeng Deng

ACKNOWLEDGEMENTS

My deepest gratitude goes first and foremost to my advisor, Dr. Haobo Jiang. During these 4 years, his rigorous academic attitude and dedicated exploration of science impressed me deeply. What he has taught enriches my mind and broadens my views. I benefited a lot from his patient guidance and instruction. He is not only an excellent knowledgeable teacher but a close friend, and even more like my third parent. “A day for the teacher, lifelong for father.”

I would like to thank my committee members, Dr. Junpeng Deng, Dr. Stephen Marek, and Dr Bruce Noden. They provide me a lot of valuable guidance and selfless assistance, which encourage me a lot in my study and research.

I would like to thank my current and previous fellow lab members, including Yang Wang, Xiaolong Cao, Udeshika Kariyawasam, Tisheng Shan, Qiao Jin, Chao Xiong, and Haibo Bao. Special thanks to Yang Wang, who is our lab manager and helped us with all kinds of research works. I appreciate the friendship with others in Stillwater. We work and live together like a big family, with everyone caring and helping each other.

Many thanks to members in Entomology and Plant Pathology department, including our previous and current department head, Dr. Phillip Mulder and Dr. Justin Talley, Graduate Program Coordinator, Dr. Nathan Walker and Dr. George Opit, the office staffs including Barbara Brown, Sharon Hillock, Lafonda Barrera, Noah Dunagan, etc. I appreciate knowledge I learned from other professors in OSU, including Dr. Astri Wayadande, Dr. Li Ma, Dr. Robert Matts, etc. I appreciate suggestions and technical supports from Dr. Steve Hartson, Janet Rogers, Dr. Lei Lei, Dr. Peter Hoyt and Dr. Hongjin Hwang in the OSU Core Facility. I thank my proud and immortal alma mater, Oklahoma State University. I will stay loyal and true, making bright shines your name.

I would like to thank the financial support from Chinese Scholarship Council (CSC) during these 4 years. Special thanks to Dr. Guolong Zhang, who supports me for the CSC application. I also appreciate the suggestions from Yongfang Li and Daijing Zhang, who are the teachers from Henan Normal University in China.

I am much grateful to my parents, they always help me to keep up my hope when I am in trouble. They are like the Guardian Angel at my side who assists me in times of fear and pain. They give me the strength and the courage to continue the fight and never give up. They made me what I am today and who I am.

Thank you, for everything, since it is fantastic – everything has its purpose and meaning.

Name: ZELONG MIAO

Date of Degree: JULY, 2022

Title of Study: IMMUNE SIGNALING PATHWAYS AND DIGESTION-RELATED PROTEINS IN THE *MANDUCA SEXTA* LARVAE

Major Field: ENTOMOLOGY

Abstract:

M. sexta has been used as a biochemistry model to study insect immunity. While it is unclear how immune genes are specifically regulated by Toll or Imd pathway. We produced the recombinant Spätzle-1 and -2 precursors, activated them using prophenoloxidase activating protease-3 (PAP3), and separately injected them into hemocoel to trigger specific up-regulation of genes controlled by the Toll pathway. *M. sexta* cell line was treated with *E. coli* DAP-PG to specifically induce the Imd pathway. RNA-seq analysis of fat body tissues and cell samples indicated that diapausins and lebocins are predominantly regulated via Toll signaling, gallerimycin, X-tox and its splicing variants are synthesized in response to DAP-PG through Imd pathway, whereas attacins, cecropins, moricins, defensins, gloverin, lysozymes, transferrins, and WAPs are induced via both. Furthermore, we separately injected *Enterococcus faecalis* or *Enterobacter cloacae* into hemocoel, it showed that most antimicrobial peptides could be induced by both bacteria, such as lebocinD, which is a Toll-specific gene. However, diapausin1, gloverin and cecropin6 were more sensitive to *Enterococcus faecalis*, X-toxs was more sensitive to *Enterobacter cloacae*. Our results showed the injection of bacteria cannot separate Toll or Imd pathways properly in *M. sexta*, which is very different from *D. melanogaster*. Thus, it confirms the complexity of host-pathogen interactions and innate immune response pathways in other non-dipteran insects.

Food digestion is vital for the survival and prosperity of insects. While digestive enzymes from pest species lacks a systematic analysis. In the genome of *Manduca sexta*, we identified 122 digestive enzymes including 85 proteases, 20 esterases, 16 carbohydrases, and 1 nuclease. We also further categorized 144 *M. sexta* serine esterases (SEs) and their homologs, 26 phospholipases and 13 thioesterases. Expression profiling of these genes in specific tissues and stages has provided insights into their functions including digestion, detoxification, hormone processing, neurotransmission, reproduction, and developmental regulation. In summary, these studies provide for the first time a holistic view of the digestion and SE-related proteins in a model lepidopteran insect and clues for comparative research in lepidopteran pests and beyond.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter I

Chapter II

Chapter III

Figure Page

Chapter IV

CHAPTER I

IDENTIFICATION OF IMMUNITY-RELATED GENES PREFERENTIALLY REGULATED BY *MANDUCA SEXTA* SPATZLE-1, -2, OR *E. COLI* PEPTIDOGLYCAN

Zelong Miao[1], Yang Wang[1], Chao Xiong[1], Tisheng Shan[1], Michael R. Kanost[2], Haobo Jiang[1]

[1] Department of Entomology and Plant Pathology, Oklahoma State University,

Stillwater, OK 74078, USA

[2] Department of Biochemistry and Molecular Biophysics, Kansas State University,

Manhattan, KS 66506, USA

Key words: insect immunity, RNA-Seq, fat body, hemocytes, Illumina sequencing

Abbreviations: CF, IF, CH and IH, control (C) and induced (I) fat body (F) and hemocytes (H); AMP, antimicrobial peptide/protein; DAP-PG: diaminopimelic acid peptidoglycan; PGRP, peptidoglycan recognition protein; MBP, microbe binding protein; βGRP, β-1,3-glucan recognition protein, HP, hemolymph (serine) protease; PO and proPO, phenoloxidase and its precursor; SP and SPH, serine protease and its homolog; PAP, proPO activating protease; Imd, immune deficiency; qPCR, quantitative real-time polymerase chain reaction.

Abstract

The immune system of *Manduca sexta* has been extensively studied to understand molecular mechanisms of insect antimicrobial responses. While evidence supports the existence of major immune signaling pathways in this species, it is unclear how induced synthesis of defense proteins may be specifically regulated by Toll and Imd pathways. Our recent studies indicated that diaminopimelic acid-type peptidoglycans (DAP-PGs) from Gram-negative and certain Gram-positive bacteria, rather than Lys-PGs from most other Gram-positive bacteria, likely trigger the two pathways through membrane-bound receptors homologous to *Drosophila* Toll and PGRP-LC. In this study, we produced the recombinant Spätzle-1 and -2 precursors, activated them using prophenoloxidase activating protease-3 (PAP3), and separately injected Spätzle-1 and -2 into hemocoel to trigger specific transcriptional up-regulation of genes controlled by the Toll pathway. *M. sexta* cell line was treated with *E. coli* DAP-PG to specifically induce the Imd pathway and its target genes, under the assumption that Imd and Toll pathways are conserved at the functional level. RNA-seq analysis of fat body tissues and cell samples taken at 0, 6, and 24 h after treatment indicated that diapausins and lebocins are predominantly regulated via Toll signaling, gallerimycins, X-tox and its splicing variants are synthesized in response to DAP-PG through Imd pathway, whereas attacins, cecropins, moricins, defensins, gloverin, lysozymes, transferrins, and WAPs are induced via both. Surprisingly, transcripts of the Toll-regulated genes like lebocins peaked at about six hours and this contrasts the *Drosophila* model showing a gradual increase that peaks at 24−48 h. Different time course and specificity of the pathways call for additional investigations in other insects to test validity of these models.

1. Introduction

Insect immunity has been actively studied for over 40 years since the discovery of cecropin, an antimicrobial peptide (AMP) isolated from pupae of *Hyalophora cecropia* (Steiner et al., 1981). Toll, Imd, and other immune signaling pathways are elucidated in *Drosophila* genetic studies (Lemaitre and Hoffmann, 2007; Rämet, 2012; Kingsolver et al., 2013). The Toll pathway mainly responds to fungal and Gram-positive bacterial infection in the fruit fly, which triggers an extracellular serine protease cascade to generate active Spätzle, a ligand of the Toll receptor. On the other hand, *Drosophila* PGRP-LC, Imd, Relish, and other proteins constitute the Imd pathway to fight off Gram-negative bacteria through detection of DAP-PGs. Genome analyses demonstrated the existence of homologous genes encoding the pathway members in other insects (Christophides et al., 2002; Zou et al., 2007; Kanost et al., 2016). However, rigorously scrutinized experimental data are scarce from them regarding specificity of the putative pathways towards different groups of microbes. While overlaps of cross-talks between Toll and Imd signaling are obvious in the fly and other insects (Lindsay and Wasserman, 2014), many researchers take the oversimplified *Drosophila* model as dogma and extrapolate it to other species without validation, not realizing the risk of doing so. Less influenced by that, we characterized the major soluble PGRPs in *M. sexta* hemolymph, all of which preferentially bind to DAP-PGs on the surface of Gram-negative and some Gram-positive bacteria. Along with microbe binding protein (MBP), these PGRPs trigger the protease cascade for melanization and Toll activation (Kariyawasam et al., 2022; Wang et al., 2020). Moreover, a phylogenetic analysis of 360 *Drosophila* PGRP-SA's close homologs in twelve orders of insects suggests that the preference for DAP-PGs is far more common than for Lys-PGs in insects, in terms of

triggering the protease system and Toll signaling.

Even if pathway overlaps or cross-talks are ignored, it is still difficult to distinguish the Toll and Imd responsive genes in *M. sexta* and other species. First of all, DAP-PGs may activate both pathways with overlapping time courses. Secondly, less favored binding of Lys-PGs from most Gram-positive bacteria can still trigger innate immune responses. Thirdly, unique features of Lys-PGs from certain bacteria (*e.g.*, *Micrococcus luteus*) allow some levels of specific binding by PGRPs that prefer DAP-PGs. Structural variations may also exist in DAP-PGs rendering differential recognition by a mixture of PGRPs in insect hemolymph. These complications blurred the line between Gram-positive and -negative bacteria or, more precisely, between Lys- and DAP-type peptidoglycans. It is perhaps a time to consider testing validity of the dogma in other insects, which states "*Drosophila* recruits most of the components of the Toll pathway for antifungal and anti-Gram-positive bacterial defenses" whereas "the Imd pathway govern defense reactions against Gram-negative bacteria" (Hoffmann and Reichhart, 2002).

To distinguish the Toll and Imd responsive genes in *M. sexta*, we took advantage of its genome, transcriptome and proteome data, fat body responsive to *in vitro* activated Spätzle-1 and -2, and a cell line that is directly activated by DAP-PG of *E. coli*. RNA-Seq analysis of the tissue and cell samples taken at different times revealed striking differences between *M. sexta* and *D. melanogaster* and raised serious concerns about applicability of the simple fly model. Our results also established a closer connection between melanization and Toll pathway activation.

2. Methods and materials

2.1 Cloning and express of proSpätzle cDNAs

Primers J196 and J197, J190 and 191, J194 and J195 were designed to amplify proSpätzle-1A, 2 and 7 fragments based on the sequences from Cao et al. (2015). The fat bodies of fifth instar *M. sexta* larvae collected 24 h after injection with a mixture of *E. coli*, *M. luteus*, and curdlan (Wang et al., 2017) were used as cDNA template. The PCR products were cloned into pGEM-T vector (Promega) and confirmed by DNA sequencing. Then the EcoRI-XhoI fragments were inserted into the same sites in pMFH6 in frame with an amino-terminal secretion peptide sequence and carboxyl-terminal with 6× His tag encoded in this vector. The plasmids were used to produce bacmids to infect Sf9 cells, purified from the conditioned media, concentrated, and stored at −80 °C (Sumathipala and Jiang, 2010).

2.2 Activation of recombinant proSpätzles by PAP3

To test the ability of prophenoloxidase activating protease-3 (PAP3) to cleave proSpätzles, 40 ng PAP3 isolated from pharate pupal hemolymph (Jiang et al., 2003b) was incubated with 200 ng proSpätzle-1A, 2 or 7 in the reaction buffer (20 mM Tris, pH7.5, 5 mM $CaCl_2$, 0.001% Tween-20) at room temperature for 1 h, respectively. In the controls, 40 ng PAP3 or 200 ng proSpätzles alone were in the reaction buffer. To test whether proSpätzle-7 can be cleaved or not, 1 μL cell free hemolymph from naïve (NH), immune challenged (IH) or bar (BH) stage larva and 1 μL elicitor (a mixture of *Escherichia coli*, *Micrococcus luteus*, and curdlan) (Wang and Jiang, 2017) are incubated with 200 ng proSpätzle-7 at 37 °C for 4 h with the control of 200 ng proSpätzle-7 only. The reaction mixtures were separated by 12 % SDS/PAGE, followed by electrotransfer and immunoblotting with antibody against 6× His. The cleavage site of proSpätzle-1A and -2 were determined by Edman sequencing at Molecular Structure Facility at the University of California, Davis.

2.3 Deglycosylation of Spätzle2 and 7

To explain the reason of multiple bands in the western blot, deglycosylation experiments were performed using the Peptide: N-Glycosidase F (PNGase F) and O-Glycosidase (also known as Endo-α-N-Acetylgalactosaminidase) from New England Biolabs to catalyze the removal of N or O-linked disaccharides from glycoproteins. The protocol is followed as combine 0.2 μg proSpätzles, 1 μL Glycoprotein Denature Buffer (10X) and H_2O into 10 μL reaction mixture, heat at 100 °C for 10 mins, add 2 μL Glycobuffer 2 (10X), 2 μL10 % NP-40, 0.5 μL PNGase F or 1 μL O-Glycosidase and 1μLNeuraminidase or both for de-N-glycosylation, de-O-glycosylation, or de-N,O-glycosylation and H_2O into 20 μL total reaction at 37 °C for 2 h, with the control of proSpätzles or active Spätzles only. Immunoblot analysis was performed using 10 μL from the reaction samples and rabbit polyclonal antiserum against *M. sexta* Spätzle-2 or 7 as the primary antibody (diluted 1:1000) and goat anti-rabbit IgG-alkaline phosphatase conjugate (Bio-Rad, diluted 1:2500) as the secondary antibody, and a BCIP-NBT substrate kit (Bio-Rad) for color development.

2.4 Purification of active Spätzle-1A or 2 from reaction mixtures

To obtain purified active Spätzle-1A or 2 for injection, first use 0.8 μg active PAP3 to activate 4 μg recombinant proPAP3 (Wang et al., 2014), then incubated with 50 μg purified proSpätzle-1A or 2 at 37 °C for 2 h. After the reaction, the Spätzle-1A and PAP3 mixture was loaded onto 300 μL Ni^{2+}-nitrilotriacetic acid agarose column, following a washing step with 5 mL Ni^{2+} buffer A (50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 0.005% Tween-20, 5 % glycerol, 1 mM BZ, 10 mM imidazole), bound proteins were eluted from the column with a concentration of 50, 75, 100, and 250 mM imidazole in 6, 3, 1.5 and 6 mL Ni^{2+} buffer A, respectively. While the Spätzle-2 and PAP3 mixture was applied to Q-Sepharose FF column (200 μL), following a washing step with 2 mL Q buffer A (20 mM Tris/HCl, pH 8.0), bound

proteins were eluted from the column with a concentration of 0.05, 0.1, 0.2, 0.3, 0.4 and 1 M NaCl in 2 mL Q buffer A for each. Fractions containing active Spätzle-1A or 2 were combined and concentrated on Amicon Ultra-30 centrifugal filter devices (Millipore) using protein change buffer (20 mM Tris/HCl, pH 7.5, 50 mM NaCl). The purified and concentrated Spätzle-1A or 2 were analyzed by SDS/PAGE followed by silver staining in the presence or absence of DTT, respectively. The protein aliquots were rapidly frozen in liquid nitrogen prior to storage at -80 °C.

2.5 Injection of purified active Spätzle-1A and 2 into larva and E. coli PG into M. sexta cell line

To test a possible role of Spätzle-1A and 2 in *M. setxa* immune response, filtered (0.2 μm) Bovine Serum Albumin (BSA, 100 ng/μL × 10 μL=1 μg/larva), purified active Spätzle-1A (25 ng/μL × 40 μL=1 μg/larva) and 2 (100 ng/μL × 10 μL=1 μg/larva) were injected into each of day 1–2, fifth instar larvae. The non-injected naïve larva were as negative controls. All groups had three biological replicates and each replicate contained 3 larvae. The *M. sexta* cell line (FPMI-Ms-12, the neonate larval tissues of the tobacco hornworm, *M. sexta* (Ms), at the Forest Pest Management Institute (FPMI)) was incubated in 90 % Grace's medium with 10% heat-inactivated FBS (insect tissue culture grade) at 28 °C. The cells were distributed into 12-wells plate equally until they were full after around 3-5 days. Before adding *E. coli* PG (10 ng/μL × 10 μL=100 ng/well) into the wells, the media were removed and replaced by the fresh media. The cell wells without any treatments were as negative controls. All groups had three biological replicates and each replicate contained 3 wells. Total RNA samples were prepared from fat body tissues or *M. sexta* cells 6 and 24 h later using TRIZOL Reagent (Thermo Fisher Scientific). The cell-free hemolymph samples from larva after 6, 24, and 48

h treatment were collected and heated at 95 °C for 5 min. The samples were stored at −20 °C.

2.6 Effects of Spätzles on larva immune response

Immunoblot analysis was performed using 1 μL cell-free hemolymph samples and rabbit polyclonal antiserum against *M. sexta* lysoszyme1 or HAIP (loading control) as the primary antibody (diluted 1:2000) and goat anti-rabbit IgG-alkaline phosphatase conjugate (Bio-Rad, diluted 1:2500) as the secondary antibody, and a BCIP-NBT substrate kit (Bio-Rad) for color development. The lysozyme activity was detected using 10 μL cell free hemolymph incubate with 50 μL DQ™ lysozyme substrate (Component A, 50 ng/μL) in 1X Reaction buffer (Component B) from the EnzChek® Lysozyme Assay Kit (E-22013), with the negative and positive control of 10 μL H_2O and 1 μL lysozyme from chicken egg white (Component C, 1000 U). The fluorescence was measured in a fluorescence microplate reader using excitation/emission of ~485/530 nm for 60 minutes at 37ºC. Western blot and lysozyme activity experiments both did three biological replicates.

2.7 Library construction and Illumina sequencing

The total 30 libraries with 3 biological replicates of 10 groups including the non-injected naïve as 0 h control, BSA injected control, Spätzle-1A and 2 at 6 hour and 24 hours, *M. sexta* cell line non-treated control, and *E. coli* PG treatment at 6 hour and 24 hours were constructed by Henry Bellmon Research Center at Oklahoma State University using the KAPA mRNA HyperPrep Kit for Illumina Platforms. The challenged and control libraries were sequenced on Illumina NextSeq 500 platform (Illumina, USA) using High-Output Kit v2.5 (1 × 75 bp). FASTQ files of raw-reads were produced and sorted out by barcodes for further analysis.

2.8 Assembly of transcriptomes and identification of differentially expressed transcripts

(DETs)

RSEM (Li and Dewey, 2011) package was applied for calculation of normalized gene expression value TPM. Subsequently, DETs between the control and treatment libraries were calculated based on the significance level (p value < 0.05) using DESeq2 package in R environment (Love, Huber and Anders, 2014). These transcripts with expression base mean more than 10, $\log_2$ fold change of control vs treatment less than -1 and p adjust value lower than 0.05 were considered as up-regulated induced genes.

2.9 The selection of Toll or Imd preferential candidate genes

Among the induced genes, the minimum $\log_2$ fold change of BSA control vs Spätzle-1A or 2 at 6 or 24 h treatment was selected as the maximum induced level of this gene in Toll pathway, while the minimum $\log_2$ fold change of control vs *E. coli* PG in *M. sexta* cell line at 6 or 24 h was selected as the maximum induced level of this gene in Imd pathway. The value (a) of the minimum $\log_2$ fold change in Imd subtract the minimum $\log_2$ fold change of Toll was the ratio of the maximum induced level of Toll to Imd, the percentage (b) of Toll to Imd pathways was calculated using $2^a/(2^a+1)$, which means the percentage of this gene is regulated by Toll or Imd. The name of genes listed from high to low based on a and b valued stands for the preference of these genes are involved in from Toll to Imd pathways. In order to further show the gene isoform preference of Toll and Imd in the figure, the maximum $\log_2$ fold change of Spätzle vs BSA (Toll) is set as X value, the maximum $\log_2$ fold change of *E. coli* PG vs control (Imd) is set as Y value, 119 isoforms with X, Y values are labeled as blue dots, names of 34 AMPs are listed close to the dots. The closer to the Y or X axis, the more preferential to Toll or Imd pathway, respectively. The red dash line whose slope is 1 stands for the boundary between Toll and Imd (50 % of both pathways).

3. Results

3.1 Sequence analysis of M. sexta proSpätzles cDNAs

We identified a 1251 bp proSpätzle-2 cDNA with an 83 bp 5′-noncoding sequence, a 621 bp ORF encoded a 184 amino acid residues protein with a 22-signal peptide, and a 547 bp 3′-noncoding sequence, including an AATAAA sequence (Fig. 1). The predicted cleaved site is between R74 and Q75. It also has 7 predicted glycosylated sites, which are O-linked glycosylation. The sequence analysis of proSpätzle-1A was reported previously (An et al., 2010). The cDNA of proSpätzle-7 is 823 bp including a 143 bp 5′-noncoding sequence, a 555 bp ORF encoded a 162 amino acid residues protein with a 22 amino acid residues signal peptide, and a 125 bp 3′-noncoding sequence. The predicted cleaved site is between R53 and E54. It just has 1 predicted N-linked glycosylated sites (Fig. S1). These three Spätzles are all overall highly expressed in *M. sexta* in different tissues during various developmental stages (Cao et al., 2015).

3.2 Recombinant Spätzles is a disulfide-linked dimer

The Spätzle-C108 from Spätzle-1A is a disulfide bond-linked dimer with 7 cysctein at the C terminates (An et al., 2010). The Spätzle-2 protein contains 9 cysteine, forming 4 intermolecular disulfide bonds and 1 intramolecular disulfide bond. After cleavage, it comes to 3 intermolecular and 1 intramolecular disfide bonds. The Spätzle-7 protein also contains 9 cysteine, forming 4 intermolecular disulfide bonds and 1 intramolecular disulfide bond. After cleavage, it comes to 3 intermolecular and 1 intramolecular disulfide bonds. After we expressed and purified proSpätzles with six His residues at its C-terminus, SDS/PAGE analysis in the presence of DTT indicated that the purified proSpätzle-1A, 2 and 7 had an apparent molecular mass of 37, 27 and 23 kDa. In the absence of DTT, proSpätzle-1A, 2 and

some of 7 migrated to a position around 60, 40 and 37 kDa (Fig. 2A, B, C; Fig. S2A), suggesting that the recombinant protein is a disulfide-linked dimer.

3.3 ProSpätzle-1A and 2 are activated by protease PAP3

M. sexta clip-domain protease HP8 can cleave proSpätzle-1A to produce the active form (C108) (An et al., 2010). However, HP14, HP21, HP5, HP6 or HP8 cannot cleave proSpätzle-2 or 7 (data not shown). Another serine protease called prophenoloxidase activating protease 3 (PAP3) can cleave both prospatzle-1A and 2 but not 7 into the around 17-19 kDa active form after incubating for 1 hour at room temperature (Fig. 2B, C; Fig. S2A). This band was shown by N-terminal sequencing by Edman degradation to be identical to the Spätzle-1A and 2 sequence beginning at Leu170 (LGPQE) and Gln75(QGDPDA), indicating that it is a truncated form of proSpätzle-1A and 2, cleaved after Arg169 and Arg74. After cleavage running in non-DTT gel, it shows the cleaved Spätzle-1A and 2 are migrated to a position around 23 and 27 kDa, which means they are still in the homo-dimer form (Fig. 2B, C). However, PAP3 cannot cleave proSpätzle-7, the processing enzymes of which can be found in cell free hemolymph of the 5th or bar stage larva (Fig. S2A, B).

3.3 Spätzles are glycosylated proteins

ProSpätzle-2 has 7 O-linked but no N-linked glycosylated sites based on the sequence analysis (https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0 and https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0). After de-N, O-glycosylation experiments, the 3 bands of proSpätzle-2 are all shifted to the lower size band around 25 kDa, indicating that proSpätzle-2 has both N and O-linked glycosylation (Fig. 2D). After cleavage using PAP3, multiple bands of active Spätzle-2 are shifted into one position at around 17 kD, which means active Spätzle-2 are also got glycosylated (Fig. 2D). While

proSpätzle-7 only have N-linked glycosylation based on the prediction. Only after de-N-glycosylation experiments, the two bands of proSpätzle-7 are both shifted to the lower size band around 20 kD (Fig. S2C), indicating that proSpätzle-7 only has N-linked glycosylation, which is consistent with the prediction. The results indicate that the glycosylation mainly account for the multiple bands from Spätzles in the western blot.

3.4 M. sexta Spätzles injection stimulates immune response

After activation of proSpätzle-1A and 2, we purified the active Spätzle-1A (PI: 7.76) and 2 (PI: 6.03) from the total reaction by nickel affinity or Q-Sepharose chromatography, since the dimer form of Spätzles have two 6X His tags, which shows stronger binding affinity to nickel than PAP3 (PI: 6.82, only has one 6X His tag). The silver staining gel shows it only has active Spätzle-1A or 2, which are also in the dimer form (Fig.2 E) and can be used to inject day1–2 5th *M. sexta* larva.

To test whether the injection indeed trigger the immune pathways, the lysozyme protein levels from cell free hemolymph are detected using lysozyme specific antibody. The band intensity of lysozyme1 from cell free hemolymph after injection of purified Spätzle-1A or 2, are much higher than that of non-injected naïve or BSA-injected controls, especially at 24 hours or 48 hours (Fig. 3A). The upper band is HAIP protein that is used as a loading control, which are all similar among different treatments. The lysozyme activity is higher after active Spätzles injection at 24 or 48 h compared to 6 h and other control groups (Fig. 3B), which is consistent with the western blot results. These results from two methods means Spätzles have already activated the immune response pathways in the larva of *M. sexta*.

3.5 Transcriptome analysis upon Spätzle-1A, 2 and E. coli PG treatment

The injection of BSA can only trigger wounding response. Injection of Spätzle-1A or 2

can only trigger the Toll pathway, while *E. coli* PG treatment to *M. sexta* cell line only trigger Imd pathway. After DEseq2 analysis, we found 44 gene isoforms are involved in wounding response (Table S1). Totally, 119 gene isoforms are induced whether in Toll or Imd pathways (Table 1). Among them, 97 gene isoforms are up-regulated upon Spätzles injection (Table S2), 86 isoforms are induced upon Spätzle-1A injection (Table S4), and 87 isoforms are for Spätzle-2 (Table S5), while 48 isoforms are up-regulated upon *E. coli* PG treatment (Table S3). However, 71 isoforms are only induced by Spätzle injection, 22 isoforms are specifically responded to *E. coli* PG treatment. Ten isoforms are specifically induced by Spätzle-1A, eleven isoforms are only for Spätzle-2. Fifty-nine isoforms are expressed higher at 6 than 24 h, 27 isoforms are similar expressed at 6 and 24 h, no isoforms are higher at 24 than 6 h upon Spätzle-1A injection. 54 isoforms are expressed higher at 6 than 24 h, 30 isoforms are similar expressed at 6 and 24 h, 3 isoforms are higher at 24 than 6 h upon Spätzle-2 injection. Overall, most of gene isoforms are express highly at 6 compared to 24 h upon Spätzles injection. Ten isoforms are expressed higher at 6 than 24 h, 32 isoforms are similar expressed at 6 and 24 h, 6 isoforms are higher at 24 than 6 h upon *E. coli* PG treatment (Table S1–5).

Comparison of different effectors based on the expression heat map shows it has total 5 groups (Fig. 4A). Lysozyme1, gloverin, cecropin1 and 15 are in the first group of highly expressed even in all control groups, higher expressed after Spätzles or *E. coli* PG treatment, which can be called wounding response group. The second group has X-toxs, gallerimycin2, attacin6, etc., which have higher expression in the cell line and can be induced upon *E. coli* PG treatment but no too big differences upon Spätzles injection, which can be called Imd-specific response group. The group 3 including the most of cecropins and diapausins are

specifically responded to Spätzles but not *E. coli* PG treatment, which can be called Toll-specific response group A. The group 4 including lysozyme-like-4 protein and defensin-1, 3 are highly expressed in all groups, which can be called non-induction group. The group 5 including lebocins and most of attacins are highly responded to BSA and Spätzles but not *E. coli* PG treatment too much, which can be called Toll-specifc response group B. The 4 proteins in group 1, gallerimycin2 in group 2, attacin in group 3, attacin2 and gallerimycin1 in group 4, attacin1, 4, 7, 10 and cecropin6 in group 5 are all highly induced upon BSA, Spätzles and *E. coli* PG treatments (Fig. 4A). However, the intracellular signaling molecules are not induced too much compared to control groups, except Cactus and Tube in Toll, and IKKgamma_Kenny in Imd pathway (Fig. 4B).

3.6 The preferential index of Toll and Imd pathways among gene isoforms

In order to further analyze the preference of different gene isoforms to Toll or Imd, the ratio of the maximum of Toll induced level to that of Imd level are calculated and listed in the order of high to low, standing for their favor from Toll to Imd.

Based on the Toll to Imd list (Table 1), diapausin4 (98 %), 10, 12, 13 (100 %), cecropin2 (99 %), 4, 5 (98%), attacin5 (97 %), 8 (82 %), 11 (95 %), lebocinA (79 %), B (75 %), C (61 %), D (97 %) etc. are preferentially reacted to the Toll pathway. Among them, only diapausin4, 10, 12 have the similar pattern with drosomycin of *D. melanogaster*, whose expression level are higher at 24 than 6 h. However, most of the other AMPs are expressed higher at 6 than 24 h. X-toxs (5-tox1, 22 % and 4-tox2, 12 %) are preferentially triggered by Imd pathway with the pattern of rapidly increase at 6 hours and then reduce at 24 hours, similar with diptericin of *D. melanogaster*. While most of other AMPs including moricins, attacins, cercropins, gloverin etc., they can be triggered by both pathways. When the

maximum $\log_2$ fold change of Spätzle vs BSA (Toll) is set as X value, the maximum $\log_2$ fold change of *E. coli* PG vs control (Imd) is set as Y value, the closer to the Y or X axis, the more preferential to Toll or Imd pathway, respectively. Diapausin4, 10, 12, 13 are all on the Y axis, lebocinD is also close to the Y axis, indicating they are more Toll-pathway specific. However, 4-tox2 and galllerimycin2 are closer to the X axis, standing for it is in the Imd-specific pathway (Fig. 5). Based on this result with a clear and specific signal induction, we select diapausins and lebocins are Toll pathway specific genes, and gallerimycin and X-toxs are Imd pathway specific genes.

4. Discussions

4.1 The connection between Toll pathways and melanization

In the genome level, seven genes encoding Spätzle-like proteins in *M. sexta* were identified (Cao, et al., 2015). Spätzle-3−6 contain 10, 10, 8 and 6 Cys residues after cleavage, which might have unclear different mechanism to dimerize. Additionally, the mRNA levels of Spätzle-3−6 are very low even upon immune challenge, compared to Spätzle-1, 2 and 7, which have the highest expression level with FPKM values of 224, 760 and 564, respectively (Cao et al, 2015). Thus, Spätzle-1, 2 and 7 can stand for the overall Spätzles level in the immune system of *M. sexta*. HP6 can cleave proHP8, which can activate proSpätzle-1A to trigger the Toll pathway. However, activated HP8 cannot cleave either proSpätzle-2 or 7. Another enzyme involved in prophenoloxidase activation called PAP3 can cleave both proSpätzle-1A and 2 efficiently and correctly (Fig. 2 B and C), which means the activation of proSpätzle and prophenoloxidase can be performed by the same clip-domain protease to activate both Toll and melanization pathways. The result is similar as the SPE in *Te. molitor* (Kim et al., 2008 and Kan et al., 2008). However, the SPE to activate proSpätzle-7 is still

unknown, the future identification work can focus on the proteases with high expression in the hemolymph of 5th or bar stage of larva.

4.2 False positive result in Imd pathway-preferential genes

The gene induced level to Toll pathway is set as the maximum fold change of active Spätzle-1A or 2 at 6 or 24 h vs BSA at 6 or 24 h. Some genes like lysozyme1, golverin, attacin2 and cecropin1, 6, 15 have already highly induced upon BSA injection (wounding response can also trigger the Toll pathway in some level) (Fig. 4A), even though the expression level is higher upon Spätzle injection, the ratio is still low since the denominator is too high and the numerator has already reached the expression peak. However, for the Imd pathway induction, no wounding response control can be used, the induced fold change can be larger. Therefore, the genes including lysozyme1, golverin, attacin2 and cecropin1, 6, 15 etc. are not Imd-specific genes. To exclude the false positive genes that favor in Imd pathway, the expression heat map also needs to be used, because of wounding response on the larva injection. Another false positive example is IML5 at the end of Toll-Imd list (Table 1), the $\log_2$ Fold change value (16.612) of *E. coli* PG at 6 h vs control is very high, since the control value is 0, but the overall value in the heat map of Ms-12 cell line are almost 0. Therefore, IML5 is not an Imd pathway-preferential gene. Based on the Toll-Imd index (Table 1) and expression map (Fig. 4) together, we are confident that gallerimycin and X-tox are Imd pathway preferential candidate genes.

4.3 The comparison of M. sexta larva and cell line

There are four types of cells colonies (Ms-4, epithelial-like cells; Ms-5, dendrite-like cells; Ms-7, round cells; Ms-12, long fibro blast-like cells) observed in the neonate tissue cultures of *M. sexta* (Sohi, 1995). FPMI-Ms-12 is confirmed by using the method of cloning

and sequencing Spätzle-1A gene through the primer J196 and J197 in *M. sexta.* One amino acid (R to G) of Ms-12 and 5 amino acids of Ms-7 are mismatched with the total 276 amino acids of Spätzle-1A in the current *M. sexta* genome (data not shown). Thus, Ms-12 is selected as the cell line to trigger Imd pathway with the higher similarity with current *M. sexta* genome. The fat body is selected as the major immune organ in *M. sexta* larva, since the induction levels of genes in fat body are higher than hemocytes or midgut (data not shown). However, the variance may exist between the larva and cell lines, including the stages, tissues, and conditions of *in vivo* and *in vitro*. The comparison index between *M. sexta* larva and cell line indeed has the limitations, since the *in vitro* Imd induced level in Ms-12 cannot represent that in the real condition *in vivo*. Nonetheless, due to the complexity of Toll and Imd extracellular induction *in vivo* in *M. sexta*, the cell line with fresh media is the only way we can use to get rid of extracellular serine protease cascades and cytokines to trigger Imd pathway specifically but not Toll pathway.

4.4 Concluding remarks

In conclusion, the results show that PAP3 is an activating enzyme of proSpätzle-1A and -2, building the connection between Toll pathway and melanization. This paper also first identified 2 groups of genes (diapausins and lebocins) as Toll pathway specific gene candidate, while gallerimycin and X-tox genes as Imd pathway specific genes candidate in Lepidoptera with the clear and single signal molecules using biochemistry methods in *M. sexta*. Most of other AMPs are involved in both pathways. Further biochemical work like activation of proSpätzle-7 need to be done and genetic research is required to confirm the specificity of these candidates.

Tables

Table 1. The ratio of Toll or Imd induced level of 119 immune-related gene isoforms.

Transcript id	Spz vs BSA	EcPG vs Ctrl	log_2FC (Toll/Imd)	Toll (%)	Transcript id	Spz vs BSA	EcPG vs Ctrl	log_2FC (Toll/Imd)	Toll (%)	Transcript id	Spz vs BSA	EcPG vs Ctrl	log_2FC (Toll/Imd)	Toll (%)
Diapausin12	8.173	0.000	8.173	99.7	Domeless	2.052	0.498	1.554	74.6	IML1	1.815	2.104	-0.289	45.0
Diapausin10	8.153	0.000	8.153	99.6	LebocinB	3.221	1.672	1.549	74.5	Attacin4	3.530	3.883	-0.353	43.9
Diapausin13	7.947	0.000	7.947	99.6	Moricin1	3.801	2.277	1.523	74.2	Transferrin2	1.186	1.723	-0.536	40.8
PGRP7	9.343	2.891	6.452	98.9	Attacin1	4.777	3.258	1.519	74.1	PGRP3	2.273	2.868	-0.595	39.8
Cecropin2	6.331	0.000	6.331	98.8	KAL1 anosmin	2.168	0.731	1.437	73.0	Attacin10	4.599	5.287	-0.688	38.3
Diapausin4	5.747	0.000	5.747	98.2	Serpin3	2.230	0.841	1.389	72.4	Atg101	0.458	1.202	-0.744	37.4
Cecropin5	5.638	0.000	5.638	98.0	Pvr	1.444	0.124	1.319	71.4	Attacin7	4.480	5.243	-0.763	37.1
Cecropin4	5.405	0.000	5.405	97.7	PGRP2	5.214	3.902	1.311	71.3	SPH1b	0.237	1.030	-0.792	36.6
IML4	5.474	0.182	5.293	97.5	Toll10_2	1.730	0.435	1.295	71.0	SPH101	0.428	1.277	-0.849	35.7
βGRP2	5.508	0.280	5.228	97.4	Toll5	1.301	0.040	1.261	70.6	PGRP4	3.726	4.624	-0.897	34.9
PAP3	6.230	1.110	5.120	97.2	LLP4	1.463	0.203	1.259	70.5	Hemolin	3.143	4.106	-0.963	33.9
Attacin5	5.793	0.784	5.008	97.0	HP5	2.981	1.734	1.247	70.4	Atg14	0.880	1.977	-1.096	31.9
LebocinD	6.240	1.426	4.814	96.6	PAP2	3.261	2.038	1.223	70.0	Serpin5	1.009	2.184	-1.174	30.7
Attacin11	4.987	0.784	4.203	94.8	Cecropin12	3.466	2.265	1.200	69.7	Atg1	0.173	1.400	-1.227	29.9
GP33	4.783	0.900	3.883	93.7	Leureptin8	1.134	-0.010	1.144	68.9	Reeler1	2.586	3.897	-1.311	28.7
SP34	4.230	0.376	3.855	93.5	SPH3	2.817	1.686	1.130	68.6	LRR-TMP2	-0.201	1.134	-1.336	28.4
Cathepsinlike3	3.473	0.239	3.234	90.4	Pelle	2.014	0.980	1.034	67.2	HP21	0.096	1.461	-1.364	28.0
Scolexin A	4.507	1.638	2.869	88.0	Aop	1.074	0.044	1.030	67.1	PVF2	0.091	1.519	-1.429	27.1
Serpin2	2.831	0.081	2.750	87.1	HP19	1.994	1.014	0.980	66.4	Tab2	0.257	1.693	-1.436	27.0
Serpin11	2.944	0.224	2.720	86.8	Attacin3	6.531	5.566	0.965	66.1	HP20	1.115	2.566	-1.450	26.8
Cactus	3.226	0.508	2.718	86.8	SP58	1.972	1.069	0.904	65.2	IML2	1.061	2.621	-1.560	25.3
IML15	3.214	0.497	2.717	86.8	Serpin10	1.770	0.945	0.825	63.9	5tox1	1.857	3.681	-1.823	22.0
WAP1	3.243	0.694	2.549	85.4	ScolexinB	1.308	0.575	0.733	62.4	Cecropin15	2.521	4.377	-1.857	21.6
SPH4	3.737	1.215	2.523	85.2	Lysozymelike1	1.693	1.000	0.693	61.8	Gallerimycin1	3.408	5.351	-1.943	20.6
HP7	2.553	0.037	2.516	85.1	LebocinC	3.638	2.975	0.663	61.3	βGRP3	0.099	2.092	-1.993	20.1
CTL-X5	3.808	1.443	2.365	83.7	Caspar	1.097	0.450	0.647	61.0	SRP1	2.665	4.699	-2.034	19.6
Serpin6	3.127	0.853	2.274	82.9	Kurtz	1.515	0.927	0.589	60.1	Spätzle2	1.323	3.703	-2.380	16.1
Serpin14	2.580	0.333	2.248	82.6	CathepsinFlike	1.381	0.804	0.577	59.9	Attacin2	3.037	5.636	-2.599	14.2
Attacin8	4.614	2.383	2.231	82.4	Attacin6	3.653	3.089	0.564	59.7	4tox2	0.837	3.709	-2.871	12.0
HP17a	5.805	3.604	2.202	82.1	LRR-TMP18	1.667	1.156	0.511	58.8	MBP	0.528	3.420	-2.891	11.9
NEMO	2.042	0.013	2.029	80.3	HP22	1.766	1.302	0.464	58.0	HP9	2.417	5.311	-2.894	11.9
PAP1	2.845	0.848	1.996	80.0	Serpin12	1.392	0.981	0.412	57.1	Cecropin1	1.993	5.037	-3.044	10.8
Tube	1.874	-0.105	1.979	79.8	Relish	1.463	1.339	0.124	52.1	PGRP5	2.201	5.449	-3.248	9.5
Defensin3	2.356	0.406	1.950	79.4	PGRP10	1.288	1.250	0.038	50.7	Cecropin6	3.494	6.940	-3.447	8.4
LebocinA	2.552	0.604	1.949	79.4	Atg9	1.179	1.147	0.032	50.5	Gloverin	2.056	5.658	-3.602	7.6
Serpin4	1.655	-0.173	1.828	78.0	IKKγ_Kenny	1.231	1.297	-0.067	48.8	Lysozyme1	1.315	5.103	-3.788	6.8
SPH33	2.796	1.058	1.738	76.9	IML12	1.906	2.103	-0.197	46.6	Gallerimycin2	1.852	5.883	-4.031	5.8
SRP3	2.871	1.201	1.670	76.1	FOS	1.171	1.377	-0.206	46.4	Serpin1D	0.502	5.190	-4.688	3.7
CathBlike5	1.504	-0.114	1.618	75.4	HP14a	0.968	1.186	-0.218	46.2	PGRP1	0.983	6.062	-5.079	2.9
Leureptin3	3.967	2.413	1.554	74.6	IML1	1.815	2.104	-0.289	45.0	IML5	0.858	16.612	-15.754	0.0

Table S1. The induced levels of 44 BSA wounding response genes

Transcript_id	FB_control__vs__FB_BSA_6h|log2FoldChange	FB_control__vs__FB_BSA_6h|padj	FB_control__vs__FB_BSA_24h|log2FoldChange	FB_control__vs__FB_BSA_24h|padj
Dredd	-1.267	0.018	-0.187	0.972
FADD	-1.677	0.031	-0.660	0.769
HP17a	-1.382	0.042	-0.209	0.982
HP9	-5.055	0.013	-3.125	0.448

IML-12	-1.580	0.000	-0.685	0.329
IML-15	-1.964	0.002	-1.624	0.049
IML-2	-2.028	0.000	-0.360	0.726
PAP1	-1.720	0.000	-0.639	0.495
PAP2	-1.508	0.039	-1.020	0.393
PGRP1	-2.734	0.000	-0.787	0.332
PGRP2	-5.806	0.000	-0.701	0.971
PGRP4	-9.215	0.000	-2.135	0.757
PGRP5	-4.113	0.000	-0.572	0.675
Ref2P-C	-7.539	0.014	-5.845	0.159
Relish	-2.685	0.000	0.245	0.878
SP75	-2.960	0.000	-0.653	0.793
5-tox1	-4.301	0.020	-2.130	0.643
attacin1	-7.535	0.000	-2.802	0.108
attacin10	-6.463	0.000	-2.261	0.341
attacin11	-9.278	0.000	-3.573	0.280
attacin2	-4.514	0.000	-1.716	0.280
attacin3	-7.757	0.000	-0.599	0.999
attacin4	-6.554	0.000	-2.019	0.014
attacin5	-6.665	0.000	-2.311	0.280
attacin7	-7.076	0.000	-1.881	0.058
attacin8	-10.284	0.000	-5.045	0.002
cecropin1	-5.525	0.000	-0.790	0.492
cecropin12	-5.989	0.000	0.291	1.000
cecropin15	-5.693	0.000	-1.315	0.095
cecropin5	-5.952	0.000	-3.060	0.291
cecropin6	-7.454	0.000	-2.259	0.115
gallerimycin1	-8.888	0.000	-4.360	0.061
gallerimycin2	-4.749	0.014	-0.667	0.997
gloverin	-6.780	0.000	-3.263	0.000
hemolin	-4.403	0.000	-3.596	0.000
lebocin-A	-6.372	0.000	-4.249	0.000
lebocin-B	-7.489	0.000	-6.457	0.000
lebocin-C	-6.934	0.000	-5.170	0.000
lebocin-D	-7.286	0.000	-6.519	0.000
leureptin8	-1.283	0.006	-0.906	0.195
lysozyme1	-2.334	0.000	-0.140	1.000
moricin1	-6.472	0.000	-1.551	0.880
reeler1	-5.559	0.000	-2.045	0.005
serpin-5	-1.088	0.022	-0.132	0.991

*: Numbers shaded *pink* were $\log_2$FoldChange<-1, numbers shaded *lime* were *p*adj<0.05.

Table S2. The induced levels of 97 Spätzle induced genes

Transcript_id	FB_BSA_6h__vs_FB_1A_6h|log2FoldChange	FB_BSA_6h__vs_FB_1A_6h|padj	FB_BSA_24h__vs__FB_1A_24h|log2FoldChange	FB_BSA_24h__vs__FB_1A_24h|padj	FB_BSA_6h__vs_FB_S2_6h|log2FoldChange	FB_BSA_6h__vs_FB_S2_6h|padj	FB_BSA_24h__vs__FB_S2_24h|log2FoldChange	FB_BSA_24h__vs__FB_S2_24h|padj
diapausin10	-5.093	0.000	-8.153	0.000	-4.850	0.001	-7.512	0.000
diapausin12	-5.444	0.004	-8.173	0.000	-5.687	0.003	-7.380	0.001
diapausin13	-7.898	0.000	-5.362	0.000	-7.947	0.000	-5.190	0.000
diapausin4	-3.548	0.008	-5.747	0.000	-4.671	0.000	-4.497	0.002
Aop	-1.074	0.005	-0.462	0.702	-0.828	0.057	-0.145	0.993
Atg9	-1.179	0.013	-0.305	0.959	-0.780	0.177	-0.306	0.934
CTL-X5	-3.808	0.001	0.143	1.000	-2.525	0.053	-2.013	0.399
Cactus	-3.226	0.000	-0.268	0.736	-2.912	0.000	-0.424	0.363
Caspar	-1.097	0.000	-0.112	1.000	-0.306	0.534	0.062	1.000
Cathepsin26_29kDalike3	-1.018	0.564	2.207	0.313	-2.454	0.051	-3.473	0.008
CathepsinBlike5	-1.304	0.007	0.521	0.770	-1.504	0.002	0.126	1.000
Domeless	-1.709	0.000	-0.856	0.182	-2.052	0.000	-0.786	0.308
GP33	-4.749	0.000	-1.795	0.689	-4.783	0.000	-2.975	0.224
HP17a	-5.805	0.000	-2.172	0.000	-5.389	0.000	-1.061	0.178
HP19	-1.471	0.000	0.193	0.983	-1.994	0.000	-0.304	0.853
HP20	-0.620	0.286	-0.690	0.502	-1.087	0.026	-1.115	0.080
HP22	-1.545	0.000	-1.261	0.000	-1.766	0.000	-0.827	0.062
HP5	-2.981	0.000	-1.049	0.647	-2.920	0.000	-0.403	0.977
HP7	-1.909	0.009	1.472	0.481	-2.553	0.000	-0.588	0.900
HP9	-2.417	0.003	1.546	0.789	-1.984	0.029	-0.689	0.947
IKKgamma_Kenny	-1.231	0.008	-0.086	1.000	-1.140	0.021	-0.061	1.000
IML-1	-1.394	0.027	-0.987	0.416	-1.575	0.012	-1.815	0.010
IML-12	0.025	1.000	-1.479	0.000	0.156	0.858	-1.906	0.000
IML-15	-3.214	0.000	-2.353	0.000	-2.239	0.000	-1.262	0.017
IML-2	-0.430	0.277	-1.061	0.001	-0.434	0.293	-0.909	0.016
IML-4	-3.913	0.000	-4.370	0.000	-4.830	0.000	-5.474	0.000
KAL-1_anosmin	-2.168	0.000	-1.114	0.312	-1.588	0.014	0.427	0.913
Kurtz	-1.357	0.004	0.022	1.000	-1.515	0.001	0.567	0.717
LRR-TMP18	-0.696	0.528	0.736	0.772	-1.667	0.033	1.140	0.468
Multicystatin_procathepsinF-like	-1.381	0.000	-0.738	0.079	-1.372	0.000	-0.861	0.026
NEMO	-2.042	0.001	0.222	1.000	-1.570	0.026	-0.329	0.978
PAP1	-2.845	0.000	-1.268	0.003	-2.318	0.000	-1.026	0.042
PAP2	-3.261	0.000	-2.687	0.000	-2.981	0.000	-2.883	0.000
PAP3	-6.230	0.000	-3.685	0.000	-5.723	0.000	-3.066	0.000
PGRP2	-4.373	0.000	-5.214	0.000	-4.267	0.000	-3.697	0.001
PGRP3	-0.953	0.130	-2.273	0.000	-0.813	0.246	-1.251	0.156
PGRP4	-1.973	0.005	-2.048	0.502	-1.861	0.012	-3.726	0.020
PGRP5	-2.201	0.000	-1.311	0.006	-2.184	0.000	-2.126	0.000
PGRP7	-9.199	0.000	-1.305	0.963	-9.343	0.000	-1.644	0.879
Pelle	-1.887	0.000	-0.371	0.902	-2.014	0.000	-0.391	0.860
Pvr	-0.698	0.040	0.069	1.000	-1.444	0.000	-0.166	0.967
Relish	-1.463	0.000	-0.246	0.916	-1.118	0.000	-0.835	0.062
SP34	-4.230	0.000	-2.205	0.000	-4.091	0.000	-2.215	0.000
SP58	-1.282	0.008	-0.463	0.828	-1.972	0.000	-1.013	0.207
SPH3	-2.674	0.000	-2.817	0.000	-1.650	0.007	-2.442	0.000

SPH33	-2.796	0.000	-2.518	0.000	-1.821	0.004	-2.416	0.000
SPH4	-3.515	0.000	-3.737	0.000	-2.542	0.000	-3.158	0.000
SRP1	-2.460	0.014	1.468	0.948	-2.665	0.008	0.210	1.000
SRP3	-2.077	0.028	0.795	0.915	-2.871	0.001	-1.235	0.666
Spz2	-0.360	0.745	-0.634	0.762	-1.323	0.038	0.016	1.000
Toll10_2	-0.725	0.471	0.531	0.915	-1.730	0.018	-0.415	0.943
Toll5	-1.106	0.006	-0.945	0.090	-1.251	0.002	-1.301	0.004
Tube	-1.874	0.000	-0.499	0.309	-1.672	0.000	-0.912	0.003
WAP1	-1.791	0.154	-0.718	0.925	-3.243	0.002	-2.299	0.109
attacin1	-4.777	0.000	-2.096	0.133	-4.240	0.000	-3.680	0.000
attacin10	-4.599	0.000	-1.395	0.603	-4.036	0.000	-2.551	0.070
attacin11	-4.987	0.000	-2.941	0.006	-4.487	0.000	-3.507	0.000
attacin2	-3.037	0.000	-0.759	0.840	-1.955	0.023	-1.943	0.099
attacin3	-3.939	0.000	-4.339	0.016	-3.758	0.000	-6.531	0.000
attacin4	-3.530	0.000	-2.339	0.000	-2.967	0.000	-3.197	0.000
attacin5	-5.462	0.000	-5.793	0.000	-4.901	0.000	-4.864	0.000
attacin7	-4.480	0.000	-2.691	0.000	-3.984	0.000	-3.641	0.000
attacin8	-4.614	0.000	-2.151	0.014	-4.042	0.000	-3.066	0.000
bGRP2	-5.508	0.000	-3.786	0.000	-4.080	0.000	-2.318	0.019
cecropin1	-1.297	0.004	-1.525	0.003	-0.825	0.126	-1.993	0.000
cecropin12	-1.004	0.091	0.112	1.000	-2.398	0.000	-3.466	0.020
cecropin15	-1.042	0.064	-1.836	0.001	-1.060	0.069	-2.521	0.000
cecropin2	-5.310	0.000	-5.823	0.000	-5.051	0.000	-6.331	0.000
cecropin4	-5.405	0.000	-3.265	0.001	-5.382	0.000	-3.920	0.000
cecropin5	-2.821	0.000	-5.606	0.000	-3.017	0.000	-5.638	0.000
cecropin6	-3.195	0.000	-2.733	0.000	-2.667	0.000	-3.494	0.000
defensin3	-1.257	0.051	-1.041	0.331	-2.298	0.000	-2.356	0.000
gallerimycin1	-2.795	0.000	-3.101	0.000	-2.890	0.000	-3.408	0.000
gallerimycin2	-1.852	0.004	0.507	1.000	-1.341	0.079	0.211	1.000
gloverin	-2.056	0.001	-2.018	0.004	-1.305	0.067	-1.631	0.054
hemolin	-0.739	0.535	-1.337	0.378	-1.459	0.116	-3.143	0.000
lebocin-A	-1.404	0.004	-2.552	0.000	-1.114	0.041	-2.442	0.000
lebocin-B	-3.221	0.000	-2.392	0.048	-2.757	0.004	-2.597	0.029
lebocin-C	-2.913	0.000	-3.638	0.000	-2.553	0.000	-2.449	0.000
lebocin-D	-5.724	0.000	-6.240	0.000	-5.091	0.000	-5.543	0.000
leureptin3	-3.864	0.000	-1.517	0.218	-3.967	0.000	-1.423	0.342
leureptin8	-1.038	0.005	-1.134	0.009	-0.343	0.578	-0.394	0.780
lysozyme-like-protein-1	-1.526	0.000	-0.689	0.240	-1.693	0.000	-1.049	0.013
lysozyme-like-protein-4	-1.463	0.004	-0.490	0.824	-1.236	0.026	-0.714	0.599
moricin1	-3.615	0.000	-2.663	0.378	-3.481	0.000	-3.801	0.076
reeler1	-1.160	0.047	-2.586	0.000	-0.554	0.514	-2.264	0.000
scolexinA	-3.831	0.000	-2.407	0.100	-4.507	0.000	-3.937	0.000
scolexinB	-1.288	0.000	0.281	0.917	-1.308	0.000	-0.456	0.696
serpin-10	-1.770	0.000	0.180	1.000	-1.307	0.010	-0.118	1.000
serpin-11	-2.793	0.000	-0.940	0.016	-2.944	0.000	-1.217	0.000
serpin-12	-1.392	0.000	-0.659	0.032	-1.380	0.000	-0.776	0.006
serpin-14	-2.158	0.000	-0.159	1.000	-2.580	0.000	0.123	1.000
serpin-2	-1.901	0.005	-0.058	1.000	-2.244	0.001	-2.831	0.000
serpin-3	-2.230	0.000	-0.591	0.445	-2.031	0.000	-0.558	0.524
serpin-4	-1.646	0.000	-0.657	0.001	-1.655	0.000	-0.682	0.001

serpin-5	-0.700	0.110	-0.634	0.413	-0.595	0.221	-1.009	0.047
serpin-6	-2.766	0.000	-0.930	0.398	-3.127	0.000	-0.981	0.375

*: Seventy-one genes shaded *yellow* were only specifically induced by Spätzles.

