QUANTITATIVE TRANSCRIPTOME AND PROTEOME ANALYSES OF IMMUNITY-RELATED MOLECULES IN *MANDUCA SEXTA* LARVAL HEMOLYMPH

By

SHUGUANG ZHANG

Bachelor of Science in biological sciences

University of Science and Technology of China

Hefei, China

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
July, 2012

QUANTITATIVE TRANSCRIPTOME AND PROTEOME ANALYSES OF IMMUNITY-RELATED MOLECULES IN *MANDUCA SEXTA* LARVAL HEMOLYMPH

Thesis Approved:

Dr. Haobo Jiang

Thesis Adviser

Dr. Jack Dillwith

Dr. Deborah Jaworski

Dr. Sheryl A. Tucker

Dean of the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Most animals on this planet rely on different functional systems to survive such as respiratory system, digestive system, nervous system, circulatory system, etc. To fight against environmental pathogens, they all possess immune systems. The defense system is divided into two major branches – innate or natural immunity and adaptive or acquired immunity. Adaptive immunity is slow and, during this process, the host synthesizes antibodies which specifically recognize, kill and memorize pathogens. In contrast, innate immunity is fast but can kill a broad spectrum of pathogens. Both defense strategies are adopted by vertebrates while in most invertebrates only innate immunity is found. As one of the most successful classes in the animal kingdom, arthropods including all insects occupy almost all environmental niches on the earth. Their evolutionary success is, to some extent, attributed to innate immunity.

The first line of defense in insects is a physical barrier of cuticle lining the body surface, gut and trachea. If microorganisms surpass this barrier, they encounter pattern recognition proteins that recognize microbial surface moieties and relay danger signal to initialize the host immune process. Two types of defense are usually involved in insect immunity: cellular and humoral responses (Tzou et al. 2002). In the former, different hemocytes participate in phagocytosis, nodulation, and encapsulation to engulf, immobilize, and kill invading pathogens. Humoral defenses are mediated by plasma factors (including enzymes and inhibitors) that form large protein complexes during hemolymph clotting and melanization. The enzyme system also

generates cytokines to induce the production of antimicrobial peptides (AMPs) that kill secondary invaders. There is no distinct boundary between the two kinds of defense responses, since many humoral factors affect hemocyte function, hemocytes are an important source of plasma molecules, and cellular and humoral defense often work together to eliminate pathogens in processes such as melanotic encapsulation (Elrod-Erickson et al. 2000).

Insect immunity studies deal with mechanisms by which insects kill viruses, bacteria, fungi and parasites. The research focuses on the discovery of immunity-related proteins, their interactions with pathogens and among themselves. Such studies have a potential in medical applications to prevent vector-borne diseases as well as in agricultural applications to control pests and protect beneficial insects. The traditional way to study insect immunity is to study immunity-related proteins and their genes individually. Previous studies have discovered some system components and illustrated their functions in immune pathways. However, many immunity-related proteins are not yet identified and roles of some known proteins are controversial. The lack of knowledge on these proteins has hampered our understanding of the entire physiological process. With the development of biotechnology, life science has stepped into the 'omics' era in which scientists treat a complex life system as a whole and simultaneously study large number of transcripts (transcriptomics) and proteins (proteomics). My research is mainly focused on the transcriptomic and proteomic studies of immunity-related genes and proteins in a biochemical model insect, *Manduca sexta*, to better elucidate the immune process.

According to the central dogma, genes are transcribed to mRNAs and then translated to proteins. Thus, mRNAs become the bridge between genetic materials - genes and the major functioning components of life system - proteins. Organisms can regulate the life activity via the regulation of mRNA abundances. Compared with proteins, mRNAs are more stable and easy to sequence. As a result, transcriptomic studies are commonly conducted in life sciences. Traditional Sanger sequencing used to be the main sequencing technology in such transcriptomic studies.

Recently the next generation sequencing (NGS) technology is adopted by more and more people because of their high throughput and constantly decreasing cost. Among the NGS technologies, 454 sequencing or pyrosequencing technology has been widely utilized in transcriptomic and genomic studies, since it yields longer reads (Morozova et al. 2009).

In a living organism, proteins function in various ways such as enzymes, structural proteins, transporters, signaling proteins, motor proteins, storage proteins, and defense proteins. They are diverse in sizes, sequences, secondary structures, and post translational modifications. Protein synthesis and sequencing are usually expensive as compared with DNAs. Some native proteins are unstable under *in vitro* condition, making it difficult to purify them and study their functions. Chemical properties of proteins render the research on proteins much more difficult than that of nucleotides. However, since they directly mediate and regulate life activities, protein research is central to understanding both cells and organisms. In proteomic studies, mass spectrometry (MS) is often utilized for protein identifications. Many MS-based quantitative methods are developed and used in proteomic studies. Of these, spectral counting has been widely used because it is easier to handle and more cost-effective (Lundgren et al. 2010).

My research has four objectives: 1) profiling of transcripts from hemocytes and fat body of both naïve and immune-challenged *M. sexta* larvae using pyrosequencing technology; 2) quantitatively analyze differentially expressed genes related to immunity and tissue specificity; 3) identify plasma proteins and peptides from naïve and immune-challenged *M. sexta* larvae using mass spectrometry; 4) semi-quantitatively analyze protein and peptide expression level changes before and after the challenge.

CHAPTER II

LITERATURE REVIEW

Insect immunity

The first line of defense in insect immunity is cuticle, a physiochemical barrier lining integument, gut, and trachea, which blocks most pathogens and prevents host from infection (Tzou et al. 2000). The cuticle is largely composed of proteins, lipids, and long N-acetyl-glucosamine fibrils known as chitin made by the basal epidermal cells (Feldhaar and Gross 2008). Since the foods insects ingest contain bacteria, immunity in their digestive tract is important for them to keep healthy. The foregut and hindgut are to some extent protected by a thin layer of cuticle. In the midgut, certain specialized cells produce a peritrophic membrane consisting of a network of chitin microfibils associating with the matrix of carbohydrates and proteins. The peritrophic membrane is permeable to digestive enzymes and nutrients but prevent most gut microbes from entry.

Once the physical barrier is breached, humoral and cellular immune responses are triggered to eliminate the invading pathogens (Jiravanichpaisal et al. 2006). Cellular responses involve different types of hemocytes in phagocytosis, nodulation and encapsulation (Lavine and Strand 2002). Plasmotocytes engulf pathogens via phagocytosis. Nodulation refers to the process of multicellular aggregation to entrap large amounts of bacteria in an extracellular material which is often followed by melanization. When pathogenic invaders, such as nematodes and parasitoid

wasp eggs, are too large to be engulfed, they are often entrapped by aggregated hemocytes that attach to them in a process known as encapsulation. Cellular responses are carried out by different types of hemocytes, whereas humoral responses are implemented by complex proteins (including enzymes and inhibitors) in body fluids. The latter include the production of AMPs and reactive intermediates of oxygen or nitrogen (Jiravanichpaisal et al. 2006). Both responses start with recognition of pathogens by pattern recognition proteins (PRRs) that bind polysaccharides on the surface of microbes.

Pathogen recognition is an essential step in immunity to detect dangerous non-self. PRRs recognize pathogen-associated molecular patterns (PAMPs) of different pathogens such as β-1, 3-glucan of fungi, lipopolysaccharide (LPS) of Gram-negative bacteria, lipoteichoic acid (LTA) of Gram-positive bacteria, and peptidoglycans of both types of bacteria. β-1, 3-glucan recognition proteins (βGRPs), Gram-negative bacteria binding proteins (GNBPs), peptidoglycan recognition proteins (PGRPs), hemolin, and C-type lectins are the most common PRRs which specifically bind to different PAMPs. Many of them are strongly induced after microbial infection (Janeway and Medzhitov 2002).

After pathogens are recognized, the pathogen-PRR complexes lead to activation of hemolymph proteases (HPs) in insect plasma (Ragan et al. 2009). The complicated HP cascade plays a critical role in extracellular signal transduction by activating the Toll pathway and pro-phenoloxidase (PPO). Phenoloxidase (PO) is a key enzyme in melanization. Besides HPs, extracellular signaling often involves another family of proteins called serine protease homologues (SPHs) which have a domain similar in sequence to HPs. Their activate-site residues, such as -Ser are substituted by inactive residues, such as Gly. Functions of SPHs are not well characterized; some SPHs seem to interact with clip-domain SPs and their substrates to regulate the signaling pathway. Serine proteases are regulated not only by SPHs, but also by their

inhibitors called serpins (serine protease inhibitors). Serpins form inactive covalent complexes with their target proteases, and that reduces damage caused by overreaction of immune responses.

Currently, three immunity-related signaling pathways have been studied in the model insect *Drosophila melanogaster*. The Toll pathway is activated by Gram-positive bacterial and fungal infection (Leclerc and Reichhart 2004). Through the serine protease cascade, spatzle precursor is cleaved and the activated spatzle subsequently binds Toll. Toll is a transmembrane receptor and lead to the activation of a series of intracellular signaling molecules. Among them, transcription factors Dorsal and DIF (Dorsal-related immune factor) translocate to the nucleus to induce the expression of antimicrobial genes (Pinheiro and Ellar 2006). In the Imd pathway, the invasion of Gram-negative bacteria and certain gram positive bacteria are recognized by PGRPs which form a complex with Imd, DREDD, and dFADD (Hu and Yang 2000; Naitza et al. 2002). Then, through the Imd pathway, Relish translocates into the nucleus to stimulate expression of other antimicrobial genes (Stoven et al. 2003). The Imd pathway also regulate genes involved in wound repair and stress response (Silverman et al. 2003). The Hop pathway is the least characterized pathway that modulates and links humoral and cellular responses (Agaisse et al. 2003).

Insect antimicrobial peptides

Antimicrobial peptides (AMPs) are effector proteins in immune processes and directly kill a spectrum of pathogens. They widely exist in various organisms ranging from prokaryotes to plants, arthropods and vertebrates (Bulet et al. 2004). In insects, hundreds of different AMPs have been reported and these AMPs share some common features. They usually have a low molecule weight (less than 10 kDa), carry positive net charge under physiological conditions, and are mostly hydrophobic (Reddy et al. 2004). These peptides are usually absent or present at low concentrations in naïve insects. However, after the insects are infected by microorganisms,

association of host recognition molecules with pathogens triggers the Toll and Imd pathways which results in large increases in AMP gene expression. Fat body, a tissue analogous to human liver and adipose, is a major source of plasma AMPs that kill the invading pathogens.

AMPs are categorized into several groups: alpha-helical peptides, disulfide-stabilized peptides, proline-rich peptides, glycine-rich peptides, and others (Meister et al. 1997). In *M. sexta*, at least five AMPs have been identified including attacin, cecropin, moricin, gloverin, and lebocin (Rayaprolu et al. 2010a). Attacin was first isolated from *Hyalophora cecropia* and it increases permeability of the outer membrane of Gram-negative bacteria and controls, at the transcription level, synthesis of outer membrane proteins associated with bacterial growth (Hultmark et al. 1983). Cecropin, first isolated from *H. cecropia*, is active against both Gram-positive and Gram-negative bacteria. It penetrates the plasma membrane of pathogens, disturbs the electrochemical gradient, and kills the bacteria (Durell et al. 1992). Moricin was first isolated from *Bombyx mori*, and it is highly active against Gram-positive and Gram-negative bacteria but weakly active against certain yeasts (Hara and Yamakawa 1995). It affects permeability of bacterial cytoplasmic membrane. Like attacin, gloverin is another Gly-rich AMP against Gram-negative bacteria which interacts with lipopolysaccharide (LPS) in the bacterial outer membrane to inhibit the synthesis of outer membrane proteins (Axen et al. 1997). Lebocin, a Pro-rich AMP, first isolated from *Bombyx mori* acts on bacterial membranes and glycosylation seems vital for its antibacterial activity. It shows weak antibacterial activity under physiological conditions and may function in synergism with cecropin D (Rayaprolu et al. 2010b).

Next-generation sequencing

In the past thirty years, automated Sanger sequencing has been widely used in biological sciences (Sanger et al. 1977). Using this “first-generation” method, scientists have made many

momentous achievements, like the completion of human genome project (International Human Genome Sequencing 2004). About seven years ago, a novel sequencing technology was introduced. Compared with the conventional sequencing technology, this next-generation sequencing (NGS) technology produces an enormous volume of data in a cost effective way. Since its arrival in the marketplace, NGS has revolutionized the way we perform scientific research in life science.

Currently there are several commercially available NGS technologies including Roche/454, Illumina/Solexa, and Applied Biosystems SOLiD. These technologies use different sequencing chemicals and methods but their procedures are similar which include template preparation, sequencing and imaging, and assembly or genome alignment (Metzker 2010). Compared with the Sanger sequencing, NGS does not require the *in vivo* cloning step and it allows a large number of DNA fragments to be sequenced simultaneously in one plate, which greatly enhanced efficiency. The first NGS was introduced by Roche/454, known as "pyrosequencing technology" (Margulies et al. 2005). It utilizes emulsion PCR to amplify template DNA sequences. Amplified DNAs are then loaded on a plate containing millions of tiny wells. Pyrosequencing reactions are simultaneously carried out in each well. Compared with the other NGS, pyrosequencing provides much longer reads and the run time is relatively short (Mardis 2008; Morozova et al. 2009; Metzker 2010). However, the reagent cost and error rates for homopolymer repeats are relatively high. In 2006 Illumina/Solexa was introduced which is now the most widely used NGS platform. It amplifies the template DNA using bridge PCR and generates much more data per run at a relatively low cost. However, it renders low multiplexing capability of samples (Metzker 2010). The Applied Biosystems (ABS) SOLiD, released in October of 2007, stands for "Sequencing by Oligonucleotide Ligation and Detection". Like pyrosequencing, ABS SOLiD uses emulsion PCR to amplify template DNA. But the sequencing method is based on ligation with dye-labeled oligonucleotides. This method yields a lower sequencing rate but requires longer sequencing time.

Besides the three popular NGS methods, some single-molecule sequencing approaches that do not require template DNA amplification have been developed. These "third-generation technologies" have the potential to reduce sequencing cost more steeply than NGS (Morozova et al. 2009).

The advent of NGS has stimulated researches in variant areas including resequencing targeted regions of interest or whole genome, *de novo* genome sequencing, transcriptomic study (RNA-Seq), and metagenomics studies. NGS has greatly advanced life science research and posed fierce bioinformatics challenges.

Transcriptomic study

The transcriptome is the complete set of messenger RNA (mRNA) and noncoding RNA transcripts in an organism, tissue or even cell. Being the dynamic link between genome and proteome, transcriptome has drawn great interest from scientists conducting research in different areas of biological science. The earliest attempts to study cellular transcriptome included inspections of total cellular RNA from different tissues, or under different physiological state for the presence and quantity of transcripts of interest (Morozova et al. 2009). The first candidate gene-based studies utilized the Northern blot analysis method, a low-throughput technology that requires the use of radioactivity and a large amount of RNA samples (Alwine et al. 1977). Because of its own limitation, the Northern blot method failed to detect the rare transcripts or those with unknown sequences. With the development of the quantitative real time polymerase chain reaction (qRT-PCR) method (Becker-Andre and Hahlbrock 1989), the experimental throughput was increased while the required quantity of input RNA was reduced. However, even after decades, the throughput of such approaches does not exceed the order of hundreds of known transcripts at a time which could not satisfy the large scale transcriptomic studies (VanGuilder et

al. 2008). In the mid-1990s, the advent of microarray technology replaced the single-gene approaches by allowing simultaneous characterization of expression levels of thousands of known transcripts (Schena et al. 1995). The microarray method is based on the hybridization of labeled cDNA samples to immobilized high-density DNA probes in a collection of microscopic wells on a plate. Each DNA probe in a well represents a specific DNA sequence which is usually a unique gene. The signals after hybridization will be monitored and used to measure differential expression levels for large numbers of genes. Because of its high throughput, microarray technology has been a dominating method in transcriptomic studies ever since it was invented (Pozhitkov et al. 2007). However, even after years of improvement, microarray technology still has several limitations including its reliance on the existing knowledge of genome sequence for the probe preparation, high background noise owing to cross-hybridization, and a limited detection dynamic range due to the background noise and saturation of signals.

Compared with microarray technology, sequencing-based approaches have the main advantage of the independence of the genome sequence since these methods could directly determine the cDNA sequences. In the initial sequencing-based transcriptomic studies people were using Sanger methods to sequence cDNA or EST (expressed sequence tag) libraries (Boguski et al. 1994; Gerhard et al. 2004). But this method is relatively expensive, low throughput and does not provide quantitative information. Then tag-based methods were developed to overcome these limitations, including serial analysis of gene expression (Velculescu et al. 1995), cap analysis of gene expression (Kodzius et al. 2006), and massively parallel signature sequencing (Brenner et al. 2000). These methods solved the problem of low throughput and achieved precise quantitative analysis. However, they are still based on the Sanger method which is expensive and requires a laborious cloning procedure.

Recently the development of massively parallel sequencing technology has revolutionized the way people work on biological science. It has provided a new method for mapping and

quantifying transcriptomes called RNA-Seq (RNA sequencing) (Wang et al. 2009). This method has obvious advantage over the existing methods and it has been applied in different organisms including human and other animals, plants, and bacteria using Illumina/Solexa, Roche/454, and Applied Biosystems SOLiD sequencing technologies. In general, a set of transcripts is converted to a library of cDNA fragments with adaptors attached to one or both ends. These cDNA molecules, with or without amplification are then sequenced in a high throughput manner for short read sequences. The resulting enormous amount of read sequences are either aligned to the reference genome if it is available or *de novo* assembled if the genome has not been sequenced. Compared with the previous transcriptomic study approaches, RNA-Seq has the following advantages which make it currently the most powerful transcriptomic study method. RNA-Seq can simultaneously provide information about tens of thousands of transcripts with their sequences and expression levels in a cost-effective way. Unlike microarray, RNA-Seq does not rely on the genome sequence which makes it an ideal tool for transcriptomic studies on non-model organisms. Moreover, since the quantitative analysis for RNA-Seq is based on the numbers of mapped reads, the background noise is very low compared to the microarray method which relies on the fluorescent signal after hybridization. In addition, RNA-Seq has a much larger dynamic range of expression levels over which transcripts can be detected. In conclusion, this method offers single-base resolution for annotation and "digital" expression levels for quantitative analysis while the cost is much lower than the conventional methods. Currently, the major challenges for RNA-Seq are the cDNA library construction and the development of bioinformatics tools.

Mass spectrometry

Mass spectrometry (MS) is a well-known analytical tool for measuring mass-to-charge ratios of small charged particles. It has been widely used in both academia and industry for various

purposes such as drug discovery, diagnostics, and bio-analyses (Feng et al. 2008). Mass spectrometers are mainly composed of three parts including ionization source, mass analyzer and detector. Ionization source is the part where target materials are ionized. In the early stage of MS, there were few ionization methods which largely relied on the electron impact. For these methods, covalent bonds in the molecules could be easily broken. Thus the application of MS was limited and it was seldom used for biological samples analysis. Then "soft" ionization methods started to emerge including matrix assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp 1988) and electrospray ionization (ESI) (Fenn et al. 1989). In MALDI, the target molecules are co-crystallized with a protective matrix. A pulse of laser light is used to force molecules into gas phase and ionized them. Another commonly used ionization method, ESI is considered as the "softest" of all the ionization methods. Even non-covalent bond such as interactions between proteins could be preserved using this method. In addition, since sample molecules are introduced in solution, ESI can be coupled with liquid chromatography and capillary electrophoresis for enhanced mass analysis. After the ionization process, molecules are separated and analyzed in mass analyzers according to their mass-to-charge ratios. Most commonly used mass analyzers include quadrupole, quadrupole ion trap (QIT), linear ion trap (LIT), time of flight (TOF), Fourier transform ion cyclotron resonance (FTICR) and Orbitrap. TOF is the most straight forward method which uses an electric field to accelerate ionized molecules and measures the time for them to reach the detector. MALDI-TOF has been widely utilized for protein identification using the peptide mass fingerprint method (Pappin et al. 1993). Orbitrap is a recently developed mass analyzer and it takes the advantage of high resolving power of each detected peak and superb accuracy (Hu et al. 2005).

Tandem MS or MS/MS has been extensively used in different proteomic studies such as protein identification, peptide sequencing, and characterization of post-translational modification. Generally tandem MS contains two mass analyzers which are separated by a collision cell.

Selected molecules with certain mass-to-charge ratios from the first mass analyzer will be fragmented in the collision cell and these fragments will be further analyzed in the second mass analyzer. The sequential analysis of the ionized molecules and their fragments can be achieved either by two mass analyzers in a row (in space) or by analysis in one mass analyzer at different times (in time) (Hernandez et al. 2006).

Proteomic and peptidomic study

The proteome is the full set of proteins comprising the structural, metabolic, and regulatory machinery of a cell, tissue, or organism. With the advancement of biological technologies, proteomic studies have drawn more and more people's attention because of the important and direct roles proteins play in a living organism. Currently most proteomic studies are based on protein identification of biological samples, differential expression of proteomes under different conditions or in different tissues, and protein-protein interactions. Different methods have been developed for proteomic studies including mass spectrometry based methods, protein microarray, two-hybrid screening of libraries for protein-protein interaction, and high-throughput protein expression and structural characterization.

MS-based methods are most extensively used in variant proteomic studies such as protein identification, protein modification, differential protein expression, and protein-protein interaction. Pre-fractionation is an optional step in MS and it plays an essential role in most proteomic studies such as protein identification where target proteins are in complex mixtures. Two-dimensional (2D) gel electrophoresis is a commonly used pre-fractionation method because it is highly compatible with peptide mass fingerprinting and differential expression could be directly observed on the gel (Rabilloud 2002). However, this method is labor intensive, low-throughput, and the reproducibility is not ideal. On the other hand, liquid chromatography (LC)

turns out to be an efficient tool to separate proteins in a complex mixture. It has been widely used in proteomic studies especially after the advent of ESI since ESI allows the direct ionization of liquid. LC coupled with MS, or LC-MS, is one of the most popular configurations for protein identification and peptide sequencing (Feng et al. 2008). Multidimensional protein identification technology (MudPIT) was introduced in 2000 in which multidimensional LC was coupled in-line with ESI-MS/MS for high-throughput protein identification (Wolters et al. 2001). Soon after that several off-line pre-fractionation methods were developed and applied in proteomic studies.

Peptides are small polymers of amino acids or low molecular weight proteins. The peptidome is the whole set of peptides in a cell, tissue or organism. Peptidomic approaches are most applied in neuroscience research and biomarker discovery. There are also some peptidomic studies on AMPs, which are important effectors of immune responses (Brown et al. 2009). Compared with proteomic studies, peptidomic samples are often prepared using specific methods since peptides are too small for analysis on 2D gels and too large for *de novo* sequencing (Baggerman et al. 2004). Since large abundant proteins sometimes can mask the signals of peptides, different methods, such as heating (Ziganshin et al. 2011) or organic solvent precipitation (Merrell et al. 2004) were used to get rid of large proteins while retaining small peptides for peptidomic studies.

Quantitative proteomic study

Early proteomic studies were merely lists of proteins identified in the given biological samples. As the development of proteomics methods, protein quantitation has become an important aspect of proteomic studies. At an earlier stage of quantitative analysis, 2D gels were used to visualize the differences between protein samples from different physiological conditions or tissues. Although it is possible to visualize more than 1000 proteins on single 2D gels, protein separation and differential analysis are not coupled with identification of proteins in the spots of

interest (Schulze and Usadel 2010). Thus, it requires an additional step of MS-based identification of target proteins. In addition, accurate reproducibility has always been a problem for this 2D-gel based quantitative analysis. The advent of difference gel electrophoresis (DiGE) technology (Unlu et al. 1997) which allows the separation of proteins from two different samples on one single gel has significantly improved the quantitative accuracy of 2D gels. However, the reagent is relatively expensive and this method also requires the additional step of protein identification using MS.

Many MS-based quantitative proteomic methods have been developed which are divided into two groups: absolute and relative quantitative analysis methods (Lau et al. 2007). Absolute quantitative analysis (e.g., AQUA, QconCAT, and SISCAPA) tries to measure the absolute protein level using internal standards with known concentrations. Relative quantitative proteomic methods measure relative abundance ratios between two or more samples under different physiological conditions. Two types of methods are adopted for such analyses, including stable isotope labeling and label-free methods. For the labeling methods, samples after different treatment or from tissues are labeled with different isotopes and mixed before the chromatographic and MS analysis. Since the incorporated isotope could only change the mass of proteins in the labeled samples, proteins from different samples could be distinguished in the mass spectra. Isotopes could be incorporated metabolically into all proteins *in vivo* or a chemical reagent could be used to label the proteins *in vitro* (Schulze and Usadel 2010). Stable isotope labeling with amino acids in cell culture (SILAC) is a popular *in vivo* labeling method where cells are cultured in a medium containing a heavy amino acid, and compared to one containing standard light variant (Ong et al. 2002). The *in vitro* methods include isotope coded affinity tag (ICAT) (Gygi et al. 1999) and multiplexed isobaric tagging technology (iTRAQ) (Ross et al. 2004) where a modifying group is added to a certain amino acid side group (ICAT) or free amines (iTRAQ). While stable isotope labeling approaches enable more sensitive detection of

differential expression than label free methods, they also have certain limitations. The expense of stable isotope and expertise required for the labeling methods largely limit their widespread use. In addition, comparison between more than two samples simultaneously using labeling methods is hampered by the technical limit. Thus label-free methods have gained much popularity in recent years due to their relative ease of use and general applicability to a wide range of proteomic studies.

Currently there are two fundamentally different label-free methods widely used in proteomic studies: ion intensity and spectral counting methods (Lundgren et al. 2010). The first approach is based on the measurement of ion intensity of peptides associate with a given protein. To improve the accuracy of quantitation using ion signal intensity methods, it is necessary to run multiple sampling of chromatographic peak by survey mass spectra at the expense of MS/MS experiments. Fewer MS/MS experiments could decrease the chance of peptide identification which potentially decreases the number of identified proteins. Thus optimizing the instrument settings for protein abundance estimation using ion signal intensity method can result in the reduction of total number of identified proteins. On the other hand, the spectral count method uses the total number of fragmentation spectra that map to peptides of a certain protein for quantitative analysis. The rationale behind this method is that more abundant proteins or peptides are more frequently detected than the low abundant proteins or peptides. Since this method relies on the number of MS/MS spectra mapped to peptides, the optimization for this method also favors optimization for total protein identification (Bantscheff et al. 2007). Based on these advantages, the spectral counting method has been used in a wide range of diverse comparative proteomic studies. However, this method showed its low accuracy especially for low abundant proteins. According to previous reports, mean spectral counts of less than five are supposed to be unreliable when working with a small number of replicates (Old et al. 2005). Thus, the spectral counting method is still considered to be a semi-quantitative approach.

CHAPTER III

MATERIALS AND METHODS

Insect rearing, bacterial injection, RNA isolation, and library construction

M. sexta eggs, purchased from Carolina Biological Supply, were hatched and reared on an artificial diet as described by Dunn and Drake (1983). Each of day 2, 5th instar larvae (60) was injected with a mixture of *Escherichia coli* (2×10^7 cells), *Micrococcus luteus* (20 mg) (Sigma-Aldrich), and curdlan (20 mg, insoluble β-1,3-glucan from *Alcaligenes faecalis*) (Sigma-Aldrich) in 30 μl H_2O. Total RNA samples were extracted from induced hemocytes (IH) and fat body (IF) 24 h later using TRIZOL Reagent (Life Technologies Inc.). Control hemocyte (CH) and fat body (CF) RNA was prepared from day 3, 5th instar naïve larvae (60). PolyA+ RNA was separately purified from the total RNA samples (1.0 mg each) by binding to oligo(dT) cellulose twice in the Poly(A) Purist™ Kit (Ambion). First strand cDNA was synthesized using mRNA (5.0 mg), random dodecanucleotides (100 pmol), and SuperScript™ III reverse transcriptase (1000 U, Life Technologies Inc.). RNase H treatment, second strand synthesis, and gap joining were performed according to the published protocol (Zou et al. 2008). After shearing via nebulization, the four samples were end-repaired (Roe 2004) and ligated to double-stranded adaptor A and biotinylated adaptor B (Margulies et al. 2005).

PCR amplification, pyrosequencing, and sequence assembly

The cDNA with adaptor B attached on one or both ends was isolated using streptavidin-coated magnetic beads, end repaired, and quantified on an Agilent 2100 Bioanalyzer (Agilent Technologies). Diluted DNA molecules, individually captured by beads, were amplified using emulsion PCR with the two primers complementary parts of A and B adaptors (Margulies et al. 2005). After removal of the second strand and empty beads, the sequencing primer identical to another part of A adaptor was used for sequencing. Two full plates were run with one-half plate for each library on a 454 GSFLX pyrosequencer (Roche Applied Science) using long-read GSFLX Titanium chemistry. Reads were assembled separately for each library (CF, CH, IF, IH) and collectively (CIFH) using Newbler Assembler (Roche Applied Science) into five datasets: CF, CH, IF, IH, and CIFH (Fig. 1). To improve coverage and quality of the sequence sets, data from our previous run on a 454 GS20 (Zou et al. 2008) were assembled into two datasets (06 for the 2006 data and 06CIFH for the 2006 and 2009 data) using the updated Newbler software. The resulting contigs and singletons from the seven datasets were compared against the NCBI nr/nt and KEGG databases using BLASTN, BLASTP, and BLASTX with a maximum E-value of 1×10^{-5}. For the combined library CIHF, numbers of CH, CF, IH, and IF reads assembled into each contig were extracted from the standard Newbler Assembler output and tabulated using Microsoft Excel.

Read normalization and ratio calculation

Based on frequencies of several commonly used standards in each of the four libraries (e.g., number of rpS3 reads in CH/number of total reads in CH), a set of six ribosomal protein genes were selected as internal standards, which had high total read numbers and low coefficients of variation (i.e., SD/mean) in their frequencies. The sums of their read numbers for specific libraries, or library normalization factors (LNFs), which already reflected the differences in library sizes, were directly used to calibrate other read numbers in the corresponding libraries. For a specific contig in CIFH, its relative abundance (RA) in libraries X and Y is defined as: RAx/y =

(actual read # in library X/LNFx)/(actual read # in library Y/LNFy). In case read # in library Y is zero, adjusted read number (ARN), instead of RA, is calculated as: ARNx = actual read # in library X × LNFy/LNFx. Some of the contigs in CIFH, whose RAs or ARNs are above certain thresholds, are categorized into UP, DN, HC, and FB: UP for up-regulated genes ($RA_{IF/CF} > 5$, $RA_{IH/CH} > 8$, $ARN_{IF} > 10$ when $RN_{CF} = 0$, or $ARN_{IH} > 10$ when $RN_{CH} = 0$), DN for down-regulated genes ($RA_{CF/IF} > 10$, $RA_{CH/IH} > 10$, $ARN_{CF} > 20$ when $RN_{IF} = 0$, or $ARN_{CH} > 20$ when $RN_{IH} = 0$), HC and FB for genes preferentially expressed in hemocytes ($RA_{IH/IF} > 40$, $RA_{CH/CF} > 40$, $ARN_{IH} > 80$ when $RN_{IF} = 0$, or $ARN_{CH} > 80$ when $RN_{CF} = 0$) and fat body ($RA_{IF/IH} > 100$, $RA_{CF/CH} > 100$, $ARN_{IF} > 200$ when $RN_{IH} = 0$, or $ARN_{CF} > 200$ when $RN_{CH} = 0$), respectively.

Sequence extension, database search, and domain prediction

CIFH contigs in UP, DN, HC, and FB categories were used as queries to search local databases of 06CIFH_contigs/singletons, UK_gut - contigs by BLASTN (http://darwin.biochem.okstate.edu/blast/blast.html). The *M. sexta* midgut ESTs (i.e., UK_gut_contigs) (Pauchet et al. 2010) were kindly provided by Dr. Yannick Pauchet at University of Exeter, UK. The search results were used to extend the CIFH contigs or, in some cases, fill a gap between two contig sequences. The extended sequences were searched against NCBI using BLASTX as described above. For UP CIFH contigs lacking BLAST hits, a set of more stringent conditions was applied to select sequences for further analysis: a) $RA_{IF/CF} > 15$, $RA_{IH/CH} > 15$, $ARN_{IF} > 30$ when $RN_{CF} = 0$, or $ARN_{IH} > 30$ when $RN_{CH} = 0$, b) total read number > 70, and c) GC content ≥ 35% (i.e., coding region-including). Open reading frames in a chosen contig were examined for leader peptide using SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/), which is commonly found in proteins highly induced upon immune challenge (Jiang 2008). The polypeptide sequences were

then analyzed to detect conserved domain structures by SMART (http://smart.embl-heidelberg.de/smart/set_mode.cgi).

Insect rearing, pathogen injection, and plasma collection

M. sexta eggs, obtained from Dr. Michael Kanost's lab at Kansas State University, were hatched and reared on an artificial diet as described by Dunn and Drake (1983). Each of day 2, 5th instar larvae was injected with a mixture of *Escherichia coli* (1.3×10^7 cells), *Micrococcus luteus* (13 mg) (Sigma-Aldrich), and curdlan (13 mg, insoluble β-1,3-glucan from *Alcaligenes faecalis*) (Sigma-Aldrich) in 20 μl H_2O. As a negative control, larvae at the same stage were injected with 20 μl of sterile phosphate buffer saline. At 24 h after injection, prolegs of the insects were cut and hemolymph was collected to clean tubes containing phenylthiourea (PTU) and paraaminobenzamidine to the final concentrations of 0.01% and 1mM, respectively. The tubes were centrifuged at 4000*g* for 5 min at 4 ℃ to precipitate hemocytes, the supernatants transferred to clean tubes, and equal amount of plasma samples were pooled from three individual insects as one sample. Three biological replicates of both immunized and negative control were prepared. Forty microliters of pooled plasma samples (3 control and 3 induced) was pipetted to a clean tube mixed with 8 μl of 6 × SDS sample buffer (7.0 ml 0.5M Tris-HCl, 3.0 ml glycerol, 1.0 g SDS, 50mM TECP, 3mg bromphenol blue), respectively. The rest of plasma samples were aliquoted (100 μl/tube) and stored at -80 ℃.

Preparation of plasma samples for peptidomic analysis

The control and induced plasma samples were thawed on ice and mixed with equal amount of chilled acetonitrile by vortexing. After incubation on ice for 2h, the suspensions were

centrifuged at 10000*g* for 10 min at 4°C. The supernatants (150 µl, control and induced, each with three biological replicates) were moved to new tubes, dried, and redissolved in 8 M urea in 100 mM Tris-HCl pH-8.5 at room temperature. Protein concentrations of the samples were quantified using BCA assay (Bio-Rad). After mixing with 5 mM TCEP at room temperature for 20 min, 1/20-volume of 200 mM iodoacetamide was added to each tube and alkylation was allowed to proceed for 15 min in the dark at room temperature. The sample was then diluted with four volumes of 100 mM ammonium bicarbonate and digested with 8 µg/ml trypsin for 4 h at 37 ℃. Digested samples were acidified with 5% formic acid prior to mass spectrometry analysis.

Preparation of plasma proteins for proteomic analysis

The protein samples (three control and three induced) in 1 × SDS loading buffer were thawed and heated for 5 min at 95 ℃. Based on their protein concentrations, 60 µg of each sample was loaded into each well of a 4-15% linear gradient SDS-polyacrylamide gel (Bio-Rad). After electrophoresis at constant current of 30 mA for 45 minutes, the gel was stained with Coomassie brilliant blue (CBB) R-250 and then destained in 30% methanol and 10% acetic acid. Each lane was sliced into nine pieces and these gel slices were further destained by extensive washing using 50% acetonitrile in 50 mM ammonium bicarbonate, pH 8.0, dehydrated with 100% acetonitrile, and dried briefly. Dried gel pieces were rehydrated with 10 mM TCEP in 50 mM ammonium bicarbonate and reduced for 1 h at room temperature. After incubation, the reducing buffer was replaced with 55 mM iodoacetamide in 50 mM ammonium bicarbonate to alkylate Cys for 1 h at room temperature in the dark. Samples were then rinsed with ammonium bicarbonate, dehydrated with acetonitrile, and rehydrated/infiltrated with sequencing-grade trypsin solution containing 8 µg trypsin per ml of 50 mM ammonium bicarbonate. After digestion for 6-16 h at 37 ℃, the

trypsinolytic peptides were extracted with 1% TFA, and used for subsequent analysis by mass spectrometry.

LC-MS/MS analysis

Samples were analyzed on a hybrid LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific) coupled to a New Objectives PV-550 nanoelectrospray ion source and an Eksigent NanoLC-2D chromatography system. Peptides were analyzed by trapping on a 2.5 cm pre-column and analytical separation on a 75 μm ID fused silica column, using a vented column configuration packed in house with 10-cm of Magic C18 AQ and terminating with an integral fused silica emitter pulled in house. Peptides were eluted using a 5-40% AcCN/0.1% formic acid gradient performed over 40 min at a flow rate of 300 nL/min. During elution, samples were analyzed using Big Six methodology, consisting of one full-range FT-MS scan (nominal resolution of 60,000 FWHM, 300 to 2000 m/z) and six data-dependent MS/MS scans performed in the linear ion trap mode. MS/MS settings used a trigger threshold of 8,000 counts, monoisotopic precursor selection (MIPS), and rejection of parent ions that had unassigned charge states, were previously identified as contaminants on blank gradient runs, or that had been previously selected for MS/MS (dynamic exclusion at 150% of the observed chromatographic peak width). Column performance was monitored using trypsin autolysis fragments (m/z 421.76), and via blank injections between samples to assay for contamination.

Protein database construction

The protein database contains sequences from four sources: 1) *M. sexta* protein sequences downloaded from NCBI 2) translated DNA sequences from 06CIFH09 dataset in our

transcriptomic project; 3) translated DNA sequences from previously reported *M. sexta* gut transcriptomic sequences; 4) translated DNA sequences from *M. sexta* genome CUFFLINK sequences.

Protein identifications

Centroided ion masses were extracted using the extract_msn.exe utility from Bioworks 3.3.1 and were used for database searching with Mascot v2.2.04 (Matrix Science) and X! Tandem v2007.01.01.1 (www.thegpm.org). Searches were conducted using the following search parameters:

Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.2.04) and X! Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1)). X! Tandem was set up to search a subset of the Msexta_040612 database also assuming trypsin. Mascot was set up to search Msexta_040612 database (unknown version, 1306670 entries) assuming the digestion enzyme is trypsin. Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 5.0 PPM. S-carbamoylmethylcysteine cyclization (N-terminus) of the N-terminus, oxidation of Methionine, N-Formylation of the N-terminus, acetylation of the n-terminus, iodoacetamide derivative of cysteine and acrylamide adduct of cysteine were specified in X! Tandem and Mascot as variable modifications.

Scaffold (version Scaffold_3.4.5, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 80.0% probability as specified by the Peptide Prophet algorithm (Keller et al. 2002). Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least two identified peptides. Protein probabilities

were assigned by the Protein Prophet algorithm (Nesvizhskii et al. 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. A reversed database served as a decoy protein database to determine the false discovery rate (FDR) as described before (Elmore et al. 2012).

Sequence trimming and statistical analysis

Protein sequences from translated DNAs were manually trimmed – only the translated region from the starting Met to stop codon was kept. For those without starting Met or stop codon, the longest translated sequences were kept. BLASTP was used to determine the functions of these proteins. Based on the total spectral counts in each sample, the spectral counts for each protein were normalized. The normalized numbers were used for Student's t-test to see if there was significant difference (p-value < 0.05) between induced and control samples. We defined I/C as the ratio of the average normalized spectral counts for induced sample (I) over the counts for control sample (C). In case that the average number for C was 0, we used 1 as the divisor for the calculation. We regarded the proteins with significant changes after immune challenge and the normalized spectral counts increased at least one fold (p-value < 0.05, I/C ≥ 2) or without significant changes but spectral counts increased at least four folds (p-value ≥ 0.05, I/C ≥ 5) as up-regulated proteins. Similarly, we regarded those with p-value < 0.05 and I/C ≤ 0.5 or p-value ≥ 0.05 and I/C ≤ 0.2 as down-regulated proteins.

CHAPTER IV

RESULTS AND DISCUSSION

Identification of differentially regulated genes in *M. sexta* by pyrosequencing

In order to find immunity-related genes expressed in fat body or hemocytes based on their expression profiles, we isolated mRNA of these two tissues from naïve and bacteria-injected larvae of *M. sexta*, a lepidopteran insect whose genome sequence has not yet been determined. Using random dodecanucleotide primers that annealed to different regions of mRNA molecules, we generated four cDNA libraries: CF, CH, IF and IH. To facilitate assembly and ORF identification, we adopted long-read Titanium chemistry to sequence these libraries on a 454 GS-FLX pyrosequencer and obtained a total of 227,302 reads from CF, 647,587 reads from CH, 405,739 reads from IF, and 541,024 reads from IH (Table 1). The total number of reads from two plates (0.5 plate per library) was 1,821,652, which was 19.1-fold higher than that from one plate (95,358 reads) sequenced on a 454 GS20 in 2006 (Zou et al., 2008). There also was a substantial increase in average read length from 185 bp to 289 bp, but that was still much lower than what the manufacturer claimed (>400 bp) (http://454.com/about-454/index.asp).

We assembled the reads into five datasets: CF, CH, IF, IH, and CIFH (Fig. 1). The first four each came from its respective library, whereas the 5th dataset was assembled from the 1,821,652

reads in the four libraries sequenced in 2009. In CF, CH, IF and IH, 84.1~86.6% of the total reads were incorporated into contigs at average sizes of 764~832 bp; In CIFH, 1,677,738 (92.1%) of the 1,821,652 reads were assembled to 19,020 contigs at 923 bp per contig (Table 1). These assemblies were better than the previous one, that integrated 69,429 (72.8%) of the 95,358 reads into 7,231 contigs at an average length of 300 bp (Zou et al. 2008). To improve the transcriptome coverage, we used the latest version of Newbler to re-analyze the previously generated flowgrams, assembling 64,874 of the reads into 1,471 contigs with an average of 391 bp per contig in the 6th dataset ("06") (Table 1). Finally we assembled all the source libraries (2006, CF, CH, IF, and IH) into "06CIFH", which contained 19,504 contigs (average size: 911 bp) and 120,670 singletons.

We used numbers of CH, CF, IH, and IF reads for each CIFH contig to identify differentially regulated genes. Since read numbers depended on library sizes and needed to be normalized against control genes, we compared frequencies of commonly used internal standards in each of the four libraries and found that six ribosomal protein genes (rpS2-rpS5, rpL4 and rpL8) showed low coefficients of variation (<30%) and high total read numbers (>1,000). So, we used the sums of their read numbers 825 (CF), 3,980 (CH), 1,618 (IF), and 3,352 (IH) as library normalization factors (LNFs) to calibrate read numbers and calculate relative abundances (RAs) (Fig. 1). Based on the RA values, 920 or 4.84% of the 19,020 contigs in CIFH were categorized into four groups: UP and DN for up- and down-regulated genes upon immunization; HC and FB for genes preferentially expressed in hemocytes and fat body, respectively.

Sequence analysis and function prediction of UP genes

We discovered 528 CIFH contigs whose $RA_{IF/CF}$ or $RA_{IH/CH}$ was greater than 5 and 8, respectively, or whose adjusted number of IF (or IH) reads (ARN) was >10 when the CF (or CH) read was zero – the adjustment for IF was read # × 825/1618 and that for IH was read # ×

3980/3352. As we anticipated, these contigs encoded polypeptides either identical to immunity-related proteins previously isolated from *M. sexta* (e.g. hemolin), or similar in sequence or domain structure to defense factors found in other insects (e.g. *Spodoptera frugiperda* X-tox), or related to proteins previously not known to play a role in immune responses (e.g. carboxylesterase), or having no significant sequence similarity to known proteins. In the following, we describe these contigs in the order of their putative immune functions.

A. Recognition of molecular patterns associated with microbes

To reinforce detection of invading organisms, certain pattern recognition receptors (PRRs) are synthesized in insects at higher levels after the initial encounter of foreign entities or abnormal host components. For instance, we found an Ig-domain protein (contig 03442) had an $RA_{IF/CF}$ of 748.5 (Table 2). This protein, *M. sexta* hemolin, was reported previously as a highly inducible PRR that recognizes LPS of Gram-negative bacteria (Ladendorff and Kanost 1991). Other PRRs included *M. sexta* immulectin-2 (contig 04775, $RA_{IF/CF}$: 45.4), immulectin-4 (contig 04808, ARN_{IF}: 217.2), peptidoglycan recognition protein-1 (PGRP1) (contig 13190, ARN_{IH}: 10.7; contig 14104, $RA_{IF/CF}$: 6.3; ARN_{IH}: 15.4), PGRP2 (contig 14700, residues 1-96, ARN_{IF}: 93.3; contig 14752, residues 98-196, ARN_{IF}: 60.2), β-1,3-glucan recognition protein-2 (βGRP2) (contig 01326, $RA_{IF/CF}$: 9.7; $RA_{IH/CH}$: 9.2). These data not only confirmed the published PRR sequences but also provided information on fold increases in their transcript abundances. Contig 06630 ($RA_{IF/CF}$: 11.2), 58% identical to *M. sexta* immulectin-3 (Yu et al., 2005) in residues 1-276, represented a previously unknown immulectin discovered based on its induced expression as well as sequence similarity. Newly identified PRRs also included PGRP3 (contig 00575, $RA_{IF/CF}$: 44.0), homologs of *Bombyx mori* PGRP5 (contig 11845, $RA_{IH/CH}$: 10.1) and PGRP-S6 (contig 08467, ARN_{IF}: 57.6), homologs of *B. mori* CTL10 (contig 14515, residues 54-182, $RA_{IF/CF}$: 8.7; contig 15639, residues 233-308, $RA_{IF/CF}$: 5.6; contig 11458, residues 54-306, ARN_{IF}: 28.0), homolog of *B. mori* Gram-negative binding protein (contig 08247, $RA_{IH/CH}$: 10.7) (Tanaka et al.

2008), LPS-binding leureptin (contig 15857, $RA_{IH/CH}$: 10.7) (Zhu et al. 2010), Ig domain-containing hemicentin-1 (contig 00131, $RA_{IF/CF}$: 6.4) and -2 (contig 14278, $RA_{IF/CF}$: 8.7) (Vogel and Hedgecock 2001). Therefore, expression profiling and sequence similarity together provided a powerful tool to discover process-related genes without a priori genome sequence.

B. Extracellular signal transduction and modulation

Hemolymph proteinases (HPs) in insect plasma form enzyme cascades to detect pathogen-PRR complexes and activate precursors of defense proteins (e.g. PO, spätzle, serine proteinase homolog (SPH), and plasmatocyte-spreading peptide (PSP) by limited proteolysis (Jiang and Kanost 2000). We found eight HPs in the UP list: *M. sexta* HP7 (ARN_{IF}: 11.2), HP9 ($RA_{IH/CH}$: 28.5), HP17 (ARN_{IH}: 15.4), HP18 ($RA_{IH/CH}$: 40.4), HP19 ($RA_{IF/CF}$: 7.1), HP22 ($RA_{IF/CF}$: 5.1), proPO-activating proteinase-2 (PAP2) (ARN_{IF}: 50.0), and PAP3 (ARN_{IF}: 22.9) (Table 3). Expression profiles associated with the immune inducibility agreed well with the RT-PCR and northern blot results published earlier (Jiang et al. 2003a; Jiang et al. 2003b; Jiang et al. 2005). We also found six contigs encoding isoforms of a strongly inducible protein (scolexin) that contained all three catalytic residues of S1A proteinases but did not display any amidase activity (Finnerty et al. 1999). The high ratios and read numbers of these contigs ($RA_{IF/CF}$: 338.6 and 551.2; ARN_{IF}: 70.9, 129.5, 145.3, 169.8) suggested that primer binding and reverse transcriptase pausing were biased at certain sites of the template because, otherwise, there should not have been any gap for such a short ORF of ~1.36 kb. The exact role of scolexin in defense is still unclear.

In the reaction of proPO activation, a high molecular weight complex of SPH1 and SPH2 has to be present along with PAP and proPO to generate active PO (Gupta et al. 2005). In this study, we identified SPH1 (contig 02813, $RA_{IH/CH}$: 9.5) and SPH2 (contig 6149, $RA_{IF/CF}$: 16.7; contig 14393, $RA_{IF/CF}$: 33.7) and confirmed their induced expression (Yu et al. 2003). Besides, contig 02985 ($RA_{IF/CF}$: 27) contained a complete ORF coding for a regulatory clip domain

followed by a serine proteinase-like domain. The protein, designated *M. sexta* SPH4, is 49% and 92% identical to SPH1 in the amino- and carboxyl-terminal domains, respectively. Such a disparity in sequence alterations suggests that the selection pressures or structural constraints for these two regions differ dramatically.

Functions of serine proteinases are modulated not only by SPHs but also by their inhibitors. Particularly, some members of the serpin superfamily regulate serine proteinase activities by forming covalent complexes with their cognate enzymes (Kanost 1999). We have identified six serpins in the UP list (Table 3), five of which are known as *M. sexta* serpin-1 (contig 7639: ARN_{IH}: 16.6), serpin-2 (four contigs, ARN_{IH}: 61.7, $RA_{IH/CH}$: 13.3, 15.4, 20.2), serpin-2 homolog (four contigs, $RA_{IH/CH}$: 19.8, 38.8, 77.2, 112.8), serpin-3 (contig 2693, $RA_{IF/CF}$: 7.5), serpin-5 (three contigs, RA: 5.9, 11.9, 16.5). We have found a new serpin (contig 6215, $RA_{IH/CH}$: 9.5) and its ortholog in *B. mori*, SLP or serpin-12. The silkworm serpin was expressed in fat body of bacteria-injected larvae but not in fat body of naïve ones (Zou et al. 2009). Its transcription in hemocytes also was similar to that of the *M. sexta* serpin: the mRNA was low in naïve larvae and became higher in induced ones.

Besides serine proteinases, SPHs and serpins, we also have found other proteins that either mediate or regulate immune responses in *M. sexta* or other moths (Table 3). These include: tyrosine hydroxylase (contig 2023, $RA_{IF/CF}$: 16.8) (Gorman et al. 2007), dopa decarboxylase (contig 00940, ARN_{IF}: 106.6) (Noguchi et al. 2003), PSP-binding protein (contig 15055, $RA_{IF/CF}$: 8.2) (Matsumoto et al. 2003), and Zn proteinase (contig 0915, ARN_{IF}: 11) (Altincicek and Vilcinskas 2008). Four immunity-related proteins, Hdd1, Hdd11, Hdd13, and Hdd23 (Shin et al. 1998), are included here even though their functions remain unknown.

C. Intracellular signaling pathways and their components

Pathogen recognition and signal transduction can either go through a PRR-SP system in insect plasma (e.g. spätzle processing for Toll activation) or directly binds to PRRs on the surface

of immune tissues/cells (e.g. PGRP-LC binding for Imd activation in Drosophila). After that, intracellular proteins are mobilized to relay signals into the cell nucleus where transcriptional regulation occurs. As shown in Table 4, we have detected increase in transcript levels of the putative pathway members: Toll-like receptors (contigs 06893 and 18001, 68% and 94% similar in amino acid sequence to ABO21763) (Ao et al. 2008), cactus (contig 01044) (Furukawa et al. 2009), relish (contigs 04802 and 15532) (Tanaka et al. 2007), and eiger (contig 01139, a membrane-bound TNF homolog) (Kauppila et al. 2003). Other intracellular proteins possibly involved in signal transduction or modulation include a Ser/Thr protein kinase, GTP/GDP exchange factors, a receptor Tyr phosphatase, a protein phosphatase 2c, ankyrin repeat proteins, and vrille transcription factor.

D. Antimicrobial peptides/proteins

Overproduction of effector proteins that immobilize pathogens, block their proliferation, or directly kill them is a hallmark of insect immunity (Bulet et al. 2004). Consistent with this notion, we have detected 65 UP contigs encoding: A) antimicrobial peptides, B) low molecular weight proteinase inhibitors, C) lysozymes, and D) transferrins (Table 5). In group A, twenty-five contigs (06782, 07203, 08902, 11040, 11711, 13563, 14343, 14380, 14641, 15159, 15732, 15744, 15953, 15997, 16129, 16150, 16576, 17135, 17304, 17350, 17632, 17705, 18324, 18814, 18977) code for at least six attacins (Fig. 2), eight (03746, 14568, 15998, 16292, 17184, 18150, 18699, 18819) for at least three X-tox (Girard et al. 2008), six (04913, 07116, 10853, 13916, 17301, 17343) for four lebocin-related proteins (Rayaprolu et al. 2010a), four (12151, 13894, 14997, 15041) for three cecropins (Zhu et al. 2003), two (09484, 17439) for two moricins (Dai et al. 2008), and one (02067) for gloverin (Zhu et al. 2003). Group B consists of eight contigs (03142, 03674, 04175, 05197, 08286, 10722, 13936, 16018) encoding proteinase inhibitor-like proteins which may block proteinases released by bacteria, fungi, or parasites (Zang and Maizels 2001; Armstrong 2006). Group C has three contigs (08421, 15931, 16133) coding for two lysozymes

(Mulnix and Dunn 1994) that hydrolyze bacterial peptidoglycans. Group D includes seven contigs (02145, 11027, 14937, 16606, 17206, 18239, 18308) encoding at least two transferrins that may sequester iron and, by doing so, prevent bacteria from proliferation (Nichol et al. 2002).

E. Other up-regulated genes

Among the 528 UP contigs, 177 did not have any BLAST hits, indicating that some of them may encode polypeptides previously not known to be involved in immunity. To ensure these sequences are indeed up-regulated, we selected contigs with RA >15 (or ARN >30) and total read numbers >70. We then extended these contigs, if possible, with sequences in dataset "06" (Table 1) and in the *M. sexta* gut EST dataset (Pauchet et al. 2010). After eliminating the contigs with GC-contents <35% (hence, likely representing 5' or 3' AT-rich untranslated regions of up-regulated genes), we examined the remaining ones in greater detail (Table 6). Contigs 00327, 01714, 04720, 05523, and 07536 contain ORFs with a secretion signal peptide. The putative mature proteins (41, 61, 37, 86, 179 residues long) could be novel AMPs or in other ways involved in immunity. Contig 02467 encodes a secreted protein containing ten Cys that may tether the 139-residue polypeptide into a stable domain functioning as a proteinase inhibitor or an antifungal protein (Kanost 1999). Contigs 15852 and 17316 encode proteins with 2 and 3 Kazal-type proteinase inhibitor domains, respectively. Contigs 17537 and 17568 encode proteins with a DM9 domain. Contigs 03381 and 15910, after extension, are found to be a part of cactus and serpin-2 transcripts. The other contigs encode sequences similar to *B. mori* heat shock protein 25.4, SPH, and esterases.

Sequence analysis and function prediction of DN genes

The analysis of down-regulated genes yielded results that surprised us at first: among the 53 DN CIFH contig groups with BLAST hits, ten were closely related to immune responses (Table

7). A contig group represents a single contig in most cases but, in other times, has multiple contigs with the same BLAST hit, which may come from different genes. They include lectins (06497, 07642, 11280, 13813, 14570, 14760), lacunin (00015), HP1 (16288), and proPOs (17085 and 17958). A closer inspection of the data indicated that the decreases in mRNA levels seem to always occur in fat body instead of hemocytes. Since these genes were all expressed at much higher levels in hemocytes than fat body ($RA_{CH/CF}$ or $_{IH/IF}$ >40), we suggest the apparent down regulation in fat body were caused by unequal contamination of fat body tissue by hemocytes: somehow there was much less contamination in induced fat body of these hemocyte-specific transcripts. In hemocytes, their average $RA_{CH/IH}$ was only 2.1 – no major down-regulation was observed for these immunity-related genes in cells mainly expressing them. It is likely that similar contamination of fat body tissue by hemocytes also resulted in the observation of genes not known to be directly related to immunity, which includes 11 contig groups (00010, 00248, 00379, 00623, 00628, 03286, 03654, 07139, 08686, 10124, 13842) with $RA_{CH/CF}$ or $_{IH/IF}$ >40 (hemocyte-specific) and $RA_{CF/IF}$ >10 (fat body DN) but $RA_{CH/IH}$ <3.

After eliminating contigs whose $RA_{CH/CF}$ or $_{IH/IF}$ calculated from low read numbers, we have found four DN contigs: 02730 encodes a β-glucosidase, 11098 a Met-rich storage protein, 12848 a proteinase inhibitor, and 14781 a phosphoserine amino transferase. Follow-up studies are needed to confirm their down-regulation and explore physiological relevance of the decrease in transcript levels.

Tissue-specifically regulated genes in larval hemocytes

Using the same set of read numbers in CIFH, we found 45 contig groups representing genes preferentially expressed in hemocytes. Interestingly, this tissue-specific pattern (RA >40 or ARN_{IH} >80) was only found in the induced samples but not in the control ones (Table 8). A closer examination of the data uncovered the possible reason for this bias: although fat body was collected under the same conditions, more hemocytes attached to the control fat body tissue than

the induced one. Consequently, higher read numbers from contaminating hemocytes in control fat body led to much lower $RA_{CH/CF}$ values than their corresponding $RA_{IH/IF}$'s. While the same reason caused wrong identification of some contigs as down-regulated ones (Table 7), the skewing of RAs against the control samples (i.e. lower $RA_{CH/CF}$) did not seem to affect the correct calling of hemocyte-specificity in a qualitative term. For the entire contig groups, the sums of CF and CH reads were 2173 and 105143, respectively. The average $RA_{CH/CF}$ of 10.0 was much lower than the cutoff value of 40 but still substantially higher than 2-5, thresholds commonly used in microarray or qPCR studies to assess differential expression. In comparison, the sum of IF and IH reads were 302 and 62907, respectively, and their average $RA_{IH/IF}$ was 100.5.

The hemocyte-specific gene expression is, in several cases, supported by previous studies on *M. sexta* defense proteins such as lacunin (Nardi et al. 1999), HP1 (Jiang et al. 1999), serpin-2 (Gan et al. 2001), and proPO (Jiang et al. 1997). Lacunin is an extracellular matrix protein responsible for transforming circulating non-adhesive hemocytes to adhesive ones that aggregate on foreign surfaces (Nardi et al. 2005). Contigs 16288, 16719 and 17102 encodes clip-domain HP1; contigs 08524 and 12527 encode an HP1 homolog ~97% identical in sequence to the published one (Jiang et al. 1999). HP1 may be involved in a serine proteinase cascade that proteolytically activates proPO in plasma. Hemolymph proPO is synthesized in oenocytoids only (Jiang et al. 1997): 6 contigs encode proPO subunit-1 and 9 encode proPO subunit-2.

Based on sequence homology, we also discovered 51 contigs that were not known to be related to hemocyte-mediated immunity in *M. sexta* (Table 8). Contigs 11280, 13813, 15506, 15594, and 18551 probably encode parts of hemolectin or hemocytin, a >300 kDa protein participating in hemolymph coagulation (Lesch et al. 2007). As many as 37 contigs encodes multiple lectins that bind to carbohydrates. Contigs 05933, 08686, 13271, 15116, 15350, and 15564 encode scavenger receptor C-like proteins that could also recognize carbohydrates. Apparently, hemocytes play critical roles in the recognition of pathogens that are covered with polysaccharides on the surface. Contig 02473 encodes a protein homologous to *Drosophila* eater

that mediates bacteria phagocytosis by hemocytes (Kocks et al. 2005). Contigs 03287 and 07139 may be related to antiviral and antiparasitoid responses, respectively (Abdel-latief and Hilker 2008; Liu et al. 2010).

Inside hemocytes, proteins may relay signals in a cell-specific manner. These include contigs 00541, 00752, 03246, 06319 (G-protein coupled receptors), 00882 (GTP-binding protein) 00010 (cAMP-dependent kinase), 00839 (receptor-type Tyr-protein phosphatase), 02159 (septin for ubiquitination), 15584 (GTPase atlastin), 14248, 15111, 16917, 17058, and 17751 (serpin-2 and 2'). It is unclear how these two highly inducible, intracellular serpins may inhibit a proteinase during apoptosis. Nor is it known how the other proteins may transduce signals dependent on the immune status of hemocytes.

Specific gene expression in fat body from feeding larvae

Because hemocyte samples collected through cut prolegs of feeding larvae were unlikely contaminated with fat body tissue, the 132 fat body-specific (i.e. FB) contig groups had high RACF/CH or IF/IH values (Table 9). Moreover, since chances for such contamination were equal for hemocytes from naïve and challenged *M. sexta* larvae, there was no globally uneven distribution of RAs or ARNs between the CF/CH and IF/IH groups. In other words, the data on fat body- specific gene expression were unbiased and reliable.

Insect fat body, equivalent to combined mammalian liver and adipose tissue, is the site where most intermediary metabolism takes place (Arrese and Soulages 2010). It also is the principal source of plasma proteins, including those participating in innate immune responses (Jiang 2008). These notions are strongly supported by the identification of FB contigs and BLAST search: 61 or 46% of the 132 FB contig groups are metabolism-related, whereas 32 or 24% are immunity- related (Table 9). Since metabolism-related genes and their transcript level changes after the immune challenge will be reported elsewhere, we only discuss fat body-specific gene expression involved in antimicrobial defense responses and the UP contigs covered in

Section 3.2 are not repeated here. β-1,3-glucan recognition protein-1 (02979) (Ma and Kanost 2000), immulectin-3 (01097) (Yu et al. 2005), and leureptin (04012 and 08453) (Zhu et al. 2010) are pattern recognition receptors binds fungi and bacteria (Table 9). HAIP (02947), a chitinase-like protein, inhibits hemocyte aggregation (Kanost et al. 1994). Contig 05348 encodes a protein with at least three Ig domains. Contig 00535 encodes a thrombospodin-like protein with eight EGF-like domains and one coiled coil for protein-protein interaction. Contig 07671, after extension, is found to encode a >60 kDa protein with at least four EGF domains. Hemicentin (00465) is a cell adhesion protein containing a von Willebrand A domain (Vogel and Hedgecock 2001). Contig 08820 encodes a fibrillin-like nimrod B which may play a role in pathogen recognition and phagocytosis (Kurucz et al. 2007).

We have found six proteinase inhibitor-like proteins, including homologs of *B. mori* serpin12 (or SLP: 03776, 06215, 06531, 17814), serpin13 (02184) and serpin22 (03224) (Zou et al. 2009), two Cys-rich secreted protein (06175, 06597), and cationic protein-8 (16281, 17312) (Ling et al. 2009). Contig 02651 encodes three cytokines that may regulate cellular immune responses (Kanamori et al. 2010).

Proteomics workflow and protein identification

In order to identify *M. sexta* plasma proteins, especially those involved in immune responses, we collected hemolymph samples from the larvae injected with buffer or bacteria. After hemocyte removal, the cell-free hemolymph samples and their biological replicates were separated on a 4-15% gradient SDS-polyacrylamide protein gel which was subsequently cut into nine slices for each lane according to the staining pattern (Fig. 2). Proteins in the gel pieces were digested with trypsin and the resulting mixtures (referred to as “protein” samples) were analyzed on LTQ-Orbitrap mass spectrometer. As anticipated, some peptides had left the gel and, therefore, were undetectable by this method. To locate these small molecules, including antimicrobial peptides – effectors of the insect immune system, we used equal amount of

acetonitrile (AcCN) to precipitate large proteins in the naïve and induced plasma samples. After centrifugation, the supernatants containing small peptides and some AcCN-stable proteins were treated with trypsin. The resulting mixtures (referred to as "peptide" samples) were analyzed by LC-MS/MS on the mass spectrometer. For each "protein" or "peptide" sample, we performed at least three technical replicates.

For the "protein" samples from the gel slices, Scaffold reports showed a dataset of 1,594,513 spectra, 314866 or 19.7% of which matched those of trypsinolytic peptides derived from the protein sequences *in silico*. The matched spectra corresponded to 785 independent proteins in the database (Table 10). In the "peptide" samples treated with AcCN, we detected 202,077 spectra. