A COMPREHENSIVE ANALYSIS OF

THE MANDUCA SEXTA IMMUNOTRANSCRIPTOME

By

RAMESH THARINDU GUNARATNA

Bachelor of Science in Bioinformatics

University of Colombo

Colombo, Sri Lanka

2009

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
May, 2012

A COMPREHENSIVE ANALYSIS OF

THE MANDUCA SEXTA IMMUNOTRANSCRIPTOME

Thesis Approved:

Dr. Haobo Jiang

Thesis Adviser

Dr. Jack Dillwith

Dr. Udaya Desilva

Dr. Sheryl A. Tucker

Dean of The Graduate College

.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

CF, IF, CH and IH	: control (C) and induced (I) fat body (F) and hemocytes (H)
Alk	: anaplastic lymphoma kinase
AMP	: antimicrobial peptide
ANK	: ankyrin
aPKC	: atypical protein kinase C
ARN	: adjusted read number
AtgX	: autophagy-related protein X
BP	: biological process
CC	: cellular component
CRD	: carbohydrate recognition domain
CTL	: C-type lectin
Dscam	: Down syndrome cell adhesion molecule
ECSIT	: evolutionarily conserved intermediate in Toll pathway
EGF, NIM and EMI	: epidermal growth factor, nimrod and EMILIN
EST	: expressed sequence tag
FN	: fibronectin
GO	: gene ontology
HAIP	: hemocyte aggregation inhibitor protein
Hem	: hemipterous
HP	: hemolymph proteinase
IAP	: inhibitor of apoptosis
Ig	: immunoglobulin
IKK	: IκB kinase
IMD	: immune deficiency
IML	: immulectin
JAK-STAT	: Janus kinase-signal transducer and activator of transcription
JNK	: Jun N-terminal kinase
Jra	: Jun-related antigen
LNF	: library normalization factor

LPS	: lipopolysaccharide
LRR	: leucine-rich repeat
MAPK	: mitogen-activated protein kinase
MASK	: multiple ankyrin repeats single KH domain
MEKK	: MAP kinase kinase kinase
MF	: molecular function
MLK	: mixed-linage kinase
NFκB and IκB	: nuclear factor-κB and its inhibitor
NO and NOS	: nitric oxide and its synthase
NRN	: normalized read number
NTF	: nuclear translocation
ORF	: open reading frame
PAP	: proPO-activating proteinase
PDGF and VEGF	: platelet-derived and vascular endothelial growth factors
PG and PGRP	: peptidoglycan and its recognition protein
PIAS	: protein inhibitor of activated STAT
PO and proPO	: phenoloxidase and its precursor
POSH	: plenty of SH3 domains
PPBP	: paralytic peptide-binding protein
PSP	: plasmatocyte spreading peptide
PVR	: PDGF/VEGF receptor
RA	: relative abundance
RISC	: RNA interference silencing complex
ROS	: reactive oxygen species
SH2/3	: src homology 2/3 domain
SOCS	: suppressor of cytokine signaling
SP and SPH	: serine proteinase and its homolog
SUMO	: small ubiquitin-like modifier
TAK	: transforming growth factor β (TGFβ) activated kinase
TEP	: thioester-containing protein
TIR	: Toll/interleukin-1 receptor

TRAF	: tumor necrosis factor (TNF) receptor-associated factor
UBC	: ubiquitin-conjugating domain
VWD	: Von Willebrand disease factor
WAP	: whey acidic protein
ZnF	: zinc finger
α2M	: α2-macrogobulins
βGRP	: β-1, 3-glucanas related protein

CHAPTER I

INTRODUCTION

Insects possess a pristine form of the metazoan antimicrobial defense known as innate immunity (Hultmark, 1993), together with a facet of adaptive immunity via phagocyte-mediated immune memory (Pham *et al*., 2007). However, they lack the luxury of B and T cell-mediated adaptive immunity found in vertebrates (Agaisse, 2007; Pham *et al*., 2007). Insect immunity, comprising humoral and cellular responses, is rapid and effective in identifying and eliminating invading pathogens and parasites (Brey and Hultmark, 1998; Jiang *et al*., 2010; Lemaitre and Hoffmann, 2007). The general process of insect immunity, before deploying killing mechanisms, consists of pathogen recognition via specific binding molecules (Kurata *et al*., 2006; Sansonetti, 2006; Yu *et al*., 2002), signal transduction and modulation via plasma serine proteases and serine protease inhibitors (Gillespie *et al*., 1997; Kanost, 1999; Kanost *et al*., 2001; Marmaras and Lampropoulou, 2009), and receptor-mediated intracellular signaling via Toll (Valanne *et al*., 2011), IMD (Silverman and Maniatis, 2001), JNK (Ramet *et al*., 2002), JAK-STAT (Baeg *et al*., 2005; Hou and Perrimon, 1997; Kisseleva *et al*., 2002), and MAPK (Han *et al*., 1998; Ragab *et al*., 2011) pathways. Signal transduction regulates both humoral and cellular immune responses. The former includes various antimicrobial peptides (AMPs) (Engstrom, 1999; Jiang, 2008), complement-like molecules (Aoun *et al*., 2011), and proteins involved in enzyme cascades that regulate melanin formation (Jang *et al*., 2008; Kanost and Gorman, 2008), which are synthesized

and released into the plasma to entrap and kill invading pathogens or parasites (Gillespie *et al*., 1997; Hoffmann, 2003). In contrast, cellular immunity takes place in hemocytes and is comprised of phagocytosis, nodulation, and encapsulation (Fauvarque and Williams, 2011; Lavine and Strand, 2002; Strand, 2008; Zhuang *et al*., 2005).

Innate immunity plays a role in making insects the most diverse and abundant group of metazoans in the world (Chapman *et al*., 2006; Hultmark, 2003). This makes the immune system worth investigating in its own right. On the other hand, the common ancestry and similarities among insects and mammals make insects excellent model organisms (Hoffmann and Reichhart, 1997; Hultmark, 1993, 2003). These permit discovering evolutionary roots and features of animal immunity (Hoffmann *et al*., 1999; Khush and Lemaitre, 2000; Vilmos and Kurucz, 1998) and allow functional comparisons between diverse metazoan systems to identify shared and unique aspects of innate immunity (Khush and Lemaitre, 2000; Rolff and Reynolds, 2009; Wajant and Scheurich, 2004).

The advent of microarrays and next generation sequencing technologies coupled with bioinformatics tools has generated a large amount of immunotranscriptome data from insects with known genome sequence, such as *Drosophila* sp. (De Gregorio *et al*., 2001; Irving *et al*., 2001; Sackton *et al*., 2007), *Anopheles gambiae* (Christophides *et al*., 2002), *Apis mellifera* (Evans *et al*., 2006), *Aedes aegypti* (Waterhouse *et al*., 2007), *Tribolium castaneum* (Zou *et al*., 2007), *Bombyx mori* (Tanaka *et al*., 2008), and *Acyrthosiphon pisum* (Gerardo *et al*., 2010). Most of the immunotranscriptomic studies so far, for insects without sequenced genomes, lack quantitative levels of transcripts (Altincicek and Vilcinskas, 2007; Vogel *et al*., 2011; Zhang *et al*., 2010). As a member of economically important lepidopterans, *Manduca sexta* has been studied extensively in the field of insect physiology for decades (Jiang *et al*., 2010). Despite its prominent role, the *M. sexta* genome sequence is not yet published. Recently, transcriptomes of fat body, hemocytes, and midgut, in which many immunity-related genes are expressed, were determined using 454

pyrosequencing and Sanger sequencing technology (Pauchet *et al*., 2010; Zhang *et al*., 2011; Zou *et al*., 2008). The quantitative nature of the most recent study allowed us to analyze immune inducible and tissue specific gene expression. Although genome- and homology-independent discovery of new genes is possible, stringent thresholds set in the exploration hindered complete immunotranscriptomic analysis (Zhang *et al*., 2011). Therefore, the current work intended to extend the analysis by identifying most of the immunity-related genes, and their regulation at the level of transcription in *M. sexta*, as a step towards the annotation of its immunogenome.

CHAPTER II

REVIEW OF LITERATURE

Insect immunity

Insects face a myriad of microbes during their lifespan. The hard exoskeleton of insects acts as a protective barrier, primarily to prevent microbes from entering internal structures. When microbes are sufficiently penetrative in breaching the physiochemical defense line they will encounter a pristine form of the metazoan antimicrobial defense known as innate immunity at their disposal (Hultmark, 1993). Even though information on a facet of adaptive immunity in insects via phagocyte-mediated immune memory was recently reported (Pham *et al*., 2007), insects do not have the luxury of B and T cell-mediated adaptive immune mechanism as found in vertebrates (Agaisse, 2007; Pham *et al*., 2007). Innate immune responses in insects are structured into two main branches, known as humoral and cellular responses. Both approaches are rapid and effective in terms of identification and elimination of invading pathogens and parasites.The underlying mechanism, which triggers and deploys these responses, is via two distinctive phases (Brey and Hultmark, 1998; Lemaitre and Hoffmann, 2007).

Pathogen recognition and signaling

First, pathogens are recognized by their intrinsic patterns that are foreign to an insect system, which are known as pathogen associated molecular patterns. The identification is mediated by pathogen recognition proteins/receptors that are found in insect hemolymph as well

as on insect cellular surfaces (Kurata *et al*., 2006; Sansonetti, 2006; Yu *et al*., 2002). β-1,3-glucan recognition proteins (βGRPs) and peptidoglycan recognition proteins (PGRPs) are involved in recognizing fungi and Gram-negative bacteria, respectively, while microbe binding protein, previously known as Gram-negative binding protein could bind to fungi as well as Gram-positive bacteria (Jiang *et al*, 2010).

Secondly, the message of non-self/pathogen identification is relayed to its targets via a pathway that could be divided into the extra- and intra-cellular parts. A cascade of plasma serine proteases carry and modulate/amplify the signal from pattern recognition proteins in the insect hemolymph to their targets, while serine protease inhibitors regulate the extracellular signal transduction (Gillespie *et al*., 1997; Kanost, 1999; Kanost *et al*., 2001; Marmaras and Lampropoulou, 2009). Cell surface receptor-mediated signaling occurs via several intracellular signaling pathways. Among elucidated pathways are: Gram-positive bacteria and fungi induced Toll-receptor mediated pathway (Valanne *et al*., 2011) (Fig. 3a), Gram-negative bacteria induced IMD pathway (Silverman and Maniatis, 2001) (Fig. 3b), antiviral immunity associated JAK-STAT pathway (Baeg *et al*., 2005; Hou and Perrimon, 1997; Kisseleva *et al*., 2002) (Fig. 3d), and MAPK-p38 (Han *et al*., 1998; Ragab *et al*., 2011) and MAPK-JNK (Ramet *et al*., 2002) pathways (Fig. 3c) that are triggered by stress conditions and signal percepted by IMD pathway (Park *et al*., 2004).

The Toll pathway is activated by a cytokine known as Spätzle (LeMosy *et al*., 1999), which is activated a as result of activated serine protease cascade (Jiang *et al*., 2010). Members of the Toll pathway include; MyD88, Pelle, Tube, Cactus, Dorsal and Dif that are activated subsequently to induce antimicrobial peptide expression via nuclear translocation of transcription factors, Dorsal and Dif (Valanne *et al*., 2011). The IMD pathway is activated upon binding of a Gram-negative bacterial component known as DAP-peptidoglycan to PGRP receptors LC and LE (Kaneko *et al*., 2006). Pathway members are comprised of IMD, IAP2, Ubc13, Uev1, TAK1,

Tab2, IKKβ, IKKγ, Dredd, and the transcription factor, Relish. The removal of ankyrin repeats from Relish enables nuclear translocation of the activated Relish to induce AMP expression (Stoven *et al*., 2003). The IMD pathway member TAK1 also activates MKK4, which in turn activates JNK pathway. JNK triggers negative regulation of IMD pathway/AMP expression among other stress related reactions that are yet to be elucidated in insects (Geuking *et al*., 2009; Ragab *et al*., 2011). Along with the JNK pathway, another branch of the MAP kinase pathway, p38 is activated via MKK3 and MeKK1. This leads to another set of stress related reactions together with negative regulation of AMP production (Han *et al*., 1998; Inoue *et al*., 2001).

Execution mechanisms

Early phase of immune responses after an infection activation consists of phagocytosis, nodule formation, or encapsulation. These reactions mainly occur in plasmatocytes and granular cells in lepidopteran larvae (Jiang *et al*., 2010) immediately after pathogen recognition by receptor molecules. Binding of pathogen components to receptors changes usually non-adherent hemocytes to an adherent state, where aggregates of hemocytes/plasmatocytes are formed to entrap pathogens as seen during nodule formation and encapsulation (Lavine and Strand, 2002). Hemocyte specific integrin subunits (Levine *et al*., 2005), tetraspanin, neuroglian (Zhuang *et al*., 2007b), plasmatocyte spreading or paralytic peptide, lacunin, and hemolectin, are mainly involved in postively modulating hemocyte adhesion (Marmaras and Lampropoulou, 2009), while hemocyte aggregation inhibiting protein negatively affect hemocyte adhesion (Kanost *et al*., 1994).

Another distinct physiological observation in insects, immediately after an infection, is the sporadic patterns of melanization inside insects. Melanin entraps pathogens in nodules and capsules of aggregated hemocytes and heals wounds (Cerenius and Soderhall, 2004). Activation of proHP14 via proteins involved in pathogen recognition such as, βGRPs, PGRPs, triggers activation of a serine protease cascade, which comprises HPs, PAPs, and SPHs (Jiang *et al*.,

2010). The triggered cascade leads to the activation of PO, which catalyzes certain steps in a series of enzymatic reactions to produce melanin, which a polymer of eumelanin. Members of this series of enzymatic reactions involve Phenylalanine hydroxylase, tyrosine hydroxylase, Dopa decarboxylase, and PO (Cerenius and Soderhall, 2004). Together with aforementioned, AMPs that are predominantly produced in the fat body play a significant role in reducing a broad spectrum of pathogens and their populations during an infection. High levels of AMPs in the hemolymph are found during infections, and the high expression pattern persists for few days (Jiang *et al*., 2010).

Lepidopteran marks in insect immunity research

Despite their basic defense mechanism, insects dominate the world in number of species (diversity) as well as population (abundance) by a comfortable margin (Chapman *et al*., 2006), which makes the study of their immune system interesting in its own right. The common ancestry and similarities among insects and mammals make insects excellent model organisms to study certain aspects of complex systems (Hoffmann and Reichhart, 1997; Hultmark, 1993, 2003). In addition, these permit discovering evolutionary roots and features of animal immunity (Hoffmann *et al*., 1999; Khush and Lemaitre, 2000; Vilmos and Kurucz, 1998) and allow functional comparisons between diverse metazoan systems to identify shared and unique aspects of innate immunity (Khush and Lemaitre, 2000; Rolff and Reynolds, 2009; Wajant and Scheurich, 2004).

Studies on the immune systems of Diperans such as *Drosophila melanogaster*, *Anopheles gambiae, Aedes eqypti*, Coleopteran *Tribolium castaneum,* and Lepidopterans *Bombyx mori, Helicoverpa sp., Galleria mellonella, Manduca sexta* have contributed a wealth of knowledge and insight into innate immunity and their mechanisms (Rolff and Reynolds, 2009). Among these, the immune system of lepidopteran insects has been considered an important model system especially for cell biology of hemocytes and biochemical analyses of plasma proteins. Lepidopterans are also renowned for their voracious appetite, which categorizes them among economically most

detrimental agricultural pests. Therefore, understanding their immune systems, in terms of components of the system, and how they function, has great implications in development of applications to control their populations (Goldsmith and Marec, 2010; Jiang *et al*., 2010).

Insect innate immunity: genomic and transcriptomic studies

The advent of microarrays and next generation sequencing technologies coupled with bioinformatics tools has generated a large amount of immunotranscriptome data from insects with known genome sequence, such as *Drosophila* sp. (De Gregorio *et al*., 2001; Irving *et al*., 2001; Sackton *et al*., 2007), *Anopheles gambiae* (Christophides *et al*., 2002), *Apis mellifera* (Evans *et al*., 2006), *Aedes aegypti* (Waterhouse *et al*., 2007), *Tribolium castaneum* (Zou *et al*., 2007), *Bombyx mori* (Tanaka *et al*., 2008), and *Acyrthosiphon pisum* (Gerardo *et al*., 2010).

Most of the immunotranscriptomic studies on insects without sequenced genomes lack quantitative levels of transcripts (Altincicek and Vilcinskas, 2007; Vogel *et al*., 2011; Zou *et al*., 2008). Despite its prominent role, the *M. sexta* genome sequence is not yet published. Recently, transcriptomes of fat body, hemocytes, and midgut, in which many immunity-related genes are expressed, were determined using 454 pyrosequencing and Sanger sequencing technology (Pauchet *et al*., 2010; Zhang *et al*., 2011; Zou *et al*., 2008). The quantitative nature of the most recent study allowed analysis of immune inducibility and tissue specificity of gene expression (Zhang *et al*., 2011). However, stringent thresholds that were utilized in the study hindered complete immunotranscriptomic analysis.

CHAPTER III

METHODOLOGY

Sample preparation for the construction of sequence libraries (Zhang *et al*. 2011)

M. sexta eggs were purchased from Carolina Biological Supply. After the eggs were hatched, the larvae were reared on an artificial diet as described by Dunn and Drake (1983). When larvae reached the 2 day of the 5th instar stage of their life cycle, a batch of sixty (60) healthy larvae were selected to be infected. Each larva was injected with a mixture of Gram-positive *Escherichia coli* (2×10^7 cells), Gram-positive *Micrococcus luteus* (20 μg) (Sigma-Aldrich), and a fungal component known as curdlan (20 μg, insoluble β-1,3-glucan from the bacterium, *Alcaligenes faecalis*) (Sigma-Aldrich) in 30 μl H_2O. Total RNA samples were extracted from induced hemocytes (IH) and fat body (IF) 24 h post-infection using TRIZOL Reagent (Life Technologies Inc.). Total RNA samples were extracted from control hemocyte (CH) and fat body (CF) of day 3, 5th instar naïve (non-infected, healthy) larvae (60) using the the above protocol. Polyadenylated RNA was selectively purified from the total RNA samples (1.0 mg each) by binding to oligo(dT) cellulose twice using the Poly(A) Purist^{TM} Kit (Ambion). Random dodecanucleotide primers (100 pmol) were used in conjunction with mRNA (5.0 μg), and SuperScript^{TM} III reverse transcriptase (1000 U, Life Technologies Inc.) to synthesize first strand cDNA. RNase H treatment, second strand synthesis, and gap joining were performed according to the published protocol (Zou *et al*., 2008). This was followed by nebulization of four

cDNA samples to fragment cDNA to facilitate the ligation of double-stranded adaptor A and biotinylated adaptor B after the ends were repaired (Roe, 2004; Margulies *et al*., 2005).

***Pyrosequencing and assembly of reads* (Zhang *et al*. 2011)**

The cDNA with the attachment of adaptor B was isolated on streptavidin-coated magnetic beads, end repaired, and quantified on an Agilent 2100 Bioanalyzer (Agilent Technologies). Afterward, diluted DNA molecules, individually captured by beads, were amplified by emulsion PCR using two primers corresponding to a part of adaptors A and B (Margulies *et al*., 2005). Upon second strand removal and null bead elimination, the sequencing primer identical to another part of adaptor A was annealed to the single-stranded cDNA templates associated with the beads. Two runs (0.5 plate/library, ×4) were performed on a 454 GS-FLX pyrosequencer (Roche Applied Science) using long-read GS-FLX Titanium chemistry. Reads were separately and collectively assembled using Newbler Assembler (Roche Applied Science) resulting in five libraries of sequences: CF, CH, IF, IH, and CIFH.

Homologous sequence search, GO mapping, annotation, and InterProScan search

The 19,020 contigs, assembled from fat body (F) and hemocytes (H) of naïve (C) and microbe-injected (I) into CIFH library of sequences (Zhang *et al*., 2011), were analyzed using the BLAST2GO software (Conesa *et al*., 2005; Gotz *et al*., 2008). In search for homologous sequences, the non-redundant protein database at NCBI was searched using BLASTX (Altschul *et al*., 1990) with a cutoff E-value of 10^{-15}. The BLAST hits were mapped to their corresponding GO annotations using the gene ontology database and several additional data files (Gotz *et al*., 2008). The Gene Ontology (GO) project provides a structured set of defined terms that represent the properties of gene products. These properties span across three major categories: cellular component, which gives the location of the gene product within a cell or outside the cell; molecular function, which gives the activities of a gene product at the molecular level; and

biological process, which shows operations or sets of molecular events a gene product is involved for the functioning of integrated living units such as cells, tissues, organs, and organisms (Myhre *et al*., 2006).

Subsequent annotation of contigs, to link information on cellular component (CC), molecular function (MF), and biological process (BP), was done by applying the annotation rule to all the GO terms. However, certain evidence code weights were changed from their default values to: EXP = IDA = IPI = 5, IMP = IGI = 4, and IEP = 3. Annotations were examined to remove broad or level 1 annotation. Additionally, the GO term known as auxin biosynthesis process was removed from the list of GO terms as the process does not exist in insects. Annex-based GO term augmentation was performed afterwards to, firstly, obtain extra annotations and, secondly, further validate annotations (Gotz *et al*., 2008; Myhre *et al*., 2006). Protein domain and signal peptide were predicted using InterProScan (Quevillon *et al*., 2005), which enabled further sequence annotation (Gotz *et al*., 2008). In order to obtain more refined annotations, level 1 annotation removal and Annex-based GO term augmentation were repeated.

Local BLASTX, domain search, and multiple sequence alignment

We downloaded immunity-related genes from *D. melanogaster* (462 genes from Flybase using the keyword "immunity"), *B. mori* (205 genes from Tanaka *et al*. (2008)), and *A. mellifera* (184 genes from Evans *et al*. (2006)). Amino acid sequences of these genes were incorporated into a sequence database for local BLASTX analysis of the contigs from the four-library assembly (CIFH09). Domain prediction was performed in parallel search runs using batchwise domain search web utilities of web CD-search tool, Pfam and InterProScan. Sequence alignments and manual curation of the alignments were performed using MUSCLE (Edgar, 2004) implemented in MEGA 5 (Tamura *et al*., 2011).

Relative abundance of transcripts under immune challenge

Since each contig was assembled from reads in the four libraries, normalized read numbers (NRNs) are calculated as: actual reads number in library X × (LNF_{CF} + LNF_{IF} + LNF_{CH} + LNF_{IH})/LNF_x, where X is CF, IF, CH, or IH. Library normalization factors (LNFs) for CF (825), CH (3,980), IF (1,618), and IH (3,352) are the sums of read numbers for rpS2-rpS5, rpL4 and rpL8 in the corresponding libraries (Zhang *et al*., 2011). NRNs were used to calculate relative abundance ($RA_{x/y}$ = NRN_x/NRN_y). When a particular reads number was zero, an adjusted reads number ($ARN_{x/y}$= actual read # in library X × LNFy/LNFx) was calculated instead (Zhang *et al*., 2011). When multiple contigs encode a single gene, the particular contigs were concatenated and their NRNs summed in individual libraries for calculating RA or ARN values of the gene.

Statistical analysis

Statistical differences in numbers of contigs or total normalized reads from a GO category were analyzed by Student's t-test. For instance, *p* values of immune inducibility (IC comparison) were derived from sums of number of contigs of IF and IH versus those of CF and CH. IC comparisons were also performed using sums of NRNs of IF and IH versus those of CF and CH. Similarly, *p* values of tissue specificity (FH comparison) were derived from sums of number of contigs and sums of NRNs of CF and IF versus those of CH and IH.

Percentage increases in numbers of contigs from a GO category were calculated as: I/C = [(sum of contig numbers of IF and IH – sum of contig numbers of CF and CH) / lower of the sums of contig numbers of the two groups] × 100 and F/H = [(sum of contig numbers of CF and IF - sum of contig numbers of CH and IH) / lower of the sums of contig numbers of the two groups] × 100. Similar calculations were performed using sums of NRNs from CF, IF, CH, and IH. Percentage increases in numbers of reads from a specific tissue (*e.g.*, fat body) were calculated as I/C_F = [(sum of IF NRNs - sum of CF NRNs) / lower sum of the two NRNs] × 100.

Data generated in above steps were merged, and mining of specific data were performed using SQL scripts in Microsoft Access, and default functions in Microsoft Excel.

Phylogenetic analysis of insect attacins

Amino acid sequences of attacins found in *Antheraea mylitta, Antheraea pernyi, Bombyx mori, Drosophila melanogaster, Hyalophora cecropia, Hyphantria cunea*, and *T. castaneum* were retrieved from the genbank protein database at NCBI. The amino acid sequences were aligned using MUSCLE 3.7 (Edgar, 2004) at www.phylogeny.fr. The multiple sequence alignment was manually edited to improve the quality of the alignment. The phylogenetic tree was constructed using JTT substitution matrix in Protdist+/FastDist+ neighbor with a bootstrap replicates of 1000 (Dereeper, *et al*., 2008, Dereeper, *et al*., 2010).

CHAPTER IV

RESULTS

Distribution of M. sexta immunity-related genes

Here we report the repertoire of and changes in transcripts involved in multiple facets of innate immunity in *M. sexta*, such as pathogen recognition, signal transduction/modulation, and hemocyte adhesion. We identified 129 new immunity-related genes (*i.e*. 204 contigs) in this study, apart from 103 highly regulated genes (*i.e*. 179 contigs) found in the previous study (Zhang *et al*., 2011). Taken together, genes for intracellular signal transduction account for 32% of the entire set; extracellular signaling molecules and their modulators make up 22% (Fig. 1). Gene products for pathogen recognition constitute 16%, whereas highly induced AMPs represent 13% of the total.

Global changes in level 2 GO categories

At GO level 2, expression of immunity-related genes is variable in fat body and hemocytes from naïve (C) and injected (I) larvae in terms of cellular component (CC), molecular function (MF), and biological process (BP). Since total numbers of the identified genes in each category do not significantly change between control and induced fat body (F) or hemocytes (H) (data not shown), we took advantage of the known read numbers for each contig in our datasets (Zhang *et al*., 2011), calculated summation of normalized read numbers (NRNs) for each gene (some

concatenated from two or more contigs), and compared the sums of CF, IF, CH, and IH NRNs in each of level 2 GO categories (Fig. 2). In twelve of the thirty categories, their totals of all NRN sums were lower than 10% of the single highest NRN sum in the respective CC (51074), MF (67960), or BP (65296) group and, therefore, omitted for statistical analysis. Five of the remaining eighteen had significant differences ($t < 0.05$): enzyme regulator activity (MF, $p = 0.001$, IC), molecular transducer activity (MF, $p = 0.008$, FH), cellular component organization (BP, $p = 0.001$, FH), developmental process (BP, $p = 0.024$, FH), and signaling (BP, $p = 0.015$, FH). Differences in the following five groups are less pronounced but worth mentioning, since level 2 GO terms are so general that a higher p value (*e.g.*, 0.05~0.20) may still reflect important changes: cell (CC, $p = 0.107$, FH), extracellular region (CC, $p = 0.143$, IC), macromolecular complex (CC, $p = 0.060$, FH), immune system (BP, $p = 0.145$, IC), and response to stimulus (BP, $p = 0.161$, IC).

We further inspected percentage changes of NRNs in the eighteen level 2 GO categories. When IC and FH comparisons were performed, we observed >50% changes in the following fifteen categories: extracellular region (CC, I > C: 106%), macromolecular complex (CC, H > F: 160%), catalytic activity (MF, I > C: 93%), enzyme regulator activity (MF, I > C: 74%), molecular transducer activity (MF, H > F: 381%), biological adhesion (BP, C > I: 72%, H > F: 89%), biological regulation (BP, H > F: 65%), cell wall organization or biogenesis (BP, I > C, 300%; F > H, 767%), cellular component organization (BP, F > H, 403%), developmental process (BP, F > H: 83%), immune system (BP, I > C: 300%), localization (BP, I > C: 95%), metabolism (BP, I > C: 118%, F > H: 90%), response to stimulus (BP, I > C: 392%, F > H: 59%), and signaling (BP, H > F: 276%).

While differences were observed in more categories between fat body and hemocytes, it is perhaps more interesting from the perspective of immunity to document major increases in total NRNs in either tissue before and after the immune challenge. Therefore, we studied the dataset and detected over 50% changes in extracellular region (CC, F: 155%, H: 59%), binding (MF, F:

95%), catalytic activity (MF, F: 243%), enzyme regulator activity (MF, F: 77%; H: 71%), biological regulation (BP, F: 111%), cell wall organization or biogenesis (BP, F: 297%; H: 329%), cellular process (BP, F: 69%), immune system (BP, F: 756%, H: 99%), localization (BP, F: 265%), metabolism (BP, F: 246%), and response to stimulus (BP, F: 1015%, H: 126%). The most dramatic increases in NRN occurred in the categories of extracellular region (CC, F: 31,038, 155%), catalytic activity (MF, F: 48,149, 243%), and metabolism (BP, F: 46,433, 246%). The increase in extracellular protein transcripts was consistent with the highly induced synthesis of defense molecules (*e.g.*, AMPs) in fat body after the immune challenge.

Pathogen recognition

Pathogen detection is essential in subsequent measures taken to counteract the invasion. In insects, recognition proteins sense the pathogen presence by binding to their surface components known as pathogen-associated molecular patterns. We previously reported highly regulated β-1,3-glucan recognition proteins (βGRPs), peptidoglycan recognition proteins (PGRPs), lectins, hemicentins, leureptin, nimrod B, scavenger receptor C-like protein (Zhang *et al*., 2011). Here we show sixteen new genes for putative pattern recognition proteins: leureptin-2, Dscam, thioester-containing protein (TEP)-1 and -2, galectin-2 and -4, Nimrod A, Draper, PGRP-L2, -L5, -LC, -S5, βGRP-3 and -4, immulecttin-3a and -3b (Table 1).

M. sexta leureptin (IF/CF 2.2) contains thirteen Leu-rich repeats (LRRs) and binds lipopolysaccharide for subsequent phagocytosis and encapsulation (Zhu *et al*., 2010). Leureptin-2, encoded by contig 2194 (IF/CF: 0.7) and 56% similar in sequence to leureptin, contains a signal peptide, conserved N- and C-cap Cys residues, and eleven LRRs, together with a tail rich in Asp and Glu residues.

Dscams constitute a group of closely related proteins with immunoglobulin (Ig) and fibronectin (FN3) domains, whose vast sequence diversity comes from mutually exclusive usage of alternate exons. These molecules may govern the efficiency of phagocytic engulfment of

bacteria by hemocytes (Watson *et al*., 2005). *M. sexta* Dscam is encoded by three contigs: 7892 and 5244 contain Ig and FN3 domains with conserved interdomain connections and cytokine receptor motifs, whereas 9670 codes for a C-terminal region highly similar to its counterpart in *A. mellifera*, *B. mori*, and *D. melanogaster* Dscams.

Secreted TEPs, related to mammalian α_2-macrogobulins (α_2M), may promote phagocytosis of bacteria by plasmatocytes (Blandin and Levashina, 2004). Among the five contigs encoding *Ms*TEPs, two are highly similar to *Bm*TEP1 and three to *Bm*TEP2. Contigs 9741 and 10503 encode α_2M receptor-binding domain; Contigs 8026, 8870, and 12298 encode only N- and C-terminal portions of the domain essential for triggering phagocytosis. Contig 12298 also codes for a portion of the N-terminal region of α_2M domain whereas contig 8870 encodes an α_2M and thioester-containing domains with the canonical thioester motif (GCGEQ) and a functionally important His residue located ~100 residues after the motif (Janssen *et al*., 2005).

β-galactoside-binding galectins play a role in microbial recognition or phagocytosis (Pace and Baum, 2004). *Bm*Galectin-2 and *Ms*Galectin-2 share a highly similar N-terminal region whose function is unclear. In contrast, contig 2140 encodes two carbohydrate recognition domains identical to those in *Bm*Galectin-4 (Tanaka *et al*., 2008).

Proteins in the nimrod superfamily possess characteristic EGF-like repeats known as NIM repeats, with six conserved Cys residues and one CCxGY motif (Kurucz *et al*., 2007; Somogyi *et al*., 2008). Three members of this family, Draper, NimC1 and Eater, are involved in receptor-mediated phagocytosis and microbe binding/recognition (Fauvarque and Williams, 2011). Proteins encoded by contigs 1849 and 10216 closely resemble *Bm*Draper and *Dm*Draper, respectively. The former has an N-terminal region (nearly identical to *Bm*Draper), an EMI domain, a CCxGY motif, and NIM repeats; the latter aligns with *Dm*Draper isoform A and shows a moderate global similarity as well as conserved NIM repeats. Contig 858 encodes an incomplete ORF of *Ms*NimA with its NIM repeats highly similar to *Bm*NimA's.

PGRPs recognize peptidoglycans (PGs) and stimulate prophenoloxidase (proPO) activation system as well as Toll and IMD pathways (Dziarski and Gupta, 2006). Contigs 309, 3259, and 5684 all encode an amidase domain, including two closely spaced Cys residues, PG-binding sites, and Zn-binding residues. Contrarily, contig 151 encodes one of the three Zn-binding residues, seven of the thirteen PG-binding residues, and two of the five catalytic residues in the amidase domain. Contigs 309 and 3259 encode *Ms*PGRP-L5 and -LC that lack a signal peptide, while contigs 151 and 5684 encode secreted proteins highly similar to *Bm*PGRP-L2 and -S2, respectively.

M. sexta βGRP1, 2, and 3 (*i.e*., microbe binding protein) contain a β-1,3-glucan-binding domain and a GH16-like domain (Jiang *et al*., 2010). With the catalytic residues missing in GH16, these proteins do not have a hydrolase activity. Yet they possess an RGD motif that might be recognized by integrin receptors (Ma and Kanost, 2000). *M. sexta* βGRP4, encoded by contig 4114 contains a signal peptide and a GH16 domain with the catalytic residues (Glu, Asp, and Glu) but no RGD motif.

Immulectin (IML)-1, -2, -3, and -4 in *M. sexta* are receptors that belong to a group of C-type lectins (Yu *et al*., 1999; Yu and Kanost, 2000; Yu *et al*., 2006; Yu *et al*., 2005). They have characteristic carbohydrate recognition domains (CRDs) which binds to surface sugars of pathogens to induce phagocytosis, nodulation, encapsulation as well as melanization (Jiang *et al*., 2010). IML-3a and IML-3b are two new variants of IML-3 we identified in *M. sexta*. Both IML-3a and -3b contain two characteristic CRDs in each gene. Contigs 13452 and 16454 encode IML-3a, whereas contigs 6630, 7642, 13397, and 18062 encode IML-3b (IF/CF 2.1).

Extracellular enzymes and their regulation

Many members of the serine proteinase family have been cloned and characterized from *M. sexta*. Some comprise an extracellular enzyme system that leads pathogen recognition to killing mechanisms. These proteinases are sequentially activated and later down-regulated by

inhibitors in the plasma (Jiang *et al*., 2005). We have identified contigs encoding the signal mediators and modulators (Table 2), including nineteen hemolymph proteinases (HPs), scolexin A, a Zn protease, twelve serpins, and two other protease inhibitors, as well as proteins actively involved in melanin synthesis such as punch, Phe hydroxylase, Tyr hydroxylase, dopa decarboxylase, proPOs, proPO-activating proteinases (PAPs), and serine proteinase homologs (SPHs) (Jiang *et al*., 2010; Krishnakumar *et al*., 2000; Zhang *et al*., 2011).

Melanogenesis also plays a role in cellular immunity by killing entrapped pathogens and healing wounds (Cerenius and Soderhall, 2004). In *M. sexta*, the serine proteinase cascade produces active PO that catalyzes the synthesis of quinones and melanin (Kanost *et al*., 2004). HP14 (IF/CF 2.9) initiates the cascade by autoactivation and activating proHP21 (Wang and Jiang, 2007). HP21 then cuts proPAP2 (IF^{ARN} 50.0) and proPAP3 (IF^{ARN} 23.0) (Gorman *et al*., 2007; Wang and Jiang, 2007). The other branch stems from HP6 (IF/CF and IH/CH 1.6) that activates proHP8 (IF/CF 1.9) and proPAP1 (IF/CF 3.3). HP8 converts proSpätzle to Spätzle that activates the Toll pathway (An *et al*., 2010). PAP1, PAP2, or PAP3 cleaves proPO to form active PO in the presence of a complex of SPH1 and SPH2 (Jiang *et al*., 2010; Wang and Jiang, 2008). We identified several SPHs: SPH1a (IF/CF 18.1), SPH1b (IF/CF 1.5), SPH2 (IF/CF 0.9), and SPH4 (IF/CF 26.9).

In *M. sexta*, five serpins are known to regulate the HP cascade at multiple steps (Jiang *et al*., 2010). Different variants of serpin-1 (A, E, and J) inhibit PAP2 and HP8 (Ragan *et al*., 2010). The Serpin-1 transcript level is high (NRN >7000) in fat body and does not change after immune challenge. Serpin-3 has a 7.5-fold increase in mRNA levels in fat body, suggesting an enhanced control of the PAPs. Serpin-4 (IF/CF 4.0) inhibits HP6 and HP21; serpin-5 further attenuates HP6 (Tong *et al*., 2005). Serpin-6 (IF/CF 2.0, IH/CH 2.8) contributes to the negative regulation of PAP3 (Zou and Jiang, 2005).

PO is one member of the enzyme system for melanin synthesis – Phe hydroxylase (IF/CF 2.4) converts Phe to Tyr which is further *o*-hydroxylated to become dopa by Tyr hydoxylase or

PO (Cerenius and Soderhall, 2004). The Tyr hydroxylase needs a co-factor (tetrahydrobiopterin), whose synthesis is mediated by a GTP cyclohydrolase I named Punch (IF/CF 1.5) (Thony *et al*., 2000). Dopa decarboxylase (DDC, IF^{ARN} 106.6) converts dopa to dopamine; PO oxidizes dopamine to dopamine quinone that polymerizes to form eumelanin (Nappi and Christensen, 2005).

In addition to the aforementioned HPs and their inhibitors, we found: HP2, HP3 (IH/CH 0.5), HP4 (IH/CH 0.8), HP5 (IH/CH 5.8), HP12 (IF/CF 1.3), HP13 (IH/CH 1.0), HP15 (IH/CF 6.5), HP20 (IF/CF 4.4), serpin-7 (IF/CF 2.0) and serpin-11. Functions of these proteins are not elucidated to date.

Signal transduction via major signaling pathways

The Toll, IMD, JAK-STAT, JNK, and p38 signal transduction pathways (Table 3) govern the production of effector molecules to eliminate pathogens and, hence, have been in the limelight of insect innate immunity research (Boutros *et al*., 2002; Dostert *et al*., 2005; Han *et al*., 1998; Kallio *et al*., 2005; Kim and Kim, 2005). We have identified orthologs of the pathway members and assume their functions and modes of action are conserved among insects.

a. Toll pathway

Components of Gram-positive bacteria and fungi activate the Toll pathway (Lemaitre *et al*., 1996), which is triggered by the binding of a cytokine named Spätzle (LeMosy *et al*., 1999) (Fig. 3A). Contig 2287 (IH/CH 4.0) (Table 3) encodes Spätzle-1B that was characterized in *M. sexta* (An *et al*., 2010) – its Cys^{106} is essential for the receptor binding. Toll receptor contig (IF/CF 2.2, IH/CH 1.5) shows high similarities to portions of the C-terminal region of a Toll receptor gene found in *M. sexta* (Ao *et al*., 2008). Each contig encodes a TIR domain that Toll receptors utilize to relay intracellular signals (Takeda and Akira, 2001). A complex of MyD88, Tube and Pelle may form upon binding of Spätzle to Toll receptors (Weber *et al*., 2003). Contig

864 (IF/CF & IH/CH 1.5) encodes *Ms*MyD88 that has a TIR domain for interacting with Toll's TIR domain and a death domain for interacting with Tube. Contig 1313 encodes *M. sexta* Tube that has a death domain and a protein kinase domain. The death domain may interact with both MyD88 and Pelle, a kinase that triggers the complex activation by auto-phosphorylation (Moncrieffe *et al*., 2008). Contig 2038 encodes Pelle that has two death domains and a protein kinase domain. Activation of the above complex leads to Cactus phosphorylation, ubiquitination, and degradation in the proteasome. A homolog of *Bm*Cactus in *M. sexta* (IF/CF 10.0, IH/CH 1.7) has two sets of three ANK repeats. Phosphorylation of Cactus sets free Dorsal or Dif that moves into the nucleus and transcriptionally activates AMP genes. Contig 2384 (IF/CF 1.3, IH/CH 1.2) shares a nearly identical N-terminus with *Ms*Dorsal (Genbank accession: ADK39025) but has a very long and different C-terminus.

Sumoylation plays a regulatory role in innate immunity by posttranslational covalent modification of proteins in the NF-κB signaling pathways (Mabb and Miyamoto, 2007). Contig 4591 encodes a ubiquitin-conjugating domain (UBC) protein similar to Lesswright or Ubc9, a SUMO-conjugating enzyme that stabilizes Cactus (Table 3) (Abraham, 2007; Huang *et al*., 2005). In contrast, Smt3 (IF/CF & IH/CH 1.6) may activate Dorsal by sumoylation (Bhaskar *et al*., 2002; Xu *et al*., 2010). Uba2 and Aos1 together activate Smt3 (Bhaskar *et al*., 2000). Contigs 890 (IH/CH 1.5) and 5438 (IH/CH 1.7) are related to Uba2 and Aos1, respectively, and encode the characteristic Uba2-SUMO and Aos1-SUMO domains.

Apart from the aforementioned components, Pellino, Tollip-d, Tollip-v, TRAF, atypical PKC (aPKC), Ref(2)P, and ECSIT are associated with the Toll pathway as well (Valanne *et al*., 2011). Contig 292 (IF/CF 2.5, IH/CH 1.3) encodes a protein highly similar to *Bm*Pellino (Table 3). *Drosophila* homolog of aPKC is encoded by contigs 5708 and 7433, each covering a protein kinase domain followed by PKC C-terminal domain. Contig 5971 encodes PBI and ZZ-type zinc finger domains that are highly similar to *Dm*Ref(2)P-PB. *Ms*Tollip-d includes a CUE domain; *Ms*Tollip-v has a C2 domain before the CUE domain. ECSIT is proposed to relay signal from

TRAF6 to MEKK1 (Kopp *et al*., 1999; Xiao and Ghosh, 2005) and contig 4177 encodes a part of *M. sexta* ECSIT.

b. IMD pathway

Gram-negative bacterial components, diaminopimelic acid-PGs, activate the IMD pathway via PGRP-LC, -LE, and IMD in *Drosophila* (Choe *et al*., 2002; Gottar *et al*., 2002; Kaneko *et al*., 2006) (Fig. 3B). Contig 2368 (IF/CF 2.2) partially encodes a *Ms*IMD that contains a death domain to interact with the death domain in FADD (Table 3) (Naitza *et al*., 2002). *Ms*FADD has a death effector domain in addition to the death domain. Recruitment of *Ms*Dredd may activate down-stream signaling via the effector domain (Leulier *et al*., 2000). Contig 1615 partially encodes *Ms*Dredd (IF/CF 2.0, IH/CH 1.3) with a caspase domain, whereas contigs 14535 and 15028 cover the *Ms*Dredd's N-terminal portions with moderate similarity. Dredd's caspase activity cleaves IMD and exposes its inhibitor of apoptosis (IAP)-binding motif (Shi, 2002).

The cleaved IMD may then bind to *Ms*IAP2, which has an N-terminus nearly identical to *Bm*IAP2's and a BIR (baculoviral IAP repeat) or IAP domain (Paquette *et al*., 2010) (Table 3). IMD, ubiquitinated by Uev1 and Ubc13, may act as a scaffold to bind downstream kinases TAK1 and IKK. *Ms*TAK1 contains a protein tyrosine kinase domain (Ganesan *et al*., 2011; Paquette *et al*., 2010). *Ms*Tab2 has a CUE domain and may form a complex with *Ms*TAK1 to further activate an IKK signaling complex consisting of IKKβ and IKKγ (Iwai and Tokunaga, 2009; Kanayama *et al*., 2004; Zhuang *et al*., 2006). *Ms*IKKγ has a protein kinase domain at the C-terminus, highly similar to *Bm*IKKγ's. Dredd, upon phosphorylation of Relish by the IKK complex, may cut off Relish ANK repeats and cause it to translocate into the nucleus (Stoven *et al*., 2003). We found three contigs coding for different regions of Relish. Contig 4802 (IF/CF 6.4, IH/CH 1.9) contains a Rel homology domain in the N-terminus. Contigs 15531 and 15532 cover a 90-residue region next to the N-terminus of Relish-2A (IH/CH 1.2). This region contains nuclear localization signal (KKRK) and a PEST domain followed by a Ser-rich region.

The transcription factor Serpent activates AMP gene expression by binding to cognate *cis* regulatory elements (Petersen *et al*., 1999). Contigs 4249 and 17496 encode a ZnF-GATA domain highly similar to that in *Dm*Serpent (Table 3). Among regulators of IMD pathway signaling, Sickie, Caspar, and POSH are identified. Contigs 5128 and 7157 encode a homolog of *Dm*Sickie that positively regulates Dredd-mediated Relish cleavage (Foley and O'Farrell, 2004). Contig 2428 (IH/CH 1.9) encodes Caspar. POSH governs the IMD pathway activation and termination, as well as JNK pathway activation via regulation of TAK1 degradation (Lee and Ferrandon, 2011). *Ms*POSH has two SH3 domains. Mutations in nuclear translocation factor-2 (Ntf-2) prevent Dorsal, Dif, or Relish from translocation to the nucleus (Bhattacharya and Steward, 2002). Contigs 6033 and 8947 encode *Ms*Ntf-2 (IH/CH 2.0) with a characteristic NTF2 domain.

c. MAPK-JNK-p38 pathway

In *Drosophila*, components of the Ras/MAPK pathway (Fig. 3D) activate JNK and p38, down regulate the IMD pathway, and induce lamellocyte formation as well as hemocyte proliferation (Dong *et al*., 2002; Lee and Ferrandon, 2011; Ragab *et al*., 2011). PDGF/VEGF receptor (PVR) and Alk receptor trigger the MAPK pathway and Rac1 activation (Zettervall *et al*., 2004). Contig 222 (IH/CH 1.3) encodes *Ms*PVR and contig 8785 encodes *Ms*Alk (Table 7). Binding of an unknown ligand may trigger a series of phosphorylation events through Ras85D as well as Rac1. Ras85D (IF/CF 0.9, IH/CH 1.5) contains a Ras domain. Contig 6185 encodes another member of the kinase cascade, Dsor1. Homologs of *Dm*Licrone/MKK3 and *Dm*MEKK1 activate p38 that contains a protein kinase domain (Han *et al*., 1998; Inoue *et al*., 2001).

JNK, a branch of the MAP kinase pathway, mediates stress-related responses and controls AMP gene expression (Ragab *et al*., 2011). In *Drosophila*, TAK1, Rac1, mixed-lineage kinases (MLKs), or MKK4 could initiate the JNK pathway (Gallo and Johnson, 2002; Park *et al*., 2004; Silverman *et al*., 2003; Williams *et al*., 2006). MLK1 (IF/CF 1.4, IH/CH 1.7) is a mitogen-activated protein kinase kinase kinase (MAP3K) with PK, PTK, and SH3 domains. Contig 3605

encodes *Ms*Rac1, whereas contig 3655 encodes *Ms*MKK4 with a PK-Mkk4 domain (Table 3). Activation of Hemipterous (Hem/MKK7) triggers the JNK pathway (Geuking *et al*., 2009) and *Ms*Hem, partly encoded by contig 4608, contains a PKc-Mkk7 domain. Contig 3082 encodes a part of *Ms*JNK that has a protein kinase domain. JNK then activates both FOS and Jun-related antigen (Jra) that are transcription factors (Sluss *et al*., 1996). FOS (contig 4904: IH/CH 1.9) and Jra (IF/CF 1.5, IH/CH 1.0) both contain bZIP regions important for their interaction with DNA. The JNK pathway also activates the transcription factor, Aop (Anterior open), that mediates lamellocyte formation in *Drosophila* (Zettervall *et al*., 2004) and contig 1136 codes for *Ms*Aop. Both Cdc42 and multiple ankyrin repeats single KH domain (MASK) are involved in relaying signals in the above processes (Bokoch, 2005; Hall, 1998; Kleino *et al*., 2005). The Cdc42 homolog is encoded by contig 647 while contigs 225 and 4036 encode the MASK homolog.

d. JAK-STAT pathway

Binding of Unpaired (upd), a cytokine, to its receptor Domeless initiates the JAK-STAT pathway (Arbouzova and Zeidler, 2006) (Fig. 3D),but we did not find upd homolog in our dataset. Instead, we identified two contigs coding for a homolog of *Bm*Domeless in *M. sexta*: contig 7557 encodes a highly similar N-terminus and contig 9588 has a partial FN3 domain (Table 3). Upd-bound Domeless phosphorylates receptor-associated JAK/Hopscotch (contig 20), which then activates the transcription factor STAT. STAT, encoded by contigs 2221 and 14109, contains a partial protein-interacting domain, a partial DNA-binding domain, and an SH2 domain (Baeg *et al*., 2005). Protein inhibitor of activated STAT (PIAS) and suppressor of cytokine signaling (SOCS) negatively regulate the JAK-STAT pathway (Agaisse and Perrimon, 2004). Contig 602 encodes *Ms*PIAS that has a MIZ/SP RING. *Ms*SOCS contains an N-terminal region highly similar to *Bm*SOCS, an SH2 domain, and an identical SOCS box.

Hemocyte adhesion

During an infection, usually non-adherent hemocytes tend to aggregate to trigger cellular

immune responses against invading pathogens (Lavine and Strand, 2002). We identified three hemocyte-specific integrin α subunits (α1, α2, and α3), two β subunits (349 and 1850-13553-18269), and an integrin-related protein (Table 4). They play key roles in transforming hemocytes to an adherent state (Levin *et al*., 2005). Two other cell surface molecules, neuroglian (IF/CF 0.6, IH/CH 1.7) and tetraspanin (IF/CF 1.8, IH/CH 1.4), also contribute to integrin-mediated aggregation of hemocytes (Nardi *et al*., 2006; Zhuang *et al*., 2007a; Zhuang *et al*., 2007b). Plasmatocyte spreading peptide (PSP) precursor (IF/CF 2.6), three paralytic peptide-binding proteins (PPBP-1: IF/CF 1.5, IH/CH 1.1; PPBP-2: IH/CH 0.7; PPBP-3: IF/CF 1.1, IH/CH 2.4), thrombospondin (IF/CF 0.8), lacunin (IF/CF 0.1, IH/CH 1.0), and hemolectin (IF/CF 0.1, IH/CH 0.4) may also modulate hemocyte adhesion (Jiang *et al*., 2010; Marmaras and Lampropoulou, 2009). Hemocyte aggregation inhibiting protein (HAIP) (IF/CF 1.0) negatively regulates excessive hemocyte aggregation (Kanost *et al*., 1994).

Autophagy

Autophagy governs the lysosome-dependent turnover of proteins or organelles and plays key roles in other cellular processes as well as human diseases (Shintani and Klionsky, 2004). Among nine different autophagy-related (Atg) molecules found are: two ubiquitin-like proteins, Atg8 (IF/CF 3.1, IH/CH 0.6) and Atg12, E1-like Atg5 (IF/CF 0.2), and a Cys proteinase Atg4 (IF/CF 0.3), and Atg4-like proteins (IF/CF 0.3) (Table 5). These proteins are implicated in the process of macroautophagy (Geng and Klionsky, 2008).

Antimicrobial peptides

We previously found 25 unique AMPs encoded by 61 highly up-regulated contigs (Zhang *et al*., 2011). Despite near complete coverage in the previous study, we identified other antimicrobial molecules, namely *M. sexta* lysozyme-like protein 1 (LLP1), attacin-2 through -6, and four additional proteins containing a WAP domain (Table 6). LLP1 (IF/CF 1.5, IH/CH 1.2) contains a LYZ1 domain. Attacin-1 (IF^{ARN} 178.5), -2 (IF^{ARN} 178.5, IH/CH 161.7), -3 (IF^{ARN}

70.4), -4 (IF/CF 125.9), -5 (IF^{ARN} 479.5), and -6 (IH^{ARN} 59.4) are highly induced particularly in the fat body.

A closer look at the multiple sequence alignment of attacin-coding contigs revealed that the attacin family of AMPs comprise six members as opposed to two reported previously in *M. sexta* (Table 6, Fig. 4). There is a cluster of two attacin genes in the *B. mori* genome, closely similar to *Ms*Attacin-2. Similar gene duplications gave rise to 2-3 attacin genes in other lepidopteran insects. In *M. sexta*, a different gene expansion yielded five other genes (*Ms*Attacin-1, -3, -4, -5, and -6) in a lineage-specific way. A monophylatic group of four *D. melanogaster* attacin genes as well as *T. castaneum* attacin-1, -2 and -3 was probably generated in a similar way.

Others

This category comprises other genes involved in signaling, hemocyte proliferation and development, reactive molecular species synthesis and regulation, and gene silencing (Table 7). Focal adhesion kinase mediates signals from integrin receptors to MAPK pathway and, hence, plays a central role in regulating cellular immunity (Sieg *et al*., 1999). It has characteristic functional domains, such as FERM-M, PTK, and Focal-AT. Hematopoiesis produces circulating cells in the hemocoel, which are involved in the cellular immune response against pathogen invasion (Lavine and Strand, 2002). The ligand Serrate and its receptor Notch mediate signal transduction to control hematopoiesis (Williams, 2007). Contigs 49 and 6 encode a part of *Ms*Serrate and *Ms*Notch, respectively. A transcription factor (*Ms*Brahma) may also control hematopoiesis (Remillieux-Leschelle *et al*., 2002). Cell migration is important in detecting pathogen invasion and involves microtubule reorganization and actin cytoskeleton regulation (Marmaras and Lampropoulou, 2009). Both Cdc42 and MASK are involved in relaying signals in the above processes (Bokoch, 2005; Hall, 1998; Kleino *et al*., 2005).

ROS play a part in cytotoxic defense against microbes via activating AMPs or enhancing melanogenesis (Lavine and Strand, 2002). Nitric oxide synthase (NOS) generates nitric oxide (NO) (Nappi *et al*., 2000). Contigs 6991, 11623 and 11493 cover different portions of *Ms*NOS in the given order. These contigs contain the characteristic NO synthase, flavodoxin1 and FAD-binding domains. Thioredoxin peroxidases and peroxiredoxins regulate amount of ROSs, especially after an oxidative burst in the case of an infection, to maintain cellular homeostasis (Christensen *et al*., 2005; Nappi and Christensen, 2005). We found three thioredoxin peroxidases, each containing a PRX-Typ2cys domain, and a PRX-1cys containing peroxiredoxin. Contigs 6299, 8926 (IH/CH 1.3), and 2937 (IF/CF 2.1, IH/CH 1.1) encode thioredoxins while contig 8403 encodes the peroxiredoxin.

Homology-based gene silencing is involved in the *Drosophila* antiviral response (Wang *et al*., 2006). We identified homologs of Argonaute-1 and Dicer-2 that compose a part of the RNA Interference Silencing Complex (RISC) (Ding *et al*., 2004; Wang *et al*., 2006). Contigs 412 and 824 encode *Ms*Dicer-2 (IH/CH 1.5).

CHAPTER V

DISCUSSION AND CONCLUDING REMARKS

This extended study of the quantitative transcriptome data unveiled 105 new immunity-related genes in *M. sexta*. Along with 127 reported previously (Jiang *et al*., 2010; Zhang *et al*., 2011), the number of such genes summed up to 232. In comparison to the 205, 462 and 184 genes retrieved from *B. mori*, *D. melanogaster* and *A. mellifera*, our analysis, not based on annotated immunogenome, revealed a similar number of genes. The deep analysis of fat body and hemocyte transcriptomes did uncover a large portion of the complete set of immunity-related genes that would come from genome analysis. This is a valuable piece of information for researchers doing similar transcriptome studies in organisms that lack sequenced genomes. The extensiveness and depth of our transcriptome data are further supported by the discovery and analysis of six attacin genes in *M. sexta* (Fig. 4).

A major goal of this research was not limited to discover sequences similar to the queries; rather it was to identify genes most likely related to immunity. For instance, BLAST search using aPKC as a query revealed 34 contigs at a commonly used cutoff E-value of 10^{-5}, but there is only one ortholog (contigs 5708 and 74333) in *M. sexta*. Twenty-eight of the hits were identified because they encode a kinase domain commonly found in genes, which may not be related to immunity. As such, many studies yielded inflated lists of homologous genes with limited value in orthology-based function predictions. Contrary to that practice, we took measures to reduce false

positives, such as adopting a stringent threshold in the initial BLASTX analysis, searching for domain structures, and placing more weight on experimentally proven ontology in the GO annotation. Our initial BLASTX search (E-value $<10^{-15}$) against NCBI NR database resulted in 411 hits, dominated by lectins (80), proPO subunits (32), attacins (25), serine proteinases (30), and serine proteinase inhibitors (37). The parallel, local BLASTX analysis using known immunity-related genes from *B. mori, A. mellifera* and *D. melanogaster*, along with domain searches and multiple sequence alignments, yielded 383 highly scrutinized contigs. Although the number difference was only 18, the second list overlapped with the first only in 197 cases. Over 50% or 214 of the positives in the first list were incorrect: the use of a stringent threshold did not greatly reduce false positives; it, instead, yielded a lower number of valuable hits. Therefore, we adopted the 2nd list and improved it by merging 383 contigs into 232 groups, each of which represents one or more contigs putatively encoded by a single gene (Tables 1-7).

Based on the categorization of immune functions, we found genes for signal transduction and modulation account for 54% or 179 of the 232 genes whose products form pathways which crosstalk in multiple steps (Fig. 3). Genes for pathogen recognition and execution account for 16% and 10% of the gene set and, unlike signaling proteins, their products exert similar functions by extensively complementing each other to cope with a broad spectrum of infectious agents. The remaining 20% are involved in other processes, such as cell adhesion and autophagy. While this function classification provided a good overview of the immune system, general GO analysis at level 2 did not yield clear differences in gene counts in the I-C and F-H comparisons (data not shown). Only after we took mRNA levels into consideration, could significant differences be observed in certain categories of CC, MF, and BP at GO level 2 (Fig. 2). Six of the thirty groups are significantly different ($p < 0.20$) between fat body and hemocytes, whereas four categories are in the I-C comparison. Considering the high level of generalization in GO terms at level 2, we believe $p < 0.20$ is remarkable, especially when a large percentage of increase or decrease (>50%) is observed. The most dramatic changes occur in the categories of extracellular region (31,038,

CC), catalytic activity (48,149, MF), and metabolism (46,433, BP) in fat body after the immune challenge. Highly induced expression of AMPs and other plasma defense proteins is partly responsible for the increase in total mRNA levels of extracellular molecules. Further analysis is needed to test if the enhancements in catalytic activity and metabolism are correlated.

Phagocytosis, nodulation, and encapsulation as a result of pathogen recognition and cell adhesion comprise insect cellular immunity. Except for slightly up-regulated Nim A, Dscam and Draper, up-regulated leureptin and IML-3, and down-regulated galectin-4, other genes showed no significant change at the transcription level (Table 1), suggesting a complex regulation of phagocytosis. A majority of the genes involved in autophagy showed low transcript levels (Table 5). Atg-8, which plays a critical role in autophagy, had a 3.1-fold mRNA level increase in fat body, whereas Atg-3, -4, and -5 transcripts reduced to 1/3 in the same tissue after the immune challenge. Total NRNs of the Atg genes were 4.4-fold higher in fat body than hemocytes. Since active engulfment of microbes occurs in the latter, the regulation of autophagy seems complicated, like phygocytosis, nodulation, and encapsulation. Hemolectin had a 2.5-fold down-regulation of mRNA level in hemocytes. In contrast, the increases in Reeler (IF^{ARN} 97.9; IH/CH: 3.0) and proPSP (IF/IH: 2.6) transcripts may enhance nodulation.

Melanogenesis plays a key role in immunity by participating in killing of entrapped microbes and wound healing. In *M. sexta*, an extracellular network of serine proteinases generates active PO, PSP, and Spätzle by proteolytic processing (Jiang *et al*., 2011). Many HP-related proteins were up-regulated after the immune challenges, including HP14, HP6, PAPs, and SPH1 (Table 2). Serpin-3, 4, 5, and 6, whose mRNAs became more abundant, negatively regulate some of these HPs that activate proPO. PO catalyzes the key steps for quinone and melanin formation, whereas other enzymes (*e.g*. Phe and Tyr hydroxylases, Punch, and DDC) also contribute to melanization reactions. Substantial increases in their transcript levels (Table 2) further indicate the enzyme system for melanogenesis is highly coordinated and regulated at that level. In addition to PAP1 activation, *M. sexta* HP6 generates active HP8 that processes Spätzle precursor

(IF/CF 3.6, IH/CH 4.0) to initiate the Toll pathway (An *et al*., 2009 and 2010) for AMP induction. The increases in HP6 (IF/CF 1.6) and HP8 (IF/CF 1.9) mRNA levels indicate that RT-PCR is less quantitative than deep sequencing in detecting < 2-fold induction.

Massively parallel pyrosequencing of transcripts from larval fat body and hemocytes allowed us to identify most components of the intracellular signaling pathways and quantify changes in their transcript levels after the immune challenges (Fig. 3). Although Dif, TRAF, PGRP-LE, and Wengen are missing in our contigs, evidence for the existence of Toll, IMD, MAPK-JNK-p38, and JAK-STAT pathways is compelling. We plan to search the genome sequence for these genes and profile their expression in the two tissues in the future. The current data, however, did not show dramatic mRNA level changes except for Tube (IF/CF: 8.4), Cactus (IF/CF: 10.0), and Relish (IF/CF: 7.0). The highly induced production of Cactus and Relish are probably related to the fact they are cleaved during immune signaling and need to be replenished for a secondary response. Excluding these three genes, the averages and ranges of induction for Toll, IMD, MAPK-JNK-p38, and JAK-STAT pathways are 1.8 (1.3-2.2), 1.2 (0.3-2.0), 1.2 (0.8-1.5), and 0.6 (0.2-0.9) in fat body, as well as 1.4 (0.8-2.2), 1.1 (0.2-2.1), 1.1 (0-1.9), and 0.8 (0.5-1.1) in hemocytes, respectively. The small increases in the first three pathways appear to be sufficient for substantial induction of AMP synthesis (Table 6), whereas the low level of JAK-STAT suppression could be related to the fact that we did not use any elicitor to mimic viral infection. It would be interesting to compare effects of virulent and incompatible viruses on transcription of the genes in the antiviral signaling pathway.

The next-generation sequencing approach we adopted has yielded a set of 19,020 contigs and corresponding read numbers from control and induced larval fat body and hemocytes of *M. sexta.* The long average size (923 bp), known immunity-related genes of other insects, and extensive sequence comparisons have facilitated the identification of 232 genes (or 383 contigs), assign-ment of GO terms and immune processes, and examination of transcriptional regulation of the entire system. The results validated our previous study, uncovered genes (*e.g*., components of

signaling pathways), demonstrated the practicality of genome-independent expression profiling of a complex process, and paved the way for annotation of the immunogenome.

Table 1. A list of genes related to pathogen recognition*

Gene name	Domain(s)	Contig(s)	nCF	nCH	nIF	nIH	nIF:nCF	nIH:nCH
CTL10	Lectin-C	11458, 14515, 14516, 15639, 17942	675.4	0.0	1788.3	0.0	*2.7*	0/0
Draper	EMI, NIM repeats	1849, 10216	0.0	83.5	18.1	142.9	*1.5*	1.7
Dscam	FN-3, Ig	5244, 7892, 9670	0.0	19.6	24.2	35.0	*2.0*	1.8
Galectin-2	β-propeller repeats	7714	0.0	12.3	0.0	8.7	0/0	0.7
Galectin-4	GLECT	2140	11.8	51.6	24.2	20.4	2.0	0.4
Hemicentin 1	Ig	131, 465, 14278	1729.9	186.7	2319.9	699.9	1.3	3.8
Hemicentin 2	Ig	4353	47.4	267.7	163.1	1545.6	3.4	5.8
Hemolin	Ig	3442	11.8	29.5	8868.8	116.6	748.5	4.0
IML-2		4775	11.8	0.0	537.7	0.0	45.4	0/0
IML-3	Lectin-C	1097, 13163, 14125	1670.6	800.7	2688.4	673.6	1.6	0.8
IML-3a	Lectin-C	13452, 16454	94.8	1078.2	0.0	985.7	0.0	0.9
IML-3b	Lectin-C	6630, 7642, 13397, 18062	165.9	2149.0	350.4	860.3	2.1	0.4
IML-4		4808	0.0	0.0	2573.6	5.8	*217.2*	*2.4*
Lectin		6497	651.7	8397.2	30.2	3893.1	0.0	0.5
Lectin prec.	VWD	14570	545.0	4990.7	6.0	2143.4	0.0	0.4
Leureptin	LRR	4012, 8453, 15857	1978.7	9.8	4392.1	58.3	2.2	5.9
Leureptin2	LRR	2194	177.7	4.9	126.9	0.0	0.7	0.0
NimA	NIM repeats	858	35.5	152.3	0.0	265.4	0.0	1.7
NimB		8820	11.8	90.9	18.1	81.7	1.5	0.9
PGRP-1		13190, 14104	343.6	0.0	1752.0	64.2	5.1	*26.1*
PGRP-2		14700	0.0	0.0	1105.6	5.8	*93.3*	*2.4*
PGRP-3	Amidase 2	14752	0.0	0.0	712.9	5.8	*60.2*	*2.4*
PGRP-D	Amidase 2	575	35.5	0.0	1564.7	14.6	44.0	*5.9*
PGRP-L2	Amidase 2	151	47.4	147.4	24.2	195.4	0.5	1.3
PGRP-L5	Amidase 2	309	201.4	265.3	175.2	268.3	0.9	1.0
PGRP-LC	Amidase 2	3259	0.0	39.3	0.0	23.3	0/0	0.6
PGRP-S1	Amidase 2	11845	0.0	4.9	54.4	49.6	*4.6*	10.1
PGRP-S5	Amidase 2	8467	0.0	0.0	682.7	0.0	*57.6*	0/0
PGRP-SA	Amidase 2	5684	0.0	2.5	96.7	11.7	*8.2*	4.8
SR-C-like	Somatome-din-B, Sushi (SCR repeat), MAM	5933, 8686, 13271, 15116, 15350, 15564	462.1	4573.1	48.3	4068.1	0.1	0.9
TEP1	α2M	9741, 10503, 12298	0.0	9.8	0.0	23.3	0/0	2.4
TEP2	α2M	8026, 8870	23.7	29.5	6.0	37.9	0.3	1.3
βGRP1	GH16-βGRP	2979	580.6	0.0	555.8	11.7	1.0	*4.8*
βGRP2	GH16-βGRP	1326	11.8	22.1	114.8	204.1	9.7	9.2
βGRP3		14744, 14786	308.1	0.0	296.0	0.0	1.0	0/0
βGRP4	GH16-βGRP	4114	213.3	4.9	271.9	17.5	1.3	3.6
MBP		8247	319.9	4.9	737.1	52.5	2.3	10.7

* Genes reported by Zhang *et al*. (2011) are *underlined*. nCF, nCH, nIF and nIH are normalized read numbers and, for genes with two or more contigs, they represent the total values. When nCF = 0, adjusted NRN for nIF (*italics*) is calculated as nIF × 825/9775; when nCH = 0, adjusted NRN for nIH (*italics*) is calculated as nIH × 3980/9775.

Table 2. Hemolymph proteases, serpins, and other extracellular signaling molecules (see footnote of Table 1)

Gene name	Domain(s)	Contig(s)	nCF	nCH	nIF	nIH	nIF:nCF	nIH:nCH
HP1	Trypsin	8524, 12527, 16264, 16288, 16719, 17102, 18182	876.8	8709.1	42.3	5849.8	0.1	0.7
HP2	Trypsin	6333	35.5	24.6	18.1	40.8	0.5	1.7
HP3	Trypsin	5756	0.0	137.5	0.0	72.9	0/0	0.5
HP4	Trypsin	5654	23.7	167.0	6.0	134.1	0.3	0.8
HP5	CLIP, trypsin	2203	0.0	36.8	30.2	212.9	2.6	5.8
HP6	Trypsin	540	284.4	260.3	447.1	405.3	1.6	1.6
HP7	Trypsin	3018, 3762	11.8	159.6	175.2	256.6	14.8	1.6
HP8	CLIP, trypsin	5370, 9086	379.2	4.9	712.9	29.2	1.9	5.9
HP9	Trypsin	3989	0.0	2.5	48.3	70.0	*4.1*	28.5
HP12	Trypsin	11373	23.7	2.5	30.2	0.0	1.3	0.0
HP13	Trypsin	1094	35.5	191.6	12.1	189.6	0.3	1.0
HP14	LDLa, CCP, trypsin	3606, 3785	154.0	14.7	447.1	29.2	2.9	2.0
HP15	Trypsin	4433	0.0	4.9	18.1	32.1	*1.5*	6.5
HP17s	Trypsin	5186, 14177	0.0	0.0	78.5	55.4	*6.6*	*22.6*
HP18	Trypsin	8231	0.0	2.5	0.0	99.1	0/0	40.4
HP19	Trypsin	3199, 6539, 14093	35.5	39.3	235.6	81.7	6.6	2.1
HP20	Trypsin	4242	59.2	0.0	259.8	2.9	4.4	*1.2*
HP21	CLIP, trypsin	2439	319.9	12.3	247.7	37.9	0.8	3.1
HP22	Trypsin	2361	82.9	9.8	422.9	2.9	5.1	0.3
SPH-1a	Trypsin	6149, 14393	272.5	63.9	4941.9	125.4	18.1	2.0
SPH-1b	Trypsin	2813	1279.6	22.1	1891.0	210.0	1.5	9.5
SPH2	Trypsin	2843	876.8	19.6	809.6	8.7	0.9	0.5
SPH4	Trypsin	2985	35.5	0.0	954.5	0.0	26.9	0/0
Serpin1	Serpin	7639	7713.4	0.0	7473.2	40.8	1.0	*16.6*
Serpin2	Serpin	5255, 5821, 14248, 14456, 15111, 15910 16917, 17048, 17058, 17751, 18441	106.6	790.8	302.1	9827.5	2.8	12.4
Serpin3	Serpin	2693	248.8	17.2	1872.8	55.4	7.5	3.2
Serpin4	Serpin	4422	556.9	334.0	2235.3	615.3	4.0	1.8
Serpin5		5831, 13453, 13454	94.8	31.9	1039.1	122.5	11.0	3.8
Serpin6	Serpin	1706	82.9	83.5	163.1	236.2	2.0	2.8
Serpin7	Serpin	8071, 8076	379.2	7.4	773.3	8.7	2.0	1.2
Serpin11	Serpin	6359	0.0	4.9	36.2	32.1	*3.1*	6.5
Serpin12	Serpin	3776, 6215, 6531, 17814	2417.1	19.6	5799.8	148.7	2.4	7.6
Serpin13	Serpin	2184	628.0	4.9	670.6	2.9	1.1	0.6
Serpin22	Serpin	3224	1161.2	0.0	2881.8	0.0	2.5	0/0
PI6		8286	0.0	0.0	839.8	67.1	*70.9*	*27.3*
PI-like		3674, 10722	118.5	7.4	646.4	70.0	5.5	9.5
Trypsin inh. B	Kunitz-BPTI	13936	0.0	0.0	151.0	0.0	*12.8*	0/0
Kazal-type PI	Kazal-1, -2	5197	0.0	0.0	120.8	2.9	*10.2*	*1.2*
Punch	GTP-cyclohydro1	1029	237.0	0.0	350.4	23.3	1.5	*9.5*
Phe hydroxylase	ACT	509	864.9	41.8	2036.0	20.4	2.4	0.5
Tyr hydroxylase	Biopterin-H	2023	11.8	0.0	199.4	20.4	16.8	*8.3*
DDC	DOPA-deC	940	0.0	0.0	1262.7	20.4	*106.6*	*8.3*
PAP1	CLIP, trypsin	3070	94.8	49.1	314.2	61.2	3.3	1.3
PAP2	CLIP, trypsin	1667	0.0	17.2	592.1	96.2	*50.0*	5.6
PAP3	CLIP, trypsin	1818	0.0	63.9	271.9	192.5	*23.0*	3.0
ProPO-p1		17085	616.1	5690.6	24.2	4858.3	0.0	0.9
ProPO-p2		17958	296.2	3249.3	18.1	2741.2	0.1	0.8
Zn protease	Peptidase-M14	915	0.0	51.6	126.9	75.8	10.7	1.5
Scolexin	Trypsin	10791, 10792, 16520, 18669, 18670, 18963	23.7	0.0	16650.2	5.8	702.6	*2.4*
Hdd1-like		2382	0.0	4.9	658.5	201.2	*55.6*	41.0
Hdd13-like		5606	11.8	0.0	114.8	11.7	9.7	*4.8*
Hdd23-like		6581	0.0	0.0	78.5	29.2	*6.6*	*11.9*

Table 3. Members of the intracellular signaling pathways (see footnote of Table 1)

Gene name	Domain(s)	Contig(s)	nCF	nCH	nIF	nIH	nIF:nCF	nIH:nCH
		Toll pathway						
Spätzle		2287	11.8	68.8	42.3	274.1	3.6	4.0
Toll		5599, 6893, 14282	106.64	164.6	235.6	247.9	2.2	1.5
MyD88	Death, TIR	864	118.5	137.5	181.2	201.2	1.5	1.5
Tube	Death, PK	1313	23.7	127.7	199.4	102.1	8.4	0.8
Pelle	Death, PK	2038	0.0	17.2	60.4	37.9	*5.1*	2.2
Pellino		292	71.1	147.4	175.2	189.6	2.5	1.3
Cactus	ANK-2	1044, 3381, 15574	118.5	260.3	1190.2	452.0	10.0	1.7
Lesswright	UBC	4591	11.8	167.0	54.4	198.3	4.6	1.2
aPKC	PK, PK C-domain	5708, 7433	0.0	39.3	0.0	37.9	0/0	1.0
Ref(2)P	PBI, ZZ Zinc finger	5971	23.7	78.6	24.2	64.2	1.0	0.8
Rel/Dorsal	Rel domain	2384	82.9	198.9	108.7	242.0	1.3	1.2
ECSIT	ECSIT	4177	0.0	14.7	6.0	20.4	*0.5*	1.4
Tollip-d	CUE	4949	71.1	0.0	72.5	0.0	1.0	0/0
Tollip-v	C2, CUE	731	59.2	311.9	48.3	344.1	0.8	1.1
Smt3	Sumo	7946	59.2	167.0	96.7	262.5	1.6	1.6
Aos1	Aos1-SUMO	5438	35.5	56.5	24.2	93.3	0.7	1.7
Uba2	Uba2-SUMO	890	11.8	140.0	30.2	210.0	2.6	1.5
Misshapen	PK, CNH	289	106.6	311.9	181.2	259.5	1.7	0.8
		IMD pathway						
IMD	Death	2368	35.5	61.4	78.5	67.1	2.2	1.1
FADD	Death, DID	342	876.8	122.8	531.6	157.5	0.6	1.3
Dredd	Caspase (Pepti- dase-C14)	1615, 14535, 15028	71.1	149.8	139.0	201.2	2.0	1.3
IAP2	BIR/Inhibitor of apoptosis	1174, 7327, 8290, 9234	106.6	98.2	30.2	110.8	0.3	1.1
Ubc13/ben	UBC	2901	82.9	117.9	96.7	247.9	1.2	2.1
Uev1A	UBC	3326	154.0	368.4	217.5	411.2	1.4	1.1
TAK1	-PTK	8422	0.0	4.9	6.0	17.5	*0.5*	3.6
Tab2	CUE	1637	0.0	44.2	30.2	43.7	*2.6*	1.0
IKKβ	PK	5609	11.8	14.7	6.0	2.9	0.5	0.2
IKKγ		1049	11.8	93.3	30.2	107.9	2.6	1.2
Relish-2A	PEST	15531, 15532	11.8	61.4	96.7	72.9	8.2	1.2
Relish-2B	Rel homology domain	4802	23.7	103.2	151.0	192.5	6.4	1.9
Ntf2	NTF2	6033, 8947	118.5	169.5	157.1	332.4	1.3	2.0
Serpent	ZnF-GATA	4249, 17496	485.8	582.1	163.1	501.6	0.3	0.9
Sickie		5128, 7157	35.5	2.5	18.1	64.2	0.5	26.1
Caspar	UBX	2428	11.8	27.0	36.2	52.5	3.1	1.9
POSH	SH3	1777, 5429	47.4	338.9	66.5	309.1	1.4	0.9
		MAPK pathway with JNK and p38 branches						
Eiger	TNF	1020	497.6	9.8	380.6	78.7	0.8	8.0
Cdc42	Cdc42	647	130.3	579.6	114.8	621.1	0.9	1.1
Dsor1	PK	6185	11.8	34.4	0.0	29.2	0.0	0.9
Rac1	Rac1	3605	0.0	27.0	30.2	35.0	*2.6*	1.3
Ras85D	Ras	73, 132, 205, 1185	651.7	1350.8	555.8	1991.7	0.9	1.5
MLK1	PK,PTK, SH3	1825, 1841	59.2	117.9	84.6	198.3	1.4	1.7
MEKK1	PK	1947	11.8	49.1	0.0	20.4	0.0	0.4

Licrone/MKK3	PK-MKK3-6	2351	82.9	108.1	114.8	122.5	1.4	1.1
p38	PK	7214	23.7	63.9	42.3	93.3	1.8	1.5
MKK4	PK-MKK4	3655	35.5	63.9	30.2	29.2	0.9	0.5
Hem	PK-MKK7	4608	0.0	24.6	6.0	23.3	*0.5*	1.0
JNK	PK	3082	23.7	7.4	24.2	40.8	1.0	5.5
FOS	bZIP	4904	23.7	95.8	48.3	177.9	2.0	1.9
Jra	bZIP	13290, 13291	71.1	186.7	108.7	186.6	1.5	1.0
Aop	SAM-PNT, ETS	1136	0.0	125.3	36.2	154.6	*3.1*	1.2
MASK	ANK, KH-I	225, 4036	106.6	206.3	157.1	265.4	1.5	1.3
		JAK-STAT pathway						
Domeless	SH2	7557, 9588	11.8	27.0	30.2	17.5	2.6	0.7
JAK/Hopscotch	PTK, SH2, B41	20	272.5	498.6	241.7	250.8	0.9	0.5
STAT	SH2, Protein Interacting , STAT_bind	2221, 14109	106.6	135.1	24.2	102.1	0.2	0.8
PIAS	MIZ/SP RING	602	23.7	113.0	42.3	119.6	1.8	1.1
SOCS	SH2, SOCS box	1187, 6886	11.8	159.6	30.2	125.4	2.6	0.8
Stam	VHS	13543	0.0	19.6	6.0	20.4	*0.5*	1.0

Table 4. Hemocyte adhesion-related genes (see footnote of Table 1)

Gene name	Domain(s)	Contig(s)	nCF	nCH	nIF	nIH	nIF: nCF	nIH: nCH
HAIP	GH-18	2947	6137.5	51.6	5926.6	163.3	1.0	3.2
Hemolectin	C8, EGF, F5_F8-type-C, VWD,	11280, 13813, 14711, 14760, 15506, 15594, 18551	2772.5	34035.7	126.9	13458.1	0.1	0.4
Hemocyte-specific integrin α1	FG-GAP, integrin-α	3997, 4957, 4966	94.8	677.9	36.2	939.0	0.4	1.4
Hemocyte-specific integrin α2	Integrin- α 2	5800, 5860, 8294, 9106	142.2	0.0	175.2	8.7	1.2	*3.6*
Hemocyte-specific integrin α3	Integrin- α, FG-GAP	771	71.1	110.5	84.6	113.7	1.2	1.0
Plasmatocyte-specific integrin β1	Integrin-β, EGF-2	1850, 13553, 18269	118.5	975.0	84.6	1463.9	0.7	1.5
Integrin β1	Integrin-β, -β tail, -β cyt	349	343.6	412.6	350.4	662.0	1.0	1.6
Integrin related-1	Integrin-β-cyt	461	11.8	117.9	84.6	314.9	7.1	2.7
Integrin-linked protein kinase 2	ANK, PTK	2401	71.1	135.1	24.2	204.1	0.3	1.5
Lacunin	Ig, ADAM-spacer1, Kunitz-BPTI, PLAC	15, 2717, 15269, 15273	1410.0	12488.9	120.8	12335.4	0.1	1.0
Laminin	Laminin-G2	4	639.8	4452.8	132.9	4578.4	0.2	1.0
Neuroglian	I-set, FN3	163	165.9	365.9	96.7	609.5	0.6	1.7
Tetraspanin		92, 3559, 4331, 4687, 5512, 6843, 7644, 17721	509.5	1761.0	930.4	2502.1	1.8	1.4
Reeler1/Hdd11	Reeler	3778	0.0	27.0	1160.0	81.7	*97.9*	3.0
Paralytic peptide BP1	Lipoprotein_11	3375, 7873	296.2	852.2	447.1	974.0	1.5	1.1
Paralytic peptide BP2	Lipoprotein_11	3648, 14107, 15055	35.55	331.6	0.0	236.2	0.0	0.7
Paralytic peptide BP3	Lipoprotein_11	4042, 5938, 15696, 16051, 16074, 16096, 16613, 16615	2085.3	2512.5	2344.1	5922.7	1.1	2.4
uENF/PSP	GBP-PSP	2651	284.4	0.0	749.1	0.0	2.6	0/0
Thrombospondin	EGF-CA, TSP-C & -3	535	793.8	2.5	604.1	0.0	0.8	0.0

Table 5. Autophagy-related genes (see footnote of Table 1)

Gene name	Domain(s)	Contig(s)	nCF	nCH	nIF	nIH	nIF:nCF	nIH:nCH
Atg2	Chorein-N	15384	71.1	0.0	30.2	0.0	0.4	0/0
Atg2 like	Apt1	8264	94.8	2.5	42.3	0.0	0.5	0.0
Atg3	Autophagy-N & -C, BUD22, act-C	392	414.7	144.9	145.0	140.0	0.4	1.0
Atg4	Peptidase-C54	4329	213.3	2.5	60.4	5.8	0.3	2.4
Atg4-like	Peptidase-C54	11947	94.8	4.9	24.2	0.0	0.3	0.0
Atg5	APG5	5386, 5513	343.6	7.4	78.5	17.5	0.2	2.4
Atg6	APG6	12191, 14634	201.4	0.0	48.3	2.9	0.2	*1.2*
Atg8	Atg8	7383	59.2	105.6	181.2	64.2	3.1	0.6
Atg12	APG12	12607	106.6	2.5	12.1	2.9	0.1	1.2

Table 6. Genes of AMPs and other effector proteins (see footnote of Table 1)

Gene name	Domain(s)	Contig(s)	nCF	nCH	nIF	nIH	nIF:nCF	nIH:nCH
Attacin-1	Attacin N&C	8902, 14343	0.0	0.0	2114.5	61.2	*178.5*	*24.9*
Attacin-2	Attacin N&C	11040, 11711, 17135	0.0	49.1	1443.9	7940.7	*121.9*	161.7
Attacin-3	Attacin N&C	6782, 16576, 17705	0.0	0.0	833.7	102.1	*70.4*	*41.6*
Attacin-4	Attacin N&C	7203, 14641, 18324	23.7	7.4	2984.5	64.2	125.9	8.7
Attacin-5	Attacin N&C	13563, 17350	0.0	0.0	5207.7	0.0	*439.5*	0/0
Attacin-6	Attacin C	15159, 15744	0.0	0.0	0.0	145.8	0/0	*59.4*
Cecropin-6	Cecropin	14997	0.0	0.0	205.4	29.2	17.3	11.9
Cecropin B precursor		13894	0.0	0.0	290.0	84.6	*24.5*	*34.4*
Cecropin-like peptide B-5		12151, 15041	0.0	0.0	1141.8	0.0	*96.4*	0/0
Gallerimycin		10234	0.0	2.5	1504.3	20.4	*127.0*	8.3
Gloverin		2067	11.8	0.0	1691.6	239.1	142.8	*97.4*
Putative PI	TIL	16018	0.0	0.0	241.7	35.0	*20.4*	*14.3*
Immune related protein		15998, 17184, 18819	0.0	39.3	821.6	2519.6	*69.4*	64.1
IMPI		3142	11.8	17.2	2537.4	352.9	214.2	20.5
Lebocin-A precursor		13916, 17301, 17434	35.5	0.0	8017.0	0.0	225.5	0/0
Lebocin-B precursor		10853	0.0	0.0	682.7	2.9	*57.6*	*1.2*
Lebocin C precursor		4903	0.0	0.0	1685.6	17.5	*142.3*	*7.1*
Lebocin-D precursor		7116	11.8	9.8	5449.4	8.7	459.9	0.9
Lysozyme-like protein 1	LYZ1	1285	284.4	113.0	428.9	140.0	1.5	1.2
Lysozyme	LYZ	8421, 15931, 16133	1078.2	235.8	19640.7	2633.3	18.2	11.2
Moricin		9484	11.8	0.0	809.6	163.3	68.3	*66.5*
Moricin-like		17439	0.0	0.0	592.1	90.4	*50.0*	*36.8*
Possible AMP	Toxin_2	3746, 14568, 16292, 18150, 18699	0.0	19.6	507.5	1816.8	*42.8*	92.5
Secreted peptide 30		6597	710.9	0.0	1208.3	0.0	1.7	0/0
Salivary Cys-rich peptide	WAP	4175	0.0	17.2	241.7	131.2	*20.4*	7.6
Peptidase inhibitor precursor	WAP	12848	0.00	39.3	0.0	2.9	0/0	0.1
Peptidase inhibitor-like	WAP	14536	71.09	0.0	48.3	0.0	0.7	0/0
Antileukoproteinase precursor	WAP	15064	0.00	0.0	90.6	0.0	*7.7*	0/0
Putative WAP-2 isoform 1	WAP	13827	0.00	0.0	42.3	0.0	*3.6*	0/0
Transferrin	Transferrin	2145, 11027, 14937, 16606, 17206, 18239, 18308	1196.7	36.8	8542.6	277.0	7.1	7.5

Table 7. Other putative immunity-related genes (see footnote of Table 1)

Gene name	Domain(s)	Contig(s)	nCF	nCH	nIF	nIH	nIF:nCF	nIH:nCH
FAK	FERM-M, PK-Tyr, Focal-AT	82	11.8	365.9	42.3	355.8	3.6	1.0
Alk	SH3	8785	0.0	12.3	0.0	0.0	0/0	0.2
Pvr	Ig, PTK	222	94.8	385.6	90.6	501.6	1.0	1.3
Serrate	EGF-Lam, LDLa	49	7109.1	830.1	4428.4	793.2	0.6	1.0
Notch	ANK, EGF-CA	6	35.5	808.0	24.2	842.8	0.7	1.0
Dicer-2	ResIII, Helicase C, dsRNA binding, PAZ	412, 824	71.1	275.1	90.6	402.4	1.3	1.5
Argonaute 1	PAZ-argonaute-like	10180	0.0	7.4	6.0	17.5	*0.5*	2.4
NOS	NO synthase, flavodoxin1	6991, 11493, 11623	0.0	22.1	48.3	29.2	*1.4*	1.3
Peroxiredoxin	PRX-1cys	8403	11.8	14.7	18.1	17.5	1.5	1.2
Thioredoxin POD1	PRX-Typ2cys	6299	35.5	31.9	30.2	52.5	0.8	1.6
Thioredoxin POD2	PRX-Typ2cys	8926	82.9	498.6	48.3	650.3	0.6	1.3
Thioredoxin POD3	PRX-Typ2cys	2937	59.2	147.4	126.9	166.2	2.1	1.1
Brahma	HSA, BRK, SNF2-N, HELIC, Bromodomain, TCH	291	142.2	218.6	102.7	189.6	0.7	0.9
Cationic peptide-8 precursor	amfpi-1	16281, 17312	4241.8	0.0	3268.4	0.0	0.8	0/0
DVA-AP3		7139	248.8	1883.8	12.1	764.0	0.0	0.4
Aminoacylase	Peptidase-M20	3287	82.9	1210.8	0.0	691.1	0.0	0.6
Hypoth. protein	Destabilase	6175	130.3	0.0	314.2	2.9	2.4	*1.2*
GL21066		7671	2689.6	7.4	2718.6	8.7	1.0	1.2
~ GA16498-PA	Ig	5348	592.4	0.0	404.8	0.0	0.7	0/0
Protein PTase-2c		11311	0.0	2.5	18.1	26.2	*1.5*	10.7
Ral G-exch factor	RasGEF-N, RasGEF, RA	671	11.8	29.5	60.4	134.1	5.1	4.6
G-exch factor	RhoGEF, SH3	1970	11.8	39.3	72.5	84.6	6.1	2.2
Arf6 G-exch factor		11356	0.0	2.5	24.2	20.4	*2.0*	8.3
Vrille	βZIP-2	1390	11.8	76.1	84.6	55.4	7.1	0.7
Ankyrin repeat pr.		13966	0.0	2.5	0.0	26.2	0/0	10.7
Ankyrin domain 54		6868	0.0	2.5	6.0	32.1	*0.5*	13.1
Dipeptidyl peptidase 4	DPPIV-N	6304	11.8	0.0	66.5	2.9	5.6	*1.2*
TcasGA2-TC008649		537	11.8	78.6	60.4	93.3	5.1	1.2

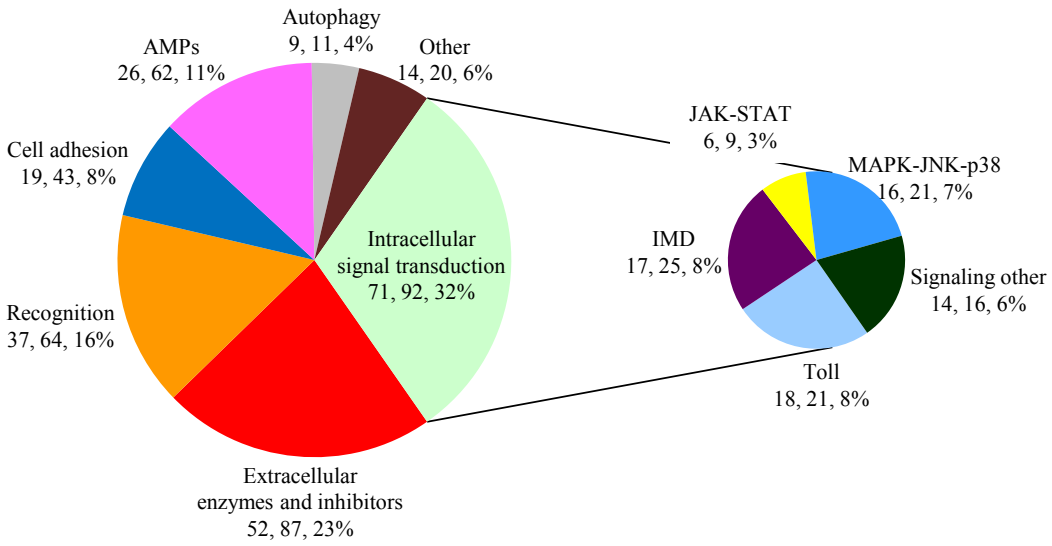
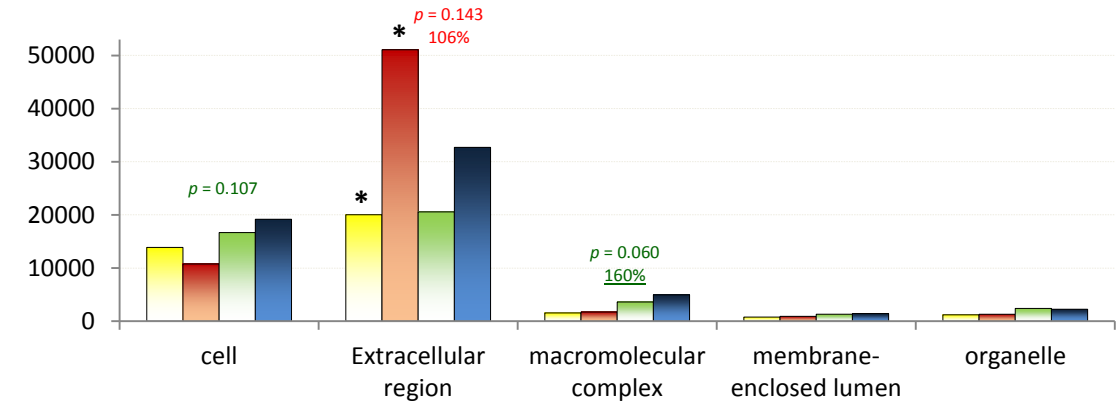
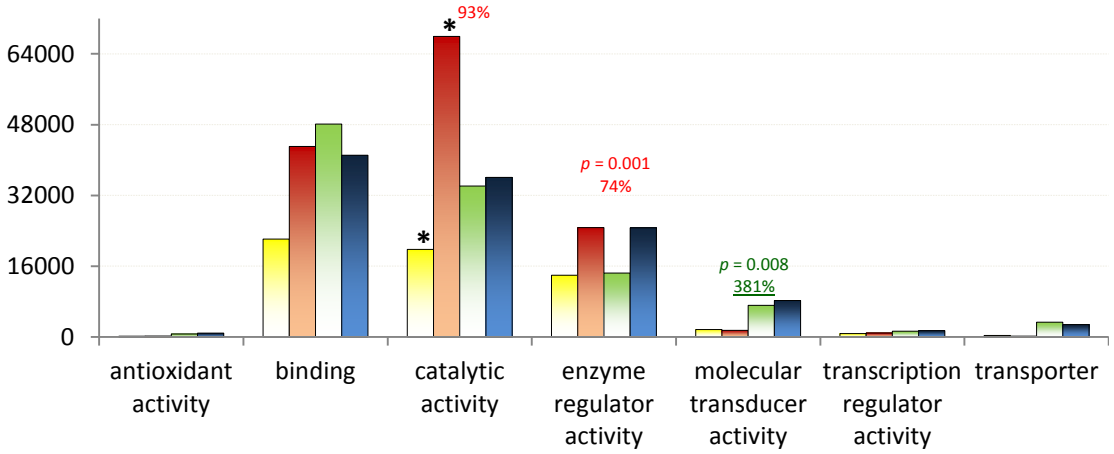


Figure 1. Distribution of 232 *M. sexta* immunity-related genes. The pie chart shows gene number, contig number, and percentage of genes in each functional category relative to the entire set.


50000
40000
30000
20000
10000
0
p = 0.107
*
p = 0.143
106%
p = 0.060
160%
cell
Extracellular region
macromolecular complex
membrane-enclosed lumen
organelle


64000
48000
32000
16000
0
*
*
93%
p = 0.001
74%
p = 0.008
381%
antioxidant activity
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
transcription regulator activity
transporter

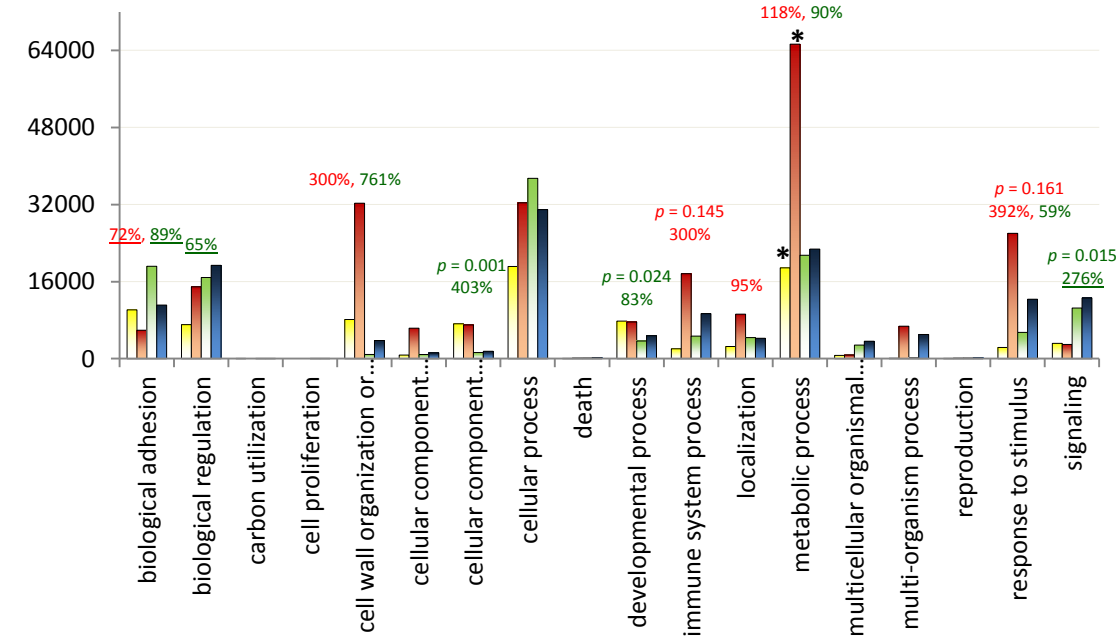

64000
48000
32000
16000
0
72%, 89%
65%
300%, 761%
p = 0.001
403%
p = 0.024
83%
p = 0.145
300%
95%
118%, 90%
*
*
p = 0.161
392%, 59%
p = 0.015
276%
biological adhesion
biological regulation
carbon utilization
cell proliferation
cell wall organization or...
cellular component...
cellular component...
cellular process
death
developmental process
immune system process
localization
metabolic process
multicellular organismal...
multi-organism process
reproduction
response to stimulus
signaling

Figure 2. Expression analysis of cellular components (CC, *top*), molecular functions (MF, *middle*), and biological processes (BP, *bottom*) at GO level 2. The bar graph is generated using data from the sum of NRNs for each annotated gene. Each GO term is comprised of four values, each from a particular library (CF: *yellow*, IF, *orange*, CH: *green*, and IH: *blue*). Bar height represents sum of sums of NRNs in a library within a specific GO group. *p* value (< 0.20) and percentage (>50%) increase or decrease (*underlined*) of immune inducibility (*red*, IF-IH vs. CF-CH) and tissue specificity (*green*, CF-IF vs. CH-IH) are indicated on the top. Of the eleven GO categories in which NRN sums have >50% differences in either or both tissues, three increase most dramatically and are marked with "*".

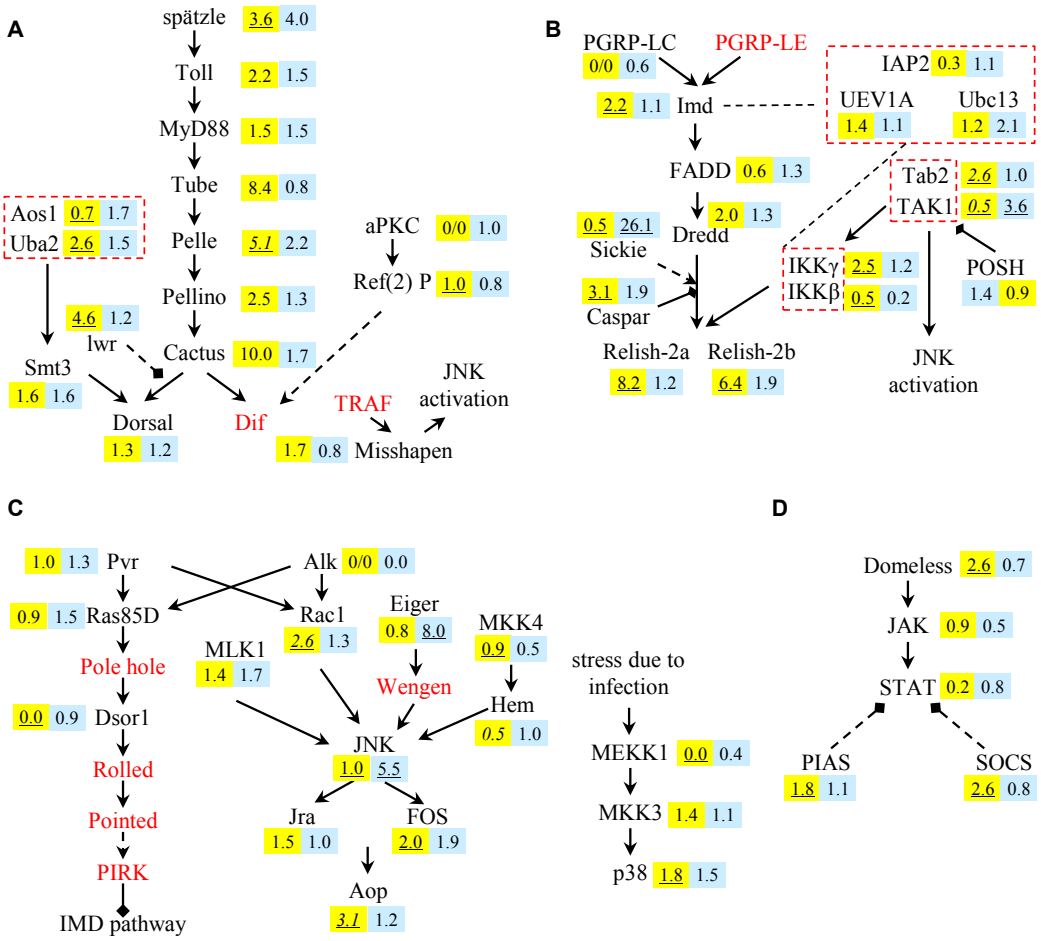


Figure 3. Identification and profiling of transcripts involved in the Toll (A), IMD (B), JAK-STAT (C) and MAPK-JNK-p38 (D) signal transduction pathways. The intracellular signaling processes, based mostly on *Drosophila* research, are described in the text assuming the pathways are conserved among insects. Genes that are not found in our dataset are shown in *red*. Immune inducibilities (*i.e.*, NRN ratios or *ARN*s) in fat body (*yellow*) and hemocytes (*blue*) are indicated near the corresponding genes. *Underlined* number or 0/0 denotes low RN_{CF} or RN_{CH} (0 ~ 4) and, hence, less reliable NRN ratio.

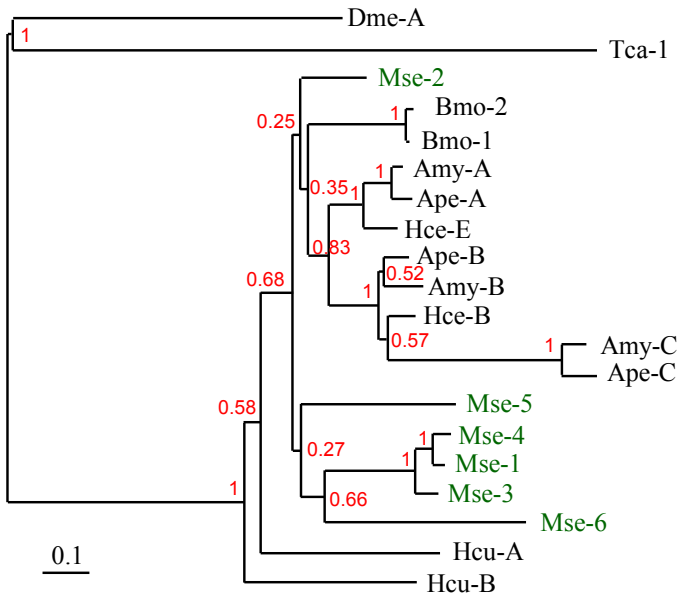


Figure 4. Phylogenetic relationships among insect attacins. Amino acid sequences of *M. sexta* attacin-1 through -6 (Mse-1, 2, 3, 4, 5, 6, green); *Antheraea mylitta* attacin-A, B, C (Amy-A, B, C), *Antheraea pernyi* attacin-A, B, C (Ape-A, B, C), *B. mori* attacin-1 and 2 (Bmo-1, 2), *D. melanogaster* attacin-A (Dme-A), *Hyalophora cecropia* attacin-B and E (Hce-B, E), *Hyphantria cunea* attacin-A and B (Hcu-A, B), *T. castaneum* attacin-1 (Tca-1) are incorporated into generating the phylogenetic tree. The bootstrap values (%) are indicated at nodes in red.

REFERENCES

Abraham, J., 2007. The role of Drosophila SUMO conjugating enzyme Lesswright in larval hematopoiesis: Effects on Cactus, Dorsal and Dorsal-related immunity factor (Dif). Ohio University, OhioLINK. http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1187275172. .

Agaisse, H., 2007. An adaptive immune response in Drosophila? Cell Host & Microbe 1, 91-93.

Agaisse, H., Perrimon, N., 2004. The roles of JAK/STAT signaling in Drosophila immune responses. Immunol. Rev. 198, 72-82.

Altincicek, B., Vilcinskas, A., 2007. Identification of immune-related genes from an apterygote insect, the firebrat Thermobia domestica. Insect Biochem. Mol. Biol. 37, 726-731.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

An, C., Jiang, H., Kanost, M.R., 2010. Proteolytic activation and function of the cytokine Spatzle in the innate immune response of a lepidopteran insect, Manduca sexta. FEBS J. 277, 148-162.

Ao, J.Q., Ling, E., Yu, X.Q., 2008. A Toll receptor from Manduca sexta is in response to Escherichia coli infection. Mol. Immunol. 45, 543-552.

Aoun, R.B., Hetru, C., Troxler, L., Doucet, D., Ferrandon, D., Matt, N., 2011. Analysis of Thioester-Containing Proteins during the Innate Immune Response of Drosophila melanogaster. Journal of Innate Immunity 3, 52-64.

Arbouzova, N.I., Zeidler, M.P., 2006. JAK/STAT signalling in Drosophila: insights into conserved regulatory and cellular functions. Development 133, 2605-2616.

Baeg, G.H., Zhou, R., Perrimon, N., 2005. Genome-wide RNAi analysis of JAK/STAT signaling components in Drosophila. Genes Dev. 19, 1861-1870.

Bhaskar, V., Smith, M., Courey, A.J., 2002. Conjugation of Smt3 to dorsal may potentiate the Drosophila immune response. Mol. Cell. Biol. 22, 492.

Bhaskar, V., Valentine, S.A., Courey, A.J., 2000. A functional interaction between dorsal and components of the Smt3 conjugation machinery. J. Biol. Chem. 275, 4033-4040.

Bhattacharya, A., Steward, R., 2002. The Drosophila homolog of NTF-2, the nuclear transport factor-2, is essential for immune response. EMBO Rep 3, 378-383.

Blandin, S., Levashina, E.A., 2004. Thioester-containing proteins and insect immunity. Mol. Immunol. 40, 903-908.

Bokoch, G.M., 2005. Regulation of innate immunity by Rho GTPases. Trends Cell Biol. 15, 163-171.

Boutros, M., Agaisse, H., Perrimon, N., 2002. Sequential activation of signaling pathways during innate immune responses in Drosophila. Dev. Cell 3, 711-722.

Brey, P.T., Hultmark, D., 1998. Molecular mechanisms of immune responses in insects. Chapman & Hall, London; New York.

Cerenius, L., Soderhall, K., 2004. The prophenoloxidase-activating system in invertebrates. Immunol. Rev. 198, 116-126.

Chapman, A.D., Australian Biodiversity Information, S., Australia. Dept. of the, E., Heritage, 2006. Numbers of living species in Australia and the world. Australian Govt., Dept. of the Environment and Heritage, Canberra, ACT.

Choe, K.M., Werner, T., Stoven, S., Hultmark, D., Anderson, K.V., 2002. Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in Drosophila. Science 296, 359-362.

Christensen, B.M., Li, J., Chen, C.C., Nappi, A.J., 2005. Melanization immune responses in mosquito vectors. Trends Parasitol. 21, 192-199.

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., Muller, 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.N., Zheng, L.B., Bork, P., Kafatos, F.C., 2002. Immunity-related genes and gene families in Anopheles gambiae. Science 298, 159-165.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674-3676.

De Gregorio, E., Spellman, P.T., Rubin, G.M., Lemaitre, B., 2001. Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proceedings of the National Academy of Sciences of the United States of America 98, 12590-12595.

Ding, S.W., Li, H., Lu, R., Li, F., Li, W.X., 2004. RNA silencing: a conserved antiviral immunity of plants and animals. Virus Res. 102, 109-115.

Dong, C., Davis, R.J., Flavell, R.A., 2002. MAP kinases in the immune response. Annu. Rev. Immunol. 20, 55-72.

Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana-Arnoux, D., Hetru, C., Hoffmann, J.A., Imler, J.L., 2005. The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of drosophila. Nat. Immunol. 6, 946-953.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, W465-469.

Dereeper, A., Audic, S., Claverie, J.M., Blanc, G., 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol. Biol. 10, 8.

Dunn, P.E., Drake, D., 1983. Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm, Manduca sexta. J Invert Pathol 41, 77-85.

Dziarski, R., Gupta, D., 2006. The peptidoglycan recognition proteins (PGRPs). Genome Biology 7.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

Engstrom, Y., 1999. Induction and regulation of antimicrobial peptides in Drosophila. Developmental and Comparative Immunology 23, 345-358.

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

Fauvarque, M.O., Williams, M.J., 2011. Drosophila cellular immunity: a story of migration and adhesion. J. Cell Sci. 124, 1373-1382.

Foley, E., O'Farrell, P.H., 2004. Functional dissection of an innate immune response by a genome-wide RNAi screen. PLoS Biol. 2, E203.

Gallo, K.A., Johnson, G.L., 2002. Mixed-lineage kinase control of JNK and p38 MAPK pathways. Nat. Rev. Mol. Cell Biol. 3, 663-672.

Ganesan, S., Aggarwal, K., Paquette, N., Silverman, N., 2011. NF-kappaB/Rel proteins and the humoral immune responses of Drosophila melanogaster. Curr. Top. Microbiol. Immunol. 349, 25-60.

Geng, J.F., Klionsky, D.J., 2008. The Atg8 and Atg12 ubiquitin-like conjugation systems in macroautophagy. Embo Reports 9, 859-864.

Gerardo, N.M., Altincicek, B., Anselme, C., Atamian, H., Barribeau, S.M., De Vos, M., Duncan, E.J., Evans, J.D., Gabaldon, T., Ghanim, M., Heddi, A., Kaloshian, I., Latorre, A., Moya, A.,

Nakabachi, A., Parker, B.J., Perez-Brocal, V., Pignatelli, M., Rahbe, Y., Ramsey, J.S., Spragg, C.J., Tamames, J., Tamarit, D., Tamborindeguy, C., Vincent-Monegat, C., Vilcinskas, A., 2010. Immunity and other defenses in pea aphids, Acyrthosiphon pisum. Genome Biology 11.

Geuking, P., Narasimamurthy, R., Lemaitre, B., Basler, K., Leulier, F., 2009. A non-redundant role for Drosophila Mkk4 and hemipterous/Mkk7 in TAK1-mediated activation of JNK. PLoS One 4, e7709.

Gillespie, J.P., Kanost, M.R., Trenczek, T., 1997. Biological mediators of insect immunity. Annu. Rev. Entomol. 42, 611-643.

Goldsmith, M.R., Marec, F., 2010. Molecular biology and genetics of the Lepidoptera. CRC Press/Taylor & Francis, Boca Raton, FL.

Gorman, M.J., Wang, Y., Jiang, H., Kanost, M.R., 2007. Manduca sexta hemolymph proteinase 21 activates prophenoloxidase-activating proteinase 3 in an insect innate immune response proteinase cascade. J. Biol. Chem. 282, 11742-11749.

Gottar, M., Gobert, V., Michel, T., Belvin, M., Duyk, G., Hoffmann, J.A., Ferrandon, D., Royet, J., 2002. The Drosophila immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. Nature 416, 640-644.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420-3435.

Hall, A., 1998. Rho GTPases and the actin cytoskeleton. Science 279, 509.

Han, Z.Q.S., Enslen, H., Hu, X.D., Meng, X.J., Wu, I.H., Barrett, T., Davis, R.J., Ip, Y.T., 1998. A conserved p38 mitogen-activated protein kinase pathway regulates Drosophila immunity gene expression. Mol. Cell. Biol. 18, 3527-3539.

Hoffmann, J.A., 2003. The immune response of Drosophila. Nature 426, 33-38.

Hoffmann, J.A., Kafatos, F.C., Janeway, C.A., Ezekowitz, R.A.B., 1999. Phylogenetic perspectives in innate immunity. Science 284, 1313-1318.

Hoffmann, J.A., Reichhart, J.M., 1997. Drosophila immunity. Trends Cell Biol. 7, 309-316.

Hou, X.S., Perrimon, N., 1997. The JAK-STAT pathway in Drosophila. Trends Genet. 13, 105-110.

Huang, L., Ohsako, S., Tanda, S., 2005. The lesswright mutation activates Rel-related proteins, leading to overproduction of larval hemocytes in Drosophila melanogaster. Dev. Biol. 280, 407-420.

Hultmark, D., 1993. Immune reactions in Drosophila and other insects: a model for innate immunity. Trends Genet. 9, 178-183.

Hultmark, D., 2003. Drosophila immunity: paths and patterns. Curr. Opin. Immunol. 15, 12-19.

Inoue, H., Tateno, M., Fujimura-Kamada, K., Takaesu, G., Adachi-Yamada, T., Ninomiya-Tsuji, J., Irie, K., Nishida, Y., Matsumoto, K., 2001. A Drosophila MAPKKK, D-MEKK1, mediates stress responses through activation of p38 MAPK. EMBO J. 20, 5421-5430.

Irving, P., Troxler, L., Heuer, T.S., Belvin, M., Kopczynski, C., Reichhart, J.M., Hoffmann, J.A., Hetru, C., 2001. A genome-wide analysis of immune responses in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 98, 15119-15124.

Iwai, K., Tokunaga, F., 2009. Linear polyubiquitination: a new regulator of NF-kappaB activation. EMBO Rep 10, 706-713.

Jang, I.H., Nam, H.J., Lee, W.J., 2008. CLIP-domain serine proteases in Drosophila innate immunity. BMB Rep 41, 102-107.

Janssen, B.J.C., Huizinga, E.G., Raaijmakers, H.C.A., Roos, A., Daha, M.R., Nilsson-Ekdahl, K., Nilsson, B., Gros, P., 2005. Structures of complement component C3 provide insights into the function and evolution of immunity. Nature 437, 505-511.

Jiang, H., 2008. The biochemical basis of antimicrobial responses in. Insect Sci. 15, 53-66.

Jiang, H., Vilcinskas, A., Kanost, M.R., 2010. Immunity in lepidopteran insects. Adv. Exp. Med. Biol. 708, 181-204.

Jiang, H., Wang, Y., Gu, Y., Guo, X., Zou, Z., Scholz, F., Trenczek, T.E., Kanost, M.R., 2005. Molecular identification of a bevy of serine proteinases in Manduca sexta hemolymph. Insect Biochem. Mol. Biol. 35, 931-943.

Kallio, J., Leinonen, A., Ulvila, J., Valanne, S., Ezekowitz, R.A., Ramet, M., 2005. Functional analysis of immune response genes in Drosophila identifies JNK pathway as a regulator of antimicrobial peptide gene expression in S2 cells. Microbes Infect 7, 811-819.

Kanayama, A., Seth, R.B., Sun, L., Ea, C.K., Hong, M., Shaito, A., Chiu, Y.H., Deng, L., Chen, Z.J., 2004. TAB2 and TAB3 activate the NF-[kappa] B pathway through binding to polyubiquitin chains. Mol. Cell 15, 535-548.

Kaneko, T., Yano, T., Aggarwal, K., Lim, J.H., Ueda, K., Oshima, Y., Peach, C., Erturk-Hasdemir, D., Goldman, W.E., Oh, B.H., 2006. PGRP-LC and PGRP-LE have essential yet distinct functions in the drosophila immune response to monomeric DAP-type peptidoglycan. Nat. Immunol. 7, 715-723.

Kanost, M.R., 1999. Serine proteinase inhibitors in arthropod immunity. Developmental and Comparative Immunology 23, 291-301.

Kanost, M.R., Gorman, M.G., 2008. Phenoloxidases in insect immunity in: Beckage, N. (Ed.), Insect immunology. Academic Press/Elsevier, San Diego, pp. 69-96.

Kanost, M.R., Jiang, H.B., Wang, Y., Yu, X.Q., Ma, C.C., Zhu, Y.F., 2001. Hemolymph proteinases in immune responses of Manduca sexta. Phylogenetic Perspectives on the Vertebrate Immune System 484, 319-328.

Kanost, M.R., Jiang, H.B., Yu, X.Q., 2004. Innate immune responses of a lepidopteran insect, Manduca sexta. Immunol. Rev. 198, 97-105.

Kanost, M.R., Zepp, M.K., Ladendorff, N.E., Andersson, L.A., 1994. Isolation and Characterization of a Hemocyte Aggregation Inhibitor from Hemolymph of Manduca-Sexta Larvae. Arch. Insect Biochem. Physiol. 27, 123-136.

Khush, R.S., Lemaitre, B., 2000. Genes that fight infection - what the Drosophila genome says about animal immunity. Trends Genet. 16, 442-449.

Kim, T., Kim, Y.J., 2005. Overview of innate immunity in Drosophila. Journal of Biochemistry and Molecular Biology 38, 121-127.

Kisseleva, T., Bhattacharya, S., Braunstein, J., Schindler, C.W., 2002. Signaling through the JAK/STAT pathway, recent advances and future challenges. Gene 285, 1-24.

Kleino, A., Valanne, S., Ulvila, J., Kallio, J., Myllymaki, H., Enwald, H., Stoven, S., Poidevin, M., Ueda, R., Hultmark, D., Lemaitre, B., Ramet, M., 2005. Inhibitor of apoptosis 2 and TAK1-binding protein are components of the Drosophila Imd pathway. EMBO J. 24, 3423-3434.

Kopp, E., Medzhitov, R., Carothers, J., Xiao, C., Douglas, I., Janeway, C.A., Ghosh, S., 1999. ECSIT is an evolutionarily conserved intermediate in the Toll/IL-1 signal transduction pathway. Genes Dev. 13, 2059.

Krishnakumar, S., Burton, D., Rasco, J., Chen, X.Y., O'Donnell, J., 2000. Functional interactions between GTP cyclohydrolase I and tyrosine hydroxylase in Drosophila. J. Neurogenet. 14, 1-23.

Kurata, S., Ariki, S., Kawabata, S., 2006. Recognition of pathogens and activation of immune responses in Drosophila and horseshoe crab innate immunity. Immunobiology 211, 237-249.

Kurucz, E., Markus, R., Zsamboki, J., Folkl-Medzihradszky, K., Darula, Z., Vilmos, P., Udvardy, A., Krausz, I., Lukacsovich, T., Gateff, E., Zettervall, C.J., Hultmark, D., Ando, I., 2007. Nimrod, a putative phagocytosis receptor with EGF repeats in Drosophila plasmatocytes. Curr. Biol. 17, 649-654.

Lavine, M.D., Strand, M.R., 2002. Insect hemocytes and their role in immunity. Insect Biochem. Mol. Biol. 32, 1295-1309.

Lee, K.Z., Ferrandon, D., 2011. Negative regulation of immune responses on the fly. EMBO J. 30, 988-990.

Lemaitre, B., Hoffmann, J., 2007. The host defense of Drosophila melanogaster. Annu. Rev. Immunol. 25, 697-743.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., Hoffmann, J.A., 1996. The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell 86, 973-983.

LeMosy, E.K., Hong, C.C., Hashimoto, C., 1999. Signal transduction by a protease cascade. Trends Cell Biol. 9, 102-107.

Leulier, F., Rodriguez, A., Khush, R.S., Abrams, J.M., Lemaitre, B., 2000. The Drosophila caspase Dredd is required to resist gram-negative bacterial infection. EMBO reports 1, 353-358.

Levin, D.M., Breuer, L.N., Zhuang, S., Anderson, S.A., Nardi, J.B., Kanost, M.R., 2005. A hemocyte-specific integrin required for hemocytic encapsulation in the tobacco hornworm, Manduca sexta. Insect Biochem. Mol. Biol. 35, 369-380.

Ma, C., Kanost, M.R., 2000. A beta1,3-glucan recognition protein from an insect, Manduca sexta, agglutinates microorganisms and activates the phenoloxidase cascade. J. Biol. Chem. 275, 7505-7514.

Mabb, A.M., Miyamoto, S., 2007. SUMO and NF-kappaB ties. Cell. Mol. Life Sci. 64, 1979-1996.

Margulies, M., Egholm, M., Altman, W.E., Bader, J.S. *et al*. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437 (7057), 376-380.

Marmaras, V.J., Lampropoulou, M., 2009. Regulators and signalling in insect haemocyte immunity. Cell. Signal. 21, 186-195.

Moncrieffe, M.C., Grossmann, J.G., Gay, N.J., 2008. Assembly of oligomeric death domain complexes during Toll receptor signaling. J. Biol. Chem. 283, 33447-33454.

Myhre, S., Tveit, H., Mollestad, T., Laegreid, A., 2006. Additional gene ontology structure for improved biological reasoning. Bioinformatics 22, 2020-2027.

Naitza, S., Rosse, C., Kappler, C., Georgel, P., Belvin, M., Gubb, D., Camonis, J., Hoffmann, J.A., Reichhart, J.M., 2002. The Drosophila immune defense against gram-negative infection requires the death protein dFADD. Immunity 17, 575-581.

Nappi, A., Christensen, B., 2005. Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity. Insect Biochem. Mol. Biol. 35, 443-459.

Nappi, A.J., Vass, E., Frey, F., Carton, Y., 2000. Nitric oxide involvement in Drosophila immunity. Nitric Oxide-Biology and Chemistry 4, 423-430.

Nardi, J.B., Pilas, B., Bee, C.M., Zhuang, S., Garsha, K., Kanost, M.R., 2006. Neuroglian-positive plasmatocytes of Manduca sexta and the initiation of hemocyte attachment to foreign surfaces. Dev. Comp. Immunol. 30, 447-462.

Pace, K.E., Baum, L.G., 2004. Insect galectins: roles in immunity and development. Glycoconj J 19, 607-614.

Paquette, N., Broemer, M., Aggarwal, K., Chen, L., Husson, M., Erturk-Hasdemir, D., Reichhart, J.M., Meier, P., Silverman, N., 2010. Caspase-mediated cleavage, IAP binding, and ubiquitination: linking three mechanisms crucial for Drosophila NF-kappaB signaling. Mol. Cell 37, 172-182.

Park, J.M., Brady, H., Ruocco, M.G., Sun, H., Williams, D., Lee, S.J., Kato, T., Jr., Richards, N., Chan, K., Mercurio, F., Karin, M., Wasserman, S.A., 2004. Targeting of TAK1 by the NF-kappa B protein Relish regulates the JNK-mediated immune response in Drosophila. Genes Dev. 18, 584-594.

Pauchet, Y., Wilkinson, P., Vogel, H., Nelson, D.R., Reynolds, S.E., Heckel, D.G., ffrench-Constant, R.H., 2010. Pyrosequencing the Manduca sexta larval midgut transcriptome: messages for digestion, detoxification and defence. Insect Mol. Biol. 19, 61-75.

Petersen, U.M., Kadalayil, L., Rehorn, K.P., Hoshizaki, D.K., Reuter, R., Engstrom, Y., 1999. Serpent regulates Drosophila immunity genes in the larval fat body through an essential GATA motif. EMBO J. 18, 4013-4022.

Pham, L.N., Dionne, M.S., Shirasu-Hiza, M., Schneider, D.S., 2007. A specific primed immune response in Drosophila is dependent on phagocytes. PLoS Path. 3.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R., 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 33, W116-120.

Roe. B., 2004. Shotgun library construction for DNA sequencing. Methods Mol Biol 255: 171-187.

Ragab, A., Buechling, T., Gesellchen, V., Spirohn, K., Boettcher, A.L., Boutros, M., 2011. Drosophila Ras/MAPK signalling regulates innate immune responses in immune and intestinal stem cells. EMBO J. 30, 1123-1136.

Ragan, E.J., An, C., Yang, C.T., Kanost, M.R., 2010. Analysis of mutually exclusive alternatively spliced serpin-1 isoforms and identification of serpin-1 proteinase complexes in Manduca sexta hemolymph. J. Biol. Chem. 285, 29642-29650.

Ramet, M., Lanot, R., Zachary, D., Manfruelli, P., 2002. JNK signaling pathway is required for efficient wound healing in Drosophila. Dev. Biol. 241, 145-156.

Remillieux-Leschelle, N., Santamaria, P., Randsholt, N.B., 2002. Regulation of larval hematopoiesis in Drosophila melanogaster: a role for the multi sex combs gene. Genetics 162, 1259-1274.

Rolff, J., Reynolds, S.E., 2009. Insect infection and immunity: Evolution, ecology, and mechanisms. Oxford University Press, Oxford.

Sackton, T.B., Lazzaro, B.P., Schlenke, T.A., Evans, J.D., Hultmark, D., Clark, A.G., 2007. Dynamic evolution of the innate immune system in Drosophila. Nat. Genet. 39, 1461-1468.

Sansonetti, P.J., 2006. The innate signaling of dangers and the dangers of innate signaling. Nat. Immunol. 7, 1237-1242.

Shi, Y., 2002. A conserved tetrapeptide motif: potentiating apoptosis through IAP-binding. Cell Death Differ. 9, 93-95.

Shintani, T., Klionsky, D.J., 2004. Autophagy in health and disease: A double-edged sword. Science 306, 990-995.

Sieg, D.J., Hauck, C.R., Schlaepfer, D.D., 1999. Required role of focal adhesion kinase (FAK) for integrin-stimulated cell migration. J. Cell Sci. 112, 2677.

Silverman, N., Maniatis, T., 2001. NF-kappa B signaling pathways in mammalian and insect innate immunity. Genes Dev. 15, 2321-2342.

Silverman, N., Zhou, R., Erlich, R.L., Hunter, M., Bernstein, E., Schneider, D., Maniatis, T., 2003. Immune activation of NF-kappa B and JNK requires Drosophila TAK1. J. Biol. Chem. 278, 48928-48934.

Sluss, H.K., Han, Z., Barrett, T., Davis, R.J., Ip, Y.T., 1996. A JNK signal transduction pathway that mediates morphogenesis and an immune response in Drosophila. Genes Dev. 10, 2745.

Somogyi, K., Sipos, B., Penzes, Z., Kurucz, E., Zsamboki, J., Hultmark, D., Ando, I., 2008. Evolution of Genes and Repeats in the Nimrod Superfamily. Mol. Biol. Evol. 25, 2337-2347.

Stoven, S., Silverman, N., Junell, A., Hedengren-Olcott, M., Erturk, D., Engstrom, Y., Maniatis, T., Hultmark, D., 2003. Caspase-mediated processing of the Drosophila NF-kappaB factor Relish. Proc Natl Acad Sci U S A 100, 5991-5996.

Strand, M.R., 2008. The insect cellular immune response. Insect Sci. 15, 1-14.

Takeda, K., Akira, S., 2001. Roles of Toll-like receptors in innate immune responses. Genes Cells 6, 733-742.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.

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.

Thony, B., Auerbach, G., Blau, N., 2000. Tetrahydrobiopterin biosynthesis, regeneration and functions. Biochem. J. 347 Pt 1, 1-16.

Tong, Y., Jiang, H., Kanost, M.R., 2005. Identification of plasma proteases inhibited by Manduca sexta serpin-4 and serpin-5 and their association with components of the prophenol oxidase activation pathway. J. Biol. Chem. 280, 14932-14942.

Valanne, S., Wang, J.H., Ramet, M., 2011. The Drosophila Toll Signaling Pathway. J. Immunol. 186, 649-656.

Vilmos, P., Kurucz, E., 1998. Insect immunity: evolutionary roots of the mammalian innate immune system. Immunol. Lett. 62, 59-66.

Vogel, H., Altincicek, B., Glockner, G., Vilcinskas, A., 2011. A comprehensive transcriptome and immune-gene repertoire of the lepidopteran model host Galleria mellonella. BMC Genomics 12, 308.

Wajant, H., Scheurich, P., 2004. Analogies between Drosophila and mammalian TRAF pathways. Prog Mol Subcell Biol 34, 47-72.

Wang, X.H., Aliyari, R., Li, W.X., Li, H.W., Kim, K., Carthew, R., Atkinson, P., Ding, S.W., 2006. RNA interference directs innate immunity against viruses in adult Drosophila. Science 312, 452-454.

Wang, Y., Jiang, H., 2007. Reconstitution of a branch of the Manduca sexta prophenoloxidase activation cascade in vitro: snake-like hemolymph proteinase 21 (HP21) cleaved by HP14 activates prophenoloxidase-activating proteinase-2 precursor. Insect Biochem. Mol. Biol. 37, 1015-1025.

Wang, Y., Jiang, H., 2008. A positive feedback mechanism in the Manduca sexta prophenoloxidase activation system. Insect Biochem. Mol. Biol. 38, 763-769.

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.B., 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.B., 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.

Watson, F.L., Puttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., Schmucker, D., 2005. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 309, 1874-1878.

Weber, A.N., Tauszig-Delamasure, S., Hoffmann, J.A., Lelievre, E., Gascan, H., Ray, K.P., Morse, M.A., Imler, J.L., Gay, N.J., 2003. Binding of the Drosophila cytokine Spatzle to Toll is direct and establishes signaling. Nat. Immunol. 4, 794-800.

Williams, M.J., 2007. Drosophila hemopoiesis and cellular immunity. J. Immunol. 178, 4711-4716.

Williams, M.J., Wiklund, M.L., Wikman, S., Hultmark, D., 2006. Rac1 signalling in the Drosophila larval cellular immune response. J. Cell Sci. 119, 2015-2024.

Xiao, C., Ghosh, S., 2005. NF-κB, an evolutionarily conserved mediator of immune and inflammatory responses. Mechanisms of Lymphocyte Activation and Immune Regulation X, 41-45.

Xu, H., Hao, W., He, D., Xu, Y., 2010. Smt3 is required for the immune response of silkworm, Bombyx mori. Biochimie 92, 1306-1314.

Yu, X.Q., Gan, H., Kanost, M.R., 1999. Immulectin, an inducible C-type lectin from an insect, Manduca sexta, stimulates activation of plasma prophenol oxidase. Insect Biochem. Mol. Biol. 29, 585-597.

Yu, X.Q., Kanost, M.R., 2000. Immulectin-2, a lipopolysaccharide-specific lectin from an insect, Manduca sexta, is induced in response to gram-negative bacteria. J. Biol. Chem. 275, 37373-37381.

Yu, X.Q., Ling, E., Tracy, M.E., Zhu, Y., 2006. Immulectin-4 from the tobacco hornworm Manduca sexta binds to lipopolysaccharide and lipoteichoic acid. Insect Mol. Biol. 15, 119-128.

Yu, X.Q., Tracy, M.E., Ling, E., Scholz, F.R., Trenczek, T., 2005. A novel C-type immulectin-3 from Manduca sexta is translocated from hemolymph into the cytoplasm of hemocytes. Insect Biochem. Mol. Biol. 35, 285-295.

Yu, X.Q., Zhu, Y.F., Ma, C., Fabrick, J.A., Kanost, M.R., 2002. Pattern recognition proteins in Manduca sexta plasma. Insect Biochem. Mol. Biol. 32, 1287-1293.

Zettervall, C.J., Anderl, I., Williams, M.J., Palmer, R., Kurucz, E., Ando, I., Hultmark, D., 2004. A directed screen for genes involved in Drosophila blood cell activation. Proc Natl Acad Sci U S A 101, 14192-14197.

Zhang, F.J., Guo, H.Y., Zheng, H.J., Zhou, T., Zhou, Y.J., Wang, S.Y., Fang, R.X., Qian, W., Chen, X.Y., 2010. Massively parallel pyrosequencing-based transcriptome analyses of small brown planthopper (Laodelphax striatellus), a vector insect transmitting rice stripe virus (RSV). BMC Genomics 11.

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.

Zhu, Y.F., Ragan, E.J., Kanost, M.R., 2010. Leureptin: A soluble, extracellular leucine-rich repeat protein from Manduca sexta that binds lipopolysaccharide. Insect Biochem. Mol. Biol. 40, 713-722.

Zhuang, S.F., Kelo, L., Nardi, J.B., Kanost, M.R., 2007a. Neuroglian on hemocyte surfaces is involved in homophilic and heterophilic interactions of the innate immune system of Manduca sexta. Developmental and Comparative Immunology 31, 1159-1167.

Zhuang, S.F., Kelo, L.S., Nardi, J.B., Kanost, M.R., 2007b. An integrin-tetraspanin interaction required for cellular innate immune responses of an insect, Manduca sexta. J. Biol. Chem. 282, 22563-22572.

Zhuang, S.F., Nardi, J.B., Kanost, M.R., 2005. An integrin, neuroglian, and a tetraspanin modulate adhesion in hemocytes of an insect, Manduca sexta. FASEB J. 19, A1061-A1062.

Zhuang, Z.-H., Sun, L., Kong, L., Hu, J.-H., Yu, M.-C., Reinach, P., Zang, J.-W., Ge, B.-X., 2006. Drosophila TAB2 is required for the immune activation of JNK and NF-kappaB. Cell. Signal. 18, 964-970.

Zou, Z., Evans, J.D., Lu, Z.Q., Zhao, P.C., Williams, M., Sumathipala, N., Hetru, C., Hultmark, D., Jiang, H.B., 2007. Comparative genomic analysis of the Tribolium immune system. Genome Biology 8.

Zou, Z., Jiang, H., 2005. Manduca sexta serpin-6 regulates immune serine proteinases PAP-3 and HP8. cDNA cloning, protein expression, inhibition kinetics, and function elucidation. J. Biol. Chem. 280, 14341-14348.

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.

VITA

Ramesh Tharindu Gunaratna

Candidate for the Degree of

Master of Science

Thesis: A COMPREHENSIVE ANALYSIS OF THE MANDUCA SEXTA IMMUNOTRANSCRIPTOME

Major Field: Entomology

Biographical:

Education:

Completed the requirements for the Master of Science in Entomology at Oklahoma State University, Stillwater, Oklahoma in May, 2012.

Completed the requirements for the Bachelor of Science in Bioinformatics at University of Colombo, Sri Lanka in Aug, 2009.

Experience:

Worked as a graduate research assistant from January 2010 through May 2012 at insect molecular genetics labs under the supervision of Dr. Haobo Jiang. Worked as a full time teaching assistant as well as a moderator and a content developer for the 'On-line certificate course in Bioinformatics' in the Department of Plant Sciences, University of Colombo, Sri Lanka from September 2009 through December 2009. Worked as a research Assistant at 'Plant Virus Indexing Center' in Sri Lanka from October 2008 through May 2009.

Professional Memberships:

Entomological Society of America
Phi Kappa Phi Honor Society
Golden Key International Honor Society
Sri Lanka Association for the Advancement of Science

Name: Ramesh Tharindu Gunaratna Date of Degree: May, 2012

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: A COMPREHENSIVE ANALYSIS OF THE MANDUCA SEXTA IMMUNOTRANSCRIPTOME

Pages in Study: 55 Candidate for the Degree of Master of Science

Major Field: Entomology

Scope and Method of Study: As a biochemical model, *Manduca sexta* substantially contributed to our knowledge on insect innate immunity. The RNA-Seq approach based on massively parallel pyrosequencing was implemented in three studies to examine tissue immunotranscriptomes of this species. With the latest and largest focusing on highly regulated process- and tissue-specific genes, we further analyzed the same set of data using BLAST2GO to explore functional aspects of the larval fat body (F) and hemocyte (H) transcriptomes with (I) or without (C) immune challenge.

Findings and Conclusions: Using immunity-related sequences from *Drosophila melanogaster*, *Apis mellifera*, and *Bombyx mori*, we identified 383 homologous contigs and compared them with those discovered based on relative abundance changes and BLASTX analyses. By concatenating the contigs, we established a repertoire of 232 immunity-related genes encoding proteins for pathogen recognition (16%), signal transduction (53%), and microbe killing (13%).

We examined their expression levels along with attribute classifications and detected prominent differences in nine of the thirty level 2 GO categories, such as enzyme regulator (IC), cellular component organization (FH), signaling (FH), and extracellular region (IC). The increase in extracellular proteins (155% or 31,038 normalized read number) was consistent with the highly induced synthesis of defense molecules (*e.g.*, antimicrobial peptides) in fat body after the immune challenge.

We identified most members of the putative Toll, IMD, MAPK-JNK-p38, and JAK-STAT pathways and detected 1.1~1.8-fold increases in the first three and ~30% average decrease in the fourth. The minor increases in the antibacterial and antifungal pathways led to dramatic elevations of transcripts for all antimicrobial peptides as well as some proteins involved in recognition, extracellular signaling, and cellular responses. Most importantly, this study sets the stage for on-going analysis of *M. sexta* immunogenome.

ADVISER'S APPROVAL: Dr. Haobo Jiang