Table S3. The induced levels of 48 *E. coli* PG induced genes

	Transcript_id	Ms12_control__vs__Ms12_EcPG_6h|log2FoldChange	Ms12_control__vs__Ms12_EcPG_6h|padj	Ms12_control__vs__Ms12_EcPG_24h|log2FoldChange	Ms12_control__vs__Ms12_EcPG_24h|padj
B	Atg1	0.034	1.000	-1.400	0.001
B	Atg101	0.004	1.000	-1.202	0.018
B	Atg14	-0.106	1.000	-1.977	0.001
C	Atg9	-0.200	1.000	-1.147	0.021
C	FOS	-0.146	1.000	-1.377	0.038
C	HP14a	-1.186	0.004	-0.217	0.842
C	HP17a	-3.604	0.001	-2.292	0.087
A	HP21	-1.461	0.000	-0.201	0.784
C	HP9	-5.311	0.000	-4.226	0.000
C	IKKgamma_Kenny	-0.680	1.000	-1.297	0.003
A	IML5	-16.612	0.000	0.000	1.000
B	LRR-TMP2	0.320	1.000	-1.134	0.011
A	MBP	-3.420	0.000	-2.045	0.000
A	PGRP1	-6.062	0.000	-5.052	0.000
C	PGRP10	-0.479	1.000	-1.250	0.041
A	PGRP2	-3.902	0.004	-1.952	0.286
C	PGRP3	-2.868	0.000	-2.194	0.000
C	PGRP4	-4.624	0.000	-3.370	0.000
C	PGRP5	-5.449	0.000	-5.354	0.000
C	PVF2	-0.761	1.000	-1.519	0.003
A	Relish	-1.339	0.000	-0.223	0.775
C	SPH101	-1.214	0.013	-1.277	0.003
C	SPH1b	-1.030	0.001	-0.877	0.004
C	SRP1	-4.699	0.000	-3.371	0.000
B	Spz2	-0.496	1.000	-3.703	0.000
B	Tab2	-0.210	1.000	-1.693	0.000
C	4-tox2	-3.709	0.001	-3.402	0.002
C	5-tox1	-3.681	0.000	-3.095	0.000
C	attacin10	-5.287	0.020	-3.487	0.148
C	attacin2	-5.636	0.000	-5.629	0.000
C	attacin3	-5.566	0.000	-3.969	0.005
A	attacin4	-3.883	0.002	-1.726	0.344
C	attacin6	-2.111	0.001	-3.089	0.000
A	attacin7	-5.243	0.000	-2.945	0.023
C	bGRP3	-1.645	0.001	-2.092	0.000
A	cecropin1	-5.037	0.000	-2.561	0.000

C	cecropin15	-4.377	0.000	-3.587	0.000
C	cecropin6	-6.940	0.000	-6.260	0.000
C	gallerimycin1	-5.092	0.000	-5.351	0.000
C	gallerimycin2	-5.091	0.000	-5.883	0.000
C	gloverin	-5.658	0.000	-5.309	0.000
C	hemolin	-4.106	0.000	-3.840	0.000
C	lebocin-C	-2.532	0.000	-2.975	0.000
C	lysozyme1	-5.103	0.000	-3.872	0.000
C	reeler1	-3.298	0.000	-3.897	0.000
A	serpin-1D	-5.190	0.005	0.000	1.000
C	serpin-5	-2.184	0.000	-1.294	0.006
C	transferrin2	-1.723	0.049	-0.472	0.755

*: Twenty-two genes shaded *blue* were only specifically induced by *E. coli* PG. 10 genes labeled with A were expressed higher at 6 than 24 h, 6 genes labeled with # were expressed higher at 24 than 6 h, 32 genes labeled with C were expressed similar at 6 and 24 h.

Table S4. The summary of the names of 86 Spätzle-1A induced genes

	Transcript_id	FB_BSA_6h__vs__FB_1A_6h|log2FoldChange	FB_BSA_6h__vs__FB_1A_6h|padj	FB_BSA_24h__vs__FB_1A_24h|log2FoldChange	FB_BSA_24h__vs__FB_1A_24h|padj
C	diapausin10	-5.093	0.000	-8.153	0.000
C	diapausin12	-5.444	0.004	-8.173	0.000
C	diapausin13	-7.898	0.000	-5.362	0.000
C	diapausin4	-3.548	0.008	-5.747	0.000
C	Aop	-1.074	0.005	-0.462	0.702
A	Atg9	-1.179	0.013	-0.305	0.959
A	CTL-X5	-3.808	0.001	0.143	1.000
A	Cactus	-3.226	0.000	-0.268	0.736
A	Caspar	-1.097	0.000	-0.112	1.000
A	CathepsinBlike5	-1.304	0.007	0.521	0.770
A	Domeless	-1.709	0.000	-0.856	0.182
A	GP33	-4.749	0.000	-1.795	0.689
A	HP17a	-5.805	0.000	-2.172	0.000
A	HP19	-1.471	0.000	0.193	0.983
C	HP22	-1.545	0.000	-1.261	0.000
A	HP5	-2.981	0.000	-1.049	0.647
A	HP7	-1.909	0.009	1.472	0.481
A	HP9	-2.417	0.003	1.546	0.789
A	IKKgamma_Kenny	-1.231	0.008	-0.086	1.000
C	IML-1	-1.394	0.027	-0.987	0.416
C	IML-12	0.025	1.000	-1.479	0.000
A	IML-15	-3.214	0.000	-2.353	0.000
A	IML-2	-0.430	0.277	-1.061	0.001
C	IML-4	-3.913	0.000	-4.370	0.000

C	KAL-1_anosmin	-2.168	0.000	-1.114	0.312
A	Kurtz	-1.357	0.004	0.022	1.000
A	Multicystatin_procathepsinF-like	-1.381	0.000	-0.738	0.079
A	NEMO	-2.042	0.001	0.222	1.000
A	PAP1	-2.845	0.000	-1.268	0.003
C	PAP2	-3.261	0.000	-2.687	0.000
A	PAP3	-6.230	0.000	-3.685	0.000
A	PGRP2	-4.373	0.000	-5.214	0.000
C	PGRP3	-0.953	0.130	-2.273	0.000
A	PGRP4	-1.973	0.005	-2.048	0.502
A	PGRP5	-2.201	0.000	-1.311	0.006
A	PGRP7	-9.199	0.000	-1.305	0.963
A	Pelle	-1.887	0.000	-0.371	0.902
A	Relish	-1.463	0.000	-0.246	0.916
A	SP34	-4.230	0.000	-2.205	0.000
A	SP58	-1.282	0.008	-0.463	0.828
C	SPH3	-2.674	0.000	-2.817	0.000
C	SPH33	-2.796	0.000	-2.518	0.000
C	SPH4	-3.515	0.000	-3.737	0.000
A	SRP1	-2.460	0.014	1.468	0.948
A	SRP3	-2.077	0.028	0.795	0.915
C	Toll5	-1.106	0.006	-0.945	0.090
A	Tube	-1.874	0.000	-0.499	0.309
A	attacin1	-4.777	0.000	-2.096	0.133
A	attacin10	-4.599	0.000	-1.395	0.603
A	attacin11	-4.987	0.000	-2.941	0.006
A	attacin2	-3.037	0.000	-0.759	0.840
A	attacin3	-3.939	0.000	-4.339	0.016
A	attacin4	-3.530	0.000	-2.339	0.000
A	attacin5	-5.462	0.000	-5.793	0.000
A	attacin7	-4.480	0.000	-2.691	0.000
A	attacin8	-4.614	0.000	-2.151	0.014
C	bGRP2	-5.508	0.000	-3.786	0.000
A	cecropin1	-1.297	0.004	-1.525	0.003
A	cecropin15	-1.042	0.064	-1.836	0.001
C	cecropin2	-5.310	0.000	-5.823	0.000
C	cecropin4	-5.405	0.000	-3.265	0.001
C	cecropin5	-2.821	0.000	-5.606	0.000
A	cecropin6	-3.195	0.000	-2.733	0.000
A	gallerimycin1	-2.795	0.000	-3.101	0.000
A	gallerimycin2	-1.852	0.004	0.507	1.000
A	gloverin	-2.056	0.001	-2.018	0.004
C	lebocin-A	-1.404	0.004	-2.552	0.000
C	lebocin-B	-3.221	0.000	-2.392	0.048

A	lebocin-C	-2.913	0.000	-3.638	0.000
C	lebocin-D	-5.724	0.000	-6.240	0.000
A	leureptin3	-3.864	0.000	-1.517	0.218
C	leureptin8	-1.038	0.005	-1.134	0.009
A	lysozyme-like-protein-1	-1.526	0.000	-0.689	0.240
C	lysozyme-like-protein-4	-1.463	0.004	-0.490	0.824
A	moricin1	-3.615	0.000	-2.663	0.378
A	reeler1	-1.160	0.047	-2.586	0.000
C	scolexinA	-3.831	0.000	-2.407	0.100
A	scolexinB	-1.288	0.000	0.281	0.917
A	serpin-10	-1.770	0.000	0.180	1.000
A	serpin-11	-2.793	0.000	-0.940	0.016
C	serpin-12	-1.392	0.000	-0.659	0.032
A	serpin-14	-2.158	0.000	-0.159	1.000
A	serpin-2	-1.901	0.005	-0.058	1.000
A	serpin-3	-2.230	0.000	-0.591	0.445
A	serpin-4	-1.646	0.000	-0.657	0.001
A	serpin-6	-2.766	0.000	-0.930	0.398

*: Ten genes shaded *red* were only specifically induced by Spätzles-1A. 59 genes labeled with A were expressed higher at 6 than 24 h, no gene were expressed higher at 24 than 6 h, 27 genes labeled with C were expressed similar at 6 and 24 h.

Table S5. The summary of the names of 87 Spätzle-2 induced genes

	Transcript_id	**FB_BSA_6h__vs__FB_S2_6h|log2FoldChange**	**FB_BSA_6h__vs__FB_S2_6h|padj**	**FB_BSA_24h__vs__FB_S2_24h|log2FoldChange**	**FB_BSA_24h__vs__FB_S2_24h|padj**
C	diapausin10	-4.850	0.001	-7.512	0.000
C	diapausin12	-5.687	0.003	-7.380	0.001
A	diapausin13	-7.947	0.000	-5.190	0.000
C	diapausin4	-4.671	0.000	-4.497	0.002
A	Cactus	-2.912	0.000	-0.424	0.363
C	Cathepsin26_29kDalike3	-2.454	0.051	-3.473	0.008
A	CathepsinBlike5	-1.504	0.002	0.126	1.000
A	Domeless	-2.052	0.000	-0.786	0.308
A	GP33	-4.783	0.000	-2.975	0.224
A	HP17a	-5.389	0.000	-1.061	0.178
A	HP19	-1.994	0.000	-0.304	0.853

C	HP20	-1.087	0.026	-1.115	0.080
C	HP22	-1.766	0.000	-0.827	0.062
A	HP5	-2.920	0.000	-0.403	0.977
A	HP7	-2.553	0.000	-0.588	0.900
A	HP9	-1.984	0.029	-0.689	0.947
A	IKKgamma_Kenny	-1.140	0.021	-0.061	1.000
C	IML-1	-1.575	0.012	-1.815	0.010
B	IML-12	0.156	0.858	-1.906	0.000
A	IML-15	-2.239	0.000	-1.262	0.017
C	IML-4	-4.830	0.000	-5.474	0.000
A	KAL-1_anosmin	-1.588	0.014	0.427	0.913
A	Kurtz	-1.515	0.001	0.567	0.717
C	LRR-TMP18	-1.667	0.033	1.140	0.468
A	Multicystatin_procathepsinF-like	-1.372	0.000	-0.861	0.026
A	PAP1	-2.318	0.000	-1.026	0.042
C	PAP2	-2.981	0.000	-2.883	0.000
A	PAP3	-5.723	0.000	-3.066	0.000
A	PGRP2	-4.267	0.000	-3.697	0.001
A	PGRP4	-1.861	0.012	-3.726	0.020
A	PGRP5	-2.184	0.000	-2.126	0.000
A	PGRP7	-9.343	0.000	-1.644	0.879
A	Pelle	-2.014	0.000	-0.391	0.860
A	Pvr	-1.444	0.000	-0.166	0.967
A	Relish	-1.118	0.000	-0.835	0.062
A	SP34	-4.091	0.000	-2.215	0.000
A	SP58	-1.972	0.000	-1.013	0.207
B	SPH3	-1.650	0.007	-2.442	0.000
B	SPH33	-1.821	0.004	-2.416	0.000
C	SPH4	-2.542	0.000	-3.158	0.000
A	SRP1	-2.665	0.008	0.210	1.000

A	SRP3	-2.871	0.001	-1.235	0.666
A	Spz2	-1.323	0.038	0.016	1.000
A	Toll10_2	-1.730	0.018	-0.415	0.943
C	Toll5	-1.251	0.002	-1.301	0.004
A	Tube	-1.672	0.000	-0.912	0.003
C	WAP1	-3.243	0.002	-2.299	0.109
A	attacin1	-4.240	0.000	-3.680	0.000
A	attacin10	-4.036	0.000	-2.551	0.070
A	attacin11	-4.487	0.000	-3.507	0.000
A	attacin2	-1.955	0.023	-1.943	0.099
A	attacin3	-3.758	0.000	-6.531	0.000
A	attacin4	-2.967	0.000	-3.197	0.000
A	attacin5	-4.901	0.000	-4.864	0.000
A	attacin7	-3.984	0.000	-3.641	0.000
A	attacin8	-4.042	0.000	-3.066	0.000
C	bGRP2	-4.080	0.000	-2.318	0.019
A	cecropin1	-0.825	0.126	-1.993	0.000
A	cecropin12	-2.398	0.000	-3.466	0.020
A	cecropin15	-1.060	0.069	-2.521	0.000
C	cecropin2	-5.051	0.000	-6.331	0.000
C	cecropin4	-5.382	0.000	-3.920	0.000
C	cecropin5	-3.017	0.000	-5.638	0.000
A	cecropin6	-2.667	0.000	-3.494	0.000
C	defensin3	-2.298	0.000	-2.356	0.000
A	gallerimycin1	-2.890	0.000	-3.408	0.000
C	hemolin	-1.459	0.116	-3.143	0.000
C	lebocin-A	-1.114	0.041	-2.442	0.000
C	lebocin-B	-2.757	0.004	-2.597	0.029
A	lebocin-C	-2.553	0.000	-2.449	0.000
C	lebocin-D	-5.091	0.000	-5.543	0.000

A	leureptin3	-3.967	0.000	-1.423	0.342
C	lysozyme-like-protein-1	-1.693	0.000	-1.049	0.013
C	lysozyme-like-protein-4	-1.236	0.026	-0.714	0.599
A	moricin1	-3.481	0.000	-3.801	0.076
A	reeler1	-0.554	0.514	-2.264	0.000
C	scolexinA	-4.507	0.000	-3.937	0.000
C	scolexinB	-1.308	0.000	-0.456	0.696
A	serpin-10	-1.307	0.010	-0.118	1.000
C	serpin-11	-2.944	0.000	-1.217	0.000
C	serpin-12	-1.380	0.000	-0.776	0.006
A	serpin-14	-2.580	0.000	0.123	1.000
C	serpin-2	-2.244	0.001	-2.831	0.000
A	serpin-3	-2.031	0.000	-0.558	0.524
A	serpin-4	-1.655	0.000	-0.682	0.001
C	serpin-5	-0.595	0.221	-1.009	0.047
A	serpin-6	-3.127	0.000	-0.981	0.375

*: Eleven genes shaded green were only specifically induced by Spätzle-2. 54 genes labeled with A were expressed higher at 6 than 24 h, 3 gene were expressed higher at 24 than 6 h, 30 genes labeled with C were expressed similar at 6 and 24 h.

Table S6. The primers used for proSpätzle-1A, 2 and 7 cloning

Gene	forward primer	reverse primer
proSpätzle-1A	J196: CCATGGGAATTCACAAATGCAAAGACTGCT	J197: CTGCAGTTACTCGAGTATGATTGTCAATTTGGC
proSpätzle-2	J190: CCATGGGAATTCAAACTACAGAATACGTGCCT	J191: CTGCAGTTACTCGAGCTTGTTATCGTTCAT
proSpätzle-7	J194: CCATGGGAATTCGTAAATTGGACATTCC	J195: CTGCAGTTACTCGAGATCAAATTCTGAAGTG

Figures

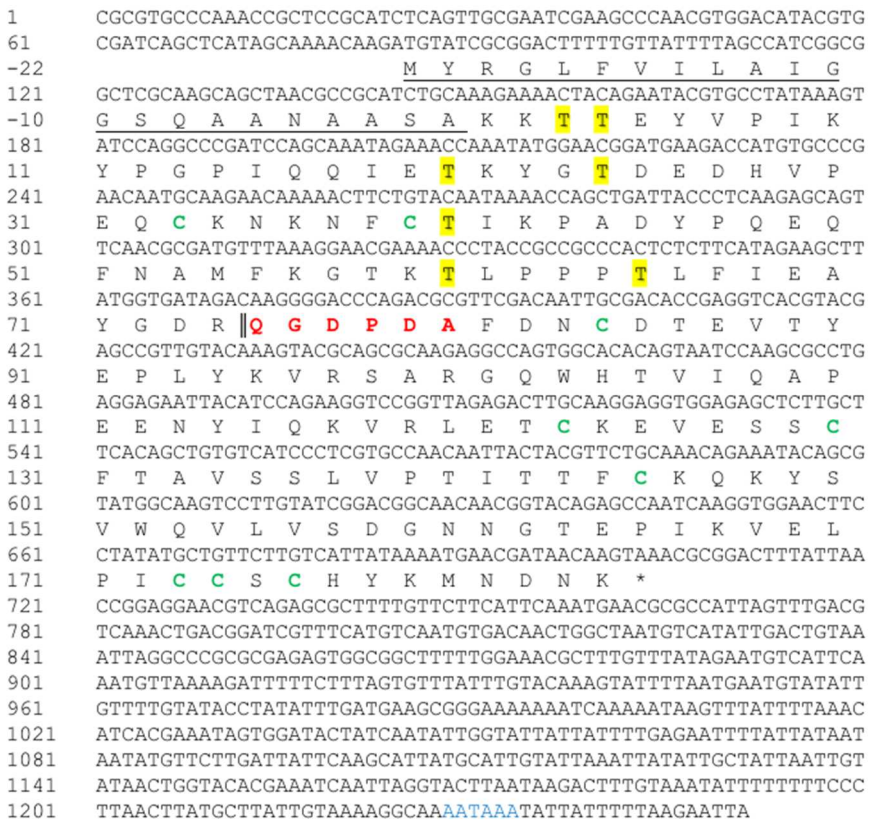


Fig. 1. cDNA and deduced amino acid sequences of *M. sexta* proSpätzle-2. The one-letter code for each amino acid is aligned with the second nucleotide of the corresponding codon. The stop codon is marked with '*'. The predicted secretion signal peptide is underlined. The proteolytic activation site is indicated with '‖'. The N-terminal sequence, determined by Edman degradation, of the activated form of Spätzle-2 after cleavage by PAP3 is shown in red. Putative N and O-linked glycosylation sites are shaded in yellow. The cysteine amino acids for linking disulfide bond are shown in green. AATAAA sequences (shown in blue) near the end of the 3′-UTR are potential polydenylation signals.

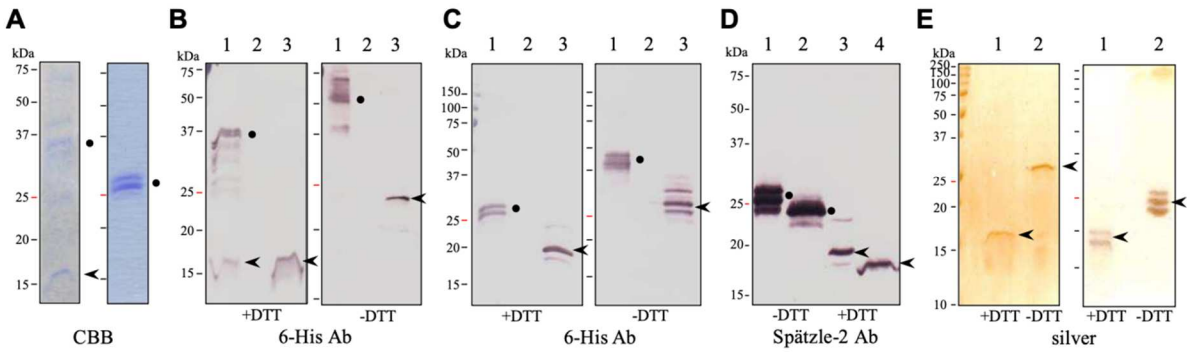


Fig. 2. SDS-PAGE and immunoblot analysis of *M. sexta* Spätzle-1A and -2.

(A) The 12% SDS/PAGE analysis of the purified proSpätzle-1A (left) and -2 (right) from Sf9 cells. Aliquots of the protein (1.0 µg for staining) were treated by SDS sample buffer and detected by Coomassie brilliant blue (CBB). Positions and sizes of the Mr makers are indicated. Immunoblot analysis of the PAP3 proteolytic products of *M. sexta* Spätzles precursors following 12% SDS (B) Spätzle-1A (C) Spätzle-2. The purified proSpätzle-1A or 2 (200 ng/µL, 1 µL) were separately incubated with PAP3 (40 ng/µL, 1 µL) in 22 µL reaction buffer (0.001% Tween-20, pH 7.5, 20 mM Tris-HCl, 5 mM CaCl2) at 37 °C for 1 h. The reaction mixtures and controls (40 ng PAP3 or 200 ng proSpätzle only) were distributed into 2 tubes (12 µL/tube) and treated with 1 × SDS sample buffer with (left) or without (right) DTT at 95 °C for 5 min. After electrotransfer, immunoblot analysis was performed using 1:1000 diluted antibody against the hexahistidine tag. (Lane 1, 100 ng proSpätzle-1A or -2; lane 2, 20 ng PAP3; lane 3, both Spätzle-1A or -2 and PAP3). (D) Deglycosylation of Spätzle-2. 0.5 µL PNGase F (500,000 U/mL), 1 µL O-Glycosidase (40,000,000 U/mL) and 1 µL Neuraminidase (20,000 U/mL) were added to 200 ng proSpätzle-2 (lane 3) and 200 ng proSpätzle-2 already processed by 40 ng PAP3 (lane 4) for de-N and O-glycosylation at 37 °C for 2 h, with the control of proSpätzle-2 (lane 1) or active Spätzle-2 (lane 2) only without the enzymes involved in deglycosylation. After electrotransfer, immunoblot analysis was

performed using 1:2000 diluted antibody against the Spätzle-2. The Spätzle before and after glycosylation are marked in blue and red, respectively. (E) Silver stain gel of active Spätzle-1A (left) or 2 (right) after purification. 500 ng/lane treated with 1 × SDS sample buffer with (lane 1) or without (lane 2) DTT running in 15 % (Spätzle-1A) or 12 % (Spätzle-2) SDS gel. The Spätzle precursor and active form are marked with circles and triangles, respectively.

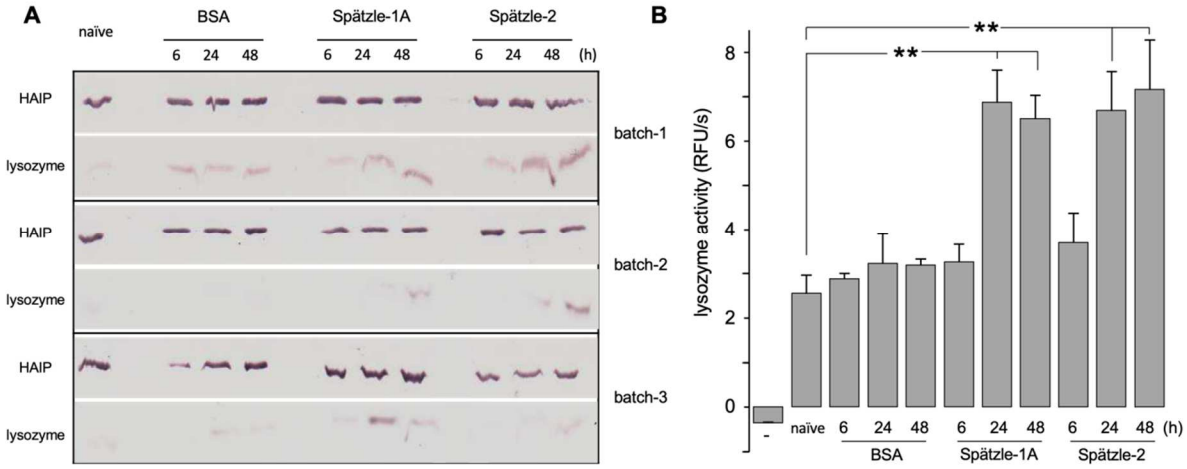


Fig. 3. Effects of injected Spätzle on lysozyme levels. Fifth instar, day 1–2 larvae were injected with BSA, or activated Spätzle-1A or 2. Six, twenty-four and forty-eight hours later, injection and non-injection naïve hemolymph was collected. **(A)** Immunoblot analysis of *M. sexta* lysoszyme1 or HAIP (loading control) protein level in 1 μL cell-free hemolymph samples. **(B)** Lysozyme activity of cell-free hemolymph between non-injection naïve and injection groups. H_2O is the negative control. Western blot and lysozyme activity experiments both did three biological replicates (**, P<0.01, compared to the naïve control group).

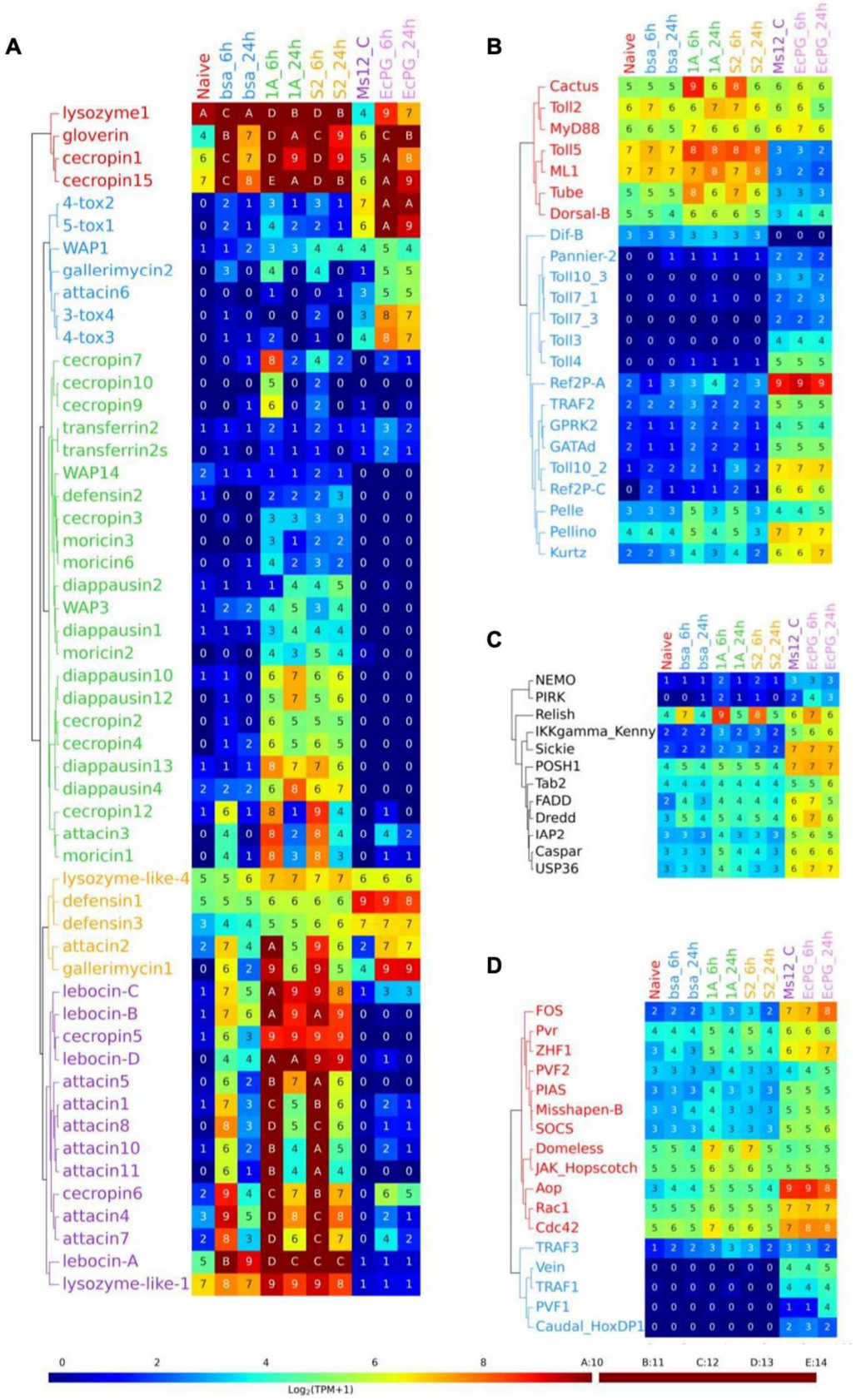

A
Naive
bsa_6h
bsa_24h
1A_6h
1A_24h
S2_6h
S2_24h
Ms12_C
EcPG_6h
EcPG_24h
lysozyme1
gloverin
cecropin1
cecropin15
4-tox2
5-tox1
WAP1
gallerimycin2
attacin6
3-tox4
4-tox3
cecropin7
cecropin10
cecropin9
transferrin2
transferrin2s
WAP14
defensin2
cecropin3
moricin3
moricin6
diappausin2
WAP3
diappausin1
moricin2
diappausin10
diappausin12
cecropin2
cecropin4
diappausin13
diappausin4
cecropin12
attacin3
moricin1
lysozyme-like-4
defensin1
defensin3
attacin2
gallerimycin1
lebocin-C
lebocin-B
cecropin5
lebocin-D
attacin5
attacin1
attacin8
attacin10
attacin11
cecropin6
attacin4
attacin7
lebocin-A
lysozyme-like-1
B
Cactus
Toll2
MyD88
Toll5
ML1
Tube
Dorsal-B
Dif-B
Pannier-2
Toll10_3
Toll7_1
Toll7_3
Toll3
Toll4
Ref2P-A
TRAF2
GPRK2
GATAd
Toll10_2
Ref2P-C
Pelle
Pellino
Kurtz
C
NEMO
PIRK
Relish
IKKgamma_Kenny
Sickie
POSH1
Tab2
FADD
Dredd
IAP2
Caspar
USP36
D
FOS
Pvr
ZHF1
PVF2
PIAS
Misshapen-B
SOCS
Domeless
JAK_Hopscotch
Aop
Rac1
Cdc42
TRAF3
Vein
TRAF1
PVF1
Caudal_HoxDP1
0
2
4
6
8
A:10
B:11
C:12
D:13
E:14
Log2(TPM+1)

Fig. 4. Expression heat map of 53 induced effectors **(A)**, 23 Toll **(B)**, 12 Imd **(C)** and 17 JNK and JAK/STAT **(D)** intracellular signaling molecules. On top of the heat map, fat body of the 10 RNA-seq datasets are indicated as naïve larvae, BSA, Spätzle-1A and 2 injections at 6 or 24 h larva, Ms-12 cell line non-treated control, *E. coli* PG treatment at 6 or 24 h cell line. Log_2 (TPM+1) values for these immune-related transcripts are shown in the gradient heat map from dark blue (0) to maroon (≥10). The values of 0−0.49, 10.50−1.49, 1.50−2.49 … 8.50−9.49, 9.50−10.49, 10.50−11.49, 11.50−12.49 are labeled as 0, 1, 2 … 9, A, B, C, respectively. Relatedness in expression patterns revealed by hierarchical cluster analysis is shown on the left, with abbreviated protein names marked in different colors.

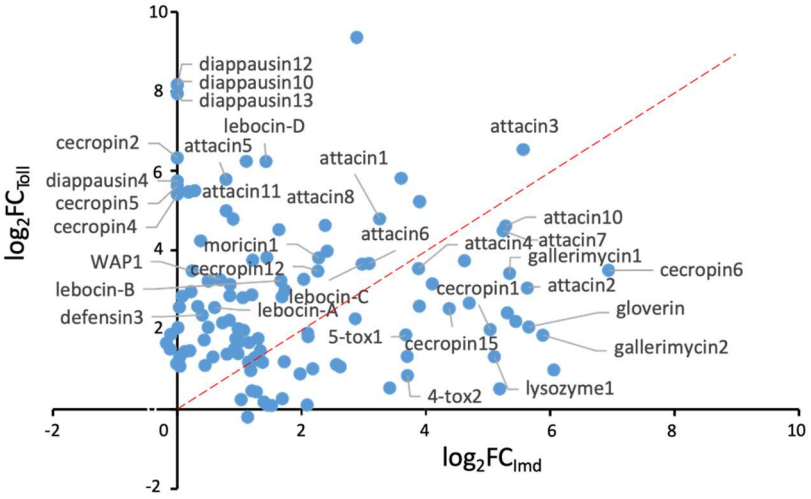


Fig. 5. Toll or Imd preference of 119 differential expression gene isoforms. The maximum log_2 fold change of Spätzle vs BSA (Toll) is set as X value, the maximum log_2 fold change of *E. coli* PG vs control (Imd) is set as Y value, 119 isoforms with X, Y values are labeled as blue dots, names of 34 AMPs are shown close to the dots. The closer to the Y (Toll induction level) or X axis (Imd induction level), the more preferential to Toll or Imd pathway, respectively. The red dash line whose slope is 1 stands for the boundary between Toll and Imd (50 % of both pathways).

```
1      AAAATAAGCAATTTAGTTATTATTTAAGGCGCTGCTCTAAGTCCTGACAGCAGTAGTCCT
61     CAAGCCTCGGCCAGTGTTGATGCTCGTGCCACATCAAAAAGTAGAGTTATTGTGAATATT
121    GCAACGTAGAACATAAAAATAAAATGGCGTCTGTAACTTTAATGTTTACGGTGGTGTTGG
-22                            M  A  S  V  T  L  M  F  T  V  V  L
181    TAGCCTTGAGTCGATCTGCAGTTACAGCAAGTAAATTGGACATTCCCCCAGAATGTAAAG
-10    V  A  L  S  R  S  A  V  T  A  S  K  L  D  I  P  P  E  C  K
241    ACAAAATGTTTTGCGACGTCGAGCCCGCGTTCTACAACGAATACAGAGAAGACATCGAAC
11     D  K  M  F  C  D  V  E  P  A  F  Y  N  E  Y  R  E  D  I  E
301    TCAAGTTTAAACCCATTAAAAGTATGTGGAATTTGGCCGATAATTATAACTATGATTATC
31     L  K  F  K  P  I  K  S  M  W  N  L  A  D  N  Y  N  Y  D  Y
361    AAGCGAGGGAATATACTTTTGAGAATAACTGCAACTATGTTAAAGAGATGAGTACGCCGT
51     Q  A  R ‖E  Y  T  F  E  N  N  C  N  Y  V  K  E  M  S  T  P
421    ACCTCGTGCACTCGAACGGGACAGCAAAGATCATCGTGCAGACCTCTTTTTACAAGCAAC
71     Y  L  V  H  S  N  G  T  A  K  I  I  V  Q  T  S  F  Y  K  Q
481    GTTTCATGACAACGAGATGCGTGGACTCGTTCTCCAATGCGGCCGATGGTGAGCAATGCT
91     R  F  M  T  T  R  C  V  D  S  F  S  N  A  A  D  G  E  Q  C
541    TCAAAGGCATTTTAAGTGGGAACTCAAGAGGCAGCTGCAAGACCAAATATACAAACATTA
111    F  K  G  I  L  S  G  N  S  R  G  S  C  K  T  K  Y  T  N  I
601    TTTTGTATGCGTACGACGAAGCGAACAAGAGAATCCAACGACTTCCTATTGAGGTGCCCG
131    I  L  Y  A  Y  D  E  A  N  K  R  I  Q  R  L  P  I  E  V  P
661    TGTGCTGCTACTGCGGCATCACTTCAGAATTTGATTAGTTCCGACGGAGAATACGTTTTT
151    V  C  C  Y  C  G  I  T  S  E  F  D  *
721    AAATTTAAGTCTTAATGAAACTAAATAGTGAATTAGACATTAATTGGTATTTGTGTATAA
781    ATTGAAACCAAAGAATCTGTTATGATAATAAATATTAAGCATC
```

Fig. S1. cDNA and deduced amino acid sequences of *M. sexta* proSpätzle-7. The one-letter code for each amino acid is aligned with the second nucleotide of the corresponding codon. The stop codon is marked with '*'. The predicted secretion signal peptide is underlined. The proteolytic activation site is indicated with '‖'. Putative N-linked glycosylation sites are shaded in yellow. The cysteines for linking disulfide bond are shown in green.

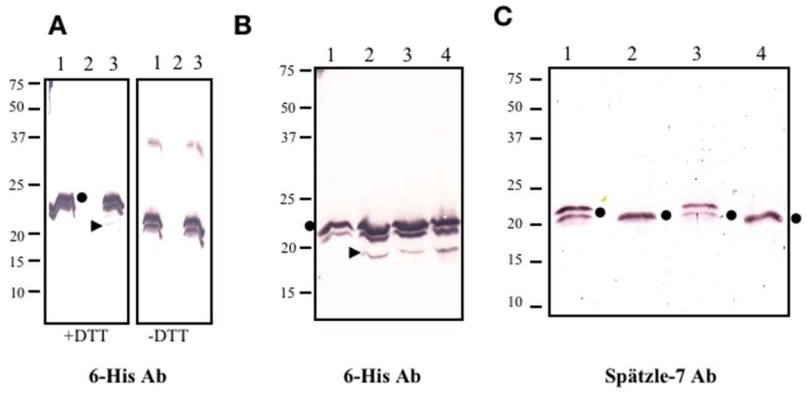


Fig. S2. SDS-PAGE and immunoblot analysis of *M. sexta* Spätzle-7. **(A)** Immunoblot analysis of the PAP3 proteolytic products of *M. sexta* Spätzle-7 precursors following 12% SDS. The reaction method is the same as Spätzle-1A and 2 (Fig.2). **(B)** The purified

proSpätzle-7 (200 ng/μL, 1 μL) were separately incubated with 1μLcell free hemolymph from naïve (CH), immune challenged (IH) or bar (BH) stage larva, and 1 μL elicitor (a mixture of *Escherichia coli*, *Micrococcus luteus*, and curdlan) are incubated together at 37 °C for 4 h with the control of 100 ng proSpätzle7 only. After electrotransfer, immunoblot analysis was performed using 1:1000 diluted antibody against the hexahistidine tag. **(C)** Deglycosylation of proSpätzle-7. 0.2 μg proSpätzle-7, with 0.5 μL PNGase F, or 1 μL O-Glycosidase and 1 μL Neuraminidase, or both for de-N-glycosylation, de-O-glycosylation or de-N, O-glycosylation at 37 ℃ for 2 h, with the control of proSpätzle-7 only without the enzymes involved in deglycosylation. After electrotransfer, immunoblot analysis was performed using 1:2000 diluted antibody against the Spätzle-7. The Spätzle before and after glycosylation are marked in blue and red, respectively.

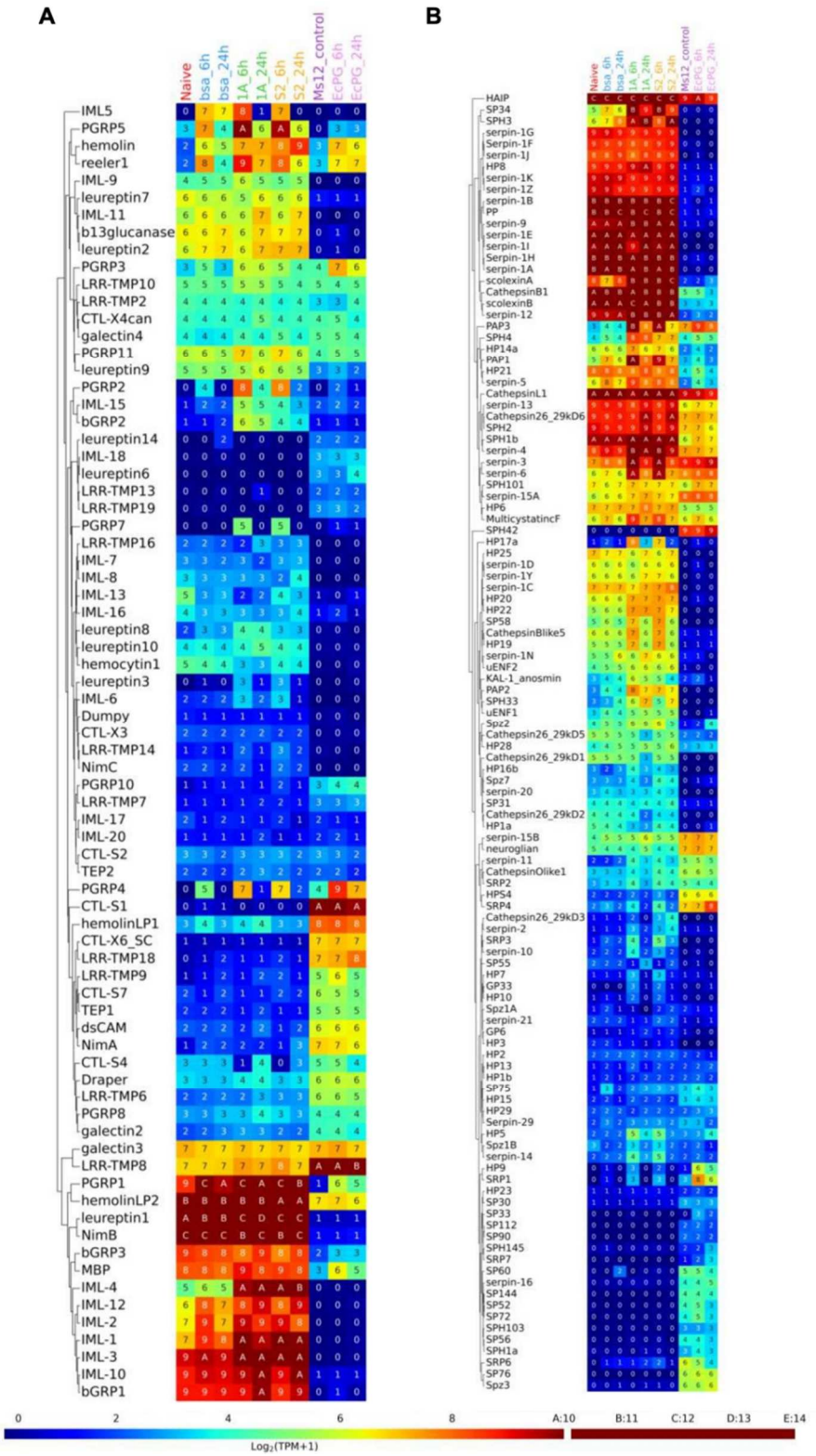


Fig. S3 Expression heat map of induced **(A)** 75 PRRs and **(B)** 114 extracellular signaling molecules. Labels are the same as Fig. 4.

References

An, C., Jiang, H., Kanost, M. R. 2010. Proteolytic activation and function of the cytokine Spätzle in the innate immune response of a lepidopteran insect, *Manduca sexta*. *FEBS j.*, 277(1), 148-162.

Cao, X., He, Y., Hu, Y., Wang, Y., Chen, Y-R., Bryant, B., Clem, R.J., Schwartz, L.M., Blissard, G.W., Jiang, H., 2015. The immune signaling pathways of *Manduca sexta. Insect Biochem. Mol. Biol*. 62, 64−74.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N.T., Hoffmann, J.A., Kanzok, S.M., Letunic, I., Levashina, E.A., Loukeris, T.G., Lycett, G., Meister, S., Michel, K., Moita, L.F., Müller, H.M., Osta, M.A., Paskewitz, S.M., Reichhart, J.M., Rzhetsky, A., Troxler, L., Vernick, K.D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., Kafatos, F.C., 2002. Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 298, 159–165.

Hoffmann, J. A., Reichhart, J. M., 2002. Drosophila innate immunity: an evolutionary perspective. *Nat Immunol.*, 3(2), 121-126.

Jiang, H., Wang, Y., Yu, X. Q., Zhu, Y., Kanost, M., 2003. Prophenoloxidase-activating proteinase-3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domain serine proteinase regulated by serpin-1J and serine proteinase homologs. *Insect Biochem. Mol. Biol*., 33(10), 1049-1060.

Kan, H., Kim, C. H., Kwon, H. M., Park, J. W., Roh, K. B., Lee, H., Park, B. J., Zhang, R., Zhang, J., Söderhäll, K., Ha, N.C., Lee, B. L., 2008. Molecular Control of Phenoloxidase-induced Melanin Synthesis in an Insect. *J Biol Chem*, 283(37), 25316-25323.

Kanost, M.R., Arrese, E.L., Cao, X., Chen, Y-R, Chellapilla, S., Goldsmith, M., Grosse-Wilde, E., Heckel, D.G., Herndon, N., Jiang, H., Papanicolaou, A., Qu, J., Soulages, J.L. Vogel, H., Walters, J., Waterhouse, R.M., Ahn, S.J., Almeida, F.C., An, C., Aqrawi, P., Bretschneider, A., Bryant, W.B., Bucks, S., Chao, H., Chevignon, G., Christen, J.M., Clarke, D.F., Dittmer, N.T., Ferguson, L.C., Garavelou, S., Gordon, K.H., Gunaratna, R.T., Han, Y., Hauser, F., He, Y., Heidel-Fischer, H., Hirsh, A., Hu, Y., Jiang, H., Kalra, D., Klinner, C., König, C., Kovar, C., Kroll, A.R., Kuwar, S.S., Lee, S.L., Lehman, R., Li, K., Li, Z., Liang, H., Lovelace, S., Lu, Z., Mansfield, J.H., McCulloch, K.J., Mathew,

T., Morton, B., Muzny, D.M., Neunemann, D., Ongeri, F., Pauchet, Y., Pu, L.L., Pyrousis, I., Rao, X.J., Redding, A., Roesel, C., Sanchez-Gracia, A., Schaack, S., Shukla, A., Tetreau, G., Wang, Y., Xiong, G.H., Traut, W., Walsh, T.K., Worley, K.C., Wu, D., Wu, W., Wu, Y.Q., Zhang, X., Zou, Z., Zucker, H., Briscoe, A.D., Burmester, T., Clem, R.J., Feyereisen, R., Grimmelikhuijzen, C.J., Hamodrakas, S.J., Hansson, B.S., Huguet, E., Jermiin, L.S., Lan, Q., Lehman, H.K., Lorenzen, M., Merzendorfer, H., Michalopoulos, I., Morton, D.B., Muthukrishnan, S., Oakeshott, J.G., Palmer, W., Park, Y., Passarelli, A.L., Rozas, J., Schwartz, L.M., Smith, W., Southgate, A., Vilcinskas, A., Vogt, R., Wang, P., Werren, J., Yu, X.Q., Zhou, J.J., Brown, S.J., Scherer, S.E., Richards, S., Blissard, G.W., 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem. Mol. Biol*. 76, 118−147.

Kim, C. H., Kim, S. J., Kan, H., Kwon, H. M., Roh, K. B., Jiang, R., Yang, Y., Park, J. W., Lee, H. H., Ha, N. C., Kang, H. J., Nonaka, M., Söderhäll, K., Lee, B. L., 2008. A three-step proteolytic cascade mediates the activation of the peptidoglycan-induced toll pathway in an insect. *J Biol Chem*, 283(12), 7599-7607.

Kingsolver, M.B., Huang, Z., Hardy, R.W., 2013. Insect antiviral innate immunity: pathways, effectors, and connections. *J. Mol. Biol*. 425, 4921−4936.

Lemaitre, B., Hoffmann, J., 2007. The host defense of *Drosophila melanogaster*. *Ann. Rev. Immunol.* 25, 697−743.

Li, B., Dewey, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12(1), 1-16.

Lindsay, S.A., Wasserman, S.A., 2014. Conventional and non-conventional *Drosophila* Toll signaling. *Dev. Comp. Immunol*. 42, 16−24.

Love, M. I., Huber, W., Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15(12), 1-21.

Rämet, M., 2012. The fruit fly *Drosophila melanogaster* unfolds the secrets of innate immunity. *Acta Paediatr*. 101, 900−905.

Sohi, S. S., 1995. Development of lepidopteran cell lines. *Methods Mol Biol.*, 39, 397-412.

Steiner, H., Hultmark, D., Engström, Å., Bennich, H., Boman, H. G., 1981. Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature*, 292(5820), 246-248.

Sumathipala, N., Jiang, H., 2010. Involvement of *Manduca sexta* peptidoglycan recognition protein-1 in the recognition of bacteria and activation of prophenoloxidase system. *Insect Biochem. Mol. Biol*., 40(6), 487-495.

Wang, Y., Lu, Z., Jiang, H., 2014. *Manduca sexta* proprophenoloxidase activating proteinase-3 (PAP3) stimulates melanization by activating proPAP3, proSPHs, and proPOs. *Insect Biochem. Mol. Biol*., 50, 82-91.

Wang, Y., Jiang, H., 2017. Prophenoloxidase activation and antimicrobial peptide expression induced by the recombinant microbe binding protein of *Manduca sexta*. *Insect Biochem. Mol. Biol*., 83, 35-43.

Wang, Y., Yang, F., Cao, X., Zou, Z., Lu, Z., Kanost, M. R., Jiang, H., 2020. Hemolymph protease-5 links the melanization and Toll immune pathways in the tobacco hornworm, *Manduca sexta*. *Proc Natl Acad Sci U S A.*, 117(38), 23581-23587.

Zou, Z., Evans, J.D., Lu, Z., Zhao, P., Williams, M., Sumathipala, N., Hetru, C., Hultmark, D. and Jiang, H., 2007. Comparative genomic analysis of the *Tribolium* immune system. *Genome Biol*, 8(8), 1-16.

CHAPTER II

IN SEARCH FOR IMMUNITY-RELATED GENES SPEFICALLY REGULATED BY THE TOLL AND IMD PATHWAYS IN THE TOBACCO HORNWORM, *MANDUCA SEXTA*

Zelong Miao[1], Xiaolong Cao[1], Chao Xiong[1], Michael R. Kanost[2], Haobo Jiang[1]

[1] Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA

[2] Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS 66506, USA

Key words: insect immunity, RNA-Seq, fat body, hemocytes, midgut, Illumina sequencing

Abbreviations: CF, IF, CH, IH, CG and IG, control (C) and induced (I) fat body (F), hemocytes (H), and midgut (G); AMP, antimicrobial peptide/protein; DAP-PG: diaminopimelic acid peptidoglycan; PGRP, peptidoglycan recognition protein; MBP, microbe binding protein; βGRP, β-1,3-glucan recognition protein, HP, hemolymph (serine) protease; PO and proPO, phenoloxidase and its precursor; SP and SPH, serine protease and its homolog; PAP, proPO activating protease; Imd, immune deficiency; qPCR, quantitative real-time polymerase chain reaction.

Abstract

Sequencing and annotation of the genome of *Manduca sexta* have greatly enhanced its role as a model for lepidopteran insects and beyond. While pattern recognition receptors, nondigestive serine proteases, serpins, cytokines, intracellular signal transducers, and antimicrobial effectors are identified based on homology, it is unclear how induction of defense proteins is regulated by specific immune pathways. Beginning to address these questions, we performed a transcriptome analysis of fat body, hemocytes and midgut at 6 and 24 h after larvae had been injected with phosphate buffered saline or a mixture of *Escherichia coli*, *Micrococcus luteus* and curdlan, an insoluble β-1,3-glucan from *Alcaligenes faecalis*. Under this condition, the immunity-related genes whose mRNA levels peaked at 6 h were far more than those at 24 h in these three tissues. According to the expression time courses found in *Drosophila*, the former (pattern-1) may be preferentially controlled by the Imd-Relish pathway whereas the latter (pattern-2) by the Toll-Dorsal pathway. A total of 215 differentially expressed genes were identified and few of them showed consistent expression pattern-1 or -2 in all three tissues. Injection of *Enterococcus faecalis* or *Enterobacter cloacae* induced the attacin1, 2, 10, moricin1, lebocinD and gallerimycin1 transcription to similar levels in the same patterns. However, diapausin1, gloverin and cecropin6 are more sensitive to *Enterococcus faecalis*, X-toxs is more sensitive to *Enterobacter cloacae*. Injection of *Beauveria bassiana* induced much lower levels of the nine genes. Our results show the injection of bacteria cannot separate Toll or Imd pathways properly in *M. sexta*, which is very different from *D. melanogaster*. Thus, it confirms the complexity of host-pathogen interactions and innate immune response pathways in other non-dipteran insects.

1. Introduction

Insect immunity has been an active subject of entomological research for over four decades, especially in in *Drosophila*. With the identification of Toll, Imd, and other signaling pathways, significance of the genetic research goes far beyond insects (Lemaitre and Hoffmann, 2007; Capo et al., 2016; Rämet, 2012; Valanne et al., 2011, Ramet et al., 2002; Baeg et al., 2005; Ragab et al., 2011; Kingsolver et al., 2013), as similar pathways are active in innate immune responses of human (Hoffmann and Reichhart, 2002). In the fruit fly, Toll pathway primarily respond to fungal and Gram-positive bacterial infection whereas Imd, Relish and other intracellular proteins form another major pathway for fighting off Gram-negative bacteria. Aseptic wounding causes small and temporary increases in antimicrobial peptide (AMP) expression; DAP-PGs induce through the Imd pathway larger increases in AMP production that peaks at 6−10 h, followed by a sharp decrease; Through a serine protease cascade and the Toll pathway, Lys-PGs of Gram-positive bacteria and β-1, 3-glucans of fungi lead to steady increases in AMP synthesis over a period of 24 h or longer (Lemaitre and Hoffmann, 2007). Some AMP genes (*e.g*., DIM1) are Toll-specific, others (*e.g*., Diptericin) are Imd-specific, and Drosomycin is up-regulated through Toll mainly and Imd modestly. More AMP genes are controlled by both pathways, due to overlaps of or cross-talks between them (Lindsay and Wasserman, 2014).

The simple *Drosophila* model has greatly impacted immunological research in other insects. It provides a framework for placing orthologous proteins in corresponding pathways in genome annotation (Christophides et al., 2002; Zou et al., 2007; Cao et al., 2015). On the negative side, most researchers treat the model as dogma without proper

validation. Solid data are scarce regarding the specificity of signaling pathways toward different microbial groups. Neither is it rigorously tested whether the temporal activation of Toll and Imd pathways is true in other insect systems. Researchers generally believe that genes regulated by the Imd pathway show an acute phase profile of rapid rise and fall of mRNA, whereas target genes of the Toll pathway exhibit a late and sustained expression pattern. Considering pathway specificity and time course as open questions for other insects, we designed experiments to address the concerns in *Manduca sexta*.

M. sexta has long been used as a model to study the biochemistry of insect immunity. The genome analysis (Kanost et al., 2016) greatly enhanced its role in biochemical and molecular investigations, especially those require substantial quantities of tissues, cells, and body fluids. Predicted genes related to immunity are among the best annotated ones in this species, encoding proteins involved in pathogen recognition (CTLs and other PRRs) (Rao et al., 2015; Zhang et al., 2015), signal transduction and modulation (SPs, SPHs, serpins, cytokines, intracellular signal transducers) (Cao et al., 2015a and b; Li et al., 2018), and effectors that kill or sequester microbes (*e.g.*, AMPs) (He et al., 2015). Expression profiles for the 583 genes in 52 tissue samples collected at various life stages provide an overview of their transcriptional regulation in relation to development and tissue specificity. Prior to that, two EST projects yielded useful information on innate immunity in fat body, hemocytes, midgut, trachea, and integument before and after an immune challenge (Zou et al., 2008; Zhang et al., 2011). The later genome-independent, quantitative RNA-seq analyses revealed process- and tissue-specific expression of immunity-related genes and -unrelated, such as metabolic enzymes. Due to historical reasons (*e.g.*, technical limitations), those genome-independent studies lack the

throughput needed for accurate calculation of induction ratios, biological replicates for assessing statistical significance and data from midgut, a tissue actively involved in the balance of commensal and pathogenic microbial populations. In this study, we first did RNA-Seq analysis of the three tissues at two different times after injection of a mixture of killed bacteria and its component (curdlan) to identify Imd or Toll-specific genes based on their predicted time courses. Then, we used quantitative real-time PCRs to quantify mRNA levels of the candidate genes in a new set of tissue samples from larvae separately challenged with killed *E. faecalis*, *E. cloacae*, and *B. bassiana*, representing Gram-positive and -negative bacteria and fungi, to larvae and studied tissue-specific expression of selected genes during immune induction. Lessons learned during this difficult project truthfully reflect the huge differences between the fruit fly and the tobacco hornworm. Implications of the findings in other insects are discussed along with our recent findings on pattern recognition receptors at the level of molecular mechanisms.

2. Methods and materials

2.1. Insect rearing, bacterial injection, RNA isolation, library construction, and Illumina sequencing

M. sexta eggs, purchased from Carolina Biological Supply, were hatched and reared on an artificial diet (Dunn and Drake, 1983). Forty-eight day 2, 5th instar larvae were injected with 20 μL phosphate-buffered saline (PBS) as wounding controls and a mixture of *Escherichia coli* (2×10^7 cells), *Micrococcus luteus* (20 μg) (Sigma-Aldrich), and curdlan (20 μg, insoluble β-1,3-glucan from *Alcaligenes faecalis*) (Sigma-Aldrich) in 20 μL H_2O per larva (Wang and Jiang, 2017). Total RNA samples were extracted from PBS injected fat body (CF), hemocyte (CH), and midgut (CG) and induced fat body (IF),

hemocytes (IH), and midgut (IG) 6 and 24 h later using TRIZOL Reagent (Thermo Fisher Scientific). PolyA+ RNA was separately purified from the total RNA samples (1.0 mg each) by binding to oligo(dT) cellulose twice in the Poly(A) Purist^{TM} Kit (Ambion). First strand cDNA was synthesized using mRNA (5.0 μg), random dodecanucleotides (100 pmol), and SuperScriptTM III reverse transcriptase (1000 U, Life Technologies Inc.). RNase H treatment, second strand synthesis, and gap joining were performed according to the published protocol (Zou et al., 2008). After shearing via nebulization, the samples were end-repaired (Roe, 2004) and ligated to double-stranded adaptor A and biotinylated adaptor B (Margulies et al., 2005). Another control group including naïve fat body (NF), hemocyte (NH), and midgut (NG) RNA were prepared from day 2, 5th instar naïve larvae (12). The 9 naïve libraries were constructed by Henry Bellmon Research Center at Oklahoma State University using the KAPA mRNA HyperPrep Kit for Illumina Platforms. The challenged and control libraries were sequenced on Illumina NextSeq 500 platform (Illumina, USA) using Mid-Output Kit (2 × 75 bp). FASTQ files of raw-reads were produced and sorted out by barcodes for further analysis.

2.2. Assembly of transcriptomes and identification of differentially expressed transcripts (DETs)

RSEM (Li and Dewey, 2011) package was applied for calculation of normalized gene expression value TPM. Subsequently, DETs between the control and treatment libraries were calculated based on the significance level (p value < 0.05) using DESeq2 package in R environment (Love, Huber and Anders, 2014). These transcripts with expression base mean more than 10, fold change more than 2 or less than 0.5 folds were considered as up- and down-regulated genes, respectively. For principal component analysis (PCA), PCs,

PC variances, and PC scores were calculated using eigen function in R language (Chen et al., 2015).

2.3. Quantitative real-time PCR analysis

Day 2, fifth instar larvae of *M. sexta* were injected with 20 μL dead *E. cloacae* ($A_{600} \approx 1$), *E. faecalis* ($A_{600} \approx 1$) or *B. bassiana* (2×10^7 conidia/mL) suspension, separately. *E. cloacae* and *E. faecalis* were at stationary phase in LB broth (A600 ≈ 1.0), *B. bassiana* was cultured on potato dextrose agar (PDA) plates then suspended it into YEPD (yeast extract peptone dextrose) medium at 25 °C 200 rpm 48h until it had 2×10^7 conidia/mL, all the three pathogens were autoclaved at 121 °C 20 mins separately, which were used for killed pathogen suspension injection (20 μL/larva). Fat body (FB), hemocytes (HC), and midgut (MG) were dissected for RNA isolation. cDNA synthesis and qRT-PCR analysis of antimicrobial peptide mRNA levels were performed using specific primers (Table S1) for them and rpS3. Amplification efficiencies of the cDNA fragments were determined as described before (Schrag et al., 2019). Relative antimicrobial peptide mRNA levels were calculated as $(1 + E_{rpS3})^{Ct,\ rpS3} / (1 + E_x)^{Ct,\ x.}$

3. Results

3.1. Overview of Illumina transcriptome data

In the genome level of 15542 structural genes, we totally identify 2841 genes up-regulated after injection of PBS buffer in at least one tissue, indicating that they may be involved in the wounding reaction. In addition, 2749 genes get induced and 3380 genes are down-regulated after injection of the dead mixture of pathogens in at least one tissue at 6 or 24 hours compared to both naïve and PBS wounding group. In fat body, 1231 genes get up-regulated at 6 h, 558 genes at 24 h, 1290 genes are down-regulated at 6 h,

619 genes at 24 h. 287 genes are up-regulated and 199 gens are down-regulated at both 6 and 24 h. 1417, 326, 1392, 341, 714, 373, 1016, 372 gens are up-regulated at 6 and 24 h, down-regulated at 6 and 24 h in hemocytes, midgut, respectively (p<0.05) (Table 1).

3.2 Tissue-specific immune genes

In annotated IMRGIs, 215 (31 %) isoforms are up-regulated whose expression levels are shown in Fig. 3, and 94 (14 %) isoforms are down-regulated in either FB, HC or MG (baseMean>10, Log_2FC<-1 or >1, $p$adj<0.05). Among them, 163 (24 %) isoforms are induced in FB at 6 or 24 h. Among them, 19 isoforms are in the Toll signaling pathway, 6 are in the Imd pathway. AMPs such as 7 diapausins, 4 X-toxs, 4 WAPs, all 11 attacins except attacin9, all 15 cecropins except cecropin8, 11, 13, 14, 2 defensins, all 3 gallerimycins, 1 gloverin, all 4 lebocins, lysozyme1, all 6 moricins are all induced either 6 or 24 h in FB. Ninety isoforms are up-regulated in FB at both 6 and 24 h, for example, diapausin1, 5, 9, 10, 12, 13, Cactus, HP5, Pelle, Relish,4 X-toxs, attacin1-4, 6-8, 10, cecropin1-7, 9, 10, 15, defensin3, 4, all 4 lebocins and 6 moricins, etc. While only 36 (5 %) isoforms are down-regulated at either 6 or 24 h in FB, including serpin1A, 19-21. Six isoforms (Atg3, Cathepsin26_29kDalike2, HP3, IML-13, Tollip-1, serpin-20) are down-regulated in FB at both 6 and 24 h. In HC, 147 (21 %) isoforms are up-regulated at 6 or 24 h. Among them, 18 molecules are involved in Toll signaling, while 7 are in Imd signaling pathway. The AMPs induced in HC includes diapausin1, 13, 4 X-toxs, WAP7, 11, 12, attacin1-8, 10, 11, cecropin1-7, 9, 10, 12 15, defensin4, gallerimycin1, 2, gloverin, 4 lebocins, lysozyme1 and 6 moricins. Fifty-eight isoforms are induced in HC at both 6 and 24 h, including MBP, PGRP1, PIRK, 4 X-toxs, WAP11, 12, attacin1-8, 10, cecropin1, 5-7, 9, 10, 15, defensin4, gallerimycin1, 2, glovein, lebocinB-D, lysozyme1

and moricin1, 3-6. While only 55 (8 %) isoforms are down-regulated in HC at 6 or 24 h. Three genes (Spz6, Toll7_2, and WAP13) are down-regulated in HC at both 6 and 24 h significantly. In MG, 115 (14 %) isoforms are induced at 6 or 24 h, including 14 Toll and 7 Imd signaling components. AMPs such as diapausin9, 12 , 13, 4 X-toxs, WAP3, 11, 12, attacin1-8, 10, 11, cecropin1, 4-6, 9, 12, 15, defensin3, all 3 gallerimycins, gloverin, all 4 lebocins, lysozyme1 and moricin1, 3, 6. Fifty-nine isoforms are up-regulated in MG at both 6 and 24h, including diapausin13, Cactus, MBP, PGRP1, all 4 X-toxs, WAP11, 12, attacin1-4, 6-8, 10, cecropin1, 4-6, 12, 15, defensin3, gallerimycin1-3, gloverin, lebocinA-D, lysozyme1, moricin1, 3. While only 30 (4 %) isoforms are down-regulated at 6 or 24 h in MG, including serpin-17 and 20. Five isoforms (Cathepsin26_29kDalike1, GP69, HP1a, IML-7, and Multicystatin) are down-regulated at both 6 and 24 h (Table 2). Seventy-two isoforms have no tissue-specificity, which are up-regulated in all three tissues, including diapausin13, 8 Toll and 6 Imd signaling pathway components, all 4 X-toxs, WAP11, 12, attacin1-8, 10, 11, cecropin1, 4-6, 9, 12, 15, gallerimycin1, 2, gloverin, all 4 lebocins, lysozyme1, moricin1, 3, 6. However, 30 isoforms are specifically induced in FB, such as diapausin4, 5, and 10, HP5, Spz6, Toll5, and WAP1. Another 30 isoforms like HP8, IMD, SPH101, SPH1a, SPH1b, WAP7, 10 isoforms of serpin1 are only specifically induced in HC. Seventeen isoforms are MG specifically induced including Dif, Dorsal and caspase genes (Fig. 1A). PCA of OGS2 shows the genes in FB and HC are more similar, which have much more variance with MG (Fig. 2A). The lower percentage of "defense" genes are up-regulated in MG (only 6 and 14 % genes are up-regulated in OGS2 and IMRGIs). An alternative explanation is that the mixture of dead pathogens is injected into hemocoel, the gut cells indirectly receive the signal of invasion.

In summary, most of diapausins can found in FB, HP8 and SPH1 variants can be induced only in HC, and caspase genes are MG-specifically induced.

3.3 Early and late immune response genes

Among the 163 induced isoforms in FB, 85 (52 %) isoforms are induced higher at 6 than 24 h (early response). Only 7 isoforms are induced higher at 24 than 6 h (late response), which are diapausin1, 4, 9, 12, WAP3, cecropin9, and 10. Among 147 induced isoforms in HC, 96 (65 %) isoforms are early response to the pathogens. Only 4 isoforms (SPH1a, WAP7, 11, serpin-2) are late response patterns. Fifty-one (44 %) of 115 induced isoforms in MG are higher expressed at 6 than 24 h. While only 6 isoforms (diapausin9, 12, Cathepsin26_29kDalike3, LRR-TMP9, cecropin9, scolexinA) are higher expressed at 24 than 6 h. Twenty-eight isoforms are expressed higher at 6 than 24 h in all 3 tissues, such as Cactus, HP14a, PAP1, PGRP1, PIRK, Relish, SRP1, SRP3, 3-tox4, attacin2, 10, 11, cecropin1, 15, gallerimycin2, gloverin and moricin1. However, no genes are higher expressed at 24 than 6 h in all 3 tissues, indicating that a lot more genes are up-regulated at 6 than 24 h (Fig. 1B). PCA of IMRGIs shows that the difference between induced at 6 h and control groups is much more significant than that between induced at 24 h and control groups in all 3 tissues (Fig. 2B, C, D). Overall, most of diapausins are late-response (expression level at 24 higher than 6 h) genes, but most of other AMPs can be induced at a very early time at 6 h after immune challenge, and then reduced at 24 h.

3.4 Dead pathogens induction

In order to find pathogen preference, we inject three dead pathogens separately. Diapausin1 and lebocinD are selected as Toll pathway preferential genes, while gallerimycin1 and X-toxs are selected as Imd pathway specific candidate genes.

Diapausin1 are more sensitive to the Gram-positive bacteria with the expression pattern of 24 h higher than 6 h, but lebocinD are equally reacted to both kinds of bacteria, the expression pattern of which is also 24 h higher than 6 h (Fig. 4). The expression pattern of these two Toll pathway specific candidate genes are similar with drosomycin in *D. melanogaster* (Hoffmann and Reichhart, 2002). X-toxs are more sensitive to the Gram-negative bacteria with the pattern of 24 higher than 6 h (Fig. 4), which is similar with diptericin in *D. melanogaster* (Hoffmann and Reichhart, 2002) but with a different time pattern. Some other AMPs like gloverin and cercropin6 are more sensitive to Gram-positive bacteria with the patter of 24 h higher than 6 h, but other AMPs like moricin1, gallerimycin1 and attacin1, 2, and 10 are equally responded to both bacteria (Fig. 4). The expression levels of moricin1 are similar at 6 and 24 h, while that of gallerimycin1, attacin1, 2, and 10 are higher at 6 than 24 h. For the killed *B. bassiana* fungi injection, the induced level are almost 10 folds lower than the bacteria injections. All the other eight genes are expressed higher at 6 h then reduced at 24 and 48 h, except diapausin1, which expressed higher at 24 than 6 h. All the genes got highly induced, but all go back to normal level at 48 h after 3 killed pathogen injections except diapausin1, which is a very slow response to the pathogens and whose peak is at 24 or 48 h (Fig. 4). Therefore, Gram-negative and positive bacteria can induce both Toll and Imd pathways non-specifically in *M. sexta*. Using bacteria injection cannot separate Toll or Imd pathways properly in *M. sexta*, which is very different from *D. melanogaster*.

4. Conclusion

The 215 immunity-related transcripts discovered are involved in recognition, signal transduction and modulation, and execution mechanisms. Moreover, we find that most

genes displayed mRNA profiles of early response to the pathogens, which are expressed highly at 6 h and reduced at 24 h. We also found the order of the tissue induced level is fat body, hemocytes, and midgut. Diapausins are more preferential to be induced in fat body, some of extracellular serine proteases are specific in hemocytes expression, and caspase are midgut-special genes that get induced upon the immune challenge. Additionally, our results also indicate Gram negative and positive bacteria can induce both Toll and Imd pathways in *M. sexta*. The Toll pathway-specific candidate genes, diapausin1 are more sensitive to the gram-positive bacteria, but lebocinD are equally reacted to two kinds of bacteria. The Imd pathway-specific candidate gene, X-toxs are more sensitive to the gram-negative bacteria, whereas gallerimycin1 are equally reacted to two kinds of bacteria. Some other antimicrobial peptides like gloverin and cercropin6 are more sensitive to gram positive bacteria, most of other antimicrobial peptides are equally responded to both bacteria. Therefore, using bacteria injections cannot separate Toll or Imd pathways properly in *M. sexta*, which is very different from *D. melanogaster*.

Taken together, this study provided a view of immune-related gene expression profiles in response to distinct bacterial and fungal challenges in a lepidopteran insect. Compared to the *D. melanogaster* work, some results are similar, but most of our work challenged their dogma, which cannot be applied to all the other species in the same way, thereby helping improve the understanding of host-pathogen interactions and innate immune response pathways in insects.

Tables

Table 1. The numbers of up and down regulated genes upon immune challenge against

OGS2 (15542 genes)

Tissues	6h up	24h up	6 or 24h up	6 and 24h up	6h down	24h down	6 or 24h down	6 and 24h down
FB	1231	558	1502 (10%)	287 (2%)	1290	619	1710 (11%)	199 (1%)
HC	1417	326	1528 (10%)	215 (1%)	1392	341	1606 (10%)	127 (0.8%)
MG	714	373	924 (6%)	163 (1%)	1016	372	1225 (8%)	163 (1%)

Table 2. The numbers of up and down regulated isoforms upon immune challenge against annotated IMRGIs (684 isoforms)

Tissues	6h up	24h up	6 or 24h up	6 and 24h up	6h down	24h down	6 or 24h down	6 and 24h down
FB	157	96	163 (24%)	90 (13%)	27	15	36 (5%)	6 (0.9%)
HC	142	63	147 (21%)	58 (8%)	46	12	55 (8%)	3 (0.4%)
MG	104	70	115 (14%)	59 (9%)	26	9	30 (4%)	5 (0.7%)

Table S1. The oligonucleotides primer used in real-time qPCR

Gene	Forward primer	Reverse primer
attacin1,10	GCAGGCGACGACAAGAAC	ATGCGTGTTGGTAAGAGTAGC
attacin2	TCTTGGTCTGCCTTCTCGTT	ACTCCAGCAGTCGCAGAACT
moricin1	TGCTTTCTTTAACCTTTGTCCTC	TATTCTAACACAGCCTATAATGCG
cercropin6	CCGTGTTTTATTCTTCGTCTTC	AATCCTTTGACCTGCACCC
gloverin	GCAAGTCGGCAACAATGG	ACCCTGTCCTGTCAGTTTG
diapausin1	AGTTCGCTAACATCGCCTTGA	GGCAGCACTCGTCCTTCTCA
lebocinD	CCTGTACTGCCCTCGATCAT	TTGTATCCCGGGTGAGTAGC
X-toxs	GAAAGACTGCAACTTCCAAGCC	GATCCTATGATGGTCAGGGATACTG
gallerimycin1	AGTGCTCCTAATTTCGGTAGCCT	CTCGTAGAAGACGCATCTTGGTAA
rpS3	CTCAGGCCGAGTCTTTGAGATACA	ACTTCATGGACTTGGCTCTCTGAC

Figures

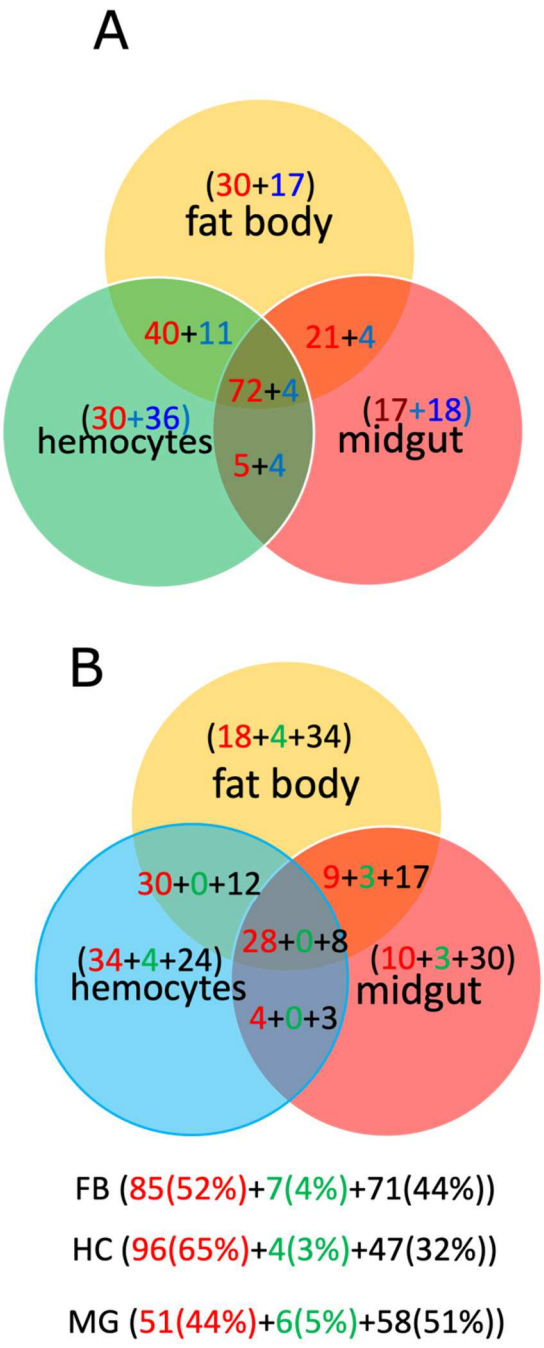


Fig. 1. A. Tissue distribution of up and down-regulated IMRGIs (215 and 94 up- and down-regulated isoforms). **B**. Numbers of induced IMRGIs involved in different time patterns.

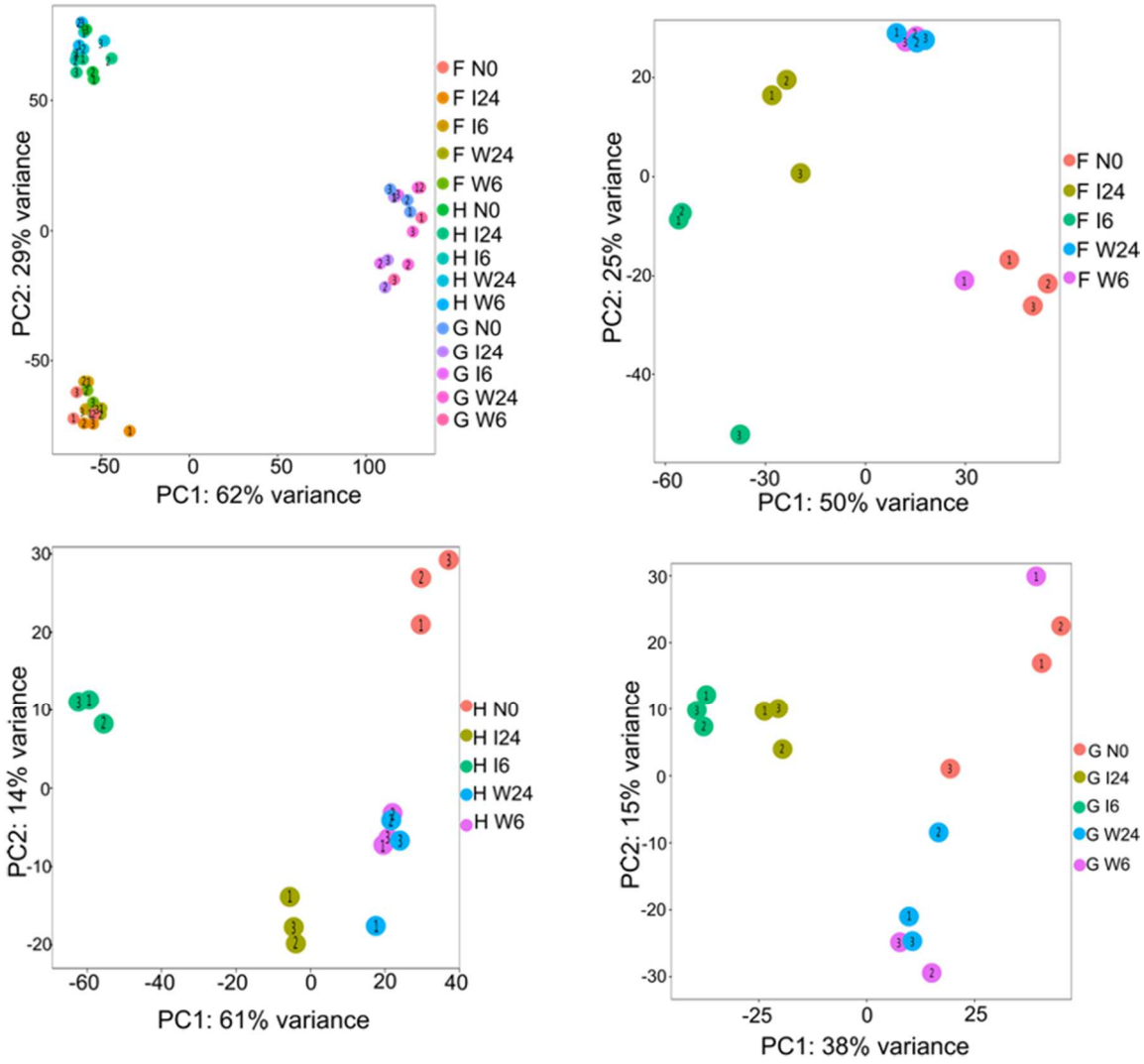


Fig. 2. **A**. Principal component analysis (PCA) of global gene expression of all 15542 OGS2 genes. PCA of 684 annotated IMRGIs expression in FB (**B**), HC (**C**) and MG (**D**).

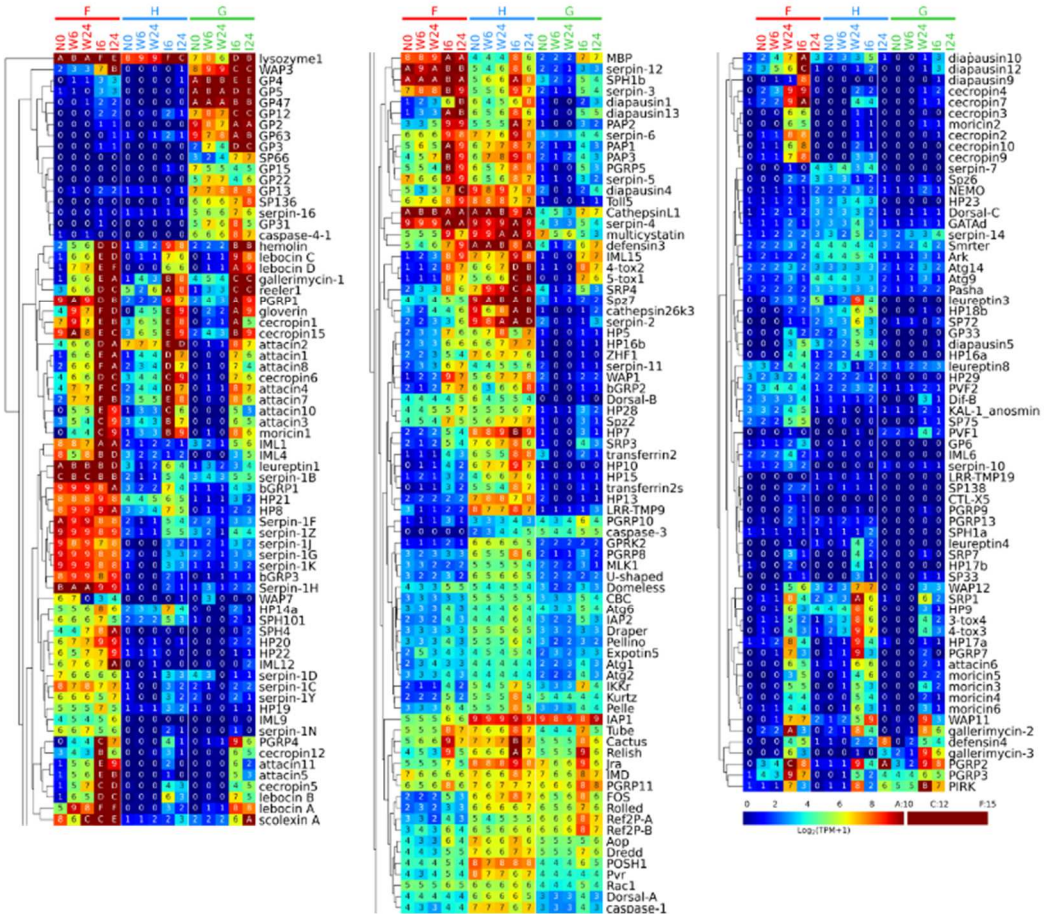


Fig. 3. Expression heat map of all 215 induced IMRGIs in FB, HC or MG. On top of the heat map, the 15 RNA-seq datasets are indicated as naïve larvae (N0), PBS at 6 and 24 h (W6 and 24), immune challenge at 6 and 24 h (I6 and 24) in fat body (F), hemocytes (H) and midgut (G). Log_2 (TPM+1) values for these immune-related transcripts are shown in the gradient heat map from dark blue (0) to maroon (≥10). The values of 0−0.49, 10.50−1.49, 1.50−2.49 … 8.50−9.49, 9.50−10.49, 10.50−11.49, 11.50−12.49 are labeled as 0, 1, 2 … 9, A, B, C, respectively. Relatedness in expression patterns revealed by hierarchical cluster analysis is shown on the left, with abbreviated protein names shown on the right.

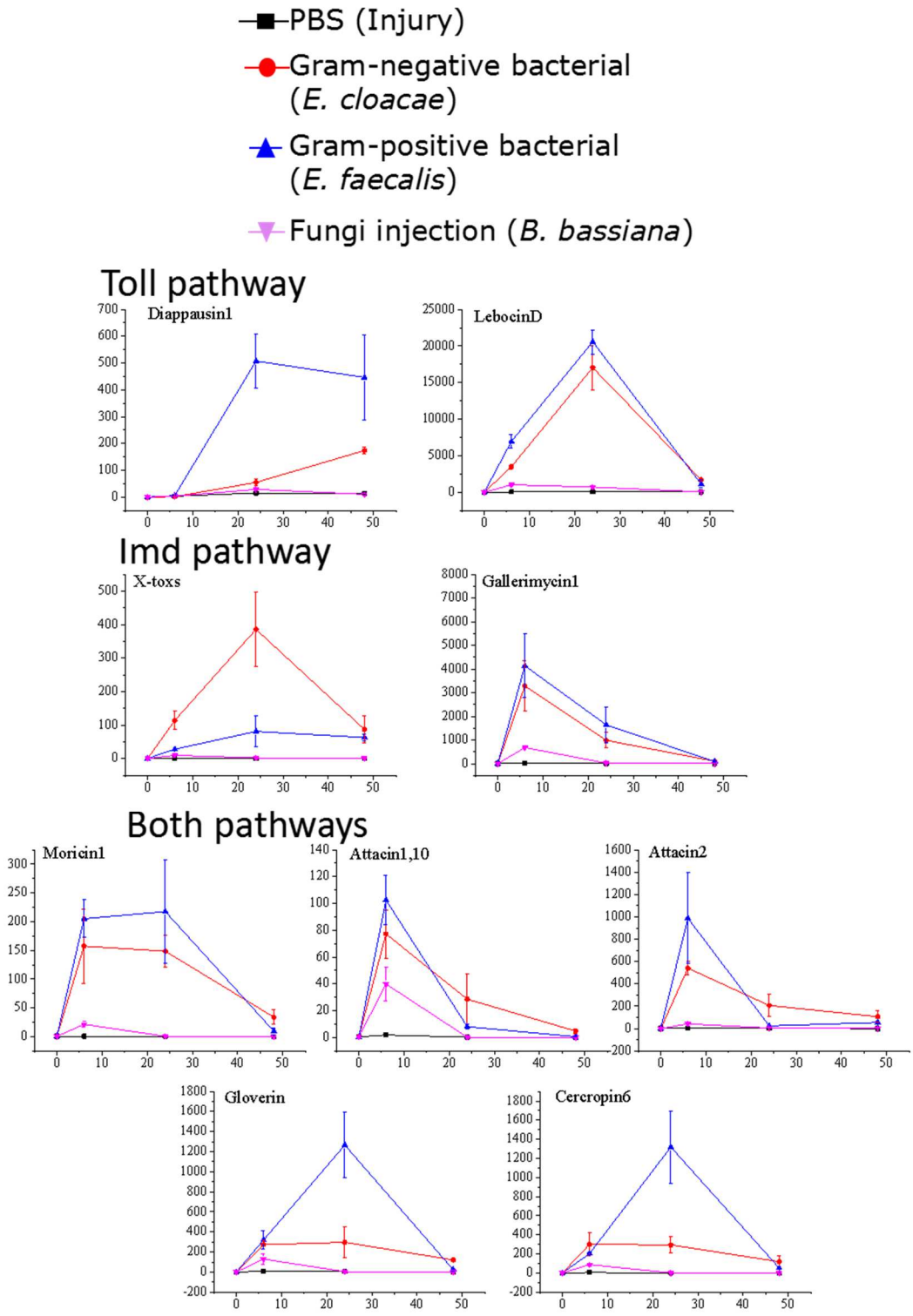


Fig. 4. Nine selected AMP gene expression patterns after dead pathogens injection using qRT-PCR. **A**. Two Toll-pathway candidate genes after different pathogens treatment. **B**.

Two Imd-pathway candidate genes after different pathogens treatment. Injection of 20 μL dead *B. bassiana* (2×10^7 conidia/mL), *E. cloacae* ($A_{600} \approx 1$), or *E. faecalis* ($A_{600} \approx 1$) suspension.

References

Cao, X., He, Y., Hu, Y., Wang, Y., Chen, Y-R., Bryant, B., Clem, R.J., Schwartz, L.M., Blissard, G.W., Jiang, H., 2015. The immune signaling pathways of *Manduca sexta. Insect Biochem. Mol. Biol*. 62, 64−74.

Cao, X., He, Y., Hu, Y., Zhang, X., Wang, Y., Zou, Z., Chen, Y., Blissard, G.W., Kanost, M.R., Jiang, H., 2015. Sequence conservation, phylogenetic relationships, and expression profiles of nondigestive serine proteases and serine protease homologs in *Manduca sexta. Insect Biochem. Mol. Biol*. 62, 51−63.

Cao, X., Jiang, H., 2017. An analysis of 67 RNA-seq datasets from various tissues at different stages of a model insect, *Manduca sexta. BMC Genomics* 18, 796.

Capo, F., Charroux, B., Royet, J., 2016. Bacteria sensing mechanisms in *Drosophila* gut: local and systemic consequences. *Dev. Comp. Immunol*. 64, 11−21.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N.T., Hoffmann, J.A., Kanzok, S.M., Letunic, I., Levashina, E.A., Loukeris, T.G., Lycett, G., Meister, S., Michel, K., Moita, L.F., Müller, H.M., Osta, M.A., Paskewitz, S.M., Reichhart, J.M., Rzhetsky, A., Troxler, L., Vernick, K.D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., Kafatos, F.C., 2002. Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 298, 159–165.

Evans, J.D., Aronstein, K., Chen, Y.P., Hetru, C., Imler, J.L., Jiang, H., Kanost, M.R., Thompson, G.J., Zou, Z., Hultmark, D., 2006. Immune pathways and defense mechanisms in honey bees *Apis mellifera*. *Insect Mol Biol*. 15, 645–656.

Gunaratna, R., Jiang, H., 2013. A comprehensive analysis of the *Manduca sexta* immunotranscriptome. Dev. Com. Immunol. 39, 388−398.

He, Y., Cao, X., Li, K., Hu, Y., Chen, Y., Blissard, G.W., Kanost, M.R., Jiang, H., 2015.

A genome-wide analysis of antimicrobial effector genes and their transcription patterns in *Manduca sexta. Insect Biochem. Mol. Biol*. 62, 23−37.

He, Y., Cao, X., Zhang, S., Rogers, J., Hartson, S., Jiang, H., 2016. Changes in the plasma proteome of *Manduca sexta* larvae in relation to the transcriptome variations after an immune challenge: evidence for high molecular weight immune complex formation. *Mol. Cell. Proteomics* 15, 1176−1187.

Hoffmann, J. A., Reichhart, J. M., 2002. Drosophila innate immunity: an evolutionary perspective. *Nat Immunol.*, 3(2), 121-126.

Hori, A., Kurata, S., Kuraishi, T., 2018. Unexpected role of the Imd pathway in *Drosophila* gut defense against *Staphylococcus aureus*. Biochem. Biophys. Res. Commun. 495, 395−400.

Jiang, H., Vilcinskas, A., Kanost, M.R., 2010. Immunity in lepidopteran insects. In "Invertebrate Immunity" (K. Söderhäll ed.), *Adv. Exp. Med. Biol*. 708, 181−204.

Kanost, M.R., Arrese, E.L., Cao, X., Chen, Y-R, Chellapilla, S., Goldsmith, M., Grosse-Wilde, E., Heckel, D.G., Herndon, N., Jiang, H., Papanicolaou, A., Qu, J., Soulages, J.L. Vogel, H., Walters, J., Waterhouse, R.M., Ahn, S.J., Almeida, F.C., An, C., Aqrawi, P., Bretschneider, A., Bryant, W.B., Bucks, S., Chao, H., Chevignon, G., Christen, J.M., Clarke, D.F., Dittmer, N.T., Ferguson, L.C., Garavelou, S., Gordon, K.H., Gunaratna, R.T., Han, Y., Hauser, F., He, Y., Heidel-Fischer, H., Hirsh, A., Hu, Y., Jiang, H., Kalra, D., Klinner, C., König, C., Kovar, C., Kroll, A.R., Kuwar, S.S., Lee, S.L., Lehman, R., Li, K., Li, Z., Liang, H., Lovelace, S., Lu, Z., Mansfield, J.H., McCulloch, K.J., Mathew, T., Morton, B., Muzny, D.M., Neunemann, D., Ongeri, F., Pauchet, Y., Pu, L.L., Pyrousis, I., Rao, X.J., Redding, A., Roesel, C., Sanchez-Gracia, A., Schaack, S., Shukla, A., Tetreau, G., Wang, Y., Xiong, G.H., Traut, W., Walsh, T.K., Worley, K.C., Wu, D., Wu, W., Wu, Y.Q., Zhang, X., Zou, Z., Zucker, H., Briscoe, A.D., Burmester, T., Clem, R.J., Feyereisen, R., Grimmelikhuijzen, C.J., Hamodrakas, S.J., Hansson, B.S., Huguet, E., Jermiin, L.S., Lan, Q., Lehman, H.K., Lorenzen, M., Merzendorfer, H., Michalopoulos, I., Morton, D.B., Muthukrishnan, S., Oakeshott, J.G., Palmer, W., Park, Y., Passarelli, A.L., Rozas, J., Schwartz, L.M., Smith, W., Southgate, A., Vilcinskas, A., Vogt, R., Wang, P., Werren, J., Yu, X.Q., Zhou, J.J., Brown, S.J., Scherer, S.E., Richards, S., Blissard, G.W., 2016.

Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem. Mol. Biol*. 76, 118−147.

Kingsolver, M.B., Huang, Z., Hardy, R.W., 2013. Insect antiviral innate immunity: pathways, effectors, and connections. *J. Mol. Biol*. 425, 4921−4936.

Lemaitre, B., Hoffmann, J., 2007. The host defense of *Drosophila melanogaster*. *Ann. Rev. Immunol.* 25, 697−743.

Li, M., Christen, J.M., Dittmer, N.T., Cao, X., Zhang, X., Jiang, H., Kanost, M.R., 2018. The *Manduca sexta* serpinome: analysis of serpin genes and proteins in the tobacco hornworm. *Insect Biochem. Mol. Biol*. 102, 21−30.

Lindsay, S.A., Wasserman, S.A., 2014. Conventional and non-conventional *Drosophila* Toll signaling. *Dev. Comp. Immunol*. 42, 16−24.

Rämet, M., 2012. The fruit fly *Drosophila melanogaster* unfolds the secrets of innate immunity. *Acta Paediatr*. 101, 900−905.

Rao, X., Cao, X., He, Y., Hu, Y., Zhang, X., Chen, Y., Blissard, G.W., Kanost, M.R., Yu, X-Q., Jiang, H., 2015. Structural features, evolutionary relationships, and transcriptional regulation of C-type lectin-domain proteins in *Manduca sexta. Insect Biochem. Mol. Biol*. 62, 75−85.

Tanaka, H., Ishibashi, J., Fujita, K., Nakajima, Y., Sagisaka, A., Tomimoto, K., Suzuki, N., Yoshiyama, M., Kaneko, Y., Iwasaki, T., Sunagawa, T., Yamaji, K., Asaoka, A., Mita, K., Yamakawa, M., 2008. A genome-wide analysis of genes and gene families involved in innate immunity of Bombyx mori. Insect Biochem. Mol. Biol. 38, 1087−1110.

Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z.Y., Alvarez, K.S., Bartholomay, L.C., Barillas-Mury, C., Bian, G.W., Blandin, S., Christensen, B.M., Dong, Y.M., Jiang, H., Kanost, M.R., Koutsos, A.C., Levashina, E.A., Li, J.Y., Ligoxygakis, P., MacCallum, R.M., Mayhew, G.F., Mendes, A., Michel, K., Osta, M.A., Paskewitz, S., Shin, S.W., Vlachou, D., Wang, L.H., Wei, W.Q., Zheng, L., Zou, Z., Severson, D.W., Raikhel, A.S., Kafatos, F.C., Dimopoulos, G., Zdobnov, E.M., Christophides, G.K., 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 316, 1738–1743.

Zhang, S., Zhang, X., Gunaratna, R., Najar, F., Wang, Y., Roe, B., Jiang, H., 2011.

Pyrosequencing-based expression profiling and identification of differentially regulated genes from *Manduca sexta*, a lepidopteran model insect. *Insect Biochem. Mol. Biol*. 41, 733−746.

Zhang, X., He, Y., Cao, X., Gunaratna, R.T., Chen, Y., Blissard, G.W., Kanost, M.R., Jiang, H., 2015. Phylogenetic analysis and expression profiling of the pattern recognition receptors: insights into molecular recognition of invading pathogens in *Manduca sexta. Insect Biochem. Mol. Biol*. 62, 38−50.

Zou, Z., Evans, J., Lu, Z., Zhao, P., Williams, M., Sumathipala, N., Hetru, C., Hultmark, D., Jiang, H., 2007. Comparative genome analysis of the *Tribolium* immune system. *Genome Biol*. 8, R177.

Zou, Z., Najar, F., Wang, Y., Roe, B., Jiang, H., 2008. Pyrosequence analysis of expressed sequence tags for *Manduca sexta* hemolymph proteins involved in immune responses. *Insect Biochem. Mol. Biol*. 38, 677−682.

CHAPTER III

DIGESTION-RELATED PROTEINS IN THE TOBACCO HORNWORM, *MANDUCA SEXTA*

Zelong Miao [a, †], Xiaolong Cao [a, †], Haobo Jiang [a, ‡]

[a] Department of Entomology and Plant Pathology, Oklahoma State University,

Stillwater, OK 74078, USA

Key words: phylogenetic analysis, insect immunity, midgut, serine protease, lipase, esterase, carbohydratase, RNA-Seq

Abbreviations: AL and ALH, acidic lipase and homolog; AP and APH, aminopeptidase and homolog; αNAGalase, α-*N*-acetyl galactosaminidase; βNAGase, β-*N*-acetylglucosaminidase; CP and CPH, carboxypeptidase and homolog; DPP, dipeptidyl peptidase; FE, feruloyl esterase; FPKM, fragments per kilobase of transcript per million mapped reads; GH, glycoside hydrolase; GP and GPH, gut serine protease and homolog; JHE, juvenile hormone esterase; LFQ, label-free quantification; LLP, lysozyme-like protein; NL and NLH, neutral lipase and homolog; PGRP, peptidoglycan recognition protein; SE and SEH, serine esterase and homolog; SP and SPH, serine protease and homolog; ssDNase, single stranded DNA 3'-5' exonuclease; TAG, triacylglycerol; TPM, transcripts per kilobase million; ZnCP, zinc carboxypeptidase; ZnP, zinc protease.

Abstract

Food digestion is vital for the survival and prosperity of insects. Research on insect digestive enzymes yields knowledge of their structure and function, and potential targets of antifeedants to control agricultural pests. While such enzymes from pest species are more relevant for inhibitor screening, a systematic analysis of their counterparts in a model insect has broader impacts. In this context, we identified a set of 122 digestive enzyme genes from the genome of *Manduca sexta*, a lepidopteran model related to some major agricultural pests. These genes encode hydrolases of proteins (85), lipids (20), carbohydrates (16), and nucleic acids (1). Gut serine proteases (62) and their noncatalytic homologs (11) in S1A subfamily are encoded by abundant transcripts whose levels correlate well with larval feeding stages. Aminopeptidases (10), carboxypeptidases (10), and other proteases (3) also participate in dietary protein digestion. A large group of 11 lipases as well as 9 esterases are probably responsible for digesting lipids in diets. The repertoire of carbohydrate hydrolases (16) is relatively small, including two amylases, three maltases, two sucreases, two α-glucosidases, and others. Lysozymes, peptidoglycan amidases, and β-1,3-glucanase may hydrolyze peptidoglycans and glucans to harvest energy and defend the host from microbes on plant leaves. One alkaline nuclease is associated with larval feeding, which is likely responsible for hydrolyzing denatured DNA and RNA undergoing autolysis at a high pH of midgut. Proteomic analysis of the ectoperitrophic fluid from feeding larvae validated at least 131 or 89% of the digestive enzymes and their homologs. In summary, this study provides for the first time a holistic view of the digestion-related proteins in a model lepidopteran insect and clues for comparative research in lepidopteran pests and beyond.

1. Introduction

The evolutionary success of diverse insects depends on their abilities to acquire nutrients from

various food sources, detoxify xenobiotics, and kill pathogens. Understanding the biochemistry and molecular biology of insect digestion, detoxification, and defense is, therefore, crucial for the development of novel strategies that control agricultural pests and disease vectors (Terra and Ferreira, 2012; Zhu-Salzman and Zeng, 2015; Saraiva et al., 2016). Genome and gut transcriptome data are available from pests including *Plutella xylostella* (You et al., 2013; Lin et al. 2018), *Spodoptera frugiperda* (Brioschi et al., 2007; Kakumani et al., 2014), *Helicoverpa armigera* (Kuwar et al., 2015; Pearce et al., 2017), *Tenebrio molitor* (Oppert et al., 2018), *Leptinotarsa decemlineata* (Schoville et al., 2018), *Mayetiola destructor* (Zhang et al., 2010) and *Locusta migratoria* (Spit et al., 2016), and from vectors such as *Anopheles gambiae* (Dennison et al., 2016) and *Aedes aegypti* (Canton et al., 2015). Similar data are also available from model insects like *Drosophila melanogaster* (Dutta et al., 2015), *Tribolium castaneum* (Morris et al., 2009) and *Manduca sexta* (Pauchet et al., 2010). In comparison with medically important insects, agricultural pests feed on crops and, hence, food digestion is better studied in many of these species. Digestive proteases and pest responses to protease inhibitors or proteolytically activated bacterial toxins have received special attentions due to close associations with practical problems and biotechnological applications (Khajuria et al., 2010; Srp et al., 2016; Oppert et al., 2012).

The tobacco hornworm *M. sexta* represents a large group of agricultural pests in the order of Lepidoptera and should contribute to the research on insect digestive enzymes on the basis of its well-studied genome and transcriptomes (Kanost et al., 2016; Cao and Jiang, 2017). Thirteen of the 67 RNA-seq datasets are from midgut tissues of *M. sexta* ranging from 2nd instar larvae to late adults. While nondigestive serine proteases (SPs, 107) and their homologs (SPHs, 18) in the S1A subfamily were well annotated (Cao et al., 2015), gut (serine) proteases (GPs) and their noncatalytic homologs (GPHs) in the same group have not yet been reported so far. Prior to the genome analysis, a study of the midgut transcriptome (Pauchet et al., 2010) provided three key functional perspectives of this tissue during larval feeding stages, digestion, detoxification and defense. That

research reported cDNA sequences but not mRNA levels. Lacking gene transcription details, it is unclear which gene products are responsible for digesting major food ingredients or how their roles may vary in distinct life stages. In comparison to SP-like proteins in S1A, less is known about other digestive enzymes that hydrolyze proteins, lipids, carbohydrates, and nucleic acids. Although overview of these enzymes is available for a few agricultural pests (Pearce et al., 2017; Spit et al., 2016; Oppert et al., 2018), in-depth analysis of a complete set of digestion-related proteins is not yet available for any insect to the best of our knowledge. In fact, there is no exquisite separation between digestive enzymes and midgut-specific proteins, between enzymes and their noncatalytic homologs, between extracellular and intracellular proteins, and between feeding and nonfeeding stages. To address these problems, we examined GP(H) sequences, found new SP(H) genes in the genome, and identified other groups of digestive enzymes. We then explored their gene expression patterns, proposed tentative names and functions for these proteins, and identified proteins in digestive juice from feeding larvae to validate our predictions. The systematic analysis in a biochemical model should facilitate research on similar proteins in agricultural pests and development of inhibitors that cause indigestion.

2. Materials and methods

2.1. Identification of M. sexta digestive enzymes and new S1A SP(H)s

Genes of putative digestive enzymes were first selected from expression group-9, -10 and -11, representing genes preferentially expressed in midgut tissues (Cao and Jiang, 2017). Genes with FPKM values lower than 100 in all the 67 libraries were excluded from that study. A BLASTP search was performed against NR database of NCBI in a local supercomputer with "hit-table" used as the output format. Subject sequences with at least one hit (identity >25%, length >50, E-value $<10^{-6}$, and bit score >100) were considered as homologs of the queries. Using Python scripts, query genes were first named after the best matched homologs based on hit-linked information. Highly

expressed genes were manually examined to ensure accuracy. Queries without names or with but unrelated to digestion were eliminated prior to further analysis. Gut (serine) proteases (GPs) and their noncatalytic homologs (GPHs) in the S1A subfamily were set aside from other digestion-related proteins for more extensive examination along with the S1A SP(H)s. Domain structures of protein sequences in MCOT 1.0 (Cao and Jiang, 2015) and *Manduca* Official Gene Set 2.0 (Kanost et al., 2016) were predicted using InterProScan 5 (v5.17). Proteins containing a chymotrypsin-like domain were extracted for comparison with the 193 SP(H)s (Cao et al., 2015) to find new ones. Their gene models were manually improved by crosschecking Oases 3.0, Trinity 3.0, Cufflinks 1.0, and *Manduca* genome contigs (Cao and Jiang, 2015; Kanost et al., 2016).

2.2. Expression profiling

Fifty-two cDNA libraries were sequenced by Illumina technology, each representing a sample of whole larvae, organs, or tissues at various life stages (Kanost et al., 2016). The number of reads mapped onto each transcript in the list of digestive enzymes and their noncatalytic homologs was used to calculate transcripts per kilobase million (TPM) in the libraries by RSEM (Li and Dewey, 2011). Hierarchical clustering of the $\log_2$(TPM +1) values was performed using Seaborn, a Python package with the Euclidean-based metric and average linkage clustering method.

2.3. Sequence properties of M. sexta digestion-related proteins

For S1A SPs and SPHs, sequences were categorized by examining the presence of a His-Asp-Ser catalytic triad. If all three residues existed in the conserved TAAHC, DIAL and GDSGGP motifs, the proteins were considered as SPs or GPs. Sequences lacking one or more of the residues were designated SPHs or GPHs. GP(H)s were first identified in midgut (Pauchet et al., 2010) but GP6, GP33, GPH35 and GPH46 were expressed at similar levels in other tissues (Cao et al., 2015).

For other digestion-related proteins, structural features were revealed by multiple sequence alignment of homologous enzymes whose catalytic mechanisms are known. Protease families are named based on a BLAST search of MEROP peptidase database (https://www.ebi.ac.uk/merops/). Signal peptide was predicted by SignalP 4.1 (Petersen et al., 2011). Domain structure and catalytic residues of each enzyme was predicted using InterProScan (Jones et al., 2014). Residues 190, 216 and 226 (chymotrypsin numbering) (Perona and Craik, 1995), which determine the primary substrate-binding pocket, were identified from the aligned protease domain sequences for predicting substrate specificities of the GPs. Residue annotation of Zn proteases, amino- or carboxy-peptidases, lipases, esterases, and other hydrolases were performed using InterProScan (v5.35).

2.4. Multiple sequence alignment and phylogenetic analysis

A multiple sequence alignment of the 80 entire S1A SP(H)s specifically expressed in midgut were performed using MUSCLE, one module of MEGA 7.0 (http://www.megasoftware.net) under the default settings with maximum iterations changed to 1000. The aligned sequences were converted to NEXUS format by MEGA (Kumar et al., 2016). Phylogenetic analysis was conducted using MrBayes v3.2.6 (Ronquist et al., 2012) under the default model with the setting "nchains=12". MCMC (Markov chain Monte Carlo) analyses were terminated after the standard deviations of two independent analyses was <0.01 or after 10 million generations (SD < 0.01). FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) was used to display the phylogenetic tree.

2.5. Insect rearing and collection of digestive fluid

M. sexta eggs were ordered from Carolina Biological Supply and larvae were reared on an artificial diet (Dunn and Drake, 1983). Whole guts, dissected from 3–5 day 1 or 2, 5th instar larvae, were rinsed three times with sterile phosphate-buffered saline (4.3 mM Na_2HPO4, 1.4 mM KH_2PO4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) and once with sterile 50 mM Tris-HCl, pH 7.0 to

remove hemolymph. After blotting with a dry filter paper, a scission was made through the midgut epithelial layer to expose peritrophic membrane. Upon removal of the membrane and its content, central 3/5 of the whole midgut was excised and transferred to a microfuge tube containing 100 μL of sterile 50 mM Tris-HCl, pH 7.0 and, after gentle mixing, centrifuged at 15,000×*g* for 2–3 min. About 150 μL of the supernatant was collected as a sample of ectoperitrophic fluid or digestive juice and aliquoted for protein concentration measurement, for storage at -80 °C, and for treatment with 1/5 volume of 6×SDS sample buffer at 95 °C for 5 min.

2.6. SDS-PAGE separation, peptide sample preparation, and LC-MS/MS analysis

Proteins in the digestive juice (6 μg) were separated by electrophoresis on a 4−15% Mini-PROTEAN TGX precast protein gel (Bio-Rad). After Coomassie staining, the protein lane was cut into four gel slices and processed as described before (He et al., 2016). Extracted trypsinolytic peptides were dissolved in 45 μL of mobile phase A (0.1% HCOOH in H_2O), and 10 μL were injected onto a heated 75 μm×50 cm Acclaim PepMap RSLC C18 column (Thermo Fisher) using a vented trap-column. Peptides were separated via an HPLC gradient of 4−35% mobile phase B (0.1% HCOOH in 20% H_2O and 80% ACN) developed over a 120-min period. Peptides were ionized in a Nanospray Flex ion source using a charged stainless-steel needle and analyzed by MS and MS/MS in an Orbitrap Fusion quadrupole mass spectrometer (Thermo Fisher) in DNA/Protein Resource Facility at Oklahoma State University. Intact ion m/z's were measured in the FT sector at nominal 120,000 resolution, while MS/MS fragments were fragmented by HCD at 34% energy and their fragments analyzed in the ion trap sector. Ions were selected for MS/MS using the quadrupole (1.6 m/z width), monoisotopic precursor selection, charge state screening of ions +2 to +6, and minimum ion intensities of 5×10^4. Dynamic exclusion (45 sec) was used to minimize repetitive MS/MS sampling, and the scan cycle was repeated every 3 sec.

2.7. Database construction and protein identification

For peptide identification, a database named "Manduca_121119_v3" was assembled from OGS 2.0 (Kanost et al., 2016), MCOT 1.0 (Cao and Jiang, 2015), NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/262/585/GCF_000262585.1_Msex_1.0/GCF_000262585.1_Msex_1.0_protein.faa.gz), and a list of 545 immunity- and digestion-related genes. Peptides were identified by searching this database with MaxQuant (v1.6.10.43) (Cox and Mann, 2008) using the default settings, but with the variable modifications of Cys by acrylamide or iodoacetamide, oxidized Met, cyclization of Gln to pyroglutamate, and acetylation of protein N-termini. Protein groups were assembled by MaxQuant using principles of parsimony.

3. Results

3.1. Identification of digestive enzymes and related proteins in M. sexta, an overview

Digestive enzymes are synthesized by midgut epithelial cells and secreted into ectoperitrophic space and gut lumen to hydrolyze food ingredients during feeding stages of insects. Based on this definition, we examined a list of 1,323 genes preferentially expressed in the midgut (Cao and Jiang, 2017). Of these, group-9 genes are favorably transcribed from the 2nd instar to pre-wandering stage of the 5th instar larvae (*i.e.*, larval feeding stages); group-10 genes are specially expressed from pre-wandering to the end of wandering stage; group-11 represents genes that are transcribed at higher levels in the pupal and adult stages. Among the genes that produce transcripts at considerable levels (*i.e.*, FPKM >100), only a small portion encode digestive enzymes to hydrolyze the bulk of food during larval feeding stages (Fig. 1). Some genes in group-10 and -11 could be related to digestion as well, since their products may participate in midgut tissue recycling in pupae or nectar consumption in adults. Others are also expressed in midgut at lower but substantial levels in feeding larvae. For simplicity, we focus on digestion-related proteins in larval feeding stages. From the BLAST search results, we have identified 174 candidate genes in these groups and classified their protein products into hydrolases of proteins, lipids, carbohydrates, and nucleic acids.

3.2. GPs, GPHs, and other SP-related proteins in M. sexta

The tobacco hornworm has 53 more SP(H) genes in the S1A group than reported before (Cao et al., 2015). In the previous work, we uncovered 193 SP(H)s and divided them into 68 digestive and 125 nondigestive based on the preliminary data of expression profiling. The former greatly overlapped with the SP(H)s uncovered in the transcriptome study of 15,451 OGS2.0 genes (Cao and Jiang, 2017; Table S1). Transcripts of eleven genes (GP16, 23, 29, 31, 54, 64, 67, GPH25, SP67, 95 and 137) in a group named "<9" had similar patterns to those in group-9 but were less abundant (FPKM <100 in all the 67 libraries). We further searched MCOT1.0 (Cao and Jiang, 2015) and *Manduca* OGS2.0 (Kanost et al., 2016) to identify additional S1A SP-like proteins based on the domain features and found 53 new SP(H)s (SPH201–SP253, Table S1, Supplemental text). OGS2.0 models of these genes were improved using MCOT1.0 models based on Msex_1.0 genome assembly. Ten of the 53 genes lack OGS2.0 models. SPH226, 233, 235 and SP252 belong to group-9 or -<9, whereas SP251 is a group-11 member. Totally, 246 S1A SP(H) genes exist in the *M. sexta* genome, close to 257 in *D. melanogaster* (Cao and Jiang, 2018). Eighty of them (66 SPs and 14 SPHs) are in group-9 (48 and 9), -<9 (11 and 2), -10 (1 and 0), and -11 (6 and 3), respectively. In other words, nearly one third of the family members function in the midgut of *M. sexta* at various developmental stages.

3.2.1. Features of the 80 SP(H)s preferentially expressed in the midgut

Except for GP28, 52, GPH50, 70 and SPH235, entire sequences of the 80 SP(H)s are 244−309 residues long (average: 276 residues) (Table S1). Almost all of them are predicted to contain a signal peptide that suggests extracellular functions. Over half of the 66 SPs have trypsin-like specificity, more than chymotrypsins (18) and elastases (13). Predicted proteolytic activation sites of the 66 SPs are almost all located between Arg and Ile and, thus, a minute amount of trypsin is anticipated to rapidly generate a mixture of trypsins, chymotrypsins, and elastases.

While 59 of these gut SPs are expressed during larval feeding stages, transcripts of 14 SPHs are detected in group-9 (GPH7, 41, 50, 60, 70, SPH128, 135, 226, 233), -<9 (GPH25, SPH235), -11 (SPH79, 96, 139). Except for GPH50, 70, SPH135 and 233, all the gut SPHs have a cleavage site between Arg and Ile/Val-Val/Ala/Ile-Gly-Gly (Table S1), indicative of functional importance because, otherwise, these sites would have deteriorated beyond recognition before long. Cleavage-induced conformational change occurs during zymogen activation of S1A SPs (Shibata et al., 2018). The liberated N-terminal hydrophobic residues move from the surface to an internal site where its cationic amino group forms an ion pair with the invariant Asp before the active site Ser, leading to proper formation of the enzyme specificity pocket and oxyanion hole. We suspect that similar conformational change in the SPHs, induced by cleavage at the conserved site and insertion of the new N-terminus, likely associates with their functions (unrelated to catalysis).

3.2.2. Phylogenetic relationships of the 80 GP(H)s and other related findings

Multiple sequence alignment and phylogenetic tree construction led to division of the 80 SP(H) genes into six clades (Fig. 2): GP1 through GPH25 in clade A (20), GP12 through SPH135 in B (11), GP27, SP252, SPH235 and SP251 in C (4), GPH7 through GP32 in D (12), GP8 through GP16 in E (26), and GP28 through SPH233 in F (7). In each clade, most members possess similar features including exon number, enzyme specificity, codon of active site Ser, and genome location. For instance, 19 and 22 genes in the clades A and E contain 4 and 5 exons, respectively. While 15 clade A genes encode trypsin-like SPs, 22 clade E genes code for elastases or chymotrypsins (none for trypsin). Specificities of elastases and chymotrypsins overlap but not with trypsins. Apart from GP55, 21 SPs in clades B, C and D all have trypsin-like specificity and they are more similar to the trypsins in clade A than to members of clades E and F. In clade F, SP95 gene has 7 exons, the other six SP(H)s have 8, and they encode four chymotrypsins, one elastase and two SPHs. Of the 80 SP(H)s, seven use AGY (Y: C/T) codon for the active site Ser residue (5 in clade F; 2 in clade B) and the other 73 use TCN (N: A/T/C/G) codon (or its derivatives in SPHs). We identified nine gene

clusters including GP12-13, GP44-63-SP136-137, and GP1-17-18-22-21 (Fig. 2). Residing in the genome contigs 9016–9021 (data not shown), the largest cluster consists of 18 genes coding for GP2−5, 37, 39, 40, 43, 47, 54, 61, 64, 69, SP65, 66, 63, 62, and 47. The last four had low FPKM (<10) in all the midgut libraries and were thus classified as nondigestive SPs in expression groups A and D (Cao et al., 2015). Probability values of 84−100% provided strong support for the phylogenetic relationships among members of the gene clusters (Fig. 2). While GPH50-70 and SPH79-96 pairs arose from recent gene duplications, their ancestors and the other ten SPH genes are located close to the root. In other words, they derived from SPs so long ago that mutation should have already ruined the R*IVGG region unless it has a cleavage-induced change needed for noncatalytic functions.

3.2.3. Expression profiles of the 80 GP(H)s

GPs and GPHs fall into two expression groups with similar characteristics (Fig. 3A). The group of high expression (GP36 to GP45) consists of 34 genes whose TPM values were in the range of 2^9 and 2^{13} in the midgut tissues from 2nd instar feeding larvae to 5th instar wandering larvae. Their mRNA levels were so high that similar TPM values were observed in whole larvae from the 1st to 3rd instar. The transcripts of GP34, 36, 57, 52, SPH226, and 233 were detected in late embryo. In another midgut sample from wandering larvae, a major drop of mRNA levels to 2^2 and 2^6 was observed in 32 of the 34 genes. While sequencing methods differ, sample difference is more likely responsible for the drop. Around the time larvae cease to feed, a small difference in sampling time appears to have greatly impacted the results. Expression of these genes is generally low in the pupal and adult stages. This expression pattern strongly correlates with the feeding behavior of *M. sexta*. In the group of low expression (GP2 to SP252, 38 genes), TPM values are mostly in the range of 2^4 and 2^8. Their preferential expression in midgut is still obvious in the feeding stages.

3.3. Proteases and their homologs from other families

We identified 36 other protease-related genes in the expression groups 9−11 (Table S2), 23 encoding proteases. Serine proteases include a CP (standing for carboxypeptidase rather than Cys protease) in S10 family and a dipeptidyl peptidase in S9B subfamily; Cysteine proteases are three C14A caspases and one C85 ubiquitin thioesterase; Metalloproteases include eight M1 aminopeptidases (APs) and two homologs, two M24B Xaa-Pro APs, four M12A zinc endopeptidases, nine M14 zinc CPs and two homologs, one M20 dipeptidase, one M28 CP, and one M49 dipeptidyl peptidase. No aspartic protease is found in the midgut groups. The C14A, C85, M20, and M49 peptidases lack a signal peptide and, hence, are not directly related to food digestion. Midgut expression of caspase-4-1 and -4-2 in the feeding through adult stages (Fig. 3B) is likely involved in apoptosis of damaged epithelial cells (Napoleão et al., 2019).

Structural features and expression patterns provided functional insights into these proteins. The finding of M1 and M14 homologs indicates that loss of catalytic residue(s) is not limited to S1A SPs. Noncatalytic homologs may participate in digestion by means other than peptide hydrolysis (*Section 4.2*). ZnCP9, one of the eight M14s, contains all the key residues for catalytic activity (Table S2) but, unlike the other seven, is not produced in the larval feeding stages. It is highly expressed in late pupae and adults perhaps for tissue remodeling and/or digestion of proteins in nectar. In addition to these CPs, Zn proteases 1–4 in subfamily M12A and AP1, 4–10 in family M1 may hydrolyze dietary proteins.

3.4. Lipases, serine esterases, and their noncatalytic homologs

Thirty-five genes in expression groups 9–11 encode proteins for lipid hydrolysis. These include nine neutral lipases (NLs) and three homologs (NLHs), four acidic lipases (ALs) and two homologs (ALHs), twelve serine esterases (SEs) and five homologs (SEHs) (Table S2). As a special group of SEs, ALs may only hydrolyze triacylglycerols whereas NLs account for most of the lipase activity for phospholipid and galactolipid digestion in lepidopteran midgut (Christeller et al., 2011). Other

SEs include esterases of ferulates, acetylcholine, juvenile hormones, and other carboxylic acid esters. Neutral and acidic lipases hydrolyze water-insoluble long-chain triacylglycerols; esterases act on water-soluble short acyl chain esters; their homologs lack the catalytic residue(s) for cleavage of carboxyl ester bonds but may somehow affect lipid binding or digestion. All these proteins adopt an α/β hydrolase fold comprising 8 β-strands and 6 α-helices surrounding the β-sheet (Casas-Godoy et al., 2018). While thirty-one genes belong to group-9 that well associates with digestion, AL2, NL9, α-esterase-3, and feruloyl esterase FE4-7 in group-11 displayed no or low preferential expression in midgut of the feeding larvae (Fig. 3B). Lacking a signal peptide, α-esterase-1 and -2 may not digest dietary esters but may detoxify allelochemicals or insecticides inside epithelial cells (Wang et al, 2015; Guillemaud et al, 1997). The preferential transcription of α-esterase-1 in feeding larvae is more pronounced than α-esterase-2 or -3. In general, lipase mRNAs are several folds higher than other esterases'. This seems consistent with a higher content of neutral lipids and phospholipids than short-chain flavor esters in plant leaves (Lin and Oliver, 2008; Roughan and Batt, 1969).

M. sexta neutral lipases 1−4, 7−9, 11, 12, *H. armigera* NL, and *B. mori* NL2 (Fig. S2A) contain a G(F/H/Y)S**L**G region conserved in pancreatic lipases (*e.g. Equus caballus* NL) of mammals. Ser residue in the motif forms a catalytic triad with Asp and His at the conserved sites. Substitution of the Ser with Gly in NLHs 5, 6 and 10 likely abolishes the lipase activity. *M. sexta* acidic lipases 1−4 and *H. armigera* AL (Fig. S2B) contain a G(H/F)S**Q**G motif found in gastric lipases (*e.g.* human AL1) of mammals. While the active site Ser exists, *M. sexta* ALH5 and 6 lack the Asp and His residues for catalysis to occur, due to a truncation at the carboxyl-terminus.

In a phylogenetic tree of the lipases (Fig. 4A), NLs and ALs along with respective homologs belong to two distinct groups, suggesting their divergence occurred long before the separation of vertebrates and invertebrates, since insect NLs and ALs are closer to mammalian pancreatic and gastric lipases, respectively. Yet, both groups of lipases have the same oxyanion hole (Casas-Godoy

et al., 2018) and are in the same class (http://www.led.uni-stuttgart.de). The other SE-related proteins in *M. sexta* display complex phylogenetic relationships (Fig. 4D) and some levels of sequence conservation around their active sites (Fig. S3). The thirty-five lipase- and SE-like proteins account for a small portion of the SE homologs encoded by the genome (data not shown).

3.5. Carbohydrate hydrolases and their noncatalytic homologs

We identified a total of 32 carbohydratase-related proteins in group-9−11 (Table S2), including two α-amylases, two α-glucosidases, three maltases, two sucrases, one trehalase, two lysozymes, three peptidoglycan recognition proteins (PGRPs), and one β-1,3-glucanase. Based on expression profiles of these genes (Fig. 3B), sixteen encode digestive enzymes (Fig. 1).

As members of glycoside hydrolase family 77 (GH77), α-amylases hydrolyze α-1,4-glycosidic bonds of glycogen, starch, and related polysaccharides. Over-production of amylases may protect insects from plant amylase inhibitors (Franco et al., 2002). *M. sexta* α-amylase-1 mRNA levels are about ten-fold higher than α-amylase-2's during feeding stages (Fig. 3B). Both proteins contain a $(\beta/\alpha)_8$-barrel domain A with a catalytic triad of Asp, Glu and Asp (Fig. S4), an N-terminal Ca^{2+}-binding domain B, and a C-terminal domain C with a Greek key motif.

M. sexta α-glucosidase-2 mRNA levels are twice as high as α-glucosidase-1's. Both α-glucosidase homolog-1 and -2 lack signal peptide and at least one catalytic residue. The former is widely distributed in other tissues, the latter is mainly limited to midgut, and both are favorably produced in midgut cells perhaps for carbohydrate transport, a process that may concur with food digestion in gut lumen and ectoperitrophic space. Maltases, named after their preferred substrate maltose, are the most abundant form of insect α-glucosidases. *M. sexta* maltase-1 mRNA levels are four-fold higher than maltase-2's (Fig. 3B). As maltase-3 TPM value in the midgut of feeding larvae is so low but reaches 2^{5-8} in late pupae and young adults, we suspect it is an adult-specific enzyme that digests sucrose in nectar. Sucrase-1 and -2 hydrolyze sucrose to glucose and fructose.

Their transcript levels are comparable in larval feeding stages, but sucrase-2 mRNA levels are maintained in the pre-wandering and wandering stages. As a result of midgut alkalinity, sucrases are not inhibited by alkaloid sugars (Daimon et al, 2008).

Trehalase hydrolyzes trehalose, a nonreducing sugar of two glucose units linked by an α-1,1-glycosidic bond. Apical trehalase might have a role in food digestion while basal trehalase utilizes trehalose in hemolymph (Terra and Ferreira, 2012). *M. sexta* trehalase expression profile cannot distinguish these roles (Fig. 3B) but the enzyme is not detected in the midgut digestive juice (Table S2). While the trehalase transcripts in midgut are considerably more abundant than in fat body and muscles during the larval feeding stages, the mRNA peak was reached in midgut of wandering larvae and new pupae. The detection of its mRNA in head and several other tissues in various life stages suggests that trahalase is required by all these tissues (including midgut) to utilize trehalose in the hemolymph.

Lysozymes cleave β-1,4-glycosidic bonds in the glycan strands of bacterial peptidoglycans (PGs); certain PGRPs hydrolyze the lactylamide bond between N-acetylmuramic acid and *L*-Ala in the stem peptide (Mellroth et al., 2003). β-1,3-glucanase hydrolyzes plant callose as well as fungal cell wall. *M. sexta* lysozymes and PGRP2−4 may digest cell wall of bacteria (He et al., 2015; Zhang et al., 2005). We think PGRP2, PGRP3, and β-1,3-glucanase are digestive enzymes, based on their elevated mRNA levels in midgut of feeding larvae (Fig. 3B). It is possible that killing pathogens via digestion is a natural defense mechanism of insects.

Myrosinases are β-thioglucosidases that hydrolyze glucosinolates, compounds plants make to fight herbivores and diseases (Winde and Wittstock, 2011). *M. sexta* myrosinase-1 and -2 are made in all larval feeding stages; myrosinase-3 mRNA exists in midgut of 1^{st}–4^{th} instar; myrosinase-4 is produced in adult midgut and Malpighian tubules (Fig. 3B). While some hydrolytic products of glucosinolates are toxic, *M. sexta*, as a specialist, must have developed ways to detoxify or at least

tolerate isothiocyanates as it harvests glucose from their precursors. While catalytic residues of the plant myrosinases are Gln and Glu (Fig. S5), the equivalent residues in the cabbage aphid and tobacco hornworm are both Glu (Barrett et al., 1995; Jones et al., 2002; Bhat et al., 2019).

3.6. Nucleases and alkaline midgut

Four putative nucleases are identified in the three gut-specific groups (Table S2). Two lack signal peptide, RNase Oy does not display any expression association with larval feeding, but the last one is expressed in midgut of larvae in the feeding stages (Fig. 3B). Its ortholog in the silkworm hydrolyzes single- and double-stranded DNA and RNA (Arimatsu et al., 2007a). Since DNA strands separate and RNA molecules undergo autolysis at a high pH, its natural substrate is likely single stranded DNA. The putative nuclease and its orthologs in the lepidopterans form a tight monophyletic group distinct from the homologs in other orders of insects (Fig. 4B).

3.7. Identification of digestion-related proteins in midgut juice of M. sexta feeding larvae

To test if the 122 hydrolases and 25 noncatalytic homologs are digestion-related, we collected fluid in the ectoperitrophic space of 5th instar feeding larvae. SDS-PAGE analysis showed most of the proteins are smaller than 50 kDa (Fig. S6A) and no distinct band was detected at the positions of storage proteins or lipophorins, major hemolymph proteins in this stage. Thus, contamination of hemolymph was minimal. We identified 131 of the 147 proteins in the protein sample by LC-MS/MS analysis (Tables S1 and S2). For the ones with transcriptome and proteome data, a low correlation was observed between $\log_2$(TPM+1) and $\log_2$(LFQ+1) values (Fig. S6B). A similar discordance was observed and examined in detail between plasma protein levels and their mRNA levels in hemocytes and in fat body (He et al., 2016). Of the sixteen proteins not identified in the gut juice (GP5, 13, 15, 16, 23, 29, 53, SP95, 252, SPH235, ZnP1, AL1, AL3, AL4, PGRP3, and myrosinase-3), TPM values of GP5, AL1, AL3, and AL4 gene transcripts were >2^6 in the midgut sample at the corresponding stage of feeding larvae (Fig. 3). Low expression of the other genes

may account for their protein levels below the detection limit. The acidic lipases may digest lipids taken up by gut epithelial cells (*Section 4.3*). In summary, preferential genes expression in midgut of larval feeding stages and examination of structural features (*e.g*., signal peptide) yielded a rate of >95% success in predicting digestion-related proteins.

4. Discussion

4.1. Elucidation of a midgut digestive enzyme system in feeding larvae of M. sexta

Digestive enzymes account for a small number of proteins made by midgut epithelial cells. In addition to the gene products that maintain basic cellular functions, proteins are produced for gut-specific processes, such as digestion, defense, and detoxification (Pauchet et al., 2010). While the epithelial cells secret soluble proteins or enzymes into gut lumen to build up peritrophic membrane and hydrolyze food ingredients, these cells may also contribute to the pool of hemolymph proteins via basal lamina (Palli and Locke, 1987). As such, when genome sequence and gut transcriptome data are available, cautions need to be practiced to correctly identify digestive enzymes.

We adopted a combined approach to define a complete set of digestive enzymes from *M. sexta*. Our transcriptome analysis yielded three groups of gut-specific genes expressed in various life stages (Cao and Jiang, 2017). With the first group closely linked to the larval feeding stages, the other two are much less correlated with food digestion. A BLAST search allowed us to select 174 digestion-related genes for further characterization (Fig. 1). By examining SP(H)s from different sources, we identified gut SPs as candidates of major endopeptidases of dietary proteins. We also examined expression patterns, signal peptides, and catalytic residues of other peptidases, lipases, nucleases and carbohydrate hydrolases. The LC-MS/MS analysis finally validated the existence of 107 digestive enzymes and 24 noncatalytic homologs in the gut secretion from the feeding larvae. Together, the results revealed major components of the entire system of digestive enzymes in *M. sexta*. This framework of information from a model species constitutes a foundation for studying

their counterparts in lepidopteran pests and for future development of inhibitors for crop protection.

High-sensitivity mass spectrometry provided insights into other aspects of the ectoperitrophic fluid proteome. Of the 3,428 hits, 1,971 had LFQ values higher than 1.02×10^6, the lowest intensity in the group of 131 digestive enzymes and homologs. Judged from their names, the 1,840 other proteins came from cytosol (>90), mitochondria (>121), lysosomes (>3), and peroxisomes (>5). This indicates the sample contained intracellular proteins and subcellular organelles, released from lysed epithelial cells undergoing apoptosis and/or damaged tissues during dissection. Additionally, 54 abundant hemolymph proteins were observed in the digestive juice at levels much lower than those in plasma (data not shown). In contrast, we did not detect AL1, AL3 and AL4 (TPM $>2^6$) in the digestive juice. In summary, a part of the proteome complexity may come from contamination.

4.2. A possible role for noncatalytic homologs of digestive enzymes

One surprising result from this study is the detection of proteins that resemble S1A gut serine proteases, M1 aminopeptidases, M14 carboxypeptidases, lipases, JHEs, and other serine esterases, but lack one or more of their catalytic residues. Do these proteins have functions unrelated to substrate hydrolysis? Expression profiles of the 6 GPH, 5 SPH, 2 APH, 2 CPH, 3 NLH, 2 ALH, 2 JHEH and 3 SEH genes displayed close association with feeding, their mRNA levels comparable to the homologous enzymes (Fig. 3), and their proteins detected in the digestive juice (Fig. S6). Early divergence of the enzymes and respective homologs, as exhibited in the phylogenetic trees (Fig. 4), suggests that the homologs are functional because, otherwise, their genes would deteriorate and become inactive pseudogenes. In the case of SPHs, conservation of the proteolytic activation sites (R*IVGG) in most GPHs and fewer SPHs indicates that cleavage at these locations is likely needed for their functions. Based on the conformational change of SP zymogen activation (Shibata et al., 2018), we indicate that a similar change may occur in cleaved GPHs to enable binding of normal substrates or poorly hydrolyzed molecules (*e.g.*, protease inhibitors). While preferential

association with inhibitors by GPHs is not yet confirmed, a certain level of inhibitor sequestration likely protects GPs from inactivation and host from indigestion.

4.3. Possible adaptation to alkaline pH of midgut

In holometabolous insects, neutral and acidic lipases are mainly responsible for hydrolyzing tri-, di-, and mono-acylglycerols, phospholipids, and galactolipids (Horne et al, 2009). But it is not clear how both groups of lipases may efficiently hydrolyze lipids at a particular pH. In lepidopteran insects, NLs may adapt to a high pH but how may ALs change their optimal pH? We first examined whether there is a global down-regulation of AL transcription and found, except for AL2, the average mRNA levels of ALs or ALHs are comparable to those of NL(H)s (Fig. 3B). Remarkably, none of the four ALs was detected in the gut juice but LFQs of ALH5 and ALH6 were 2.31×10^7 (Table S2). In contrast, LFQs of eight NLs and three NLHs encompass 1.02×10^6 and 2.38×10^9. Only NL9, a group-11 member like AL2, was not found in the larval digestive fluid. We suggest that AL1, AL3 and AL4 function as digestive enzymes but, instead of working at an unfavorable pH, they may be transported into acidic lysosomes to digest lipids taken up by midgut epithelial cells through endocytosis. It would be interesting to test this hypothesis and compare the results with insects that have a neutral or acidic midgut.

Another example of adaptation to alkaline midgut could be the nuclease whose expression well correlates with larval feeding. While natural substrates are unclear for this enzyme, we predict it is a single-stranded deoxyribonuclease (ssDNase) based on behaviors of DNA and RNA at a high pH of 10.5 and nutritional value of nucleic acids to insects. Our premise is to some extent supported by two studies, one in *Bombyx mori* (Arimatsu et al., 2007a) and the other in *Choristoneura fumiferana* (Schernthaner et al., 2002). The 41 kDa silkworm enzyme hydrolyzes dsRNA, ssRNA, but not dsDNA at a pH of 7–10. However, knowing the enzyme has an optimal pH >11, the authors did not report ssDNA as a potential substrate. cDNA cloning indicates that it is a pre-pro-enzyme

activated by cleavage between Arg[65] and Ser[66], homologous to a dsRNase of *Serratia marcescens* (Arimatsu et al., 2007b), and has close orthologs in other lepidopterans (Fig. 4B). A digestive DNase with a pH optimum of 10–10.5 was isolated from digestive tract of the spruce budworm (Schernthaner et al., 2002), which cleaves single- and double-stranded DNA. The purified enzyme has a molecular mass of 23 kDa and an N-terminal sequence highly similar to *C. fumiferana* trypsin-1. Recombinant expression of the *M. sexta* nuclease in insect cells is needed to characterize the putative ssDNase in terms of optimal pH and substrates.

Acknowledgements

We would like to thank Dr. Estela Arrese in the Department of Biochemistry and Molecular Biology at Oklahoma State University for an insightful discussion on insect lipases and suggestion of a proteomics study of digestive juice from feeding larvae. Dr. Steve Hartson and Mrs. Janet Rogers kindly provided supports for the LC-MS/MS analysis. This study was supported by NIH grants GM58634 and AI139998. The paper was approved for publication by the Director of Oklahoma Agricultural Experimental Station and supported in part under project OKL03054.

Table S1. Sequences and other features of the 128 *M. sexta* SPs and SPHs in S1A subfamily*

Name	OGS2 ID	GenBank ID	Exp Pat	Exp Grp	DER index	TPM	LFQ	Cleavage	Specificity	Ser	Exon #	Domain
GP1	Msex2.11834-RA	XP_030035912.1	G	9	A++++5	34.8	1.10E+08	PEARIVGG	T(GGD)	TCC	≥5	SP
GP2	Msex2.09655-RB	XP_030032027.1	G	9	8++++5	60.97	7.06E+07	NPQRIVGG	T(DGG)	TCC	4	SP
GP3	Msex2.09651-RA	XP_030032014.1	G	9	D++++5	0	1.29E+06	NPQRIVGG	T(DGG)	TCC	4	SP
GP4	Msex2.09652-RA	XP_030032016.1	G	9	B++++5	208.06	NaN	NPQRIVGG	T(DGG)	TCC	4	SP
GP5	Msex2.09653-RA	XP_030032027.1	G	9	A++++5	394.47	ND	NPQRIVGG	T(DGG)	TCC	4	SP
GPH7	Msex2.04837-RA	XP_030024046.1	G	9	A+++-4	1231.58	1.38E+09	PESRIVGG	-	-	5	SPH
GP8	Msex2.02892-RA	XP_030020947.1	G	9	8++++5	123.25	1.04E+08	DGSRIFNG	E(NGV)	TCT	5	SP
GP9	Msex2.02891-RA	XP_030020943.1	G	9	8++++5	150.6	1.02E+07	DGSRIFNG	E(NGA)	TCT	5	SP
GP10	Msex2.15510-RC	XP_030020942.1	G	9	7++++5	133.69	1.74E+07	DGSRIFNG	E(NGV)	TCT	5	SP
GP11	Msex2.02894-RA	XP_030020952.1	G	9	9++++5	712.5	3.70E+08	DGSRIFNG	E(NGS)	TCT	5	SP
GP12	Msex2.04839-RA	XP_030024073.1	G	9	8++++5	43.53	0.00E+00	TDSRIVGG	T(DGG)	AGT	8	SP LC
GP13	Msex2.04840-RA	XP_030024073.1	G	9	8++++5	45.2	ND	TDSRIIGG	T(DGG)	AGT	8	SP LC
GP15	Msex2.00542-RA	XP_030027026.1	G	11	7++++5	129.08	ND	NVSRIVGG	T(GGD)	TCT	6	SP
GP16	Msex2.13966-RA	XP_030038318.1	G	~9	8++++5	13.47	ND	IASRIVGG	E(GVS)	TCA	1	SP
GP17	Msex2.11838-RA	XP_030035909.1	G	9	B++++5	1968.93	4.27E+08	PNARIVGG	T(GGD)	TCC	6	SP
GP18	Msex2.11835-RA	XP_030035910.1	G	9	B++++5	389.27	1.58E+09	QNARIVGG	T(GGD)	TCT	5	SP

GP21	Msex2.14353-RB	XP_030035914.1	G	9	B++++5	1748.94	1.39E+08	VNARIVGG	T(GGD)	TCC	6	SP
GP22	Msex2.15537-RC	XP_030035916.1	G	9	A++++5	51.14	5.35E+08	QNARIVGG	T(GGD)	TCC	6	SP
GP23	Msex2.02889-RA	XP_030020945.1	G	~9	5++++5	30.95	ND	DGARITGG	C(GGA)	TCC	5	SP
GP24	Msex2.05199-RA	XP_030024878.1	G	9	D++++5	4547.8	7.50E+09	AQSRIVGG	C(SGA)	TCT	5	SP
GPH25	Msex2.14473-RA	n.a.	G	~9	7+++-4	4.13	5.30E+07	GTERISGG	-	-	≥4	SPH
GP26	Msex2.03758-RA	XP_030022107.1	G	9	9++++5	161.84	8.17E+08	NPQRIVGG	T(DGG)	TCC	4	SP
GP27	Msex2.02231-RA	XP_030041329.1	G	11	6++++5	3.35	NaN	TSARIVNG	T(DGG)	TCT	4	SP
GP28	Msex2.01606-RA	XP_030040163.1	G	9	B++++5	3284.6	1.64E+08	SQSRIVSG	C(SVA)	AGC	8	SP
GP29	Msex2.02888-RA	XP_030020944.1	G	~9	5++++5	15.48	ND	DGARITGG	C(GVA)	TCT	5	SP
GP30	Msex2.01182-RA	XP_030037447.1	G	9	9++++5	366.92	7.04E+06	QADRIIGG	C(GAS)	TCT	5	SP
GP31	Msex2.01183-RB	XP_030037456.1	G	~9	5++++5	16.13	7.04E+06	LADRIVGG	C(GAS)	TCT	5	SP
GP32	Msex2.14895-RA	CAM84318.1	G	9	D++++5	3812.46	6.62E+09	RNARIVGG	T(GGD)	TCC	5	SP
GP34	Msex2.09542-RA	XP_030032144.1	G	9	C++++5	2908.89	1.61E+10	SVSRIVGG	T(DGG)	TCC	4	SP
GP36	Msex2.14787-RA	XP_030039303.1	G	9	B++++5	977.31	2.64E+07	SSARIVGG	T(GGD)	TCC	5	SP
GP37	Msex2.09649-RA	XP_030032011.1	G	9	C++++5	118.84	4.57E+09	NPQRIVGG	T(DGG)	TCC	4	SP
GP38	Msex2.05193-RA	XP_030024875.1	G	9	A++++5	1659.79	7.98E+08	LGSRIIGG	E(GRS)	TCT	5	SP
GP39	Msex2.09640-RA	XP_030032023.1	G	9	8++++5	58.96	1.12E+08	PQARISGG	C(GGS)	TCT	4	SP
GP40	Msex2.09639-RA	XP_030032017.1	G	9	B++++5	653.42	1.12E+08	NNARIIGG	T(DGG)	TCC	4	SP
GPH41	Msex2.02890-RB	XP_030020946.1	G	9	C+++-4	2762.35	3.91E+09	IGARVVGG	-	-	5	SPH
GP42	Msex2.01184-RA	XP_030037464.1	G	9	B++++5	854.59	4.73E+08	ALGRIVGG	C(GAS)	TCT	≥5	SP
GP43	Msex2.09656-RA	XP_030032028.1	G	9	6++++5	37.52	3.62E+07	NPQRIVGG	T(DGG)	TCC	4	SP
GP44	Msex2.03224-RA	XP_030021313.1	G	9	B++++5	609.78	1.76E+10	APGRIVGG	T(DGG)	TCC	4	SP
GP45	Msex2.05187-RA	XP_030024857.1	G	9	B++++5	2327.74	2.99E+09	SGQRIVGG	E(NGA)	TCC	5	SP
GP47	Msex2.15528-RC	XP_030032027.1	G	9	8++++5	40.98	NaN	NPQRIAGG	C(GGG)	TCC	4	SP
GPH50	Msex2.03221-RA	XP_030021371.1	G	9	B+++-4	1592.19	4.80E+09	RDPGVISH	-	-	5	LCs SPH
GP51	Msex2.09024-RA	XP_030030990.1	G	10	9++++5	210.16	2.73E+07	PISRIVGG	C(GGS)	TCC	6	SP
GP52	Msex2.01582-RA	XP_030040254.1	G	9	B++++5	1518.72	1.83E+09	ADSRIVSG	E(SVA)	AGT	8	SP LC
GP53	Msex2.01583-RB	XP_030040258.1	G	9	7++++5	4.18	ND	SGSRIVSG	E(SVA)	AGT	8	SP LC
GP54	Msex2.09638-RB	XP_030032015.1	G	~9	7++++5	15.33	1.12E+08	QLERIVGG	T(DGG)	TCC	4	SP
GP55	Msex2.09541-RA	XP_030032142.1	G	9	B++++5	1825.85	9.05E+09	TETRVVGG	C(GTG)	TCT	4	SP
GP56	Msex2.05198-RA	XP_030024879.1	G	9	B++++5	974.43	9.40E+07	GGSRIVGG	C(SGA)	TCT	5	SP
GP57	Msex2.05197-RA	XP_030024873.1	G	9	B++++5	607.62	2.81E+08	VQSRIVGG	C(GGS)	TCT	5	SP
GP58	Msex2.05188-RA	XP_030024857.1	G	9	B++++5	2106.44	2.99E+09	SNSRIVGG	C(GGA)	TCT	5	SP
GP59	Msex2.05196-RA	XP_030024880.1	G	9	C++++5	4640.53	1.05E+06	KVSRIIGG	C(GGS)	TCT	5	SP
GPH60	Msex2.05186-RA	XP_030024847.1	G	9	9+++-4	155.47	1.85E+08	TGQRVVGG	-	-	5	SPH
GP61	Msex2.09644-RA	XP_030032018.1	G	9	C++++5	997.02	8.45E+09	VPQRIVGG	T(DGG)	TCC	4	SP
GP63	Msex2.03225-RA	XP_030021319.1	G	9	9++++5	279.79	6.10E+08	TASRIVGG	T(DGG)	TCT	5	SP
GP64	Msex2.09645-RB	XP_030032019.1	G	~9	6++++5	29.11	3.41E+09	VPQRIVGG	T(DGG)	TCC	4	SP
GP66	Msex2.10710-RA	XP_030033636.1	G	9	A++++5	145.62	1.25E+10	NPQRIVGG	T(DGG)	TCC	4	SP
GP67	Msex2.15531-RC	XP_030033636.1	G	~9	5++++5	35.4	1.25E+10	NPQRIVGG	T(DGG)	TCC	4	SP
GP68	Msex2.05194-RA	XP_030024872.1	G	9	B++++5	1531.77	3.30E+09	AQSRIVGG	C(SGA)	TCT	5	SP
GP69	Msex2.09641-RA	XP_030032022.1	G	9	8++++5	147.69	4.87E+08	PQGRIVGG	E(GTS)	TCC	4	SP
GPH70	Msex2.03222-RA	XP_030021372.1	G	9	B+++-4	628.94	3.66E+10	RDPGVISH	-	-	5	LCs SPH
SP46	Msex2.05195-RA	n.a.	F	11	D--++3	5.87	ND	ATSRIVGG	C(GGS)	TCT	6	SP

SP59	Msex2.08320-RA	XP_030030001.1	F	11	E--++3	18.37	ND	ATDRIVGG	T(DGG)	TCT	4	SP
SP65	Msex2.09646-RA	XP_030032013.1	G	9	7++++5	55.38	7.01E+07	VPQRIVGG	T(DGG)	TCC	4	SP
SP67	Msex2.05198-RD	XP_030024881.1	G	~9	7++++5	974.43	1.56E+07	GGSRIAGG	E(SNA)	TCT	5	SP
SPH79	Msex2.11243-RA	XP_030034521.1	F	11	0--+-1	0	ND	VMGRVAG G	-	-	5	SPH
SP84	Msex2.11797-RB	XP_030035544.1	B	9	A--++3	4.87	ND	GSSRIVGG	T(GGD)	TCC	5	SP
SP93	Msex2.13453-RA	XP_030037568.1	G	9	B++++5	447.69	2.65E+08	TAQRIYAA	E(GHS)	TCT	5	SP
SP95	Msex2.01181-RA	XP_030037591.1	G	~9	6++++5	10.94	ND	SIGRIIGG	E(GVS)	TCG	7	SP
SPH96	Msex2.11243-RA	XP_030038331.1	F	11	D--+-2	0	ND	VMGRVAG G	-	-	5	SPH
SP100	Msex2.14030-RA	XP_030039702.1	F	11	D--++3	0	ND	PGPRIVGG	T(GGD)	TCC	6	SP
SP127	Msex2.01578-RA	XP_030040182.1	G	9	B++++5	816.69	3.53E+08	GPSRIVAG	C(SGS)	AGC	8	SP
SPH128	Msex2.01584-RA	XP_030040206.1	G	9	8+++-4	202.51	6.18E+07	GQSRIVSG	-	AGT	8	SPH
SPH135	Msex2.03223-RA	n.a.	G	9	8+++-4	135.41	2.20E+09	QDLGTPVT	-	-	5	SPH
SP136	Msex2.03227-RA	XP_030021320.1	G	9	8++++5	90.73	5.34E+08	TASRIVGG	T(DGG)	TCT	5	SP
SP137	Msex2.03226-RC	XP_030021314.1	G	~9	6++++5	11.02	4.34E+06	TASRIVGG	T(DGG)	TCT	5	SP
SPH139	Msex2.03764-RB	n.a.	F	11	5--+-2	0.43	ND	VSPRIVGG	-	-	5	SPH
SPH201	Msex2.05941-RA	XP_030026125.1	tmf	0	0--+-1	0	ND	SARRLIAG	-	-	≥7	LCs SPH
SPH202	Msex2.04726-RC	XP_030024397.1	t	0	1--+-1	0	ND	RRKRLTTE	-	-	13	SPH
SPH203	Msex2.04725-RA	XP_030024394.1	t	0	2--+-1	0	ND	ISEESNED	-	-	≥5	SPH
SPH204	Msex2.10688-RA	XP_030033719.1	t	0	0--+-1	0.59	ND	TMVTDSEI	-	-	4	SPH
SPH205	MCOT.W18045.0.0.WXX2W	XP_030037817.1	tmf	0	0--+-1		ND	PNLRIRGG	-	-	≥7	LC SPH
SPH206	Msex2.06955-RA	XP_030027378.1, XP_030027393.1, etc.	tmf	0	0--+-1	0	ND	TNRRIYKG	-	-	7	LCs SPH LCs
SPH207	Msex2.06954-RA	XP_030027391.1	tmf	0	0--+-1	0	ND	NGRRIYRG	-	-	7	LC, SPH LCs
SPH208	Msex2.06957-RA	XP_030027379.1	tmf	0	0--+-1	0	ND	RYRRIYNP	-	-	≥10	LC SPH
SPH209	Msex2.06951-RA	XP_030027389.1, XP_030027400.1	tmf	0	0--+-1	0	ND	DGRRLYKG	-	-	≥8	LCs SPH LCs
SP210	Msex2.13476-RA	XP_030037999.1	t	0	0--+-1	0	ND	GTRRVFGG	C(GGA)	-	3	LC SPH
SPH211	Msex2.01361-RB	XP_030039899.1	t	0	0--+-1	0	ND	LNRKVFKG	-	-	8	SPH LC
SP212	MCOT.C02806.0.0.CPP1W	XP_030021322.1	t	0	0--++2		ND	VQPTIVNG	T(DGT)	TCT	1	SP
SPH213	Msex2.11214-RA	XP_030034878.1	tmf	0	0--+-1	0	ND	ENYQVNAV	-	-	4	SPH
SP214	Msex2.00987-RA	XP_030031293.1	l	0	0--+-1	0	ND	ATKRIVGG	T(DGG)	-	7	frizzle, 2LDLa, LCs SP
SP215	Msex2.07676-RA	XP_030028859.1	l	0	1--++2	0	ND	RNGRIVGG	T(DGG)	-	7	LC SPH
SPH216	Msex2.14792-RA	XP_030039310.1	t	0	0--+-1	1.77	ND	ADQTVPEG	-	-	13	SPH LCs
SPH217	Msex2.05665-RA	XP_030025441.1	t	0	0--+-1	0	ND	SEKVLAGG	-	-	6	SPH LCs
SP218	Msex2.05438-RA	XP_030025104.1	mf	0	0--++2	0	ND	MEGFIVGG	E(GVS)	TCC	3	SP
SP219	Msex2.02731-RA	XP_030020549.1	e	0	0--+-1	0	ND	IDSRIAGG	T(DAS)	-	7	clip, LCs SP-CLIPB
SP220	Msex2.04042-RB	XP_030022754.1	t	0	0--++2	0.25	ND	IVPRVVNG	T(DGG)	TCG	6	SP
SPH221	Msex2.08793-RB	XP_030031090.1	h	0	4--+-1	4.83	7.58E+06	KPYGNRNE	-	-	15	LCs 2 clips, LCs SPH-CLIPA LC
SPH222	Msex2.12043-RA	XP_030036157.1	h	0	4--+-1	6.82	ND	FKPDESDG	-	-	≥4	LCs SPH
SPH223	Msex2.00204-RA	XP_030040280	l	0	1--+-1	0	ND	AVVPTTGQ	-	-	7	LCs clipA SPH-CLIPA
SPH224	Msex2.11880-RC	XP_030036365.1	t	0	1--+-1	0.33	ND	DKNSITSG	-	-	7	LCs SPH
SPH225	Msex2.01640-RB	XP_030040456.1	mf	0	1--+-1	0	ND	TKAEVTEG	-	-	≥7	SPH LCs
SPH226	Msex2.02219-RB	XP_030041352.1	G	9	A+++-4	669.78	7.18E+07	MQGRIANG	-	-	8	SPH LC
SPH227	Msex2.04936-RA	XP_030024599.1	mf	0	0--+-1	0	ND	NTRRIIIG	-	-	10	SPH LCs
SPH228	MCOT.C04542.0.0.TOO5W	XP_030024614.1	mf	0	2--+-1		ND	DTRRILYS	-	-	12	SPH LCs

SPH229	Msex2.04942-RB	XP_030024595.1	mf	0	1--+-1	0.16	ND	CQRKIING	-	-	9	SPH LC
SPH230	Msex2.04945-RD	XP_030024600.1	mf	0	0--+-1	0.15	ND	CSRRIVGG	-	-	9	SPH LCs
SPH231	Msex2.06355-RB	XP_030026414.1	t	0	1--+-1	0	ND	LLYDFVNR	-	-	5	LCs SPH
SPH232	Msex2.04266-RB	XP_030023149.1	l	0	3--+-1	0.16	ND	GQLRIIKG	-	-	6	SPH (LCs)
SPH233	Msex2.09028-RA	XP_030030987.1	G	9	B+++-4	1457.86	1.06E+10	VQGRSAGV	-	-	8	SPH
SPH234	Msex2.12379-RA	n.a	tmf	0	0--+-1	0	ND	ALRKYISS	-	-	≥7	SPH
SPH235	Msex2.06277-RA	XP_030026347.1	G	~9	7+++-4	8.34	ND	ENLAIVVS	-	-	≥8	LCs SPH
SPH236	Msex2.05437-RA	XP_030025103.1	m	0	0--+-1	0	ND	IQHLIYKG	-	-	3	SPH
SPH237	Msex2.09948-RD	XP_030032389.1	mf	0	1--+-1	3.96	ND	FKPRDGSL	-	-	8	SPH LCs
SPH238	MCOT.C04545.1.0.OOO5W	XP_030024610.1	mf	0	0--+-1		ND	QTRRILGG	-	-	10	SPH LC
SPH239	Msex2.04946-RB	XP_030024603.1	m	0	0--+-1	0	ND	CQRRIVNG	-	-	8	SPH LC
SPH240	Msex2.05948-RA	XP_030026188.1	t	0	0--+-1	0	ND	ALNENVSL	-	-	1	SPH
SPH241	MCOT.C12683.0.0.TOO2W	n.a.	t	0	0--+-1		ND	SSEHIGAG	-	-	7	SPH LCs
SPH242	Msex2.01504-RA	XP_030040534.1	t	0	2--+-1	0.16	ND	EIDVSTGN	-	-	11	SPH LCs
SPH243	MCOT.C13147.0.0.TNN1W	XP_030039767.1	t	0	2--+-1		ND	SQLHEIRT	-	-	1	LC SPH LC
SPH244	Msex2.04935-RA	XP_030024594.1	mf	0	1--+-1	0	ND	NTRRIING	-	-	10	SPH LC
SPH245	MCOT.C10628.1.1.COO1O	XP_030035792.1	l	0	3--+-1	0	ND	LGNIIFGD	-	-	10	SPH LCs
SPH246	MCOT.C01082.2.3.TOO1O	XP_030037375.1	l	0	4--+-1	0	ND	KDAQIGGA	-	-	8	LCs SPH
SPH247	Msex2.00416-RA	XP_030022159.1	t	0	1--+-1	0	ND	DDGDSAPF	-	-	7	SPH
SPH248	Msex2.11068-RA	n.a.	l	0	1----0	0	ND	AWARIRCE	-	-	6	SPH
SP249	Msex2.03212-RA	XP_030021323.1	t	0	0--++2	0.18	ND	VKPTILNG	T(DGT)	TCC	1	SP
SP250	Msex2.04455-RA	XP_030023657.1	h	0	6---+2	19.83	ND	YELRVIQG	T(DGS)	TCC	5	CC-SP
SP251	MCOT.C12910.0.0.TOO5O	XP_030039503.1	G	11	3+-++3	0	ND	EDVRIVGG	T(DGG)	AGC	5	SP
SP252	MCOT.M13955.0.0.MWOXO	XP_030040859.1	G	~9	6++-+4	7.78	ND	RVKRVIKG	T(DGG)	TCA	13	SP
SP253	Msex2.14365-RA	XP_030038892.1	t	0	0---+1	0	ND	EDYKIRAG	T(DGG)	AGT	4	SP

*: The enlisted SP(H)s include 73 digestion-related GP(H)s, 53 newly identified SP(H)s (SPH201 to SP253), and 7 SP(H)s with expression pattern B or F (Cao et al., 2015). GP(H)s are SP(H)s first identified in the midgut EST project (Pauchet et al., 2010). GP6, 33, GPH35, 46 are not gut-specific but some SP(H)s are, including SP84 (pattern B or group-9), 46, 59, 100, SPH79, 96, 139 (pattern F or group-11), SP251 (group-11), SP252, SPH226, 233 and 235 (group-9 or <9). **Exp Pat**: expression patterns B, F, and G (Cao et al., 2015); tissues "e" for embryo, "f" for fat body, "m" for Malpighian tubules, "t" for testis, and "l" for low and "h" for high in most libraries; **Exp Grp**: newly defined expression groups 0 (not detected in midgut), 9, <9 (similar to group 9 but FPKM <100), 10, and 11 (*Section 3.1*); **LFQ**: ND for undetected, NaN for "not a number", or a value for normalized protein level; **TPM**: transcripts per kilobase million in the RNA-Seq dataset of day 1, 5th instar midgut; Digestive enzyme-relatedness index (**DERI**): expression level (5−E: "+"; 0−4: "-" 1st position), expression pattern (G: "+"; others: "-", 2nd position), expression group (9 or <9: "+"; others: "-", 3rd position), signal peptide ("+" or "-", 4th position), and catalytic activity (SP: "+"; SPH: "-", 5th position), and the total number of "+"'s in positions 1−5 (5: digestive enzyme; 4: digestion-related; 3–0: digestion-unrelated). **Cleavage:** predicted activation cleavage site with the preceding residue marked in *red* font; **Specificity**: predicted enzyme specificity and its key determinants (in parentheses), T for trypsin, C for chymotrypsin, E for elastase, "-" for inapplicable; **Ser**: TCN or AGY codon of the active site serine residue; **Domain**: CC for coiled coil, LC for low complexity, SP and SPH for S1A protease domain and its noncatalytic homologous domain.

Table S2. Features of selected non-SP(H) proteins produced in *M. sexta* midgut cells*

Substrate group	Enzyme group	Enzyme names	GenBank ID	Msex ID	Catalytic residues	Structural features	Group	DER index	LFQ	TPM (L5d1)
Proteins	serine protease	S1A in Table 1	…	…	…	…	…	…	…	…

		serine CP, S10	XP_030030694.1	Msex2.08351-RB		18-436 S10 Ser CP	9	A	NaN	1.08E+02
		dipeptidyl peptidase -4 (DPP4) S9B	XP_030022993.1	Msex2.04209-RA		102-563dipeptidylpeptidase IV	11	A	3.63E+07	1.54E+02
	cycteine protease	ubiquitin thioesterase C85	XP_030020936.1	Msex2.02898-RA	9K, 42E, 44V, 45C, 74T, 76I	8-69 ubiquitin-related D; 106-230 OTU D	10	D	NaN	1.73E+01
		caspase-4-1 C14A	XP_030031593	Msex2.09469-RB		202-458 caspase-like; 218-335 C14 p20 CD; 357-454 C14 p10 NCD	9	D	5.10E+06	1.24E+02
		caspase-1 C14A	XP_030024758.1	Msex2.05494-RA	182G, 203P	40-293 caspase-like; 56-179 C14 p20 CD; 198-293 C14 p10 NCD	10	D	NaN	6.90E+00
		caspase-4-2a C14A	XP_030031595.1	Msex2.09470-RA		195-448 caspase-like; 211-328 C14 p20 CD; 351-447 C14 p10 NCD	11	D	NaN	1.36E+01
	metalloprotease	Zn protease-1 (ZnP1), M12A	XP_030038102.1	Msex2.13842-RA		94-289 M12A metallopeptidase	9	A	ND	0.00E+00
		Zn protease-2 (ZnP2), M12A	XP_030037219.1	Msex2.12971-RA		91-290 M12A metallopeptidase	9	A	1.30E+08	3.34E+02
		Zn protease-3 (ZnP3), M12A	XP_030022633.1	Msex2.03344-RA		65-262 M12A metallopeptidase	11	D	ND	0.00E+00
		Zn protease-4 (ZnP4), M12A	XP_030022631.1	Msex2.03345-RA		80-278 M12A metallopeptidase	11	D	ND	0.00E+00
		aminopeptidase-1 (AP1) M1	XP_030032717.1	Msex2.10148-RA		36-428 M1 Ala N-peptidase; 573-884 ERAP1 C-term D	9	A	1.13E+08	5.10E+02
		aminopeptidase homolog-2 (APH2) M1	XP_030032718.1	Msex2.10149-RA		93-443 M1 Ala N-peptidase; 611-928 ERAP1 C-term D	9	B	NaN	5.06E+01
		aminopeptidase homolog-3 (APH3) M1	XP_030032720.1	Msex2.10150-RA		88-529 M1 Ala N-peptidase; 602-916 ERAP1 C-term D	9	B	1.46E+06	5.86E+01
		aminopeptidase-4 (AP4) M1	XP_030032733.1	Msex2.10152-RA		49-447 M1 Ala N-peptidase; 594-894 ERAP1 C-term D	11	A	8.08E+07	1.08E+02
		aminopeptidase-5 (AP5) M1	XP_030032731.1	Msex2.10153-RA		41-447 M1 Ala N-peptidase; 589-890 ERAP1 C-term D	9	A	1.46E+07	2.03E+02
		aminopeptidase-6 (AP6) M1	XP_030032728.1	Msex2.10154-RB		47-451 M1 Ala N-peptidase; 595-897 ERAP1 C-term D	9	A	NaN	9.90E+01
		aminopeptidase-7 (AP7) M1	XP_030032730	Msex2.10155-RB		41-440 M1 Ala N-peptidase; 583-885 ERAP1 C-term D	9	A	2.88E+07	3.06E+02
		aminopeptidase-8 (AP8) M1	XP_030032722	Msex2.10156-RA		44-452 M1 Ala N-peptidase; 597-897 ERAP1 C-term D	9	A	1.08E+08	1.08E+02
		aminopeptidase-9 (AP9) M1	XP_030032721.1	Msex2.10157-RA		64-465 M1 Ala N-peptidase; 611-913 ERAP1 C-term D	9	A	2.46E+07	9.16E+01
		aminopeptidase-10 (AP10) M1	XP_030032725.1	Msex2.10158-RA		64-405 M1 Ala N-peptidase; 611-915 ERAP1 C-term D	9	A	2.46E+07	1.62E+02
		Xaa-Pro aminopeptidase-2 (XP AP2) M24B	XP_030030742.1	Msex2.08599-RA		36-186 164-348 N-peptidase P N-term D; 364-635 M24 C-term D	9	A	4.22E+08	6.17E+01
		Xaa-Pro aminopeptidase-3 (XP AP3) M24B	XP_030038221.1	Msex2.13728-RA		46-144 N-peptidase P N-term D; 164-251 M24 str. D; 276-338 M24 C-term D	9	A	5.05E+07	4.45E+01
		zinc carboxypeptidase-1 (ZnCP1) M14	XP_030023343.1	Msex2.04135-RA		19-109 PI D; 128-427 M14 CP	9	A	1.59E+06	1.68E+03
		zinc carboxypeptidase-2 (ZnCP2) M14	XP_030023345.1	Msex2.04136-RA		17-105 PI D; 122-414 M14 CP	9	A	2.86E+08	2.31E+02
		zinc carboxypeptidase-3 (ZnCP3) M14	XP_030023344.1	Msex2.04137-RA		19-106 PI D; 128-418 M14 CP	9	A	1.81E+07	1.41E+03
		zinc carboxypeptidase-4 (ZnCP4) M14	XP_030023346.1	Msex2.04138-RA		17-105 PI D; 122-414 M14 CP	9	A	4.20E+09	4.79E+02

		zinc carboxypeptidase homolog-5 (ZnCPH5) M14	XP_030023347.1	Msex2.04140-RA		19-105 PI D; 120-397 M14 CP	9	B	1.63E+06	8.27E+01
		zinc carboxypeptidase-6 (ZnCP6) M14	XP_030023348.1	Msex2.04141-RA		18-103 PI D; 123-412 M14 CP	9	A	6.98E+08	7.20E+02
		zinc carboxypeptidase-7 (ZnCP7) M14	XP_030023351.1	Msex2.04142-RA		17-107 PI D; 123-421 M14 CP	9	A	2.31E+06	3.60E+02
		zinc carboxypeptidase-8 (ZnCP8) M14	XP_030023303.1 and 305.1	Msex2.04366-RB		22-104 PI D; 121-409 M14 CP	9	A	2.36E+07	2.14E+02
		zinc carboxypeptidase-9 (ZnCP9) M14	XP_030024203.1	Msex2.05027-RA		24-105 PI D; 116-396 M14 CP	11	D	ND	0.00E+00
		zinc carboxypeptidase-10 (ZnCP10) M14	XP_030028978.1	Msex2.07772-RB		20-101 PI D; 119-417 M14 CP	9	A	2.87E+07	2.77E+02
		zinc carboxypeptidase homolog-11 (ZnCPH11) M14	XP_030028973.1	Msex2.07776-RA		19-108 PI D; 119-400 M14 CP	9	B	5.07E+06	2.43E+02
		carboxypeptidase Q (CP-Q) M28	XP_030022999.1	Msex2.04211-RA	protease or protease-like domain interface 215H, 216T, 217G	196-293 PA D; 319-498 M28	11	A	1.70E+07	1.82E+02
		dipeptidyl peptidase-3 (DPP3) M49	XP_030026731.1	Msex2.06104-RC		51-74 dipeptidylpeptidase III	10	D	4.74E+08	5.73E+01
		dipeptidase (DP), cytosolic nonspecific M20	XP_030028043.1	Msex2.06873-RA		97-468 M20; 211-385 dimer D	9	D	3.05E+09	2.04E+02
Lipids	lipase	neutral lipase-1 (NL1)	XP_030029556.1	Msex2.00709-RA	184S, 208D, 272H	52-334 AB-hydrolase TAG lipase	9	A	4.56E+07	2.45E+02
		neutral lipase-2 (NL2)	XP_030029542.1	Msex2.00710-RA	183S, 207D, 269H	53-331 AB-hydrolase TAG lipase	9	A	1.02E+06	1.15E+03
		neutral lipase-3 (NL3)	XP_030029362.1	Msex2.00711-RB	184S, 208D, 270H	50-332 AB-hydrolase TAG lipase	9	A	3.53E+08	4.07E+02
		neutral lipase-4 (NL4)	XP_030022432.1	Msex2.03896-RA	185S, 209D, 270H	51-332 AB-hydrolase TAG lipase	9	A	4.38E+07	9.96E+01
		neutral lipase homolog-5 (NLH5)	XP_030022470.1	Msex2.03905-RA	186G, 210D, 270H	6-250 AB-hydrolase TAG lipase homolog	9	B	1.09E+06	2.90E+02
		neutral lipase homolog-6 (NLH6)	XP_030022462.1	Msex2.03906-RA	142G, 166D, 227H	23-297 AB-hydrolase TAG lipase homolog	9	B	2.56E+09	1.26E+03
		neutral lipase-7 (NL7)	XP_030022474.1	Msex2.03908-RA	185S, 209D, 270H	56-346 AB-hydrolase TAG lipase	9	A	6.43E+07	7.20E+02
		neutral lipase-8 (NL8)	XP_030022480.1	Msex2.03909-RA	185S, 209D, 270H	60-332 AB-hydrolase TAG lipase	9	A	7.70E+06	3.55E+02
		neutral lipase-9 (NL9)	XP_030032052.1	Msex2.09694-RA	170S, 198D, 272H	27-329 AB-hydrolase TAG lipase	11	D	ND	1.81E+01
		neutral lipase homolog-10 (NLH10)	XP_030032286.1	Msex2.09846-RA	145G, 173D, 236H	27-298 AB-hydrolase TAG lipase homolog	9	B	1.71E+08	7.85E+02
		neutral lipase-11 (NL11)	XP_030033676.1	Msex2.10726-RA	185S, 209D, 269H	49-333 AB-hydrolase TAG lipase	9	A	2.58E+08	4.44E+02
		neutral lipase-12 (NL12)	XP_030035750.1	Msex2.12025-RA	181S, 205D, 263H	45-327 AB-hydrolase TAG lipase	9	A	2.38E+09	2.29E+03
		acidic lipase-1 (AL1)	XP_030033175.1	Msex2.10332-RA	193S, 368D, 399H	43-421 AB-hydrolase lipase	9	A	ND	6.05E+02
		acidic lipase-2 (AL2)	XP_030041138.1	Msex2.02034-RA	184S, 358D, 389H	112-411 AB-hydrolase lipase	11	D	ND	2.40E-01
		acidic lipase-3 (AL3)	XP_030022206.1	Msex2.03554-RA	189S, 361D, 392H	42-414 AB-hydrolase lipase	9	A	ND	3.82E+02
		acidic lipase-4 (AL4)	XP_030029283.1	Msex2.07865-RA	191S, 365D, 396H	42-418 AB-hydrolase lipase	9	A	ND	1.39E+03
		acidic lipase homolog-5 (ALH5)	XP_030038008.1	Msex2.13484-RA	165S, na, na	22-260 AB-hydrolase lipase homolog	9	B	2.31E+07	1.07E+02
		acidic lipase homolog-6 (ALH6)	XP_030037848.1	Msex2.13487-RA	165S, na, na	22-260 AB-hydrolase lipase homolog	9	B	2.31E+07	1.82E+00
	serine esterase	α-esterase-1 (αE1)	XP_030021129.1	Msex2.00350-RA	183S, 317E, 440H	3-532 B-type CE	9	D	4.89E+08	9.98E+02
		α-esterase-2 (αE2)	XP_030028867.1	Msex2.07667-RB	204S, 333E, 441H	19-534 B-type carboxylesterase	9	D	3.07E+08	1.16E+02
		α-esterase-3 (αE3)	XP_030029533.1	Msex2.07963-RB	199S, 331E, 446H	8-531 B-type CE	11	A	NaN	3.35E+01
		α-esterase-4 (αE4)	XP_030022822.1	Msex2.03978-RB	206S, 328E, 442H	5-531 B-type CE	9	A	NaN	2.15E+02

		α-esterase homolog-5 (αEH5)	XP_030023018.1	Msex2.04204-RA	202D, 325E, 425H	1-515 B-type CE	9	B	NaN	2.13E+02
		α-esterase homolog-6 (αEH6)	XP_030030744.1	Msex2.08585-RA	197D, 311E, 418H	8-506 B-type CE	9	B	NaN	9.92E+02
		α-esterase-7 (αE7)	XP_030027603.1	Msex2.06988-RB	204S, 326E, 440H	8-532 B-type CE	9	A	4.85E+06	6.32E+01
		SEH4-1	XP_030026304.1	Msex2.06294-RA	208K, 337E, 463S	16-564 B-type CE	9	B	2.41E+06	1.45E+02
		SEH4-2	XP_030026301.1	Msex2.06295-RA	211E, 342E, 469T	5-566 B-type CE	9	B	3.71E+07	4.66E+02
		SEH4-3	XP_030026302.1	Msex2.06296-RA	210R, 341E, 467T	7-575 B-type CE	9	B	1.47E+07	2.03E+02
		esterase FE4-1	XP_030034366.1	Msex2.11175-RB	201S, 349E, 445H	2-535 B-type CE	9	A	1.61E+07	1.64E+02
		esterase FE4-2	XP_030040069.1	Msex2.01458-RA	202S, 322E, 426H	7-513 B-type CE	9	A	1.77E+06	3.55E+02
		esterase FE4-3	XP_030040068.1	Msex2.01457-RA	202S, 322E, 428H	8-515 B-type CE	9	A	4.91E+07	5.83E+02
		esterase FE4-4	XP_030025623.1	Msex2.05782-RA	201S, 331E, 445H	2-534 B-type CE	9	A	1.61E+07	7.10E+01
		esterase FE4-5	XP_030027701.1	Msex2.07029-RB	202S, 330E, 443H	2-534 B-type CE	9	C	ND	5.18E+00
		esterase FE4-6	XP_030035718.1	Msex2.12089-RB	201S, 331E, 445H	22-551 B-type CE	9	A	1.61E+07	8.03E+01
		esterase FE4-7	XP_030037025.1	Msex2.13096-RA	201S, 331E, 445H	2-533 B-type CE	11	A	1.61E+07	9.44E+01
Carbohydrates	α-amylase	α-amylase-1	XP_030040954.1	Msex2.00124-RA	211D, 248E, 313D	20-405 GH13CD; 409-499 α-amylase, C-term. D, all β, IPR006046	9	A	8.40E+09	2.06E+03
		α-amylase-2	XP_030024430.1	Msex2.05241-RC	210D, 247E, 312D	1-117 RTD; 125-530 GH13CD; 514-604 α-amylase, C-term. D, all β, IPR006046	9	A	1.06E+09	1.78E+02
	β-glucanase	β-1,3-glucanase	XP_030023420.1	Msex2.04401-RE	188E, 190D, 193E	28-375 GH16	9	A	1.72E+08	2.92E+02
	chitinase	β-N-acetylglucosaminidase (βNAGase1)	XP_030029670.1	Msex2.08131-RA		63-208β-hexosaminidase D2; 208-594 GH20 CD	9	D	3.98E+07	1.22E+01
		β-N-acetylglucosaminidase (βNAGase2)	XP_030030429.1	Msex2.08667-RA		15-103 β-hexosaminidase D2; 102-405 GH20 CD; 225-285 PI-specific PLC, XD	11	D	6.08E+06	1.20E-01
		chitinase-8 (CHT8)	XP_030034142.1	Msex2.10627-RA	120D, 122E	11-374 GH18 CD; 1118-1179 CBD	11	D	ND	0.00E+00
	lysozyme	lysozyme-1	XP_030024172.1	Msex2.04822-RA	50E, 68D	19-134 GH22 lysozyme	11	C	NaN	1.45E+02
		lysozyme-2	XP_030024171.1	Msex2.04820-RA	51E, 69D	21-134 GH22 lysozyme	10	C	NaN	1.24E+01
		lysozyme-like protein (LLP2)	XP_030024170.1	Msex2.04821-RA	51E, 69N	27-134 GH22 lysozyme-like	9	D	ND	0.00E+00
	catalytic PGRP	PGRP2	XP_030021983.1	Msex2.03607-RC	60H, 96Y, 170H, 176T, 178C	31-195 PGRP CD	9	A	1.42E+07	1.03E+03
		PGRP3	XP_030021968.1	Msex2.03605-RA	65H, 101Y, 175H, 181T, 183C	34-201 PGRP CD	9	A	ND	1.44E+01
		PGRP4	n.a.	Msex2.03606-RB	60H, 95Y, 169H, 175T, 177C	29-194 PGRP CD	10	C	ND	3.58E+00
	α-glucosidase	α-glucosidase-1	XP_030026573.1	Msex2.06470-RB	400D, 458D	236-536 GH31	9	A	1.54E+07	1.18E+02
		α-glucosidase-2	XP_030026572.1	Msex2.06471-RA	374D, 432D	221-625 GH31	9	A	7.63E+07	2.73E+02
		α-glucosidase homolog-1 (α-glucosidase H1)	XP_030028824.1	Msex2.07554-RE	687D, 744P	547-885 GH31	11	D	ND	2.15E+01
		α-glucosidase homolog-2 (α-glucosidase H2)	XP_030027477.1	Msex2.06949-RA		76-125 P-type trefoil domain; 303-770 GH31	9	D	ND	2.69E+02
	β-thioglucosidase	myrosinase-1	XP_030037280.1	Msex2.13243-RA	186E,398E	22-486 GH1 CD	9	A	2.92E+08	3.00E+02
		myrosinase-2	XP_030026263.1	Msex2.06063-RA	185E,396E	35-497 GH1 CD	9	A	8.61E+07	2.84E+02
		myrosinase-3	XP_030026260.1	Msex2.06067-RA	186E,398E	39-503 GH1 CD	9	A	ND	4.30E+00
		myrosinase-4	XP_030026264.1 C	Msex2.06061-RA	185E,397E	21-486 GH1 CD	9	C	1.43E+07	6.72E+01
	trehalase	trehalase	XP_030022193.1	Msex2.03542-RA		22-554 GH37 CD	10	C	1.53E+07	3.59E+01

	acetylhexoaminidase	α-N-acetyl galactosaminidase-1 (αNAGalase-1)	XP_030032307.1	Msex2.09968-RA	154D, 215D	15-308 GH27; 311-405 GH13, all β	11	A	4.47E+06	1.63E+00
		α-N-acetyl galactosaminidase-2 (αNAGalase-2)	XP_030028468.1	Msex2.07265-RB	163D, 224D	25-317 GH27/36; 318-419 GH13	11	C	ND	3.00E-01
	β-fructofuranosidase	sucrase-1	XP_030033238.1	Msex2.10409-RB	41D, 164D, 219E	22-337 GH32 N-term. D; 309-452 GH32 C-term. D	9	A	2.11E+06	5.03E+00
		sucrase-2	XP_030029844.1 N	Msex2.08181-RA	62D, 180D, 234E	52-349 GH32 N-term. D; 321-478 GH32 C-term. D	9	A	4.09E+07	1.33E+02
	α-L-fucosidase	α-L-fucosidase-1	XP_030040079.1	Msex2.00174-RA		34-368 GH29; 410-479 GH13 C-term. all β. GH95	11	C	1.73E+07	1.28E+00
	maltase	maltase-1	XP_030031300.1	Msex2.09215-RA	224D, 298E, 364D	31-515 GH13 CD; 493-596 GH13 C-term. all β, GH31, IPR 006047	9	A	1.61E+08	6.62E+02
		maltase-2	XP_030035812.1	Msex2.12170-RA	226D, 299E, 365D	48-585 GH13 CD; 565-658 GH13 C-term. all β GH31, IPR 006047	9	A	NaN	9.14E+01
		maltase 3	XP_030034371.1	Msex2.10850-RA	220D. 296E, 361D	21-326 GH13 CD; 304-426 GH13 C-term D, GH31, IPR006047	11	C	ND	0.00E+00
	glucosylceramidase	glucosylceramidase	XP_030029249.1	Msex2.07973-RA		1-90 GH30; 123-488 GH13 CD; 482-535 GH13 C-term. allβ	9	A	1.30E+08	1.72E+02
	α-mannosidase	α-mannosidase	XP_030037312.1	Msex2.12639-RA	170D, 292D	1-69 GH38/57 N-term. D; 72-161 GH38/57 central D	11	C	4.61E+07	1.46E+00
	β-glucuronidase	β-glucuronidase	XP_030027258.1	Msex2.06626-RA	336 - 631	47-239 galactose-binding D; 240-340 IgD; 341-639 GH2 CD	11	D	2.72E+06	4.13E+00
Nucleic acids	nucleases	nuclease-1	XP_030023583.1	Msex2.04563-RA		185-429 DNA/RNA non-specific endonuclease	9	A	1.25E+08	3.06E+02
		exonuclease 3'-5' domain protein	XP_030023602.1	Msex2.04535-RA		101-226 3'-5' exonuclease domain, 50-274 Rnase H superfamily	9	D	ND	1.65E+02
		ribonuclease Oy	XP_030024160.1	Msex2.04812-RA		45-235 Rnase Rh domain, T2 superfamily, 74W, 75I, 76V,77H, 78G, 79V,80W, 81P,126F, 127W, 128I,129H,130E, 131W, 132S,133K,134H, 135G, 136T, 137C	11	C	2.62E+06	2.16E+01
		ribonuclease H2B	XP_030027675.1	Msex2.07022-RA	heterotrimeric interface 29L, 30V, 31R, 44T, 68V, 76R, 77S, 78W, 79F, 80I	39-97 Rnh202 triple barrel domain, 97-224 Rnase H2 subunit B, wHTH domain	11	D	ND	2.16E+01

*: **Substrate group**: proteins, lipids, carbohydrates, or nucleic acids. **Enzyme group**: Ser, Cys and metallo proteases, lipases, serine esterases, nucleases, amylases, and others. **Exp. group**: G for preferential expression in mid*g*ut and "?" for unclear; **Group**: 9, 10 or 11; Digestive enzyme-relatedness index (**DERI**): A for digestive enzyme, B for digestion-related protein, C not specific for digestion, and D for unrelated to digestion in larvae. Evaluation was based on similar criteria described in Table S1 footnotes; **LFQ**: ND for undetected, NaN for "not a number", or a value for normalized protein level; **TPM**: transcripts per kilobase million in the RNA-Seq dataset of day 1, 5th instar midgut. **Abbreviations**: AL and ALH, acidic lipase and homolog; AP and APH, aminopeptidase and homolog; βNAGase, β-N-acetylglucosaminidase; CP and CPH, carboxypeptidase and homolog; DPP, dipeptidyl peptidase; FE, feruloyl esterase; GH, glycoside hydrolase; GP and GPH, gut serine protease and homolog; JHE and JHEH, juvenile hormone esterase and homolog; LLP, lysozyme-like protein; NL and NLH, neutral lipase and homolog; PGRP, peptidoglycan recognition protein; SE

and SEH, serine esterase and homolog; SP and SPH, serine protease and homolog; ZnCP and ZnCPH, zinc carboxypeptidase and homolog; ZnP, zinc protease.

Table S3. Features of the newly identified homologous proteins

Substrate group	Enzyme group	NCBI_ID	Msex/MCOT_ID	DER index	LFQ	TPM (L5d1)	Catalytic residues	Homolog in Table S2
Proteins	cycteine protease	XP_030031595.1	Msex2.15524-RE	D	NAN	45.57		caspase-4-1 C14A
	metalloprotease	XP_030029518.1	Msex2.07488-RA	D	ND	0.00		Zn protease-1 (ZnP1), M12A
		XP_030023350.1	MCOT.C03871.5.0.OOO5B	A	5.03E+08	4295.31		zinc carboxypeptidase-6 (ZnCP6) M14
Lipids	lipase	XP_030022443.1	Msex2.03907-RA	A	6.43E+07	65.23	184S,208D,269H	neutral lipase-4 (NL4)
		XP_030022463.1	MCOT.C03303.4.0.COO4B	B	2.29E+07	3327.12	142G,166D,227H	neutral lipase homolog-6 (NLH6)
	serine esterase	XP_030021117.1	Msex2.00349-RA	D	3.22E+06	89.48	183S,317E,441N	αE1
		XP_030029535.1	Msex2.07962-RB	D	ND	8.22	196S,328E,443H	αE3
		XP_030034367.1A	Msex2.11171-RA	A	1.61E+07	62.66	216S,346E,460H	esterase FE4-6
		XP_030034367.1B	Msex2.11174-RA	A	1.61E+07	43.52	201S,331E,445H	esterase FE4-4
		XP_030025623.1A	Msex2.05781-RA	D	ND	1.65	201S,331E,445H	esterase FE4-4
		XP_030025623.1B	Msex2.12087-RB	A	1.61E+07	3.07	221S,351E,465H	esterase FE4-4
		XP_030025623.1C	Msex2.12088-RB	A	1.61E+07	1.11	201S,331E,445H	esterase FE4-4
		XP_030027742.1	Msex2.07025-RA	A	9.46E+05	13.57	201S,329E,442H	esterase FE4-5
		XP_030027740.1	Msex2.07026-RA	C	NAN	56.94	218S,346E,459H	esterase FE4-5
		XP_030027699.1	Msex2.07027-RA	C	4.83E+07	62.00	202S,330E,443H	esterase FE4-5
		XP_030027704.1	Msex2.07028-RC	C	9.46E+05	8.30	204S,332E,445H	esterase FE4-5
Carbohydrates	α-amylase	XP_030040921.1	Msex2.00123-RA	C	1.14E+06	0.74	210D,247E,312D	α-amylase-2
	β-thioglucosidase	XP_030026258.1	Msex2.06062-RA	A	1.43E+07	60.83	185E,397E	myrosinase-4
		XP_030026255.1	Msex2.06064-RA	A	NAN	4.27	185E,397E	myrosinase-4
		XP_030026264.1N	MCOT.C05365.10.0.COB4B	A	1.43E+07	187.31	185E,397E	myrosinase-4
	β-fructofuranosidase	XP_030029844.1C	Msex2.08182-RA	C	ND	1.22	62D, 180D, 234E	sucrase-2
	α-L-fucosidase	XP_030031341.1	Msex2.00175-RB	C	4.61E+06	0.17		α-L-fucosidase-2
Nucleic acids	nuclease	XP_030023618.1	Msex2.04564-RA	C	ND	63.51		nuclease-2

Figures

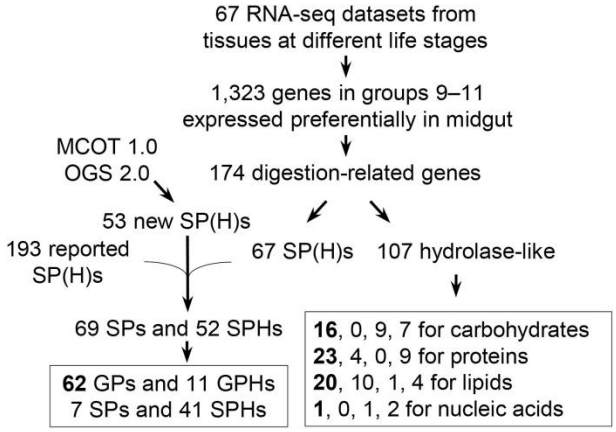


Fig. 1. Scheme for identification of the 122 digestive enzymes in *M. sexta*. As described in *Section 2.1*, the transcriptome analysis revealed three groups of genes favorably expressed in midgut at different developmental stages (Cao and Jiang, 2017). BLAST search indicated that 174 genes encode SPs, SPHs, and other hydrolase-related proteins. A comparison with SPs and SPHs in the MCOT and OGS 2.0 assemblies (Cao and Jiang, 2015) and the 125 nondigestive and 68 digestive SPs and SPHs (Cao et al., 2015) led to the identification of 69 SPs and 52 SPHs beyond the 125 nondigestive SP(H)s. The 121 SP(H)s were divided into GP(H)s and nonGP(H)s based on their expression profiles. The other 107 hydrolase-like proteins are classified into: 1) digestive enzymes, 2) digestion-related noncatalytic proteins, 3) not digestion-specific, and 4) digestion-unrelated. Their numbers are listed in columns 1−4, respectively, with those in *bold* font standing for the digestive enzymes in larvae.

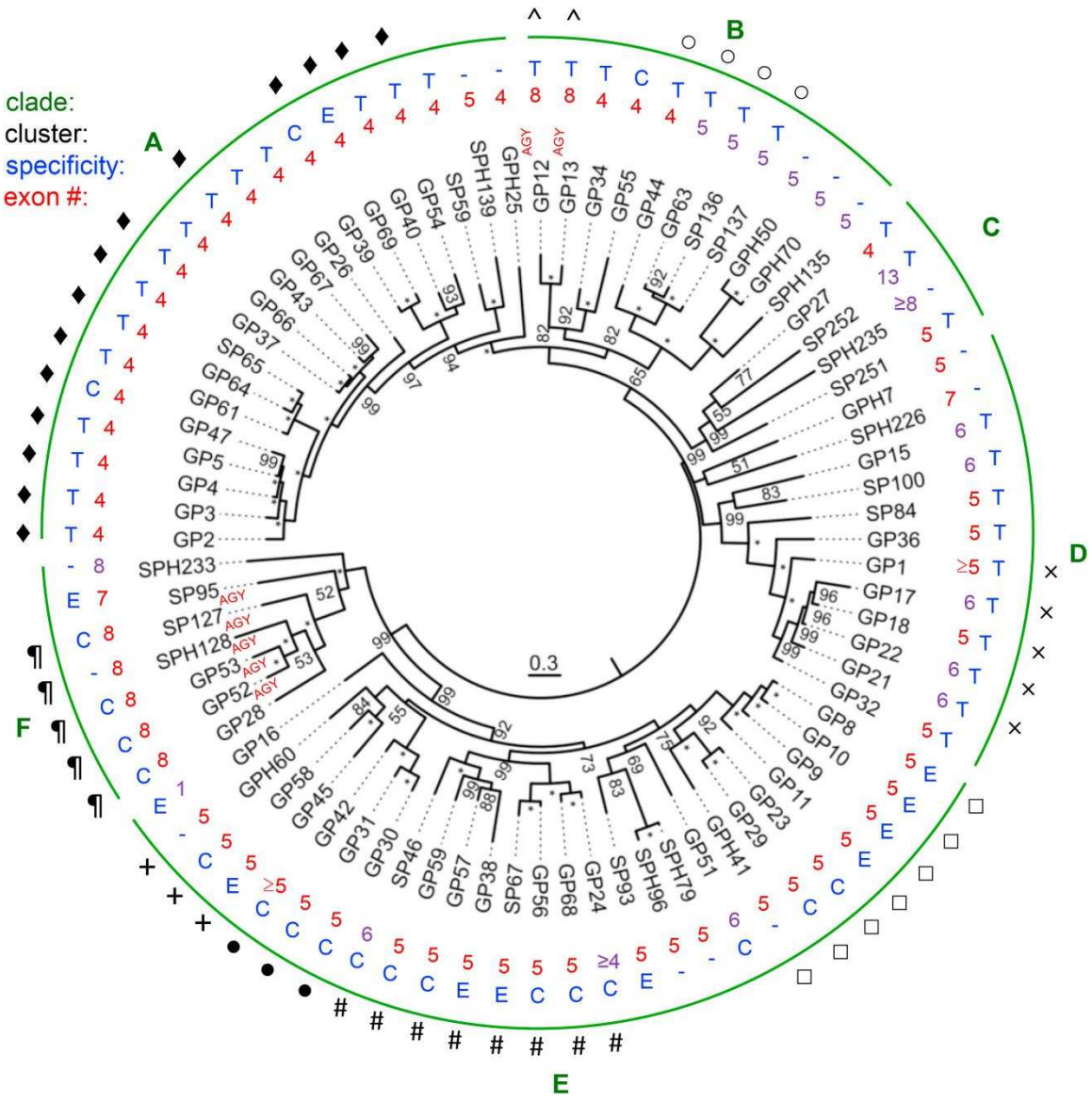
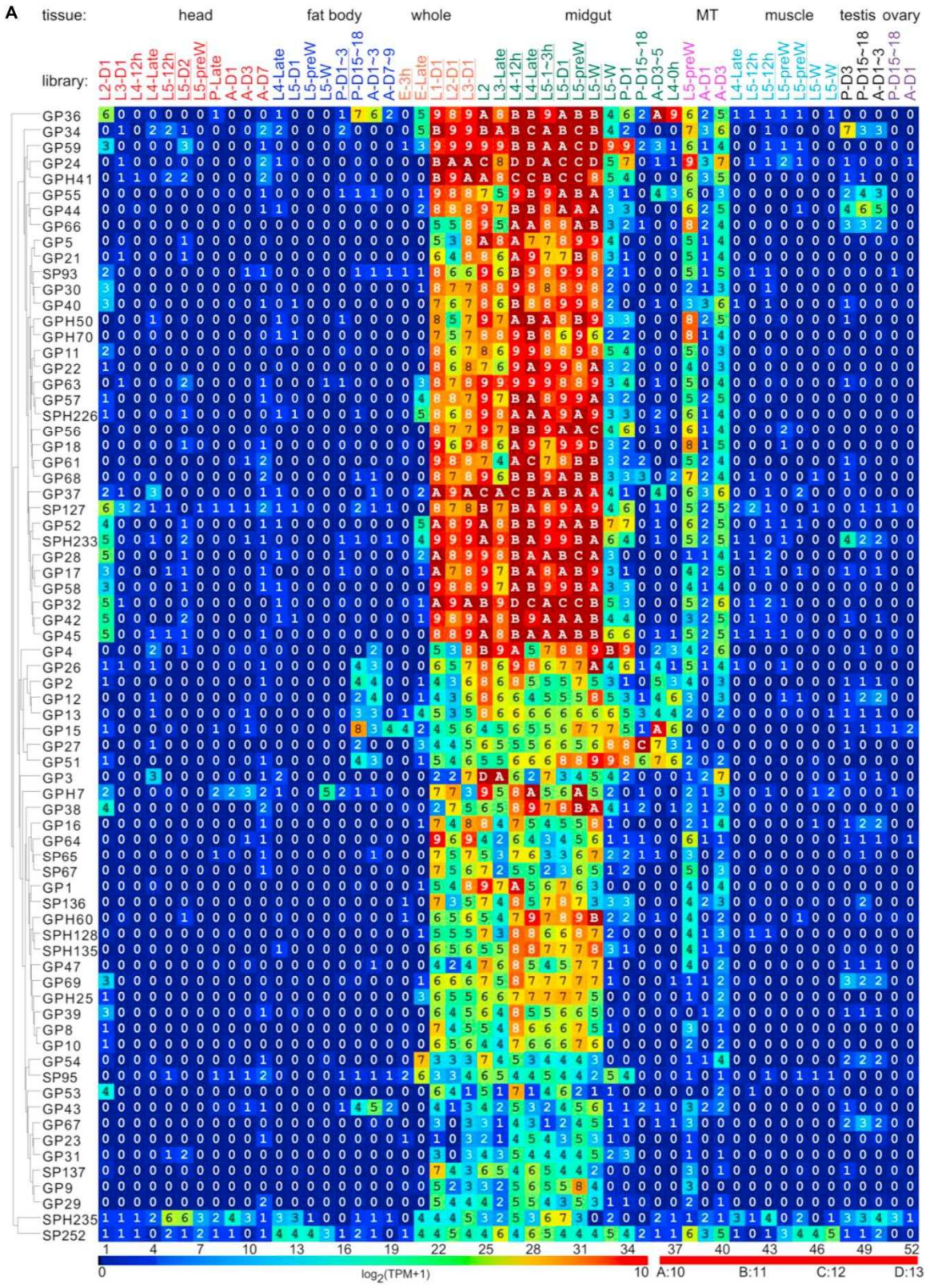


Fig. 2. Phylogenetic tree from MrBayes analysis of the 80 *M. sexta* SP(H) preferentially expressed in midgut. As described in *Section 2.4*, entire protein sequences of the SP(H)s were aligned for constructing the tree. Percentages of Bayesian posterior probabilities are shown at the nodes, with "*" representing 100. Branch lengths represent expected substitution rates per site. AGY (Y: C/T) codons for the active site Ser are indicated whereas TCN (N: A/C/G/T) codons for this residue are not shown. Numbers of exons for each gene are labeled in *red* font. Predicted enzyme specificity of SPs is indicated in *blue* font: "T" for trypsin, "C" for chymotrypsin, "E" for elastase, and "-" for SPHs that lack an amidase activity. Members of a gene cluster are marked with ♦, ^, ○, ×, □, #, ●, +, or ¶. Six clades (A–F) are indicated in *green* font.


A
tissue:
head
fat body
whole
midgut
MT
muscle
testis
ovary
library:
L2-D1
L3-D1
L4-12h
L4-Late
L5-12h
L5-D2
L5-preW
P-Late
A-D1
A-D3
A-D7
L4-Late
L5-D1
L5-preW
L5-W
P-D1~3
P-D15~18
A-D1~3
A-D7~9
E-3h
E-Late
L1-D1
L2-D1
L3-D1
L2
L3-Late
L4-12h
L4-Late
L5-1~3h
L5-D1
L5-preW
L5-W
L5-W
P-D1
P-D15~18
A-D3~5
L4-0h
L5-preW
A-D1
A-D3
L4-Late
L5-12h
L5-12h
L5-preW
L5-preW
L5-W
L5-W
P-D3
P-D15~18
A-D1~3
P-D15~18
A-D1
GP36
GP34
GP59
GP24
GPH41
GP55
GP44
GP66
GP5
GP21
SP93
GP30
GP40
GPH50
GPH70
GP11
GP22
GP63
GP57
SPH226
GP56
GP18
GP61
GP68
GP37
SP127
GP52
SPH233
GP28
GP17
GP58
GP32
GP42
GP45
GP4
GP26
GP2
GP12
GP13
GP15
GP27
GP51
GP3
GPH7
GP38
GP16
GP64
SP65
SP67
GP1
SP136
GPH60
SPH128
SPH135
GP47
GP69
GPH25
GP39
GP8
GP10
GP54
SP95
GP53
GP43
GP67
GP23
GP31
SP137
GP9
GP29
SPH235
SP252
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52
0
log2(TPM+1)
10
A:10
B:11
C:12
D:13

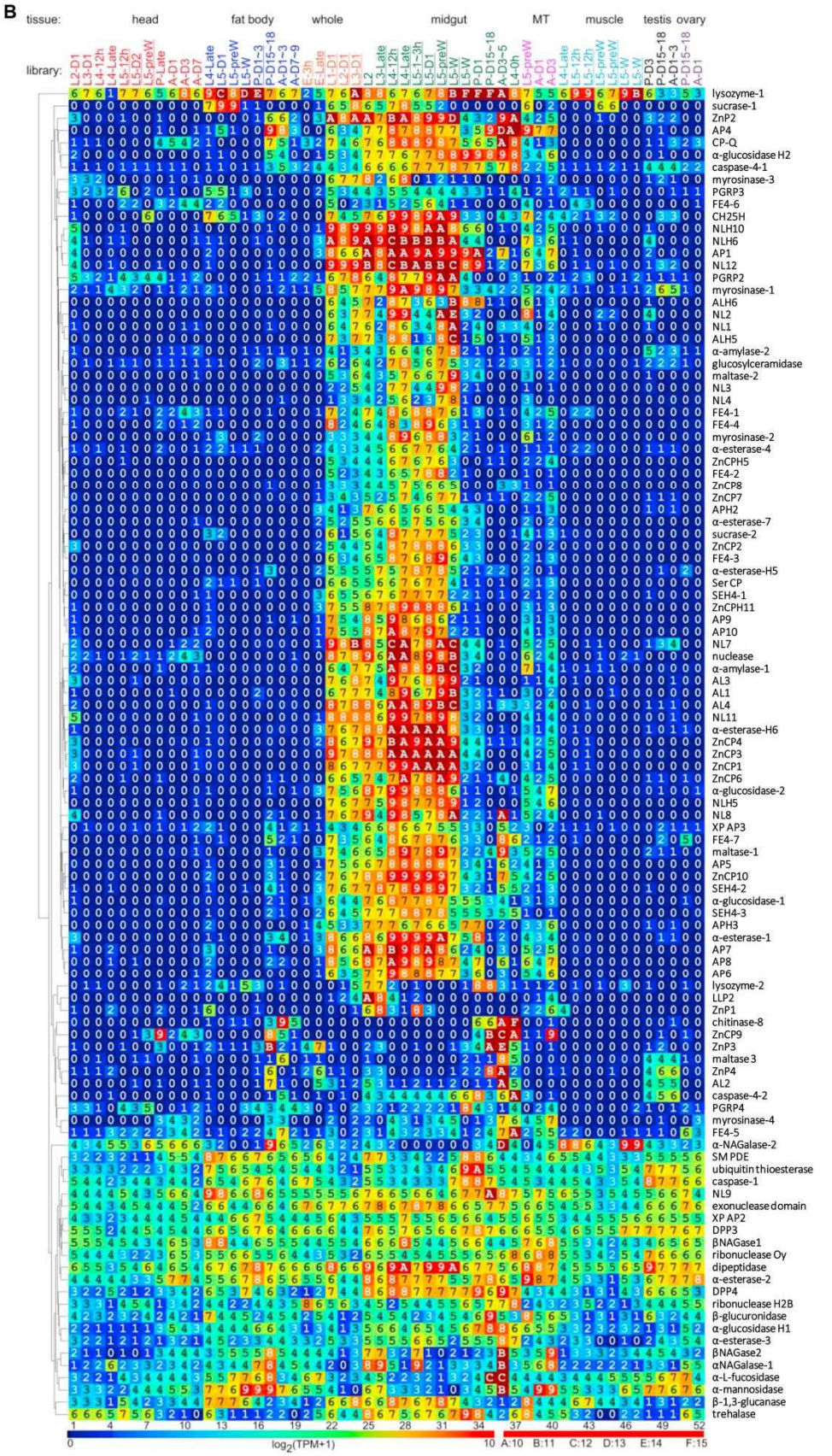

B
tissue:
head
fat body
whole
midgut
MT
muscle
testis
ovary
library:
L2-D1
L3-D1
L4-12h
L4-Late
L5-12h
L5-D2
L5-preW
P-Late
A-D1
A-D3
A-D7
L4-Late
L5-D1
L5-preW
L5-W
P-D1~3
P-D15~18
A-D1~3
A-D7~9
E-3h
E-Late
L1-D1
L2-D1
L3-D1
L2
L3-Late
L4-12h
L4-Late
L5-1~3h
L5-D1
L5-preW
L5-W
P-D1
P-D15~18
A-D3~5
L4-0h
L5-preW
A-D1
A-D3
L4-Late
L5-12h
L5-preW
L5-preW
L5-W
L5-W
P-D3
P-D15~18
A-D1~3
P-D15~18
A-D1
lysozyme-1
sucrase-1
ZnP2
AP4
CP-Q
α-glucosidase H2
caspase-4-1
myrosinase-3
PGRP3
FE4-6
CH25H
NLH10
NLH6
AP1
NL12
PGRP2
myrosinase-1
ALH6
NL2
NL1
ALH5
α-amylase-2
glucosylceramidase
maltase-2
NL3
NL4
FE4-1
FE4-4
myrosinase-2
α-esterase-4
ZnCPH5
FE4-2
ZnCP8
ZnCP7
APH2
α-esterase-7
sucrase-2
ZnCP2
FE4-3
α-esterase-H5
Ser CP
SEH4-1
ZnCPH11
AP9
AP10
NL7
nuclease
α-amylase-1
AL3
AL1
AL4
NL11
α-esterase-H6
ZnCP4
ZnCP3
ZnCP1
ZnCP6
α-glucosidase-2
NLH5
NL8
XP AP3
FE4-7
maltase-1
AP5
ZnCP10
SEH4-2
α-glucosidase-1
SEH4-3
APH3
α-esterase-1
AP7
AP8
AP6
lysozyme-2
LLP2
ZnP1
chitinase-8
ZnCP9
ZnP3
maltase 3
ZnP4
AL2
caspase-4-2
PGRP4
myrosinase-4
FE4-5
α-NAGalase-2
SM PDE
ubiquitin thioesterase
caspase-1
NL9
exonuclease domain
XP AP2
DPP3
βNAGase1
ribonuclease Oy
dipeptidase
α-esterase-2
DPP4
ribonuclease H2B
β-glucuronidase
α-glucosidase H1
α-esterase-3
βNAGase2
αNAGalase-1
α-L-fucosidase
α-mannosidase
β-1,3-glucanase
trehalase
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52
0
log2(TPM+1)
10 A:10 B:11 C:12 D:13 E:14 F:15

Fig. 3. Gene expression profiles of the 61 GPs and 11 GPHs (**A**) and 107 other hydrolases and homologs (**B**) in various tissues and stages of *M. sexta*. The first part of the library names indicates major stages of the insect, embryo (E), 1st to 5th instar larvae (L1−L5), pupae (P), and adults (A). In the second part, "D" stands for day, "h" for hour, "preW" for pre-wandering, and "W" for wandering. As shown on the *left*, order of the proteins represents their relatedness in expression pattern, as revealed by the cluster analysis. Log_2(TPM+1) values for the transcripts are shown in the gradient heat map from dark blue (0) to red (≥10). The values of 0~0.49, 0.50~1.49, 1.50~2.49, … 8.50~9.49, 9.50~10.49, …, 14.50~15.49 are labeled as 0, 1, 2, … 9, A, …, F, respectively. The other hydrolase-related genes are listed as abbreviations on the right of panel B (see Table S2 for full names).

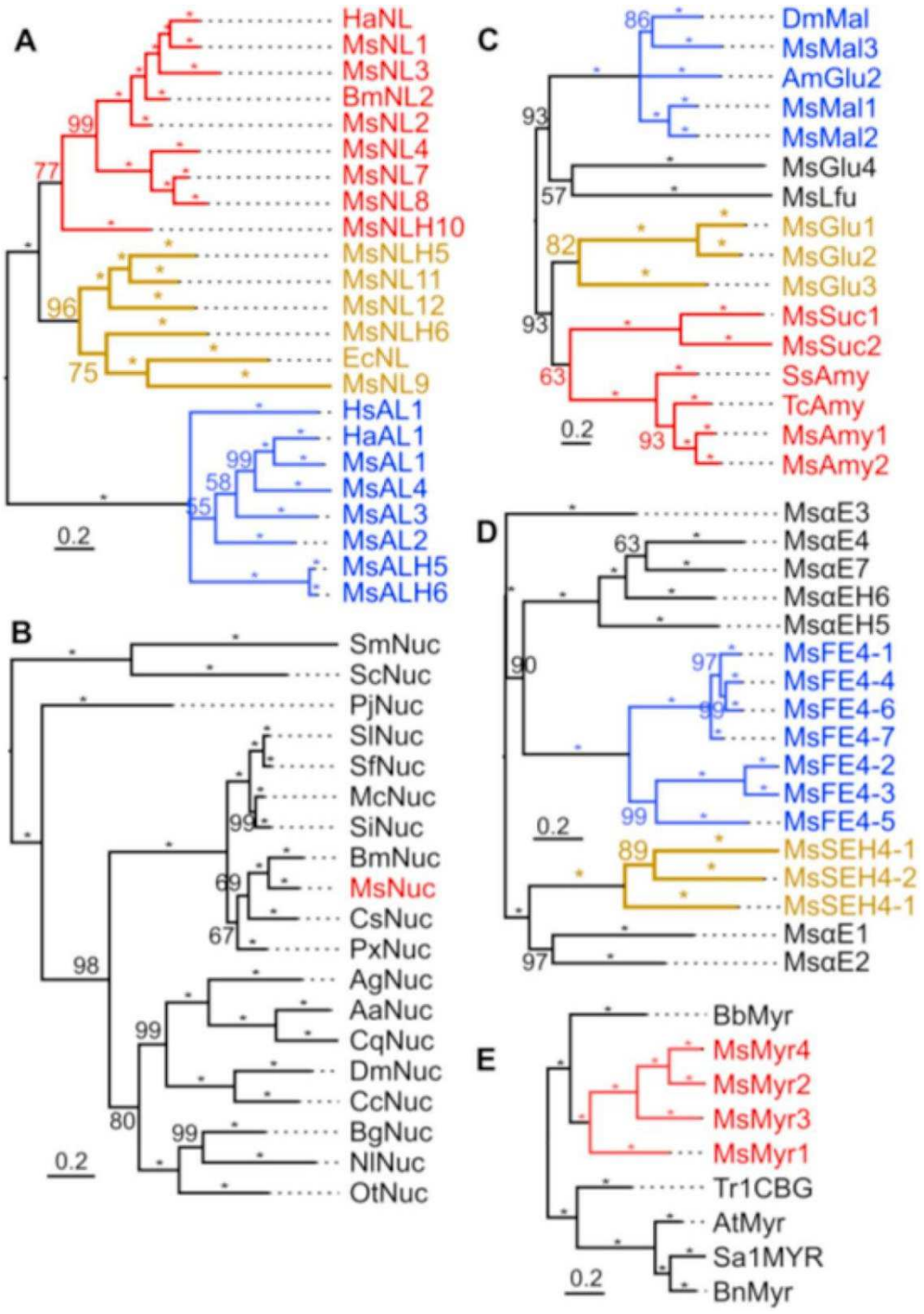


Fig. 4. Phylogenetic trees of lipase-related proteins (**A**), nucleases (**B**), carbohydratases (**C**), serine esterase-related proteins (**D**), and myrosinases (**E**) from *M. sexta* and other organisms. Based on the sequence alignments (Figs. S2–S5), phylogenetic trees were constructed using MrBayes v3.2.6. Probability values are indicated on top of branches; major taxonomic or protein groups are shown in different colors. Abbreviations are indicated in the corresponding supplemental figures.

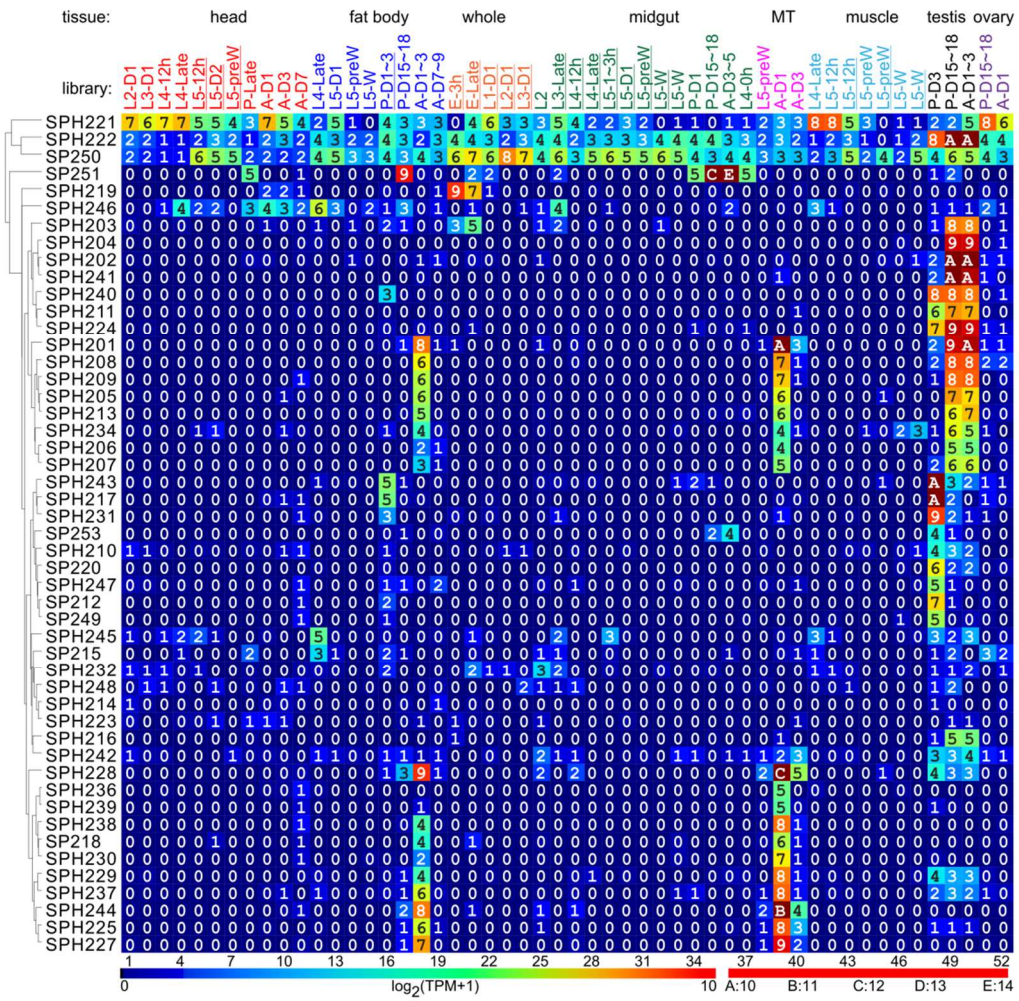


Fig. S1. Expression profiles of the 49 newly identified SP(H) genes unrelated to food digestion. Tissues, stages, and library IDs (1−52) are color coded as described in Fig. 3 legend.

A

B

Fig. S2. Aligned sequences of the neutral (**A**) and acidic (**B**) lipase-related proteins from *M. sexta* and other animals. *Equus caballus* neutral or pancreatic lipase (EcNL, NP_001157421.1), *H. armigera* NL (XP_021186848.1), *B. mori* NL2 (XP_004929630.1), *M. sexta* NL(H)s, *Homo sapiens* acidic or gastric lipase-1 (HsAL1, NP_001185758.1), *H. armigera* AL1

(XP_021181637.1), and *M. sexta* AL(H)s. *Green* font denotes residues of signal peptide; *red* font denotes conserved catalytic residues.

Fig. S3. Sequence alignment of the serine esterase-related proteins expressed in *M. sexta* midgut. Sequences of α-esterases (αEs), juvenile hormone esterases (JHEs), feruloyl esterases (FEs), and their homologs are aligned. *Green* font denotes signal peptides; *red* font denotes conserved catalytic residues.

Fig. S4. Aligned sequences of α-amylases, α-glucosidases, maltases, sucrases, and α-*L*-fucosidase from *M. sexta* and other animals. *Sus scrofa* pancreatic α-amylase (XP_003125929.2), *T.*

castaneum α-amylase (NP_001107848.1), *D, melanogaster* maltase A1 (NP_476627.3), *Apis mellifera* α-glucosidase-2 (NP_001035326.1) are aligned with their *M. sexta* homologs. *Green* font denotes signal peptides; *red* font denotes conserved catalytic residues.

```
SalMYR                              DEEITCQENNPFTCGNTDGLNSSSFEADFIFGVASSAYQIEGTI---GRGLNIWDGFTHRYPDKSGPDHGNGDTTCDSFSYWQKDIDVLDELNATGYRFS
BnMyrosinase   MKLLHGLALVFLLAAASCKADEEITCEENNPFTCSNTDILSSKNFGKDFIFGVASSAYQIEGGR---GRGVNVWDGFSHRYPEKAGSDLKNGDTTCESYTRWQKDVDVMGELNATGYRFS
AtMyrosinase   MKLLM-LAFVFLLALATCK-GDEFVCEENEPFTCNQTKLFNSGNFEKGFIFGVASSAYQVEGGR---GRGLNVWDSFTHRFPEKGGADLGNGDTTCDSYTLWQKDIDVMDELNSTGYRFS
Tr1CBG                                    FKPLPISFDDFSDLNRSCFAPGFVFGTASSAFQYEGAAFEDGKGPSIWDTFTHKYPEKIK-DRTNGDVAIDEYHRYKEDIGIMKDMNLDAYRFS
BbMyrosinase                                    MDYKFPKDFMFGTSTASYQIEGGWNEDGKGENIWDRLVHTSPEVIK-DGTNGDIACDSYHKYKEDVAIIKDLNLKFYRFS
MsMyrosinase1                          MASWAIIGLLAVCHASYITAYTKFPEGFTFGVATAAHQIEGAWNVSGKSENVWDRLSHTRPEMIA-DGTNGDIACDSYNRYREDVEELAHLGVDFYRFS
MsMyrosinase2                           MALARVVLFSLVGAVACFQDLDFPPGFQFGAASSSYQTEGAWNTSDKSESIWDRFVHEKPHLIG-DGSTGDIACDSYHLWRRDIEMVSELGLHFYRFS
MsMyrosinase3                          MLFFGLLVLSATASASSCPIDPTFPSCFKLGASTAAYQIEGAWNVSDKSVSIWDNFTHRVPRVTV-DNSTGDIACDSYNLWQRDVEMAAEMGLDYYRFS
MsMyrosinase4                           MISSNVVLLSLLAVTVTCENLDFPPGFQFGAASAAYQVEGAWNVSDKSENIWDRLVHEKPHLIS-DRSTGDVACDSYHLWRRDIEMAADLGLHFYRMS

SalMYR          IAWSRIIPRGKRSRGVNQKGIDYYHGLIDGLIKKGITPFVTLFHWDLPQTLQDEYEGFLDPQIIDDFKDYADLCFEEFGDSVKYWLTINQLYSVPTRGYGSALDAPGRCSPTVDP--SCY
BnMyrosinase    FAWSRIIPKGKVSRGVNQGGLDYYHKLIDALLEKNITPFVTLFHWDLPQTLQDEYEGFLDRQIIQDFKDYADLCFKEFGGKVKHWITINQLYTVPTRGYAIGTDAPGRCSPMVDTKHRCY
AtMyrosinase    IAWSRLLPKGKRSRGVNPGAIKYYNGLIDGLVAKNMTPFVTLFHWDLPQTLQDEYNGFLNKTIVDDFKDYADLCFELFGDRVKNWITINQLYTVPTRGYALGTDAPGRCSPKIDV--RCP
Tr1CBG          ISWPRVLPKGKLSGGVNREGINYYNNLINEVLANGMQPYVTLFHWDVPQALEDEYRGFLGRNIVDDFRDYAELCFKEFGDRVKHWITLNEPWGVSMNAYAYGTFAPGRCSDWLKLN--CT
BbMyrosinase    ISWARIAPSG-VMNSLEPKGIAYYNNLINELIKNDIIPLVTMYHWDLPQYLQDLG-GWVNPIMSDYFKEYARVLFTYFGDRVKWWITFNEPIAVCKG-YSIKAYAPN------------
MsMyrosinase1   LSWSRILPTG-FIDELNPDGVRYYNDLLDALAEHNIEPLVTLFHWDLPQSLQDLG-GWANPKMIDYFRDYADVCFKQFGDKIKSWITFNEPYEICEDAYGDEKKAPA------------
MsMyrosinase2   ISWTRLVPNG-FATHISEDGRTYYDNLINGLLESGVEPMVTLYHWELPQKLQDLG-GWTNPLIVDWFGDYARVVYSLFGDRVKTFLTINEPLVMCDLPYNTGTLAPG------------
MsMyrosinase3   ISWPRLLPKG-YDYYVSKDGAEYYNNLINGLLERGIEPVVTLYHWDLPDILQQLG-GWTNPLTADWFANYARIVFDLYGDRVKTWITINEPIVVCDAAYAAGYTAPG------------
MsMyrosinase4   IAWTRLLPNG-FANKISEDGLNYYNNLIDGLLEKGIQPMVTLYHWDLPQSLQDLG-GWTNPLIVDWFGDYARVAFSLFGDRVKTWITINEPLVICDVPYNTASMAPA------------

SalMYR          AGNSSTEPYIVAHHQLLAHAKVVDLYRKNYT-HQGGKIGPTMITRWFLPYNDTDRHSIAATERMKQFFLGWFMGPLTN--GTYPQIMIDTVG----------ARLPTFSPEETNLVKGSY
BnMyrosinase    GGNSSTEPYIVAHNQLLAHATVVDLYRTKYK-FQKGKIGPVMITRWFLPFDESDPASIEAAERMNQFFHGWYMEPLTK--GRYPDIMRQIVG----------SRLPNFTEEEAELVAGSY
AtMyrosinase    GGNSSTEPYIVAHNQLLAHAAAVDVYRTKYKDDQKGMIGPVMITRWFLPFDHSQ-ESKDATERAKIFFHGWFMGPLTE--GKYPDIMREYVG----------DRLPEFSETEAALVKGSY
Tr1CBG          GGDSGREPYLAAHYQLLAHAAAARLYKTKYQASQNGIIGITLVSHWFEPASKEK-ADVDAAKRGLDFMLGWFMHPLTK--GRYPESMRYLVR----------KRLPKFSTEESKELTGSF
BbMyrosinase    LNLKTTGHYLAGHTQLIAHGKAYRLYEEMFKPTQNGKISISISGVFFMPKNAESDDDIETAERANQFERGWFGHPVYK--GDYPPIMKKWVDQKSKEEGLPWSKLPKFTKDEIKLLKGTA
MsMyrosinase1   VDSHGVGNYLCSDTLLKAHAEAYHLYNETYRPVQNGKIMISINSIWYEPSDPENAEQVALAEVANQFKFGWFAHPIFTEEGGYPAVMVENIARQSASEGLNKPRLEQFDDYWIERIKGTS
MsMyrosinase2   IVD-VVATYMCNKNVLLAHAKAWRIYDEEFRPKYHGEVSITNQLLWFEAATPEDE---ELAEHAREWMAGRYSHAIFSAEGGWPPVVEKAMAEDSKKKGYPRSTLPAFTKEEIEFVRGTF
MsMyrosinase3   IKEPYSGVFVCTKNVLLGHAKAWRIYDKEFRHLYEGKVSISNHMIWFEPKTKADE---ALNELAMQYINGRYAHAIYSEEGGWPSSLEKYMAEYSAKQGYPRSRLPVFTEYEKTLVRGTF
MsMyrosinase4   ILDPDVGAHLCAKNIIMAHAKAWRIYDKEFRTKYHGELSLANHLIWYEPFTPEDE---EITNAAREYMVGRYSHPIYSKEGGWPPVIEKMMAENSKAKGYRKSNLPAFTKEEIEFVRGTF

SalMYR          DFLGLNYYFTQYAQPSPNPVNATNHTAMMDAGAKLTYINASGHYIGPLFESDGGDGSSNIYYYPKGIYSVMDYFKNKYYNPLIYVTENG---ISTPGSENRKESMLDYTRIDYLCSHLCF
BnMyrosinase    DFLGLNYYVTQYAQPKPNPYPSETHTAMMDAGVKLTYDNSRGEFLGPLFVEDKVN--GNSYYYPKGIYYVMDYFKTKYGDPLIYVTENG---FSTPSSENREQAIADYKRIDYLCSHLCF
AtMyrosinase    DFLGLNYYVTQYAQNNQTIVPSDVHTALMDSRTTLTSKNATGHAPGPPFNAA-------SYYYPKGIYYVMDYFKTTYGDPLIYVTENG---FSTPGDEDFEKATADYKRIDYLCSHLCF
Tr1CBG          DFLGLNYYSSYYAAKAPR-IPNARPAIQTDSLINATFEHN-GKPLGPMAASS------WLCIYPQGIRKLLLYVKNHYNNPVIYITENGRNEFNDP-TLSLQESLLDTPRIDYYYRHLYY
BbMyrosinase    DFYALNHYSSRLVT-FGSDP---NPNFNPDASYVTSVDEAW--LKPNETPYI--------IPVPEGLRKLLIWLKNEYGNPQLLITENG--------YGDDGQ-LDDFEKISYLKNYLNA
MsMyrosinase1   DFLGINHYTTHLITGAGVDPIAKSPSWLKDIGAVTTMNVG----GDSASEWL--------RVVPTGFANLLRWCKSSYNDPPIYITENG--------FSDRGT-LQDYGRIQYFNEYLSV
MsMyrosinase2   DFYAVNHYSSRLVRKAKEGE-ELHPYPLGDIPELAAKLEPHPHWRPTATAIF--------WLYPEGLHHQLVWLKKRYGDIRFVITENG--------VATPAG-RNDMDRIEYYTEYLKQ
MsMyrosinase3   DFIAVNYYTTKLIRPAREGE-VVSDLFVSGIPDLNAVMESPPTATYGYSTLM--------SIYPKGLRRILGWLKDTYGDHEMMLTENG--------FSSSGSELNDQERIDFIRNHLEQ
MsMyrosinase4   DFYAVNHYTSRLIRRAKEGE-EFGAYPLGDVPDVGAKYEINPDWKSTTSSWF--------FVYPEGLRNQILWLRKQYGDMKFIVTENG--------YARLKG-VDDTDRIEYYTEYLKQ

SalMYR          LNKVIKEKDVNVKGYLAWALGDNYEFNNGFTVRFGLSYINWNNVT-DRDLKKSGQWYQKFISP*
BnMyrosinase    LRKVIKEKGVNVRGYFAWALGDNYEFCKGFTVRFGLSYVNWEDLD-DRNLKESGKWYQRFINGTVKNAVKQDFLRSSLSSQSQ-KKRFADA*
AtMyrosinase    LSKVIKEKNVNVKGYFAWSLGDNYEFCNGFTVRFGLSYVDFANITGDRDLKASGKWFQKFINVTDEDSTNQDLLRSSVSSKNRDRKSLADA*
Tr1CBG          VLTAIG-DGVNVKGYFAWSLFDNMEWDSGYTVRFGLVFVDFKNNL-KRHPKLSAHWFKSFLKK*
BbMyrosinase    TLQAMYEDKCNVIGYTVWSLLDNFEWFYGYSIHFGLVKIDFNDPQRTRTKRESYTYFKNVVSTGKP*
MsMyrosinase1   ILNVIYDDGVRVLGYTAWTLMDNFEWRAGFSERFGFYYVDITHPDRPRTPKLSVDYYRQLIANRELPQDERFKDPAVRHKTCFKP*
MsMyrosinase2   VLLAIK-EGVNVTGYTAWCLMDNFEWSDGYKTKMGLYEVDFSSPERTRTQKASARFYKNLIQKHSLNFKVYADEL*
MsMyrosinase3   LLLAMKVDHVNVSGYTHWSLMDNFEWLDGYTSKFGLYEVNFTDPMRERTPRASANYYSKIIKQRKLDICC*
MsMyrosinase4   LLRAIKEDGVNITGYTAWTLMDNFEWMDGYISRFGLYEVDFSSPERTRTPKASAEFYKNLIRHNSLS---IYDEL*
```

Fig. S5. Multiple sequence alignment of myrosinases from *M. sexta* and other species. Myrosinases of *Sinapis alba* (P29092.2), *Arabidopsis thaliana* (P37702), *Trifolium repens* (P26205), *Brassica napus* (Q00326), and *Brevicoryne brassicae* (AAL25999.1) are aligned with their homologs in *M. sexta*. *Green* font denotes signal peptides; *red* font denotes conserved catalytic residues.

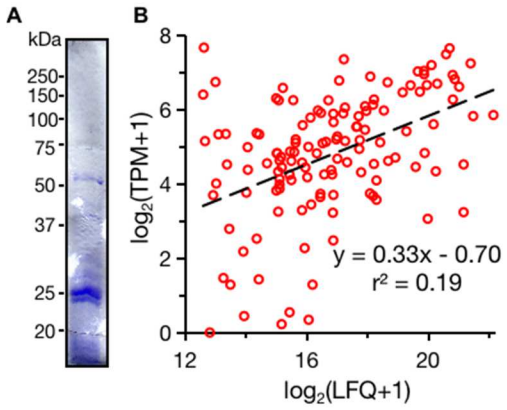


Fig. S6. SDS-PAGE and LC-MS/MS analysis of digestive fluid from feeding larvae of *M. sexta*. (**A**) Coomassie staining. Proteins (3 μg) in midgut juice from day 1, 5th instar larvae were treated with SDS-sample buffer, separated on a 10% polyacrylamide gel, and stained with Coomassie blue. Positions and sizes of the M_r makers are indicated on the *left*. (**B**) Correlation of mRNA and protein levels. Values of $\log_2$(TPM+1) and $\log_2$(LFQ+1), representing relative mRNA and protein levels, respectively, are plotted and subjected to a linear regression analysis. The equation and correlation coefficient (r^2) are shown.

Supplemental text

Information on the other 48 nondigestive SP(H)s

Among the 53 newly identified SP(H)s, five (SPH226, 233, 235, SP251and SP252) belong to group-<9, -9 and 11; SP212, 214, 215, 219, 220, 249, 250 and 253 have a trypsin-like specificity and do not belong to the midgut groups (Table S1). Neither do chymotrypsin-like SP210, elastase-like SP218, or the 38 SPHs. Fourteen of the 48 proteins (201−308 residues) are in a size range similar to the 75 GP(H)s (244−309 residues). Average size of the remaining 34 SP(H)s is 506 (316−1632 residues) and therefore most of them include other domain structures or low complexity regions absent in GP(H)s. SPH92 (Cao et al., 2015), SP219 and SPH223 each has one clip domain; SPH221 has two; SP214 has one fizzle domain, two LDLa repeats, and one trypsin-like SP domain. If a scavenger receptor domain were present next to the LDLa repeats, SP214 would have a domain organization identical to *D. melanogaster* Corin, *A. gambiae* SP214, *A. mellifera* SP30, and *T. castaneum* mSP15 (Cao and Jiang, 2018). Unlike gut-specific SPHs (*Section 3.2.1*), 25 of the 38 SPHs have two or more residues in the R*IVGG region mutated to residues with drastically different physiochemical properties, suggesting that functioning of these non-digestive SPHs does not require this motif in most cases.

The non-digestive SP-related genes display unique patterns of expression in tissues and stages (Fig. S1). Greatly differing from the GP(H)s, they are not favorably expressed in midgut during larval feeding stages. While SPH221, 222 and SP250 are expressed in all the samples, differential expression of the other genes in tissue(s) or life stage(s) is remarkable. SPH219 is highly expressed in embryos only; SPH201−211, 213, 217, 224, 231, 234, 240, 241, 243, 247, SP212, 220, 249, and 253 in pupal and adult testis; SPH201, 205−209, 213, 225, 227−230, 234, 236−239, 244, and SP218 in adult fat body and Malpighian tubules.

References

Arimatsu, Y., Furuno, T., Sugimura, Y., Togoh, M., Ishihara, R., Tokizane, M., Kotani, E., Hayashi, Y., Furusawa, T., 2007a. Purification and properties of double-stranded RNA-degrading nuclease, dsRNase, from the digestive juice of the silkworm, *Bombyx mori. J. Insect Biotechnol Sericol*. 76, 57−62.

Arimatsu, Y., Kotani, E., Sugimura, Y., Furusawa, T., 2007b. Molecular characterization of a cDNA encoding extracellular dsRNase and its expression in the silkworm, *Bombyx mori. Insect Biochem Mol Biol*. 37, 176−183.

Barrett, T., Suresh, C.G., Tolley, S.P., Dodson, E.J., Hughes, M.A., 1995. The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* 3, 951–960.

Bhat, R., Vyas, D., 2019. Myrosinase: insights on structural, catalytic, regulatory, and environmental interactions. *Crit Rev Biotechnol*. 39, 508–523.

Brioschi, D., Nadalini, L.D., Bengtson, M.H., Sogayar, M.C., Moura, D.S., Silva-Filho, M.C., 2007. General up regulation of *Spodoptera frugiperda* trypsins and chymotrypsins allows its adaptation to soybean proteinase inhibitor. *Insect Biochem Mol Biol*. 37, 1283−1290.

Canton, P.E., Cancino-Rodezno, A., Gill, S.S., Soberón, M., Bravo, A., 2015. Transcriptional cellular responses in midgut tissue of *Aedes aegypti* larvae following intoxication with Cry11Aa toxin from *Bacillus thuringiensis*. *BMC Genomics*. 16, 1042.

Cao, X., He, Y., Hu, Y., Zhang, X., Wang, Y., Zou, Z., Chen, Y., Blissard, G.W., Kanost, M.R., Jiang, H., 2015. Sequence conservation, phylogenetic relationships, and expression profiles of nondigestive serine proteases and serine protease homologs in *Manduca sexta*. *Insect Biochem Mol Biol.* 62, 51−63.

Cao, X., Jiang, H., 2015. Integrated modeling of protein-coding genes in the *Manduca sexta* genome using RNA-Seq data from the biochemical model insect. *Insect Biochem. Mol. Biol*. 62, 2–10.

Cao, X., Jiang, H., 2017. An analysis of 67 RNA-seq datasets from various tissues at different stages of a model insect, *Manduca sexta. BMC Genomics* 18, 796.

Cao, X., Jiang, H., 2018. Building a platform for predicting functions of serine protease-related

proteins in *Drosophila melanogaster* and other insects. *Insect Biochem Mol Biol*. 103, 53–69

Casas-Godoy, L., Gasteazoro, F., Duquesne, S., Bordes, F., Marty, A., Sandoval, G., 2018. Lipases: an overview. *Methods Mol Biol*. 1835, 3–38.

Cox, J., Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 26, 1367–1372.

Daimon, T., Taguchi, T., Meng, Y., Katsuma, S., Mita, K., Shimada, T., 2008. β-fructofuranosidase genes of the silkworm, *Bombyx mori*: insights into enzymatic adaptation of *B. mori* to toxic alkaloids in mulberry latex. *J Biol Chem*. 283, 15271–15279.

Dennison, N.J., Saraiva, R.G., Cirimotich, C.M., Mlambo, G., Mongodin, E.F., Dimopoulos, G., 2016. Functional genomic analyses of *Enterobacter*, *Anopheles* and *Plasmodium* reciprocal interactions that impact vector competence. *Malar J*. 15, 425.

Dunn, P.E., Drake, D.R., 1983. Fate of bacteria injected into naïve and immunized larvae of the tobacco hornworm *Manduca sexta*. *J Invertebr Pathol*. 41, 77–85.

Dutta, D., Dobson, A.J., Houtz, P.L., Gläßer, C., Revah, J., Korzelius, J., Patel, P.H., Edgar, B.A., Buchon, N., 2015. Regional cell-specific transcriptome mapping reveals regulatory complexity in the adult *Drosophila* midgut. *Cell Rep*. 12, 346−358.

Franco, O.L., Rigden, D.J., Melo, F.R., Grossi-De-Sá, M.F., 2002. Plant α-amylase inhibitors and their interaction with insect α-amylases. *Eur J Biochem*. 269, 397–412.

Guillemaud, T., Makate, N., Raymond, M., Hirst, B., Callaghan, A., 1997. Esterase gene amplification in *Culex pipiens*. *Insect Mol Biol*. 6, 319–328.

He, Y., Cao, X., Li, K., Hu, Y., Chen, Y., Blissard, G.W., Kanost, M.R., Jiang, H., 2015. A genome-wide analysis of antimicrobial effector genes and their transcription patterns in *Manduca sexta. Insect Biochem Mol Biol.* 62, 23−37.

He, Y., Cao, X., Zhang, S., Rogers, J., Hartson, S., Jiang, H., 2016. Changes in the plasma proteome of *Manduca sexta* larvae in relation to the transcriptome variations after an immune challenge: evidence for high molecular weight immune complex formation. *Mol Cell Proteomics* 15, 1176–1187.

Horne, I., Haritos, V. S., Oakeshott, J. G., 2009. Comparative and functional genomics of lipases in holometabolous insects. *Insect Biochem Mol Biol*. 39, 547–567.

Jones, A.M.E., Winge, P., Bones, A.M., Cole, R., Rossiter, J.T., 2002. Characterization and evolution of a myrosinase from the cabbage aphid *Brevicoryne brassicae*. *Insect Biochem Mol Biol.* 32, 275–284.

Jones, P., Binns, D., Chang, H-Y, Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S-Y., Lopez R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236−1240.

Kakumani, P.K., Malhotra, P., Mukherjee, S.K., Bhatnagar, R.K., 2014. A draft genome assembly of the army worm, *Spodoptera frugiperda*. *Genomics* 104, 134−143.

Kanost, M.R., Arrese, E.L., Cao, X., Chen, Y., Chellapilla, S., Goldsmith, M.R., Grosse-Wilde, E., Heckel, D.G., Herndon, N., Jiang, H., Papanicolaou, A., Qu, J., Soulages, J.L., Vogel, H., Walters, J., Waterhouse, R.M., Ahn, S., Almeida, F.C., An, C., Aqrawi, P., Bretschneider, A., Bryant, W.B., Bucks, S., Chao, H., Chevignon, G., Christen, J.M., Clarke, D.F., Dittmer, N.T., Ferguson, L.C.F., Garavelou, S., Gordon, K.H.J., Gunaratna, R.T., Han, Y., Hauser, F., He, Y., Heidel-Fischer, H., Hirsh, A., Hu, Y., Jiang, H., Kalra, D., Klinner, C., Konig, C., Kovar, C., Kroll, A.R., Kuwar, S.S., Lee, S.L., Lehman, R., Li, K., Li, Z., Liang, H., Lovelace, S., Lu, Z., Mansfield, J.H., McCulloch, K.J., Mathew, T., Morton, B., Muzny, D.M., McCulloch, K.J., Mathew, T., Morton, B., Muzny, D.M., Neunemann, D., Ongeri, F., Pauchet, Y., Pu, L., Pyrousis, I., Rao, X., Redding, A., Roesel, C., Sanchez-Gracia, A., Schaack, S., Shukla, A., Tetreau, G., Wang, Y., Xiong, G., Traut, W., Walsh, T.K., Worley, K.C., Wu, D., Wu, W., Wu, Y., Zhang, X., Zou, Z., Zucker, H., Briscoe, A.D., Burmester, T., Clem, R.J., Feyereisen, R., Grimmelikhuijzen, C.J.P., Hamodrakas, S.J., Hansson, B.S., Huguet, E., Jermiin, L.S., Lan, Q., Lehman, H.K., Lorenzen, M., Merzendorfer, H., Michalopoulos, I., Morton, D.B., Muthukrishnan, S., Oakeshott, J.G., Palmer, W., Park, Y., Passarelli, A.L., Rozas, J., Schwartz, L.M., Smith, W., Southgate, A., Vilcinskas, A., Vogt, R., Wang, P., Werren, J., Yu, X., Zhou, J., Brown, S.J., Scherer, S.E., Richards, S., Blissard, G.W., 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem Mol Biol*. 76, 118−147.

Khajuria, C., Zhu, Y.C., Chen, M.S., Buschman, L.L., Higgins, R.A., Yao, J., Crespo, A.L.,

Siegfried, B.D., Muthukrishnan, S., Zhu, K.Y., 2009. Expressed sequence tags from larval gut of the European corn borer (*Ostrinia nubilalis*): exploring candidate genes potentially involved in *Bacillus thuringiensis* toxicity and resistance. *BMC Genomics* 10, 286.

Kuwar, S.S., Pauchet, Y., Vogel, H., Heckel, D.G., 2015. Adaptive regulation of digestive serine proteases in the larval midgut of *Helicoverpa armigera* in response to a plant protease inhibitor. *Insect Biochem Mol Biol*. 59, 18−29.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf*. 12, 323.

Lin, J., Xia, X., Yu, X.Q., Shen, J., Li, Y., Lin, H., Tang, S., Vasseur, L., You, M., 2018. Gene expression profiling provides insights into the immune mechanism of *Plutella xylostella* midgut to microbial infection. *Gene* 647, 21−30.

Lin, W., Oliver, D.J., 2008. Role of triacylglycerols in leaves, *Plant Sci*. 175, 233−237.

Mellroth, P., Karlsson, J., Steiner, H., 2003. A scavenger function for a *Drosophila* peptidoglycan recognition protein. *J Biol Chem*. 278, 7059–7064.

Morris, K., Lorenzen, M.D., Hiromasa, Y., Tomich, J.M., Oppert, C., Elpidina, E.N., Vinokurov, K., Jurat-Fuentes, J.L., Fabrick, J., Oppert, B., 2009. *Tribolium castaneum* larval gut transcriptome and proteome: A resource for the study of the coleopteran gut. *J Proteome Res*. 8, 3889−3898.

Napoleão, T.H., Albuquerque, L.P., Santos, N.D., Nova, I.C., Lima, T.A., Paiva, P.M., Pontual, E.V., 2019. Insect midgut structures and molecules as targets of plant-derived protease inhibitors and lectins. *Pest Manag Sci*. 75, 1212−1222.

Oppert, B., Dowd, S.E., Bouffard, P., Li, L., Conesa, A., Lorenzen, M.D., Toutges, M., Marshall, J., Huestis, D.L., Fabrick, J., Oppert, C., Jurat-Fuentes, J.L., 2012. Transcriptome profiling of the intoxication response of *Tenebrio molitor* larvae to *Bacillus thuringiensis* Cry3Aa protoxin. *PLoS One* 7, e34624.

Oppert, B., Perkin, L., Martynov, A.G., Elpidina, E.N., 2018. Cross-species comparison of the gut: Differential gene expression sheds light on biological differences in closely related tenebrionids. *J Insect Physiol*. 106, 114−124.

Pauchet, Y., Wilkinson, P., Vogel, H., Nelson, D.R., Reynolds, S.E., Heckel, D.G., ffrench-

Constant, R.H. Pyrosequencing the *Manduca sexta* larval midgut transcriptome: messages for digestion, detoxification and defense. *Insect Mol Biol*. 19, 61−75.

Palli, S.R., Locke, M., 1987. The synthesis of hemolymph proteins by the larval midgut of an insect *Calpodes ethlius* (Lepidoptera:Hesperiidae). *Insect Biochem.* 17, 561−572

Pearce, S.L., Clarke, D.F., East, P.D., Elfekih, S., Gordon, K.H.J., Jermiin, L.S., McGaughran, A., Oakeshott, J.G., Papanicolaou, A., Perera, O.P., Rane, R.V., Richards, S., Tay, W.T., Walsh, T.K., Anderson, A., Anderson, C.J., Asgari, S., Board, P.G., Bretschneider, A., Campbell, P.M., Chertemps, T., Christeller, J.T., Coppin, C.W., Downes, S.J., Duan, G., Farnsworth, C.A., Good, R.T., Han, L.B., Han, Y.C., Hatje, K., Horne, I., Huang, Y.P., Hughes, D.S.T., Jacquin-Joly, E., James, W., Jhangiani, S., Kollmar, M., Kuwar, S.S., Li, S., Liu, N.Y., Maibeche, M.T., Miller, J.R., Montagne, N., Perry, T., Qu, J., Song, S.V., Sutton, G.G., Vogel, H., Walenz, B.P., Xu, W., Zhang, H.J., Zou, Z., Batterham, P., Edwards, O.R, Feyereisen, R., Gibbs, R.A., Heckel, D.G., McGrath, A., Robin, C, Scherer, S.E., Worley, K.C., Wu, Y.D., 2017. Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. *BMC Biol*. 15, x63.

Perona, J.J., Craik, C.S., 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci*. 4, 337–360.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions *Nat Methods* 8, 785−786.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61, 539–542.

Roughan, P.G., Batt, R.D., 1969. The glycerolipid composition of leaves. *Phytochem*. 8, 363−369.

Saraiva, R.G., Kang, S., Simões, M.L., Angleró-Rodríguez, Y.I., Dimopoulos, G., 2016. Mosquito gut antiparasitic and antiviral immunity. *Dev Comp Immunol*. 64, 53−64.

Schernthaner, J.P., Milne, R.E., Kaplan, H., 2002. Characterization of a novel insect digestive DNase with a highly alkaline pH optimum. *Insect Biochem Mol Biol*. 32, 255–263.

Schoville, S.D., Chen, Y.H., Andersson, M.N., Benoit, J.B., Bhandari, A., Bowsher, J.H., Brevik,

K., Cappelle, K., Chen, M.M., Childers, A.K., Childers, C., Christiaens, O., Clements, J., Didion, E.M., Elpidina, E.N., Engsontia, P., Friedrich, M., García-Robles, I., Gibbs, R.A., Goswami, C., Grapputo, A., Gruden, K., Grynberg, M., Henrissat, B., Jennings, E.C., Jones, J.W., Kalsi, M., Khan, S.A., Kumar, A., Li, F., Lombard, V., Ma, X., Martynov, A., Miller, N.J., Mitchell, R.F., Munoz-Torres, M., Muszewska, A., Oppert, B., Palli, S.R., Panfilio, K.A., Pauchet, Y., Perkin, L.C., Petek, M., Poelchau, M.F., Record, É., Rinehart, J.P., Robertson, H.M., Rosendale, A.J., Ruiz-Arroyo, V.M., Smagghe, G., Szendrei, Z., Thomas, G.W.C., Torson, A.S., Vargas Jentzsch, I.M., Weirauch, M.T., Yates, A.D., Yocum, G.D., Yoon, J.S., Richards, S., 2018. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Sci Rep*. 8, 1931.

Shibata, T., Kobayashi, Y., Ikeda, Y., Kawabata, S.I., 2018. Intermolecular autocatalytic activation of serine protease zymogen factor C through an active transition state responding to lipopolysaccharide. *J Biol Chem*. 293, 11589−11599.

Spit, J., Badisco, L., Vergauwen, L., Knapen, D., Vanden Broeck, J., 2016. Microarray-based annotation of the gut transcriptome of the migratory locust, *Locusta migratoria*. *Insect Mol Biol*. 25, 745−756.

Srp, J., Nussbaumerová, M., Horn, M., Mareš, M., 2016. Digestive proteolysis in the Colorado potato beetle, *Leptinotarsa decemlineata*: activity-based profiling and imaging of a multipeptidase network. *Insect Biochem Mol Biol*. 78, 1−11.

Terra, W.R., Ferreira, C., 2012 Biochemistry and molecular biology of digestion. In: Gilbert, L.I. (Ed.), "Insect Molecular Biology and Biochemistry". Academic Press, London, pp. 365−418.

Wang, L.L., Huang, Y., Lu, X.P., Jiang, X.Z., Smagghe, G., Feng, Z.J., Yuan, G.R., Wei, D., Wang, J.J., 2015. Overexpression of two α-esterase genes mediates metabolic resistance to malathion in the oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Insect Mol Biol*. 24, 467–479.

Winde, I., Wittstock, U., 2011. Insect herbivore counteradaptations to the plant glucosinolate-myrosinase system. *Phytochem*. 72, 1566−1575.

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M., Zhan, D., Baxter, S.W., Vasseur, L., Gurr, G.M., Douglas, C.J., Bai, J., Wang, P., Cui, K., Huang, S., Li, X., Zhou, Q., Wu, Z., Chen, Q., Liu, C., Wang, B., Li, X., Xu, X., Lu, C., Hu, M., Davey, J.W., Smith, S.M., Chen, M., Xia, X., Tang, W., Ke, F., Zheng, D., Hu, Y., Song, F., You, Y., Ma, X., Peng, L., Zheng, Y., Liang,

Y., Chen, Y., Yu, L., Zhang, Y., Liu, Y., Li, G., Fang, L., Li, J., Zhou, X., Luo, Y., Gou, C., Wang, J., Wang, J., Yang, H., Wang, J., 2013. A heterozygous moth genome provides insights into herbivory and detoxification. *Nat Genet*. 45, 220−225.

Zhang, S., Shukle, R., Mittapalli, O., Zhu, Y.C., Reese, J.C., Wang, H., Hua, B.Z., Chen, M.S., 2010. The gut transcriptome of a gall midge, *Mayetiola destructor*. *J Insect Physiol*. 56, 1198−1206.

Zhang, X., He, Y., Cao, X., Gunaratna, R.T., Chen, Y.R., Blissard, G., Kanost, M.R., Jiang, H., 2015. Phylogenetic analysis and expression profiling of the pattern recognition receptors: insights into molecular recognition of invading pathogens in *Manduca sexta*. *Insect Biochem Mol Biol*. 62, 38–50.

Zhu-Salzman K, Zeng R., 2015. Insect response to plant defensive protease inhibitors. *Ann Rev Entomol*. 60, 233−252.

CHAPTER IV

GENOME-WIDE IDENTIFICATION, CLASSIFICATION, AND EXPRESSION PROFILING OF SERINE ESTERASES AND OTHER ESTERASE-RELATED PROTEINS IN THE TOBACCO HORNWORM, *MANDUCA SEXTA*

Zelong Miao [a], Chao Xiong [a], Xiaolong Cao [a], Tisheng Shan [a], Qiao Jin [a], Haobo Jiang [a, ‡]

[a] Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA

Key words: phylogenetic analysis; esterase; carboxylesterase; lipase; phospholipase; thioesterase; RNA-Seq

Abbreviations: SE and SEH, serine esterase and homolog; AL and ALH, acidic lipase and homolog; NL and NLH, neutral lipase and homolog; HSL, hormone sensitive lipase; COE, carboxylesterase; AChE, acetylcholinesterase; AE and AEH, α-esterase and homolog; BE, β-esterase; FE and FEH, feruloyl esterase and homolog; IE and IEH, integument esterase and homolog; JHE, juvenile hormone esterase; GLI, gliotactin; GLT, glutactin; NLG, neuroligin; NRT, neurotactin; UCEH, uncharacteristic esterase homolog; PLA, PLB, PLC and PLD, phospholipase A−D; PLP, phospholipase-related protein; ATGL, adipose triglyceride lipase; TE, thioesterase; AKH, adipokinetic hormone; TAG or TG, triacylglycerol or triglyceride; TGL, triglyceride lipase; TPM, transcripts per kilobase million.

Abstract

Serine esterases (SEs) are hydrolases that catalyze the conversion of carboxylic esters into acids and alcohols. Lipases and carboxylesterases constitute two major groups of SEs. Although over a hundred of insect genomes are known, systematic identification and classification of SEs are rarely performed, likely due to large size and complex composition of the gene family in each species. Considering their key roles in lipid metabolism and other physiological processes, we have categorized 144 *M. sexta* SEs and SE homologs (SEHs), 114 of which contain a motif of GXSXG. Multiple sequence alignment and phylogenetic tree analysis have revealed 39 neutral lipases (NLs), 3 neutral lipase homologs (NLHs), 11 acidic lipases (ALs), 3 acidic lipase homologs (ALHs), a lipase-3, a triglyceride lipase, a monoglyceride lipase, a hormone-sensitive lipase, and a GDSL lipase. Eighty-three carboxylesterase genes encode 29 α-esterases (AEs), 12 AEHs (*e.g.*, SEH4-1–3), 20 feruloyl esterases (FEs), 2 FEHs, 2 β-esterases (BEs), 2 integument esterases (IEs), 1 IEH, 4 juvenile hormone esterases, 2 acetylcholinesterases, gliotactin, 6 neuroligins, neurotactin, and an uncharacteristic esterase homolog. In addition to these GXSXG proteins, we have identified 26 phospholipases and 13 thioesterases. Expression profiling of these genes in specific tissues and stages has provided insights into their functions including digestion, detoxification, hormone processing, neurotransmission, reproduction, and developmental regulation. In summary, we have established a framework of information on SEs and related proteins in *M. sexta* to stimulate their research in the model species and comparative investigations in agricultural pests or disease vectors.

1. Introduction

Carboxylic esterases (EC 3.1.1.x) are the enzymes that hydrolyze carboxylic ester bonds and produce organic acids and alcohols in the presence of water. They are also

called serine esterases (SEs), since a Ser residue at the active site plays an essential role in catalysis (Punta et al., 2012). Widely distributed in animals, plants, fungi and bacteria (Bornscheuer, 2002), SEs are composed of two major subgroups: carboxylesterases (COEs, EC 3.1.1.1) and triacylglycerol lipases (EC 3.1.1.3). Their sequences contain a catalytic triad (Ser-Asp/Glu-His) and a conserved sequence of GXSXG, where X represents any residue and S is the active site Ser (Wei et al., 1995). These hydrolases adopt a characteristic α/β fold, with eight β strands connected by α-helices and the catalytic residues located on loops (Ollis et al., 1992). Their mechanism of hydrolysis involves a nucleophilic attack of the ester's carbonyl carbon by the Ser residue to produce an alcohol and an acyl-enzyme complex. Then, a water molecule comes in, substitutes the acyl group with H^+ to release the acid, and regenerate the Ser for another round of catalysis.

Serine esterases mediate key physiological processes in insects and other living organisms (Oakeshott et al., 2010; Montella et al., 2012). Some COEs are named after their substrates to denote functions. For instance, juvenile hormone esterase (JHE) catalyzes JH hydrolysis to control metamorphosis. Acetylcholinesterase (AChE) breaks down acetylcholine, an excitatory neurotransmitter, to terminate cholinergic neurotransmission. Organophosphates and carbamates are common pesticides that inhibit AChEs (Casida and Durkin, 2013). Other COEs are involved in the general processes of digestion, reproduction, and pheromone degradation. Their noncatalytic homologs, including neuroligin, gliotactin, glutactin, and neurotactin, do not hydrolyze ester substrates but actively participate in recognition and neurodevelopment. In contrast to COEs that prefer "simple" esters with short-chain fatty acids, lipases favor water-

insoluble triglycerides composed of long-chain fatty acids (Casas-Godoy et al., 2018). In fact, lipases undergo interfacial activation, which requires a minimum concentration of substrates to form an apolar-aqueous phase, to shift to a fully active conformation (Rahman et al., 2012), whereas COEs follow Michaelis-Menden kinetics (Bornscheuer, 2002). The conformation change of a lipase involves the opening of a hydrophobic region or lid that covers the active site. Digestive lipases are a major research focus. Based on optimal reaction pH and site of expression, lipases in mammals are divided into gastric and pancreatic groups whereas insect lipases are separated into acidic and neutral lipases (ALs and NLs). In *Drosophila*, three genes (lip1–3) encode lipases with a GHSQG motif (Pistillo et al., 1998). Lipase-1 is digestive; lipase-2's role is unclear; lipase-3 is similar to lysosomal AL and may hydrolyze cholesteryl esters and triglycerides in lipoprotein particles via receptor-mediated endocytosis. *D. melanogaster* CG3635 encodes a protein similar to mammalian gastric lipases, whose mRNA level is higher in DDT resistant strains than the susceptible one (Qiu et al., 2013). This suggests that the resistance is also related to lipase-mediated lipid metabolism and mobilization. When demand for energy increases in human, a hormone-sensitive lipase (HSL) is regulated through reversible phosphorylation of Ser^{563} by catecholamines and insulin (Holm et al., 1988). The modified HSL mobilizes triglycerides via lipolysis in adipose and steroidogenic tissues and in heart and skeletal muscles. In *Bombyx mori*, there are 38 lipases containing a GXSLG motif, twelve of which are likely digestive (Shen et al., 2022). Some other lipases contain a consensus sequence of GDSL, and structural analysis of a bacterial GDSL lipase has revealed a catalytic diad (Ser-His) in the α/β-tertiary fold, rather than the common catalytic triad (Ser-Asp-His) in the α/β-hydrolase fold (Wei et al., 1995).

Sequence alignment and phylogenetic analysis has been utilized in the classification of insect serine esterases especially COEs, which lack a function-based naming system. COEs are better known to be associated with insecticide resistance and, therefore, have been studied more extensively than lipases. In *B. mori*, 76 COEs are divided into nine groups: α-esterases (AEs, 55), β-esterases (BEs, 2), JHEs (4), AChEs (2), integument esterases (IEs, 2), gliotactin (GLI), neuroligin (NLGs, 6), neurotactins (NRTs, 2), and uncharacteristic (2) (Yu et al., 2009). Their expression patterns are separated into midgut-, head and integument-, and silk gland-specific groups. *D. melanogaster*, *Apis mellifera* and *Anopheles gambiae* have 35, 24 and 51 COEs, including 13, 8 and 16 α-esterases, 3, 3 and 5 β-esterases, 2, 1 and 9 JHEs, 1, 2 and 2 AChEs, 3, 1 and 0 IEs, 4, 0 and 9 glutactins, 1, 1 and 1 gliotactin, 4, 5 and 5 neuroligins, 2, 1 and 2 neurotactin, as well as 2, 2 and 2 uncharacteristic, respectively (Oakeshott et al., 2010; Yu et al., 2009; Claudianos et al., 2006).

Manduca sexta is a biochemical model representing pest insects in the order of Lepidoptera and beyond. Although its genome sequence is known, a systematic analysis of the SEs and SEHs has not yet been performed to stimulate functional studies (Kanost et al., 2016). Some of the family members were transcriptionally regulated after an immune challenge (Zhang et al., 2011). In a study of digestion-related proteins, we identified 11 lipases, 9 COEs, and 15 non-catalytic homologs in *M. sexta* (Miao et al., 2020). A neuropeptide adipokinetic hormone (AKH) induces a fat body lipase activity through Ca^{2+} and cAMP to mobilize fatty acids from triglycerides in adults (Arrese et al., 1996 and 1999). A triglyceride lipase (TGL) was purified and cloned from *M. sexta* fat body, which is involved in mobilization of lipids (Arrese et al., 2010). The lipase contains a WWE domain, which may be responsible for hormone sensitivity. To acquire an overview of esterase-related proteins in *M. sexta*, we searched their genes in the genome, found

additional COEs, lipases and their non-catalytic homologs, and examined their expression profiles. In addition, we explored phospholipase-like proteins (PLPs) and thioesterases (TEs), which do not belong to SEs. The phylogenetic relationships and expression patterns revealed in this work have provided insights into their evolution and functions, which may assist pest control and insecticide resistance management in the future.

2. Methods

2.1. Identification of M. sexta esterase-related proteins

Domain structures of the protein sequences in MCOT 1.0 (Cao and Jiang, 2015) and *Manduca* OGS2.0 (Kanost et al., 2016) were predicted using InterProScan 5 (Jones et al., 2014). Proteins containing a lipase/vitellogenin, WWE, GDSL lipase, α/β hydrolase lipase, carboxylesterase, pectin acetylesterase, A2, B, C, D, or patatin-like phospholipase domain were identified and retrieved for manual improvement by crosschecking Oases 3.0, Trinity 3.0, Cufflinks 1.0, *Manduca* genome contigs, and JHU_Msex_v1.0 (Cao and Jiang, 2015; Kanost et al., 2016; Gershman et al., 2021). Signal peptide was predicted using SignalP 4.1 (Petersen et al., 2011). Catalytic residues in the NL and COE sequences were identified by BLASTP search of NCBI conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). For other esterase-like proteins, sequence features were revealed by multiple sequence alignments with homologous enzymes whose catalytic mechanisms are known. All the sequences we got were submitted to NCBI, the GenBank accession numbers of which from ON929093 to ON929276 were included in Table S1−4.

2.2. Phylogenetic analysis of M. sexta esterases and their non-catalytic homologs

A multiple sequence alignment of the entire 144 SE(H)s was performed using MUSCLE, one module of MEGA 11 (http://www.megasoftware.net) under the default settings with maximum iterations changed to 1000. A phylogenetic tree was constructed using the neighbor-joining (NJ)

method in which distance was estimated by Poisson model implemented in MEGA 11. The pairwise deletion option was used during tree reconstruction and accuracy of the tree topology was assessed by bootstrap analysis with 500 resampling replicates.

2.3. Analysis of the esterase mRNA levels in various tissues and life stages

Seventy-one cDNA libraries were sequenced using Illumina technology, each representing a sample of whole larvae, organs, or tissues at various life stages (Kanost et al., 2016; Cao and Jiang, 2017; Zhang et al., 2011). Number of reads mapped onto each transcript in the list of esterases and their non-catalytic homologs was used to calculate transcripts per kilobase million (TPM) in the RNA-seq datasets using RSEM (Li and Dewey, 2011). Hierarchical clustering of the $\log_2$(TPM +1) values was performed using Seaborn, a Python package with the Euclidean-based metric and average linkage clustering method.

3. Results and discussion

3.1. M. sexta lipases and their non-catalytic homologs

Lipases (E.C. 3.1.1.3) are responsible for cleavage of ester bonds in long-chain acylglycerols (≥10 carbon atoms). The preferred substrates of lipases are dietary or storage tri-, di-, and mono-glycerides. In *M. sexta*, we have identified a total of 61 lipases, greater than in *D. melanogaster* (56)*, A. gambiae* (51)*, A. mellifera* (26)*, B. mori* (29), and *T. castaneum* (54) (Horne et al., 2009). Lipases can be divided into six subfamilies, namely neutral lipases (NLs), acidic lipases (ALs), lipase-2, lipase-3, GDSL lipase, and hormone-sensitive lipase (HSL). According to the previous studies, *D. melanogaster* has 31, 21, 0, 1, 2, 1, *A. gambiae* has 28, 14, 0, 1, 7, 1*, A. mellifera* has 14, 4, 0, 1, 6, 1, *B. mori* has 11, 14, 0, 1, 2, 1, and *T. castaneum* has 25, 25, 0, 1, 2, and 1, which belong to the subfamilies of NL, AL, lipase-2, lipase-3, GDSL lipase, and HSL, respectively. In comparison, there are 39 NLs, 3 neutral lipase homologs (NLHs), 11 ALs, 3 acidic lipase homologs (ALHs), 2 lipase-2s, 1 lipase-3, 1 GDSL lipase, and 1 HSL in *M. sexta* (Fig. 1, Table

S1). As substitution of the catalytic Ser residue by Gly is anticipated to abolish the esterase activity, we further separate NLHs and ALHs out. Note that this phenomenon is not limited to *M. sexta* – some so-called NLs and ALs in other insects are catalytically inactive but may play roles not related to the catalytic activity (Miao et al., 2020).

Sequence alignment showed that the *M. sexta* NLs usually contain GHSLG or GFSLG, while most ALs contain GHSQG motif as the nucleophilic elbow (Fig. 2, Fig. S1). In mammals, lipases that clearly hydrolyze triacylglycerol/triglyceride (TAG/TG) must possess a long β9 loop (>15 residues) and a lid (>18 residues). However, these features may not be necessary for insect TAG lipases since, among the 109 NLs from the five holometabolous insects, only eight meet the two requirements (Horne et al., 2009): 3 in *D. melanogaster* and *T. castaneum*, 1 in *A. gambiae* and *A. mellifera*, and 0 in *B. mori*. *M. sexta* NL17, NL18, and NL26 have an 18, 24 and 18-residue β9 loop, and a 20, 26, and 25-residue lid, respectively (Fig. 2). Apparently, some of the eleven *B. mori* NLs must hydrolyze TAGs (a major storage form of lipids), but it is unclear how they achieve this. Further research is desirable to explore distinct requirements for TAG hydrolysis by insect NLs.

Based on the phylogenetic evidence, we named the *M. sexta* NLs and NLHs, which reside in a clade with well characterized neutral lipases from the fruit fly, dog, and other animals (Fig. 1). These include *D. melanogaster* CG4979 and CG6271, canine and squirrel pancreatic lipases, and mouse endothelial and hepatic lipases. Similarly, all the ALs and ALHs form another tight group with *Drosophila* CG17097, CG18284, CG3653, and a canine gastric lipase. *M. sexta* and *D. melanogaster* HSLs, lipase-3s, TGLs, MGL, GDSL lipases appear to have diverged from a common ancestor early in the evolution to perform conserved functions. Interestingly, the expression profiles of *M. sexta* HSL, lipase-3, TGL, and MGL are similar (Fig. 3).

M. sexta NL(H)1–8, 10–12, 19, 40 and AL(H)1, 3–6, 11 mRNA levels in midgut were high in

larval feeding stage (Fig. 3), indicating they are related to digestion (Miao et al., 2020). NL9 and AL2 transcripts were also detected in the gut, but they are not related to digestion, since NL9 and AL2 were mainly detected in pupae and adults. Moderate expression of *M. sexta* NL15, NL18, AL11, ALH13, AL14 and TGL genes were temporarily regulated in fat body, a center of lipid storage and mobilization (Arrese and Soulages, 2010). Different levels of NL14, 23, 24, 29, 33, 35, 39, 41, 42, AL12, TGL, and lipase-3 transcripts were detected in head and/or antenna and the preferential expression suggest these enzymes may perform sensory or neurodevelopmental functions in various life stages. Higher expression of NL20, NL22, NL31, AL11, and AL12 occurs in Malpighian tubules, so does NL23, NL34, and ALH13 expression in ovaries, NL27, NL40, AL7, ALH9, and AL10 expression in testes. Preferential production of NL26, 32, 37, and GDSL lipase in late embryo coincided with their expression in midgut of 2nd and 3rd instar larvae. The ovary- and testis-specific proteins may play roles in reproduction, whereas the four lipases in late embryo may prepare newly emerged 1st instar to readily utilize dietary lipids. *M. sexta* HSL may be expressed in response to AKH (Arrese et al., 1996 and 1999) and stimulate lipid hydrolysis in various tissues and cells, as their transcripts are detected in head, fat body, midgut, Malpighian tubules, testes, ovaries, and antenna. Such a wide distribution is also true for MGL, NL9, 17, 18, and 34. Additionally, NL9, 13, 17, 18, 33, 35–37, and HSL are induced in fat body and/or hemocytes after immune challenge (Fig. 3). Consistent with that, NL9 was also detected in the proteome of larval hemolymph after an immune challenge (He et al., 2016), suggesting that these non-digestive lipases play a role in producing energy and metabolites for making defense molecules to fight invading pathogens.

3.2. Carboxylesterases (COEs) and related proteins in M. sexta

The tobacco hornworm possesses 83 COE-like proteins, which can be divided into three classes based on the phylogenetic or genomic criteria (Oakeshott et al., 2010; Montella et al., 2012), viz., dietary/detoxification (I), hormone/semiochemical processing (II), and

(neuro)development (III) (Fig. 4). Class I consists of α-esterases (AEs) and feruloyl esterases (FEs) for food digestion and (de)toxification. In total, 29 AEs and 12 α-esterase homologs (AEHs, including SEH4-1–3) were separated into clades A and B. Twenty FEs and two noncatalytic homologs (FEHs) were found in *M. sexta*, which may be involved in resistance against organophosphate insecticides (Cui et al., 2015). Class II contains 2 β-esterases (BEs), 2 integument esterases (IEs), 1 homolog (IEH), and 4 juvenile hormone esterases (JHEs) for processing hormones and semiochemicals. Class III is composed of two acetylcholinesterases (AChEs) and nine catalytically inactive COEs, including 1 gliotactin (GLI), 6 neuroligins (NLGs) and 1 neurotactin (NRT) for neurotransmission and development. The remaining uncharacteristic esterase homolog-1 (UCEH1) gene may have arisen early in the evolution (Fig. 4).

Similar to the nucleophilic elbow in lipases, most *M. sexta* FEs share a GCSAG motif whereas AEs, BEs, AChEs and JHEs often contain a G(E/Q)SAG consensus (Fig. S2). One JHE was purified from hemolymph of *M. sexta*, which displayed a high JHE activity (Hinton and Hammock, 2001). Its full-length cDNA encodes a 573-residue protein identical in sequence to JHE1 identified here and in the hemolymph proteome (He et al., 2016). Catalytically inactive COEs (NLG, GLI and NRT) in the neuro/developmental class contain GHG(T/S)G, GPGAG and GHRAG, respectively. AE(H)3–7, SEH4-1–3 and FE4-1–7, 13, 21 were identified as digestion-related proteins (Miao et al., 2020). AE1 and AE2 (no signal peptide) are unrelated to digestion but AE1 was abundantly expressed in midgut during larval feeding stages (Fig. 5). AE17, 19 and 27 are moderately transcribed in fat body of pre-wandering and wandering larvae. AE30 and AEH14 are quite specifically expressed in testes and adult ovaries, respectively. AEH28, FE4-17 and FEH4-8 are preferentially produced in the midgut of feeding larvae and may assist digestion in some way. Higher expression of AE23, AEH18, FE4-9, 14, 20, and 22 occurs in antennae. Besides, AE31, IE1. IE2, JHE4, and UEH1 are synthesized in almost all the tissues during various life stages, with IE1, JHE4, and UEH1 mRNA at very high levels in head and antenna.

AChE activity was characterized in adult deutocerebrum (Homberg, et al., 1995), larval brain (Lester and Gilbert, 1987), and pupal antennae (Stengl et al., 1990) of *M. sexta*. We found both AChE1 and AChE2 are strongly expressed in brain of young adults and, while AChE1 transcripts are present at lower levels in other tissues and stages, there is no detectable AChE2 mRNA in the same datasets. Therefore, AChE1 may regulate cholinergic neurotransmission in some tissues during metamorphosis whereas the function of AChE2 appears to be limited to the adult stage. Investigations indicate that paralogous AChE1 terminates acetylcholine neurotransmission in most insects while *Tribolium castaneum* AChE2 plays an important role in female reproduction, embryo development, and growth of offspring (Lu et al., 2012; Zhao et al., 2013). NLG2, 3, 5, 6 and AE25 are expressed at low levels in adult head, whereas NLG1, 4 and NRT transcripts hardly exist in almost all tissues or stages. GLI mRNA levels are relatively high in gut tissue when compared with others. AE4, AE17 and IE1 proteins are present constitutively in hemolymph of larvae, pupae and adults (Cao et al., 2020). In contrast, *M. sexta* FE4-9, 12, 16, AE(H) 17–19, 36, and IE1, 2 mRNA levels are up-regulated in fat body and/or hemocytes after immune challenge (Fig. 5). Among them, AE17, AEH18 and IE2 proteins were detected in the hemolymph of larvae 24 h after a bacterial injection (He et al., 2016). Like HSL and the eight NLs, these enzymes may generate energy and defense molecules for fighting infection rather than growth.

3.3. Phospholipases and thioesterases in M. sexta

Phospholipases are enzymes that hydrolyze phospholipids at different ester bonds. One type consists of acylhydrolases and the other phosphodiesterases. Phospholipases A_1 (PLA1, EC 3.1.1.32) and A_2 (PLA2, EC 3.1.1.4) cleave the acyl ester bonds at sn-1 and sn-2 positions, respectively, and phospholipase B (PLB, EC 3.1.1.5, also known as lysophospholipase) hydrolyzes the acyl ester bonds at both positions. Unlike PLAs and PLB, phospholipases C (PLC, EC 3.1.4.3) and D (PLD, EC 3.1.4.4) are phosphodiesterases. PLCs hydrolyze the glycerophosphate bond and, according to their substrate preference, they can be further divided

into PC- and PI-PLCs. PC-PLCs have broader substrate specificity by working on phosphatidylcholine (PC), phosphatedyl-ethanolamine (PE), and phosphatidylserine (PS). PI-PLCs are more specific by working on phosphatidylinositol (PI), producing diacylglycerol (DAG) and inositol triphosphate (IPs). Both DAG and IP3 are second messengers that regulate intracellular Ca^{2+} flows. PLDs hydrolyze the other phosphodiester bond between phosphatidic acid and alcohol groups (*e.g*. choline, serine, ethanolamine, and inositol). Phosphatidic acid and PIs are second messengers for PLD-mediated signal transduction. Most PLDs are in the HKD family which contains a HXKXXXXD motif whereas non-HKD PLDs are less conserved (Filkin et al., 2020; Casas-Godoy 2012). PLA2s are involved in eicosanoid-mediated immune responses of insects (Kim et al., 2019; Kim and Stanley, 2021)

Based on the domain search (Table S3) and phylogenetic tree (Fig. 6A), *M. sexta* has 26 phospholipase-like proteins (PLPs). PLP1−10 and adipose triglyceride lipase (ATGL) contain a PLA2 domain (Soulages et al., 2012). The first seven are predicted to be secretory PLA2s (sPLA2); PLP8−10 and ATGL may be Ca^{2+}-independent (iPLA2) and contain a patatin-like domain. PLP11 and PLP12 belong to the PLB group. PLP13−17, 19−21, and 24 are putative PLCs, whereas PLP18, 22, 23, and 25 belong to the PLD clade (Fig. 6B). Classification of these proteins is the first step toward elucidation of their potential roles in cell signaling.

Thioesterases (TEs, E.C.3.1.2) are enzymes that hydrolyze thioester bonds and are known to terminate fatty acid synthesis. Their specific substrates include acetyl-CoA, palmitoyl-CoA, other acyl-CoAs, and acyl-linked ACPs (*i.e*., acyl carrier proteins). TEs are classified into 25 families by ThYme (Thioester-Active Enzyme Database, https://thyme.engr.unr.edu/v2.0/): 8 with the α/β hydrolase fold, 14 with double Hotdog fold, 1 each with NagB, flavodoxin, lactamase fold (Cantu et al., 2011; Cantu et al., 2010; Lenfant et al., 2013). In *M. sexta*, we found thirteen thioesterases. TE7 and TE11 are closely related to *Drosophila* palmitoyl thioesterase-1, which contains a cl21494 domain. TE1, 6, 8, 9, and 12 have a cl09938 domain whereas TE2−4, 10, and 13 contain

a cl00509 domain. Expression profiles of the thioesterases and phospholipases in *M. sexta* are shown in Fig. S3.

3.4. Concluding remarks

Serine esterases and other hydrolases of lipids constitute diverse families of proteins in each insect species. These enzymes have been investigated biochemically and genetically to elucidate functions in various physiological processes, such as food digestion, lipid transport, insecticide resistance, neural transmission, and hormonal control of development. While a framework of knowledge is available through studies on individual members of the protein families in different insects, a genome-wide identification, classification, and profiling of their genes would stimulate systematic research of the enzymes in biochemical model insects and kindle related studies in other species including agricultural pests and disease vectors. The selection of lipases, carboxylesterases, and thioesterases stemmed from our general interests in lipid metabolism in *M. sexta*. We also explored phospholipase-like proteins, because some of them (*e.g.*, PLA_2s, PLCs and PLDs) may play roles in cell signaling during immune responses. It is pleasing that 183 important genes from four different groups have been carefully annotated, along with their spatiotemporal expression patterns that suggest functions. However, it is also a daunting task to characterize them at the protein level in an organized manner, to inspire comparative studies in other insects, and to discover functions not known to exist in this taxonomic group. In the post-genomic era, this is a challenge insect biochemists must face and here we provide an improved framework of information.

Acknowledgements

We thank the anonymous reviewers for their critical comments on the manuscript. Dr. Estela Arrese in the Department of Biochemistry and Molecular Biology at Oklahoma State University provided useful information on insect lipases. The *Manduca* Genome

Project provided OGS2.0 and RNA-Seq datasets. This study was supported by NIH grants GM58634 and AI139998. The paper was approved for publication by the Director of Oklahoma Agricultural Experimental Station and supported in part under project OKL03054.

Tables

Table S1. Sequences and features of the 61 *M. sexta* lipases*

Msex ID	Enzyme names	MCOT ID	GenBank ID	GenBank accession number	Catalytic residues	Domain	Proteome support
Msex2.00709-RA	NL1	MCOT.C00649.0.0.CPO5P Msex2.00709-RA	XP_030029556.1	ON929093	184S, 208D, 272H	Lipase/vitellogenin	
Msex2.00710-RA	NL2	MCOT.C00650.2.0.COO5P Msex2.00710-RA	XP_030029542.1	ON929094	183S, 207D, 269H	Lipase/vitellogenin	
Msex2.00711-RB	NL3	MCOT.C00667.0.0.OOO5N Msex2.00711-RB	XP_030029362.1	ON929095	184S, 208D, 270H	Lipase/vitellogenin	
Msex2.03896-RA	NL4	MCOT.C03332.0.1.CPO4P Msex2.03896-RA	XP_030022432.1	ON929096	185S, 209D, 270H	Lipase/vitellogenin	
Msex2.03905-RA	NLH5	MCOT.C03337.2.0.CPP4P Msex2.03905-RB	XP_030022470.1	ON929097	186G, 210D, 270H	Lipase/vitellogenin	
Msex2.03906-RA	NLH6	MCOT.C03303.7.0.CON4P Msex2.03906-RA	XP_030022462.1	ON929098	142G, 166D, 227H	Lipase/vitellogenin	
Msex2.03908-RB	NL7	MCOT.C03305.1.0.TOP5P Msex2.03908-RB	XP_030022474.1	ON929099	185S, 209D, 270H	Lipase/vitellogenin	
Msex2.03909-RA	NL8	MCOT.C03305.3.0.CPO4P Msex2.03909-RA	XP_030022480.1	ON929100	185S, 209D, 270H	Lipase/vitellogenin	
Msex2.09694-RA	NL9	MCOT.C08450.0.0.COB5P Msex2.09694-RA	XP_030032052.1	ON929101	169S, 197D, 271H	Lipase/vitellogenin	IH (neutral lipase)/ LH (phospholipase A1-like)
Msex2.09846-RA	NLH10	MCOT.C08564.0.0.TOP5O Msex2.09846-RA	XP_030032286.1	ON929102	145G, 173D, 236H	Lipase/vitellogenin	IH (triacylglycerol lipase-2)
Msex2.10726-RA	NL11	MCOT.C09351.0.0.TOO5P Msex2.10726-RA	XP_030033676.1	ON929103	185S, 209D, 269H	Lipase/vitellogenin	
Msex2.12025-RA	NL12	MCOT.C10594.1.0.TOP5P Msex2.12025-RA	XP_030035750.1	ON929104	181S, 205D, 263H	Lipase/vitellogenin	
Msex2.00708-RA	NL13	MCOT.M13609.0.0.MWW XP Msex2.00708-RA	XP_030028994.1	ON929105	131S, 155D, 215H	Lipase/vitellogenin	
Msex2.02928-RB	NL14	MCOT.C02576.1.0.COO2P Msex2.02928-RB	XP_030020885.1	ON929106	151S, 178D, 286H	Lipase/vitellogenin	
Msex2.02929-RB	NL15	MCOT.C02577.0.0.TNN0N Msex2.02929-RB	XP_030020871.1	ON929107	147S, 174D, 247H	Lipase/vitellogenin	
Msex2.02948-RA	NL16	MCOT.M14242.0.0.MPOX N Msex2.02948-RA	XP_030020829.1	ON929108	178S, 202D, 267H	Lipase/vitellogenin	
Msex2.03090-RA	NL17	MCOT.C02910.0.2.CPP4P Msex2.03090-RA	XP_030021522.1	ON929109	194S, 222D, 305H	Lipase/vitellogenin	
Msex2.03106-RA	NL18	MCOT.C02884.2.0.CPP5O Msex2.03106-RA	XP_030021596.1	ON929110	206S, 229D, 324H	Lipase/vitellogenin	
Msex2.03907-RA	NL19	MCOT.C03304.0.1.CWN4N Msex2.03907-RA	XP_030022443.1	ON929111	184S, 208D, 269H	Lipase/vitellogenin	
Msex2.03910-RA	NL20	MCOT.C03307.0.0.COP4N Msex2.03910-RA	XP_030022456.1	ON929112	196S, 220D, 281H	Lipase/vitellogenin	

Msex2.04202-RA	NL21	MCOT.C03661.0.0.CPO3P Msex2.04202-RA	XP_030023019.1	ON929113	166S, 194D, 263H	Lipase/vitellogenin	
Msex2.05582-RA	NL22	MCOT.C04899.2.0.ONO5O Msex2.05582-RA	XP_030025375.1	ON929114	167S, 191D, 259H	Lipase/vitellogenin	
Msex2.05903-RA	NL23	MCOT.C05191.2.0.COO3P Msex2.05903-RA	XP_030025949.1	ON929115	235S, 262D, 333H	Lipase/vitellogenin	
Msex2.06056-RB	NL24	MCOT.C05229.0.0.CPP5P Msex2.06056-RB	XP_030026014.1	ON929116	159S, 186D, 261H	Lipase/vitellogenin	
Msex2.06084-RA	NL25	MCOT.C05362.0.0.CPP5P Msex2.06084-RA	XP_030026266.1	ON929117	167S, 195E, 262H	Lipase/vitellogenin	
Msex2.06108-RA	NL26	MCOT.C05599.0.0.CNN5O Msex2.06108-RA	XP_030026714.1	ON929118	230S, 253D, 341H	Lipase/vitellogenin	
Msex2.06717-RE	NL27	MCOT.C06010.0.0.COO5N Msex2.06717-RE	XP_030027522.1	ON929119	180S, 204D, 274H	Lipase/vitellogenin	
Msex2.06726-RA	NL28	MCOT.C06155.0.0.CON3P Msex2.06726-RA	XP_030027861.1	ON929120	184S, 212D, 275H	Lipase/vitellogenin	
Msex2.06777-RA	NL29	MCOT.C06069.2.0.CPP2O Msex2.06777-RA	XP_030027643.1	ON929121	174S, 202D, 266H	Lipase/vitellogenin	
Msex2.07012-RA	NL30	MCOT.C06044.0.0.TNN5P Msex2.07012-RA	XP_030027614.1	ON929122	172S, 197D, 262H	Lipase/vitellogenin	
Msex2.07856-RA	NL31	MCOT.C06924.0.0.TOO4O Msex2.07856-RA	XP_030029271.1	ON929123	151S, 176D, 244H	Lipase/vitellogenin	
Msex2.07857-RA	NL32	MCOT.C06925.0.0.TOP5P Msex2.07857-RA	XP_030029295.1	ON929124	178S, 203D, 271H	Lipase/vitellogenin	
Msex2.08211-RA	NL33	MCOT.C07156.0.0.CPP3P Msex2.08211-RA	XP_030029757.1	ON929125	220S, 243D, 329H	Lipase/vitellogenin	
Msex2.08887-RA	NL34	MCOT.C07912.0.0.TOO2O Msex2.08887-RA	XP_030031114.1	ON929126	178S, 206D, 273H	Lipase/vitellogenin	
Msex2.09693-RA	NL35	MCOT.C08449.0.0.OOO4O Msex2.09693-RA	XP_030032045.1	ON929127	166S, 194D, 267H	Lipase/vitellogenin	
Msex2.09695-RA	NL36	MCOT.C08454.0.0.CNN5O Msex2.09695-RA	XP_030032053.1	ON929128	177S, 205D, 280H	Lipase/vitellogenin	
Msex2.10218-RA	NL37	MCOT.C08927.0.0.CPP1P Msex2.10218-RA	XP_030032988.1	ON929129	200S, 227D, 302H	Lipase/vitellogenin	
Msex2.11343-RA	NL38	MCOT.C09956.0.0.TNO4O Msex2.11343-RA	XP_030034656.1	ON929130	133S, 157D, 227H	Lipase/vitellogenin	
Msex2.12486-RA	NL39	MCOT.C10844.0.0.TOO5O Msex2.12486-RA	XP_030036126.1	ON929131	95S, 123D, 196H	Lipase/vitellogenin	
Msex2.13120-RA	NL40	MCOT.C11476.0.0.OBO4O Msex2.13120-RA	XP_030037203.1	ON929132	183S, 207D, 267H	Lipase/vitellogenin	
Msex2.13991-RA	NL41	MCOT.C12091.0.0.CNO3N Msex2.13991-RA	XP_030038249.1	ON929133	253S, 281D, 343H	Lipase/vitellogenin	
Msex2.14926-RA	NL42	MCOT.C12862.0.0.TOO5O Msex2.14926-RA	XP_030039446.1	ON929134	170S, 198D, 265H	Lipase/vitellogenin	
Msex2.13978-RA/14803-RB	TGL	MCOT.C12185.1.0.TOO5O Msex2.14803-RB	XP_030038395.1	ACR61720.1	367S, NA, NA	WWE domain/Alpha/Beta hydrolase fold/DHD domain	
Msex2.12997-RA	MGL	MCOT.C11405.0.1.CPP5P Msex2.12997-RA	XP_030037098.1	ON929135	217S, NA, NA	AB-hydrolase lipase	
Msex2.01293-RA	Lipase3	MCOT.C01171.0.0.TNO5P Msex2.01293-RA	XP_030038532.1	ON929136	451S, 503D	Fungal lipase-like	
Msex2.01196-RB	HSL	MCOT.C01079.1.0.TNO5P Msex2.01196-RB	XP_037293343.1	ON929137	HSL-N pfam06350	Hormone-sensitive lipase	
Msex2.09290-RA	GDSL	MCOT.C08067.1.0.TNN5P Msex2.09290-RA	XP_030031313.1	ON929138	53S, 302D, 305H	GDSL lipase/esterase	
Msex2.10332-RA	AL1	MCOT.C09035.0.0.CPP5P Msex2.10332-RA	XP_030033175.1	ON929139	193S, 368D, 399H	AB-hydrolase lipase	
Msex2.02034-RA	AL2	MCOT.M13989.0.0.MPOXP Msex2.02034-RA	XP_030041138.1	ON929140	184S, 358D, 389H	AB-hydrolase lipase	
Msex2.03554-RA	AL3	MCOT.C03184.0.0.TOO5N Msex2.03554-RA	XP_030022206.1	ON929141	189S, 361D, 392H	AB-hydrolase lipase	
Msex2.07865-RA	AL4	MCOT.C06938.0.0.COO5P Msex2.07865-RA	XP_030029283.1	ON929142	191S, 365D, 396H	AB-hydrolase lipase	
Msex2.13484-RA	AL5	Not found in Mcot	XP_030038008.1	ON929143	165S, 337D, 367H	AB-hydrolase lipase	
Msex2.13487-RA	ALH6	MCOT.C11865.0.0.TOO5O Msex2.13487-RA	XP_030037848.1	ON929144	165S, NA, NA	AB-hydrolase lipase	

Msex2.04752-RA	AL7	MCOT.C04209.2.0.COO5P Msex2.04752-RA	XP_030024011.1	ON929145	166S, 340D, 371H	AB-hydrolase lipase	
Msex2.07261-RA	AL8	MCOT.C06495.0.0.CPP4P Msex2.07261-RA	XP_030028491.1	ON929146	262S, 436D, 467H	AB-hydrolase lipase	IH (acidic lipase)
Msex2.09125-RA	ALH9	MCOT.C08344.0.0.CPP5P Msex2.09125-RA	XP_030031845.1	ON929147	237A, 405D, 436H	AB-hydrolase lipase	
Msex2.09126-RA	AL10	MCOT.C08345.0.0.CPP4P Msex2.09126-RA	XP_030031837.1	ON929148	242S, 403D, 434H	AB-hydrolase lipase	
Msex2.10877-RA	AL11	MCOT.C09552.0.1.CPP5O Msex2.10877-RA	XP_030034035.1	ON929149	234S, 408D, 439H	AB-hydrolase lipase	
Msex2.10878-RA	AL12	MCOT.C09552.0.2.CPP5P Msex2.10878-RA	XP_030034036.1	ON929150	196S, 371S, 402H	AB-hydrolase lipase	
Msex2.12181-RA	ALH13	MCOT.C10799.0.0.COO5P Msex2.12181-RA	XP_030036032.1	ON929151	328A, 502D, 533H	AB-hydrolase lipase	
Msex2.08174-RA	AL14	MCOT.C07203.0.0.CPP4P Msex2.08174-RA	XP_030029854.1	ON929152	237S, 409D, 440H	AB-hydrolase lipase	IH (triacylglycerol lipase-1)

Table S2. Sequences and features of the 83 *M. sexta* COEs*

Msex ID	Enzyme names	MCOT ID	GenBank ID	GenBank accession number	Catalytic residues	Domain	Proteome support
Msex2.11175-RB	FE4-1	MCOT.C09773.2.0.COO 5O Msex2.11175-RB	XP_030034366.1	ON929153	201S, 331E, 445H	Carboxylesterase, type B	
Msex2.01458-RA	FE4-2	MCOT.C01284.3.0.COO 5N Msex2.01458-RA	XP_030040069.1	ON929154	202S, 322E, 426H	Carboxylesterase	
Msex2.01457-RA	FE4-3	MCOT.C01284.2.0.CON 5N Msex2.01457-RA	XP_030040068.1	ON929155	202S, 322E, 428H	Carboxylesterase	
Msex2.05782-RA	FE4-4	MCOT.C05018.2.0.COO 5O Msex2.05782-RA	XP_030025623.1	ON929156	201S, 331E, 445H	Carboxylesterase, type B	
Msex2.07029-RB	FE4-5	MCOT.C06108.15.0.COP 5P Msex2.07029-RB	XP_030027701.1	ON929157	216S, 344E, 457H	Carboxylesterase, type B	
Msex2.12089-RC	FE4-6	MCOT.C10576.0.0.CBB5 O Msex2.12089-RC	XP_030035718.1	ON929158	221S, 351E, 465H	Carboxylesterase, type B	
Msex2.13096-RA	FE4-7	MCOT.C11358.0.0.TOO5 O Msex2.13096-RA	XP_030034367.1	ON929159	201S, 331E, 445H	Carboxylesterase, type B	
Msex2.07032-RA	FEH4-8	MCOT.C06110.0.0.CPP5 N Msex2.07032-RA	XP_030027744.1	ON929160	206N, 335D, 448H	Carboxylesterase, type B	
Msex2.07027-RA	FE4-9	MCOT.C06108.13.0.CO O5N Msex2.07027-RA	XP_030027699.1	ON929161	202S, 330E, 443H	Carboxylesterase, type B	
Msex2.12086-RB	FE4-10	Not found	XP_030034366.1	ON929162	4S, 134E, 248H	Carboxylesterase, type B	
Msex2.07033-RA	FE4-11	MCOT.C06111.0.0.TOO5 O Msex2.07033-RA	XP_030027731.1	ON929163	200S, 328E, 441H	Carboxylesterase, type B	
Msex2.12087-RA	FE4-12	Not found	XP_030034367.1	ON929164	221S, 351E, 465H	Carboxylesterase, type B	
Msex2.11171-RA	FE4-13	Not found	XP_030034367.1	ON929165	216S, 346E, 460H	Carboxylesterase, type B	
Msex2.07031-RA	FE4-14	MCOT.C06109.0.0.COO 5N Msex2.07031-RA	XP_030027741.1	ON929166	205S, 334E, 447H	Carboxylesterase, type B	IH (antennal esterase CarE6)
Msex2.13095-RA	FEH4-15	Not found	XP_037295228.1	ON929167	116S, NA, NA	Carboxylesterase, type B	
Msex2.05781-RA	FE4-16	Not found	XP_030025623.1	ON929168	201S, 331E, 445H	Carboxylesterase, type B	
Msex2.11173-RA	FE4-17	Not found	XP_030034366.1	ON929169	73S, 203E, 317H	Carboxylesterase, type B	
Msex2.12088-RA	FE4-18	Not found	XP_030025623.1	ON929170	201S, 331E, 445H	Abhydrolase	
Msex2.07025-RA	FE4-19	MCOT.C06108.3.0.COO 5N Msex2.07025-RA	XP_030027742.1	ON929171	201S, 329E, 442H	Carboxylesterase	

Msex2.07028-RC	FE4-20	MCOT.C06108.2.0.OOO5O Msex2.07028-RC	XP_030027704.1	ON929172	204S, 332E, 445H	Carboxylesterase	
Msex2.11174-RA	FE4-21	MCOT.M16177.0.0.MOOXO Msex2.11174-RA	XP_030034367.1	ON929173	201S, 331E, 445H	Carboxylesterase	
Msex2.07026-RA	FE4-22	MCOT.C06108.10.0.COO5O Msex2.07026-RA	XP_030027740.1	ON929174	218S, 346E, 459H	Carboxylesterase	
Msex2.00350-RA	AE1	MCOT.C00252.3.0.OOO5O Msex2.00350-RA	XP_030021129.1	ON929175	183S, 317E, 440H	Carboxylesterase, type B	
Msex2.07667-RB	AE2	MCOT.C06717.0.0.TNP5P Msex2.07667-RB	XP_030028867.1	ON929176	204S, 333E, 441H	Carboxylesterase, type B	
Msex2.07963-RB	AE3	MCOT.C07085.0.0.CPP5N Msex2.07963-RB	XP_030029533.1	ON929177	199S, 331E, 446H	Carboxylesterase, type B	
Msex2.03978-RB	AE4	MCOT.C03553.2.0.COO5O Msex2.03978-RA	XP_030022822.1	ON929178	206S, 328E, 442H	Carboxylesterase, type B	LH (bile salt-activated lipase-like)
Msex2.04204-RA	AEH5	MCOT.C03662.0.0.CPP5P Msex2.04204-RA	XP_030023018.1	ON929179	202D, 325E, 425H	Carboxylesterase, type B	
Msex2.08585-RC	AEH6	MCOT.M15633.0.0.TOOXO Msex2.08585-RC	XP_030030744.1	ON929180	197D, 311E, 418H	Carboxylesterase, type B	
Msex2.06988-RB	AE7	MCOT.C06015.2.0.OPN5N Msex2.06988-RB	XP_030027603.1	ON929181	204S, 326E, 440H	Carboxylesterase, type B	
Msex2.12974-RB	AE8	MCOT.C11485.2.0.COO5P Msex2.12974-RB	XP_030037225.1	ON929182	214S, 337E, 452H	Carboxylesterase, type B	
Msex2.00366-RB	AE9	MCOT.C00314.2.0.COO5O Msex2.00366-RB	XP_030020800.1	ON929183	209S, 344E, 462H	Carboxylesterase, type B	IH (α-esterase 45)
Msex2.09113-RA	AEH10	MCOT.M15757.0.0.TOOXO Msex2.09113-RA	XP_030026304.1	ON929184	210G, 331E, 450N	Carboxylesterase, type B	
Msex2.07668-RA	AE11	MCOT.C06728.0.0.CPP5P Msex2.07668-RA	XP_030028871.1	ON929185	187S, 321E, 434H	Carboxylesterase, type B	
Msex2.00351-RA	AE12	MCOT.C00253.0.0.OON5N Msex2.00351-RA	XP_030022055.1	ON929186	184S, 321E, 442H	Carboxylesterase, type B	
Msex2.12348-RA	AE13	MCOT.C11069.0.0.COO5P Msex2.12348-RA	XP_030037225.1	ON929187	204S, 327E, 442H	Carboxylesterase, type B	
Msex2.10730-RA	AEH14	MCOT.C09352.0.0.COP5P Msex2.10730-RA	XP_030033686.1	ON929188	225G, 352E, 469H	Carboxylesterase, type B	
Msex2.06145-RA	AE15	MCOT.C05267.2.0.COO5P Msex2.06145-RA	XP_030026026.1	ON929189	190S, 325E, 438H	Carboxylesterase, type B	
Msex2.00912-RC	AEH16	MCOT.C00880.2.2.CNN2P Msex2.00912-RC	XP_030033544.1	ON929190	189E, 311D, 410H	Carboxylesterase, type B	
Msex2.11032-RA	AE17a(antennal esterase-1)	MCOT.M16150.0.0.MOOXO Msex2.11032-RA	XP_030034200.2	ON929191	190S, 278E, 382H	Carboxylesterase, type B	IH (antennal esterase-1)
Msex2.11032-RD	AE17d(antennal esterase-2)	MCOT.M16150.0.0.MOOXO	XP_030034201.1	ON929192	190S, 281E, 383H	Carboxylesterase, type B	IH (antennal esterase-2)/all H (SE4a)
Msex2.11565-RA	AEH18	MCOT.M16262.0.0.TNOXN Msex2.11565-RA	XP_030034911.1	ON929193	198G, 328E, 443H	Carboxylesterase, type B	IH (carboxyl/cholinesterase CCE002a, venom CE6)
Msex2.00367-RB	AE19	MCOT.M13516.0.0.TOOXO Msex2.00367-RB	XP_030020849.1	ON929194	206S, 346E, 461H	Carboxylesterase, type B	
Msex2.12347-RA	AE20	MCOT.M16429.0.0.MOOXO Msex2.12347-RA	XP_030036528.1	ON929195	160S, 285E, 400H	Carboxylesterase	
Msex2.00349-RA	AE21	MCOT.C00252.5.0.COO5O Msex2.00349-RA	XP_030021117.1	ON929196	183S, 317E, 441N	Carboxylesterase, type B	
Msex2.00369-RA	AE22	MCOT.C00316.3.0.COO5O Msex2.00369-RA	XP_030020839.1	ON929197	204S, 340E, 458H	Carboxylesterase	
Msex2.10707-RA	AE23	MCOT.C09758.0.0.CPP5P Msex2.10707-RA	XP_030034315.1	ON929198	196S, 327E, 444H	Carboxylesterase	
Msex2.00370-RB	AE24	MCOT.C00316.1.0.TNN5N Msex2.00370-RB	XP_030022077.1	ON929199	204S, 340E, 458H	Carboxylesterase	
Msex2.05291-RB	AE25	MCOT.C04595.2.0.CPP5P Msex2.05291-RB	XP_030024687.1	ON929200	281S, 414E, 533H	Carboxylesterase	
Msex2.15130-RA	AE26	MCOT.C13026.0.0.TOO5O Msex2.15130-RA	XP_030039646.1	ON929201	201S, 333E, 450H	Carboxylesterase	
Msex2.06147-RB	AE27	MCOT.C05253.0.1.CNP5O Msex2.06147-RB	XP_030026072.1	ON929202	33S, 168E, 279H	Carboxylesterase	
Msex2.03710-RA	AEH28	MCOT.C03027.0.0.CPO4P Msex2.03710-RA	XP_030021864.1	ON929203	265G, 394E, 510H	Carboxylesterase	

Msex2.12973-RA	AE29	MCOT.C11484.0.0.TOW5O Msex2.12973-RA	XP_030037217.1	ON929204	204S, 330E, 444H	Carboxylesterase	
Msex2.00843-RA	AE30	MCOT.C00850.0.0.CPO5P Msex2.00843-RA	XP_030034080.1	ON929205	197S, 329E, 444H	Carboxylesterase	
Msex2.00691-RA	AE31	MCOT.C00606.0.0.COO4N Msex2.00691-RA	XP_030027948.1	ON929206	187S, 322E, 438H	Carboxylesterase	
Msex2.06705-RD	AE32	MCOT.C05989.0.0.CPP4P Msex2.06705-RD	XP_030027489.1	ON929207	185S, 319E, 435H	Carboxylesterase	
Msex2.00911-RA	AEH33	MCOT.C00880.2.1.CPP0P Msex2.00911-RA	XP_030033317.1	ON929208	334G, 448A, 563R	Carboxylesterase	
Msex2.06293-RA	AEH34	MCOT.C05398.0.0.TNO1N Msex2.06293-RA	XP_030026299.1	ON929209	188G, 325E, 446A	Carboxylesterase	
Msex2.07962-RB	AE35	MCOT.C07080.1.0.TOO5N Msex2.07962-RB	XP_030029535.1	ON929210	196S, 328E, 443H	Carboxylesterase, type B	
Msex2.10708-RA	AE36	MCOT.M16088.0.0.TOOXO Msex2.10708-RA	XP_030034317.1	ON929211	204S, 336E, 451H	Carboxylesterase	
Msex2.10706-RA	AE37	MCOT.C09757.0.0.CPP5P Msex2.10706-RA	XP_030034314.1	ON929212	196S, 322E, 440H	Carboxylesterase	
Msex2.06146-RA	AE38	MCOT.C05267.5.0.CPP5P Msex2.06146-RA	XP_030026028.1	ON929213	190S, 325E, 438H	Carboxylesterase	
Msex2.06294-RA	SEH4-1	MCOT.C05387.0.0.COO5P Msex2.06294-RA	XP_030026304.1	ON929214	208K, 337E, 463S	Carboxylesterase	
Msex2.06295-RA	SEH4-2	MCOT.C05388.0.0.CPP5P Msex2.06295-RA	XP_030026301.1	ON929215	211E, 342E, 469T	Carboxylesterase, type B	
Msex2.06296-RA	SEH4-3	MCOT.C05389.0.0.CNN5P Msex2.06296-RA	XP_030026302.1	ON929216	210R, 341E, 468T	Carboxylesterase	
Msex2.03410-RB	BE1	MCOT.C02949.0.0.CPO4P Msex2.03410-RB	XP_030021680.1	ON929217	222S, 349E, 466H	Carboxylesterase	
Msex2.06675-RB	BE2	MCOT.C05866.0.0.CPO5P Msex2.06675-RB	XP_030027324.1	ON929218	202S, 332E, 454H	Carboxylesterase	IH (carboxyl/cholinesterase 7)
Msex2.09673-RA	AChE1	MCOT.C08535.0.0.CPN5P Msex2.09673-RA	XP_030032226.1	ON929219	311S, 437E, 551H	Carboxylesterase	
Msex2.05726-RB	AChE2	MCOT.C04988.0.0.CPP5P Msex2.05726-RB	XP_030025552.1	ON929220	266S, 395E, 509H	Carboxylesterase	
Msex2.00907-RA	JHE1a	MCOT.C00878.0.0.CNN5N Msex2.00907-RA	XP_030033511.1	ON929221	226S, 355E, 471H	Carboxylesterase, type B	IH (JH esterase)
Msex2.00907-RE	JHE1e	MCOT.M13695.0.0.MOOXO Msex2.00907-RE	XP_030033498.1	ON929222	541S, 672E, 786H	Carboxylesterase, type B	
Msex2.00908-RA	JHE2	MCOT.C00918.0.0.TOO5O Msex2.00908-RA	XP_030033523.1	ON929223	87S, 218E, 332H	Carboxylesterase, type B	
Msex2.00909-RC	JHE3	MCOT.C00879.2.0.COO5N Msex2.00909-RC	XP_030033487.1	ON929224	226S, 357E, 471H	Carboxylesterase, type B	
Msex2.00910-RB	JHE4	MCOT.C00879.1.0.TOO5P Msex2.00910-RB	XP_030033317.1	ON929225	226S, 357E, 471H	Carboxylesterase, type B	
Msex2.07504-RD	IE1d (SE4b)	Not found	XP_030028710.1	ON929226	219S, 349E, 468H	esterase E4-like	AH (SE4b)
Msex2.07504-RE	IE1e	MCOT.C06646.2.0.CPO5P Msex2.07504-RE	XP_030028705.1	ON929227	209S, 339E, 458H	Carboxylesterase, type B	
Msex2.08531-RA	IE2	MCOT.C07449.1.0.TOO5O Msex2.08531-RA	XP_030030295.1	ON929228	208S, 327E, 431H	Carboxylesterase	IH (integument esterase 2)
Msex2.10976-RA	IEH3	MCOT.C09601.0.0.COO5N Msex2.10976-RA	XP_030034116.1	ON929229	208G, 340E, 461H	Carboxylesterase	
Msex2.07931-RA	NLG1	MCOT.M15448.0.0.MBBXO Msex2.07931-RA	XP_030029367.1	ON929230	131A, 246E, 358T	Carboxylesterase, type B	
Msex2.01474-RB	NLG2	MCOT.C01291.0.0.TOO5O Msex2.01474-RB	XP_030040094.1	ON929231	250G, NA, NA	Carboxylesterase, type B	
Msex2.03169-RB	NLG3	MCOT.C02858.0.0.TOO5N Msex2.03169-RB	XP_030021468.1	ON929232	236G, 378E, 490H	Carboxylesterase, type B	
Msex2.01472-RB	NLG4	MCOT.C01290.1.1.CNN5P Msex2.01472-RB	XP_030040104.1	ON929233	352G, 511E, 636H	Carboxylesterase, type B	
Msex2.01473-RA	NLG5	MCOT.C01302.0.0.COP5P Msex2.01473-RA	XP_030040093.1	ON929234	245G, 367E, 481H	Carboxylesterase	
Msex2.01468-RA	NLG6	MCOT.C01300.0.0.TOO5P Msex2.01468-RA	XP_030040106.1	ON929235	282G, 442E, 560H	Carboxylesterase	
Msex2.06813-RA	NRT	MCOT.C06341.2.0.CPO4O Msex2.06813-RA	XP_030028152.1	ON929236	443R, NA, NA	Carboxylesterase, type B	

Msex2.00941-RA	GLI	MCOT.C00817.2.0.CPO5P Msex2.00941-RA	XP_030031978.1	ON929237	341G, 485E, 600H	Carboxylesterase, type B	
Msex2.00808-RA	UCEH1	MCOT.C00720.2.0.CNO4N Msex2.00808-RA	XP_030030362.1	ON929238	222G, 363E, 509H	Carboxylesterase, type B	

Table S3. Sequences and features of the 26 *M. sexta* PLPs*

Msex ID	Enzyme names	MCOT ID	GenBank ID	GenBank accession number	Catalytic residues	Domain	Superfamily
Msex2.00774-RA	PLP1	MCOT.C00744.0.0.TOO5P Msex2.00774-RA	XP_030030378.2	ON929239	160G, 178H, 179D, 182Y, 197Y, 231D	PLA2c	cl05417
Msex2.01838-RB	PLP2	MCOT.C01601.0.0.CPO2N Msex2.01838-RB	XP_030040762.1	ON929240	496H, 526D, 548Y	Phospholip_A2_2	cl05417
Msex2.05071-RB	PLP3	MCOT.C04483.0.0.TOO5O Msex2.05071-RB	XP_030024524.1	ON929241	87H, 117D, 139Y	PLA2_bee_venom_like	cl05417
Msex2.05073-RA	PLP4	MCOT.C04484.0.0.OOO5O Msex2.05073-RA	XP_030024475.1	ON929242	66H, 96D, 118Y	PLA2_bee_venom_like	cl05417
Msex2.08239-RA	PLP5	MCOT.C07199.0.1.COO3O Msex2.08239-RA	XP_030029837.1	ON929243	229H, 259D, 281Y	Phospholip_A2_2	cl05417
Msex2.13074-RA	PLP6	MCOT.C11454.1.0.TOO4P Msex2.13074-RA	XP_030037175.1	ON929244	284H, 315D, 337F	Phospholip_A2_2	cl05417
Msex2.01676-RA	PLP7	MCOT.M13907.0.0.MPPXO Msex2.01676-RA	XP_030040460.1	ON929245		Phospholip_A2_3 superfamily	cl29745
Msex2.12864-RB/13342-RB	ATGL	MCOT.C11609.2.0.CPP4P Msex2.12864-RB/MCOT.C11609.1.0.CPO4N Msex2.13342-RB	XP_037297526.1	AEJ33048.1		Pat_iPLA2	cl11396
Msex2.03481-RC	PLP8	MCOT.C02975.3.0.TOO5P Msex2.03481-RC	XP_030021739.1	ON929246	474G, 475G, 477R, 507S, 653D	Patatin_and_cPLA2 superfamily	PHA03095
Msex2.04865-RD	PLP9	MCOT.C04519.6.0.CPO5P Msex2.04865-RD	XP_030024560.2	ON929247	901G, 902G, 904R, 929S, 1049D	Pat_PNPLA6_PNPLA7/ CAP_ED	cl11396/cl00047
Msex2.01488-RA	PLP10	MCOT.C01510.0.0.CPP5P Msex2.01488-RA	XP_030040609.2	ON929248	233G, 234G, 236R, 266S, 410D	Patatin_and_cPLA2 superfamily	cl11396
Msex2.00157-RA	PLP11	MCOT.C00183.0.0.CPP2P Msex2.00157-RA	XP_030020376.2	ON929249	cl20281	Phospholip_B super family	cl20281
Msex2.00896-RA	PLP12	MCOT.C00874.0.0.CPO4P Msex2.00896-RA	XP_030034155.1	ON929250	pfam04916	Phospholip_B	cl20281
Msex2.02940-RA	PLP13	MCOT.C02582.2.0.COP5O Msex2.02940-RA	XP_037298096.1	ON929251	332H, 377H	PI-PLCc_beta/PH_PLC_beta	cl14615/cl17171
Msex2.05617-RB	PLP14	MCOT.C04959.3.0.CPP5P Msex2.05617-RB	XP_030025529.2	ON929252	170H, 229H	PI-PLCc_GDPD_SF superfamily	cl14615
Msex2.01388-RA	PLP15	MCOT.C01413.0.0.CPP5N Msex2.01388-RA	XP_030040309.1	ON929253	324H, 369H	PI-PLCc_gamma/PH_PLC_gamma	cl14615/cl17171
Msex2.07720-RD	PLP16	MCOT.C06943.0.0.CPP5P Msex2.07720-RD	XP_030029320.1	ON929254	50H, 121H	PI-PLCXD1c	cl14615
Msex2.10438-RA	PLP17	MCOT.C09301.1.0.TON5O Msex2.10438-RA	XP_037298980.1	ON929255	765H, 810H	PI-PLCc_GDPD_SF superfamily	cl14615
Msex2.10780-RA	PLP18	MCOT.C09727.5.0.CPP3P Msex2.10780-RA	XP_037299399.1	ON929256	84H, 126H	PI-PLCc_GDPD_SF superfamily	cl14615
Msex2.11795-RB	PLP19	MCOT.C10484.0.0.TOO5P Msex2.11795-RB	XP_030035541.1	ON929257	328H, 375H	PI-PLCc_beta/PH_PLC_beta	cl14615/cl17171

Msex2.14179-RA	PLP20	Not found in Mcot	XP_037303078.1	ON929258	385N, 409R, 438D, 440D, 445L	PI-PLCc_GDPD_SF superfamily	cl14615
Msex2.14177-RA	PLP21	Not found in Mcot	XP_037298980.1	ON929259		PI-PLCc_GDPD_SF superfamily	cl14615
Msex2.07384-RB	PLP22	MCOT.C06586.0.0.CPP5P Msex2.07384-RB	XP_030028608.2	ON929260	443H, 445K, 457T, 459N, 467N	PHA02820 superfamily/phospholipase-D-like protein	cl33698
Msex2.07588-RA	PLP23	MCOT.C06689.0.0.CPO5P Msex2.07588-RA	XP_030028781.1	ON929261	372H, 374K, 393S, 395N, 406E/153H, 155K, 167G, 169N, 180D	PLDc_PGS1_euk_1/2	cl15239
Msex2.08427-RA	PLP24	Not found	XP_030029995.2	ON929262		PLDc_SF superfamily/zf-CCHH	cl15239
Msex2.14404-RA	PLP25	MCOT.C11797.0.0.TOO1O Msex2.14404-RA	XP_037297426.1	ON929263	1083H, 1085K, 1098S, 1100N, 1113E	PLN02866 superfamily/ phospholipase D	cl33584

Table S4. Sequences and features of the 13 *M. sexta* thioesterases*

Msex ID	Enzyme names	MCOT ID	GenBank ID	GenBank accession number	Catalytic residues	Domain	Superfamily
Msex2.03583-RB	TE1	MCOT.C03064.0.1.CPO5P Msex2.03583-RB	XP_030021952.1	ON929264	C183, H315, H353/K1887, S1913, Y1926, N1930	Thioesterase, PKS, NADB, hot_dog	cl09938
Msex2.12759-RA	TE2	MCOT.C11302.0.0.TOO5P Msex2.12759-RA	XP_030036951.1	ON929265	D88, V111, Q113, D140, D141, K142, T143	Thioesterase, hot_dog	cl00509
Msex2.02857-RA	TE3	MCOT.M14225.0.0.MOWXP Msex2.02857-RA	KAG6446062.1	ON929266	L60, T88, Y95, L96, K97, A99	Thioesterase, PaaI_thioesterase, hot_dog	cl00509
Msex2.09017-RB	TE4	Not found	XP_030031121.1	ON929267	L56, S85, Y92, L93, S94, A95	Thioesterase, PaaI_thioesterase, hot_dog	cl00509
Msex2.11282-RB	TE5	MCOT.C10136.0.0.TOP5O Msex2.11282-RB	XP_037292860.1	ON929268	G9, T11, G12, A13, L14, V33, G34, K38, S57, R58, S81, V106, L128, D129, F130, L178, A179, S181, H183	Thioesterase, enoyl_red, NADB	cl16912
Msex2.11718-RB	TE6	MCOT.C10459.2.1.CPO5O Msex2.11718-RB	XP_030035513.2	ON929269	C175, H307, H345	Thioesterase, PKS, NADB, hot_dog	cl09938
Msex2.10287-RA	TE7	MCOT.C09154.0.0.CPP5P Msex2.10287-RA	XP_030033351.1	ON929270	N/A	Palm_thioest	cl21494
Msex2.05790-RB	TE8	MCOT.C05173.3.0.CPP5P Msex2.05790-RB	XP_030025895.1	ON929271	C170, H303, N339	Thioesterase,PKS, NADB, enoyl_red	cl09938
Msex2.05789-RA	TE9	MCOT.C05172.4.0.COO5O Msex2.05789-RA	XP_037293598.1	ON929272	T131, N264, H300	Thioesterase, PKS	cl09938
Msex2.03120-RA	TE10	MCOT.M14279.0.0.MWPXO Msex2.03120-RA	XP_030021606.1	ON929273	D72, L96, Q97, E124, D125, K126, T127	Thioesterase, hot_dog	cl00509
Msex2.02348-RA	TE11	MCOT.M14072.0.0.MPPXP Msex2.02348-RA	XP_030020038.1	ON929274	N/A	Palm_thioest	cl21494
Msex2.05791-RB	TE12	MCOT.C05169.7.0.OPO5O Msex2.05791-RB	XP_030025892.2	ON929275	N130, N263, H299	Thioesterase, PKS	cl09938
Msex2.03119-RA	TE13	Not found	XP_030021604.1	ON929276	N/A	Thioesterase, hot_dog	cl00509

*: Proteome verification is based on the previous data (He et al., 2016; Cao et al., 2020). Proteins shaded *red* were identified in the cell-free hemolymph from larvae 24 h after an immune challenge (He et al., 2016). Similarly, those shaded cyan were found in the proteomes of hemolymph at different developmental stages (Cao et al., 2020). Yellow means non-catalytic homolog; green represents lack of the GXSXG motif.

Figures

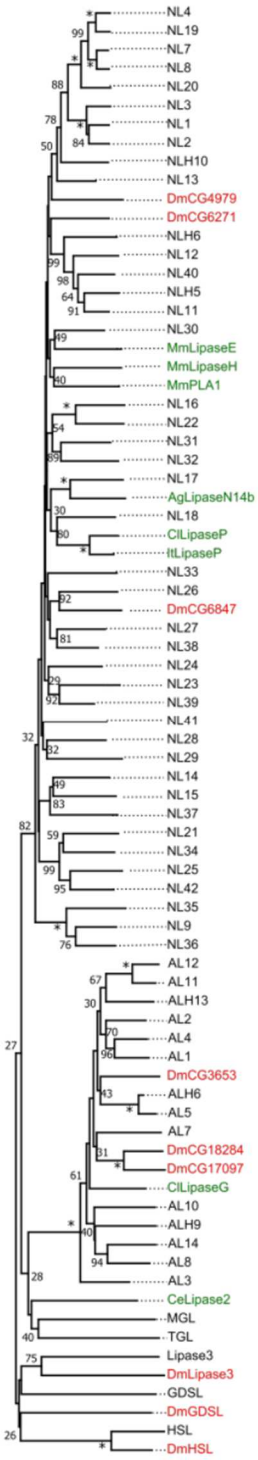


Fig. 1. A phylogenetic tree of lipase-related proteins in *M. sexta*, *D. melanogaster* and others. As described in *Section 2.2,* entire amino acid sequences of the neutral and acidic lipases were aligned with their homologs in *M. sexta* (Ms) and other organisms to construct the neighbor-joining tree. *D. melanogaster* (Dm) CG4979, CG6271, CG6847, CG17097, CG18284, CG3653,

CG33174 (lipase-3), CG11055 (HSL), CG11029 (GDSL lipase), *Anopheles gambiae* (Ag) lipase N14b (XP_310922.4), *Caenorhabditis elegans* (Ce) lipase-2 (NP_496693), *Ictidomys tridecemlineatus* (It) pancreatic lipase (AAK72259), *Canis lupus* (Cl) pancreatic (NP_001003319) and gastric (NP_001003209) lipases, and *Mus musculus* (Mm) endothelial lipase (AAD30435), lipase H (NP_001077363), and phospholipase A1 (AAL55475) are shown as abbreviations (*black* font for *Manduca*, *red* font for *Drosophila*, and *green* font for all other species). Bootstrap values greater than 25 are indicated at the nodes. Branch length represents the number of substitutions per site.

Fig. 2. A region of the aligned amino acid sequences in 42 *M. sexta* NL(H)s and 9 homologs from other organisms. *D. melanogaster* CG4979, CG6271, CG6847, *A. gambiae* lipase N14b (XP_310922), *C. lupus* pancreatic lipase (NP_001003319), *I. tridecemlineatus* pancreatic lipase (AAK72259), *M. musculus* endothelial lipase (AAD30435), lipase H (NP_001077363) and phospholipase A1 (AAL55475), and *M. sexta* NL(H)s are aligned. The catalytic residues (S, D/E and H) are in *red bold* font and the GXSXG consensus sequence around the active site serine are *highlighted yellow*. Predicted enzyme β9 loop and lid are shaded in *cyan* and *green*, respectively.

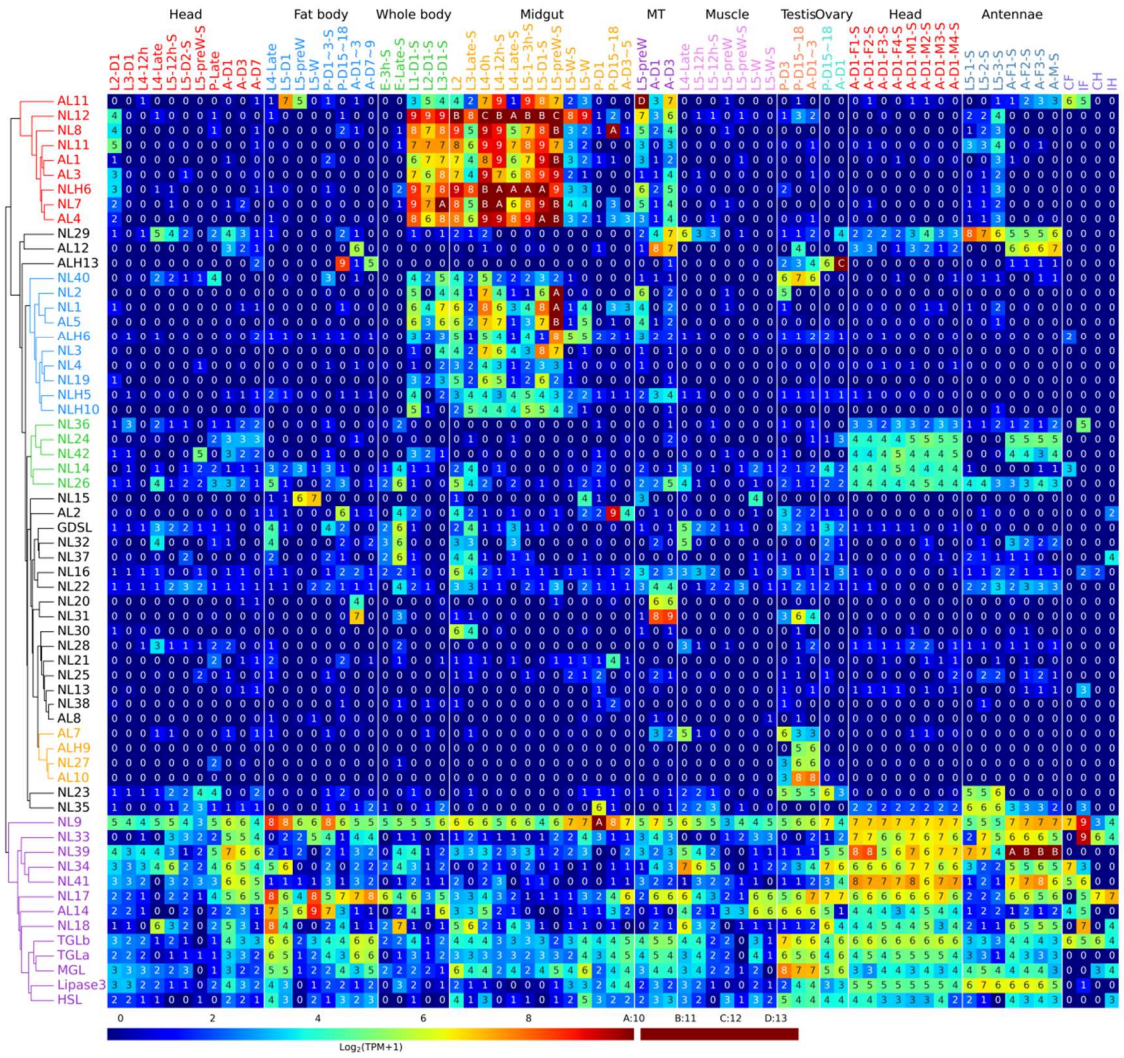


Fig. 3. Expression profiles of the 61 lipase-related genes in various tissues and stages of *M. sexta*. On *top* of the heat map, tissue sources of the 71 RNA-seq datasets are indicated as head, fat body, whole body, midgut, Malpighian tubules, muscle, testes, ovaries, antennae, control fat body (CF) and hemocytes (CH) from naïve larvae, induced fat body (IF) and hemocytes (IH) from immune challenged larvae. The first part of library names indicates major stages of the insect, *i.e.* embryo (E), 1st to 5th instar larvae (L1−L5), pupae (P), and adults (A). In the second part, "D" stands for day, "h" for hour, "preW" for pre-wandering, "W" for wandering, "M" for male, and "F" for female. In the last part, "S" indicates single-end sequencing and no "S" indicates paired-end sequencing. $Log_2(TPM+1)$ values for these lipase-related transcripts are shown in the gradient heat map from *dark blue* (0) to *maroon* (≥10). The values of 0−0.49, 0.50−1.49, 1.50−2.49, … 8.50−9.49, 9.50−10.49, 10.50−11.49, 11.50−12.49 are labeled as 0, 1, 2, … 9, A, B, C, respectively. Relatedness in expression patterns revealed by hierarchical cluster analysis is shown

on the *left*, with abbreviated protein names marked in different colors.

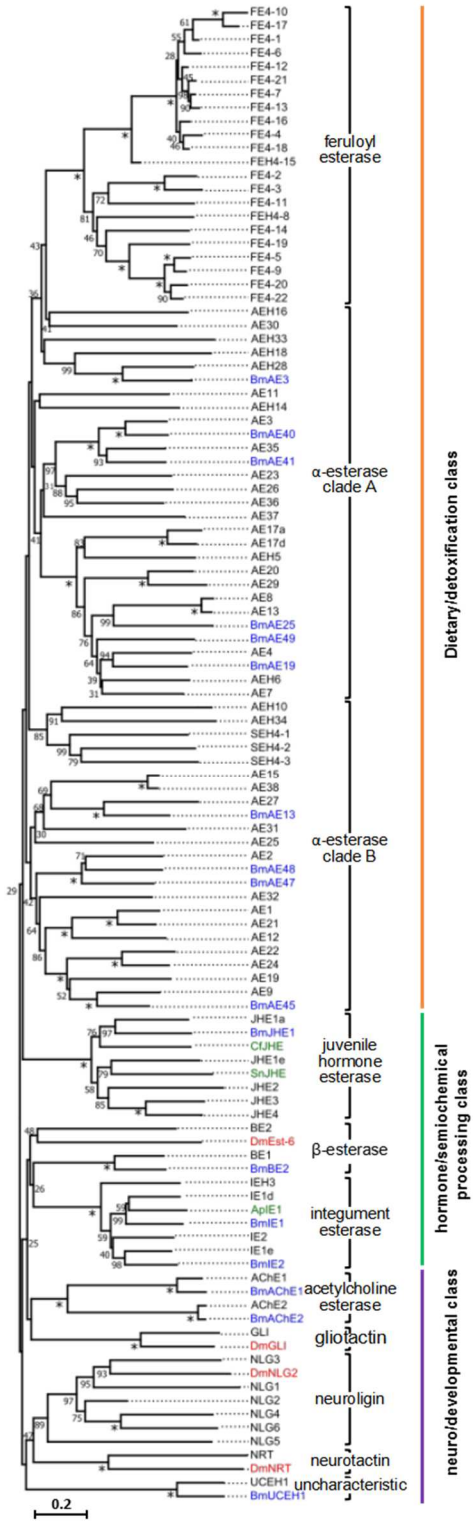


Fig. 4. Phylogenetic trees of 111 carboxylesterase-related proteins in *M. sexta* and other insects.

Entire amino acid sequences of the COEs were aligned with their homologs in *M. sexta* and other insects to construct the neighbor-joining tree. *B. mori* (Bm) α-esterase-3 (AE3, NP_001121786), 13 (NP_001040174), 19 (NP_001116501), 25 (NP_001121784), 40 (NP_001116814), 41 (NP_001124352), 45 (NP_001104822), 47 (ABK27874), 48 (NP_001040466), 49 (NP_001121785), juvenile hormone esterase-1 (JHE1, NP_001037027), integument esterase-1 (IE1, NP_001108339.1), 2 (NP_001121191), nonclassic esterase homolog-1 (UCEH1, NP_001040411), acetylcholinesterase-1 (AChE1, BAF33338), 2 (BAF33337), β-esterase-2 (BE2, NP_001124351), *D. melanogaster* (Dm) esterase-6 (Est-6, CG6917), gliotactin (GLI, CG3903), neuroligin-2 (NLG2, CG13772), neurotactin (NRT, CG9704), *Sesamia nonagrioides* (Sn) JHE (ABW24129), *Choristoneura fumiferana* (Cf) JHE (AAD34172), and *Antheraea polyphemus* (Ap) integument esterase-1 (IE1, AAM14416) are shown as abbreviations (*black* font for *Manduca*, *red* font for *Drosophila*, *blue* font for *Bombyx*, and *green* font for all other species). Bootstrap values greater than 25 are indicated at the nodes.

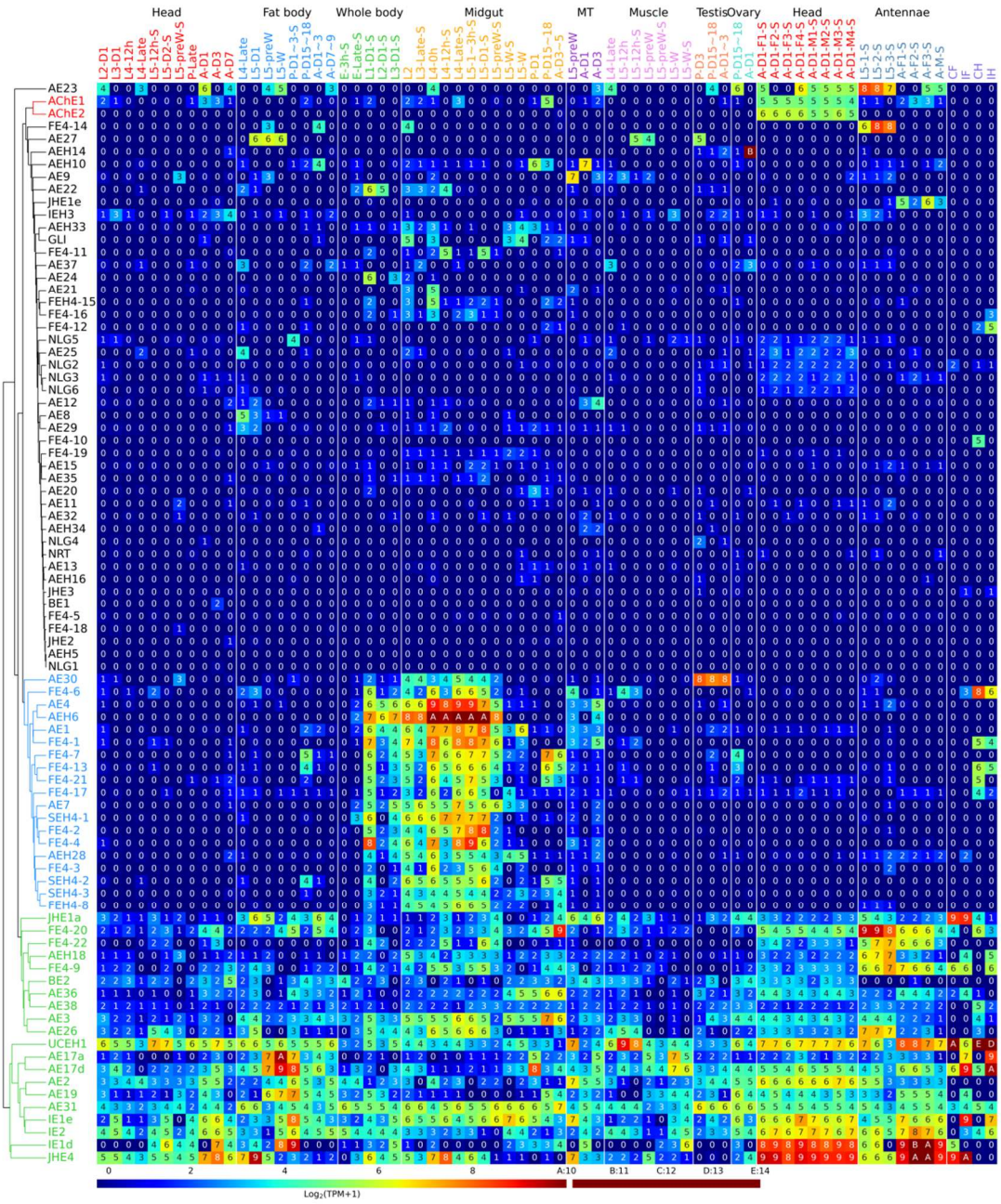


Fig. 5. Expression profiles of the 83 carboxylesterase-related genes in various tissues and stages of *M. sexta*. Tissues and stages of the RNA-seq datasets are described in the legend to Fig. 3.

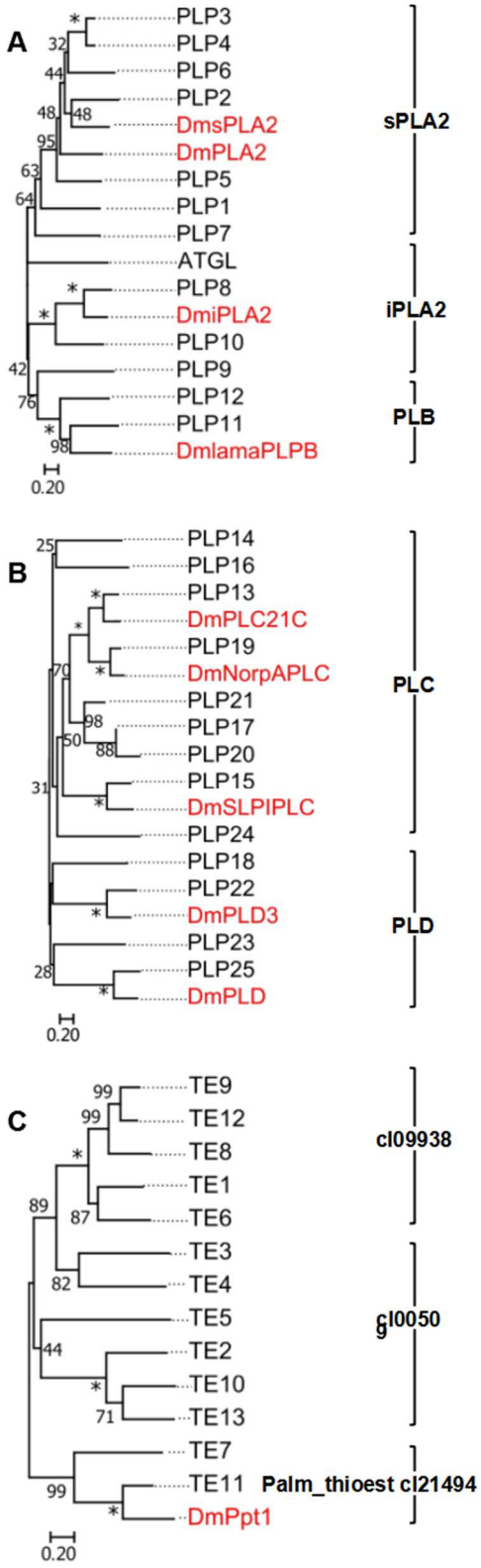


Fig. 6. Phylogenetic trees of the predicted phospholipase As and Bs (**A**), Cs and Ds (**B**), and thioesterases (**C**) in *Manduca* and *Drosophila.* Entire amino acid sequences of the

phospholipases A−D, thioesterases, and their noncatalytic homologs were separately aligned to construct the neighbor-joining trees. *D. melanogaster* (Dm) **s**ecretory PLA2s (CG11124 and CG3009), calcium-**i**ndependent PLA2 (CG6718), **lam**ina **a**ncestor PLB (CG10645), PLC at 21C (CG4574), small wing phosphoinositide PLC (CG4200), no receptor potential A PLC (CG3620), PLD3 (CG9248), PLD (CG12110), and palmitoyl-protein thioesterase-1 (CG12108) are shown as abbreviations (*black* font for *Manduca*, *red* font for *Drosophila*). Bootstrap values greater than 25 are indicated at the nodes.

Fig. S1. A region of the aligned amino acid sequences in 14 *M. sexta* AL(H)s and 4 homologs from *D. melanogaster* and *C. lupus*. *D. melanogaster* CG3653, CG17097, CG18284, *C. lupus* gastric lipase (NP_001003209), and *M. sexta* AL(H)s are aligned. The catalytic residues (S, D and H) are in *red bold* font and the GXSXG consensus sequence around the active site serine are *highlighted yellow*.

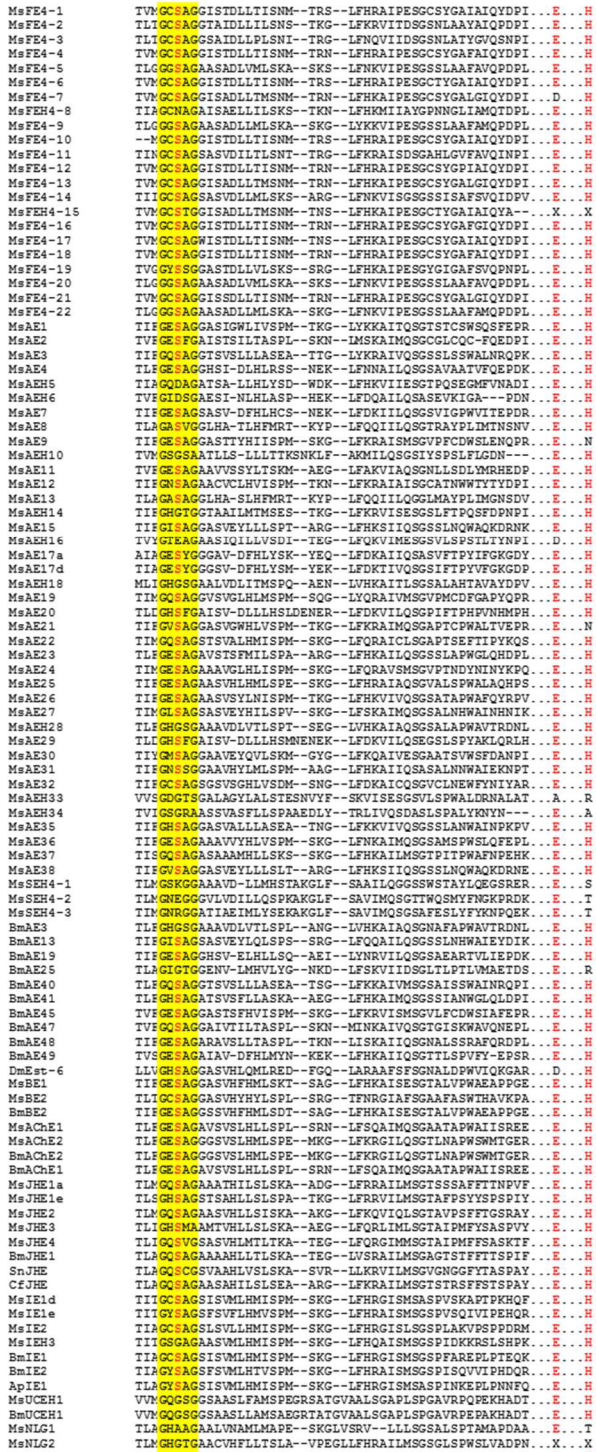


Fig. S2. A region of the aligned amino acid sequences in 111 carboxylesterase-related proteins

from *M. sexta* and other insects. As described in the legend to Fig. 4, COEs and their noncatalytic homologs in *M. sexta*, *B. mori*, *D. melanogaster*, and other species are aligned. The catalytic residues (S, D and H) are in *red bold* font and the GXSXG consensus sequence around the active site serine are *highlighted yellow*.

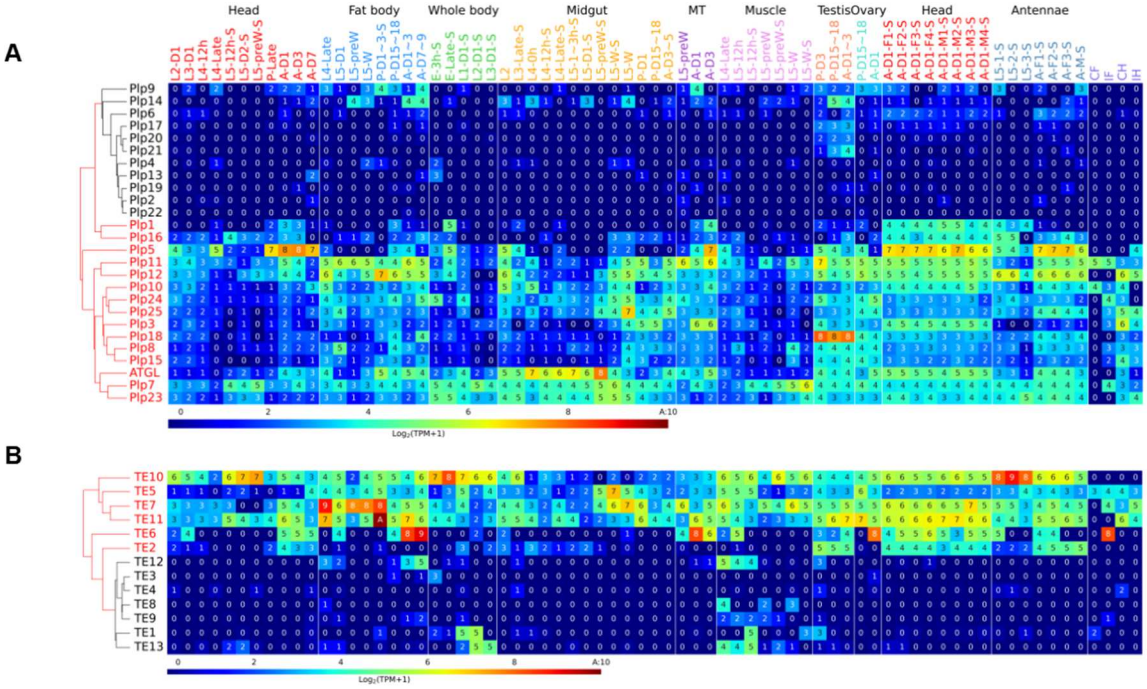


Fig. S3. Expression profiles of the 26 phospholipase- (**A**) and 13 thioesterase-related (**B**) genes in various tissues and life stages of *M. sexta*. Tissues and stages of the RNA-seq datasets are described in the legend to Fig. 3.

References

Arrese, E.L., Rojas-Rivas, B.I., Wells, M.A., 1996. The use of decapitated insects to study lipid mobilization in adult *Manduca sexta*: effects of adipokinetic hormone and trehalose on fat body lipase activity. *Insect Biochem Mol Biol.* 26, 775–782.

Arrese, E.L., Flowers, M.T., Gazard, J.L., Wells, M.A., 1999. Calcium and cAMP are second messengers in the adipokinetic hormone-induced lipolysis of triacylglycerols in *Manduca sexta* fat body. *J Lipid Res.* 40, 556–564.

Arrese, E.L., Howard, A.D., Patel, R.T., Rimoldi, O.J., Soulages, J.L., 2010. Mobilization of lipid stores in *Manduca sexta*: cDNA cloning and developmental expression of fat body triglyceride lipase, TGL. *Insect Biochem Mol Biol.* 40, 91–99.

Arrese, E.L., Soulages, J.L., 2010. Insect fat body: energy, metabolism, and regulation. *Ann Rev Entomol.* 55, 207–225.

Bornscheuer, U.T., 2002. Microbial carboxylesterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev.* 26, 73–81.

Cantu, D.C., Chen, Y., Lemons, M.L., Reilly, P.J., 2011. ThYme: A database for thioester-active enzymes. *Nucleic Acids Res.* 39, 342–346.

Cantu, D.C., Chen, Y., Reilly, P.J., 2010. Thioesterases: a new perspective based on their primary and tertiary structures. *Protein Sci*. 19, 1281–1295.

Cao, X., Jiang, H., 2015. Integrated modeling of protein-coding genes in the *Manduca sexta* genome using RNA-Seq data from the biochemical model insect. *Insect Biochem Mol Biol*. 62, 2–10.

Cao, X., Jiang, H., 2017. An analysis of 67 RNA-seq datasets from various tissues at different stages of a model insect, *Manduca sexta. BMC Genomics* 18, 796

Cao, X., Wang, Y., Rogers, J., Hartson, S., Kanost, M.R., Jiang, H., 2020. Changes in composition and levels of hemolymph proteins during metamorphosis of *Manduca sexta*. *Insect Biochem Mol Biol.* 127, 103489.

Casas-Godoy, L., Duquesne, S., Bordes, F., Sandoval, G., Marty, A., 2018. Lipases: an overview. In: Sandoval, G. (Eds) Lipases and Phospholipases. Methods in Molecular Biology, vol 1835. Human Press, New York, pp. 3–38.

Casida, J.E., Durkin, K.A., 2013. Anticholinesterase insecticide retrospective. *Chem Biol Interact*. 203, 221–225.

Claudianos, C., Ranson, H., Johnson, R.M., Biswas, S., Schuler, M.A., Berenbaum, M.R., Feyereisen, R., Oakeshott, J.G., 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Mol Biol.* 15, 615–636.

Cui, F., Li, M.X., Chang, H.J., Mao, Y., Zhang, H.Y., Lu, L.X., Yan, S.G., Lang, M.L., Liu, L., Qiao, C.L., 2015. Carboxylesterase-mediated insecticide resistance: quantitative increase induces broader metabolic resistance than qualitative change. *Pestic Biochem Physiol*. 121, 88–96.

Filkin, S.Y., Lipkin, A.V., Fedorov, A.N. 2020. Phospholipase superfamily: structure, functions, and biotechnological applications. *Biochemistry* (Moscow) 85, 177–195.

Gershman, A., Romer, T.G., Fan, Y., Razaghi, R., Smith, W.A., Timp, W., 2021. *De novo* genome assembly of the tobacco hornworm (*Manduca sexta*). *G3* (Bethesda). 11, jkaa047.

He, Y., Cao, X., Zhang, S., Rogers, J., Hartson, S., Jiang, H., 2016. Changes in the plasma proteome of *Manduca sexta* larvae in relation to the transcriptome variations after an immune challenge: evidence for high molecular weight immune complex formation. *Mol Cell Proteomics* 15, 1176–1187.

Hinton, A.C., Hammock, B.D., 2001. Purification of juvenile hormone esterase and molecular cloning of the cDNA from *Manduca sexta*. *Insect Biochem Mol Biol.* 32, 57–66.

Holm, C., Kirchgessner, T.G., Svenson, K.L., Fredrikson, G., Nilsson, S., Miller, C.G., Shively, J.E., Heinzmann, C., Sparkes, R.S., Mohandas, T., Lusis, A.J., 1988. Hormone-sensitive lipase: sequence, expression, and chromosomal localization to 19 cent-q13.3. *Science* 241, 1503–1506.

Homberg, U., Hoskins, S.G., Hildebrand, J.G., 1995. Distribution of acetylcholinesterase activity in the deutocerebrum of the sphinx moth *Manduca sexta*. *Cell Tissue Res*. 279, 249–259.

Horne, I., Haritos, V.S., Oakeshott, J.G., 2009. Comparative and functional genomics of lipases in holometabolous insects. *Insect Biochem Mol Biol.* 39, 547–567.

Jones, P., Binns, D., Chang, H-Y, Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S-Y., Lopez R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236−1240.

Kanost, M.R., Arrese, E.L., Cao, X., Chen, Y., Chellapilla, S., Goldsmith, M.R., Grosse-Wilde, E., Heckel, D.G., Herndon, N., Jiang, H., Papanicolaou, A., Qu, J., Soulages,

J.L., Vogel, H., Walters, J., Waterhouse, R.M., Ahn, S., Almeida, F.C., An, C., Aqrawi, P., Bretschneider, A., Bryant, W.B., Bucks, S., Chao, H., Chevignon, G., Christen, J.M., Clarke, D.F., Dittmer, N.T., Ferguson, L.C.F., Garavelou, S., Gordon, K.H.J., Gunaratna, R.T., Han, Y., Hauser, F., He, Y., Heidel-Fischer, H., Hirsh, A., Hu, Y., Jiang, H., Kalra, D., Klinner, C., Konig, C., Kovar, C., Kroll, A.R., Kuwar, S.S., Lee, S.L., Lehman, R., Li, K., Li, Z., Liang, H., Lovelace, S., Lu, Z., Mansfield, J.H., McCulloch, K.J., Mathew, T., Morton, B., Muzny, D.M., McCulloch, K.J., Mathew, T., Morton, B., Muzny, D.M., Neunemann, D., Ongeri, F., Pauchet, Y., Pu, L., Pyrousis, I., Rao, X., Redding, A., Roesel, C., Sanchez-Gracia, A., Schaack, S., Shukla, A., Tetreau, G., Wang, Y., Xiong, G., Traut, W., Walsh, T.K., Worley, K.C., Wu, D., Wu, W., Wu, Y., Zhang, X., Zou, Z., Zucker, H., Briscoe, A.D., Burmester, T., Clem, R.J., Feyereisen, R., Grimmelikhuijzen, C.J.P., Hamodrakas, S.J., Hansson, B.S., Huguet, E., Jermiin, L.S., Lan, Q., Lehman, H.K., Lorenzen, M., Merzendorfer, H., Michalopoulos, I., Morton, D.B., Muthukrishnan, S., Oakeshott, J.G., Palmer, W., Park, Y., Passarelli, A.L., Rozas, J., Schwartz, L.M., Smith, W., Southgate, A., Vilcinskas, A., Vogt, R., Wang, P., Werren, J., Yu, X., Zhou, J., Brown, S.J., Scherer, S.E., Richards, S., Blissard, G.W., 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem Mol Biol*. 76, 118−147.

Kim, Y., Stanley, D., 2021. Eicosanoid signaling in insect immunology: new genes and unresolved issues. *Genes* (Basel) 12, 211

Kim, Y., Ahmed, S., Stanley, D., An, C., 2018. Eicosanoid-mediated immunity in insects. *Dev Comp Immunol*. 83, 130−143.

Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., Chatonnet, A. 2013. ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: Tools to explore diversity of functions. *Nucleic Acids Res*. 41, 423–429.

Lester, D.S., Gilbert, L.I., 1987. Characterization of acetylcholinesterase activity in the larval brain of *Manduca sexta*. *Insect Biochem.* 17, 99–109.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 323.

Miao, Z., Cao, X., Jiang, H. 2020. Digestion-related proteins in the tobacco hornworm,

Manduca sexta. *Insect Biochem Mol Biol.* 126, 103457.

Montella, I.R., Schama, R., Valle, D., 2012. The classification of esterases: an important gene family involved in insecticide resistance – a review. *Mem Inst Oswaldo Cruz*, 107, 437–449.

Oakeshott, J.G., Claudianos, C., Campbell, P.M., Newcomb, R.D., Russell, R., 2010. Biochemical genetics and genomics of insect esterases. In: Gilbert, L.I., Iatrou, K., Gill, S.S. (Eds), *Comprehensive Molecular Insect Science*, Elsevier, Oxford, pp. 1–73.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J. and Sussman, J.L., 1992. The α/β hydrolase fold. *Protein Eng Des Sel*. 5, 197–211.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions *Nat Methods* 8, 785−786.

Pistillo, D., Manzi, A., Tino, A., Boyl, P.P., Graziani, F., Malva, C. 1998. The *Drosophila melanogaster* lipase homologs: a gene family with tissue and developmental specific expression. *J Mol Biol*. 276, 877–885.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. *Nucleic Acids Res.* 40, 290–301.

Qiu, X., Sun, W., Mcdonnell, C.M., Li-Byarlay, H., Steele, L.D., Wu, J., Xie, J., Muir, W.M., Pittendrigh, B.R., 2013. Genome-wide analysis of genes associated with moderate and high DDT resistance in *Drosophila melanogaster*. *Pest Manag Sci.* 69, 930–937.

Rahman, M.Z.A., Salleh, A.B., Rahman, R.N., Rahman, M.B.A., Basri, M., Leow, T.C., 2012. Unlocking the mystery behind the activation phenomenon of T1 lipase: a molecular dynamics simulations approach. *Protein Sci.* 21, 1210–1221.

Shen, Y., Chen, G., Zhao, S., Wu, X., 2021. Genome-wide identification of lipases in silkworm (*Bombyx mori*) and their spatiotemporal expression in larval midgut. *Gene* 146121.

Stengl, M., Homberg, U., Hildebrand, J.G., 1990. Acetylcholinesterase activity in

antennal receptor neurons of the sphinx moth *Manduca sexta*. *Cell Tissue Res*. 262, 245–252.

Soulages, J.L., Firdaus, S.J., Hartson, S., Chen, X., Howard, A.D., Arrese, E.L., 2012. Developmental changes in the protein composition of *Manduca sexta* lipid droplets. *Insect Biochem Mol Biol.* 42, 305–320.

Wei, Y., Schottel, J.L., Derewenda, U., Swenson, L., Patkar, S., Derewenda, Z.S., 1995. A novel variant of the catalytic triad in the *Streptomyces scabies* esterase. *Nat. Struct Biol*. 2, 218–223.

Yu, Q.Y., Lu, C., Li, W., Xiang, Z.H., Zhang, Z., 2009. Annotation and expression of carboxylesterases in the silkworm, *Bombyx mori*. *BMC Genomics* 10, 1–14.

Zhang, S., Gunaratna, R.T., Zhang, X., Najar, F., Wang, Y., Roe, B., Jiang, H., 2011. Pyrosequencing-based expression profiling and identification of differentially regulated genes from *Manduca sexta*, a lepidopteran model insect. *Insect Biochem Mol Biol*. 41, 733–746

VITA

ZELONG MIAO

Candidate for the Degree of

Doctor of Philosophy

Thesis: IMMUNE SIGNALING PATHWAYS AND DIGESTION-RELATED PROTEINS IN THE *MANDUCA SEXTA* LARVAE

Major Field: ENTOMOLOGY

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Entomology at Oklahoma State University, Stillwater, Oklahoma in July, 2022.

Completed the requirements for the Master of Science in Biology at Henan Normal University, Xinxiang, Henan, China in 2018.

Completed the requirements for the Bachelor of Science in Biological Science at Henan Normal University, Xinxiang, Henan, China in 2016.

Experience:
Doctoral Research at Oklahoma State University from 2018 to 2022.
Master and Undergraduate Research at Henan Normal University from 2014 to 2018.

Professional Memberships:
ESA (Entomological Society of America)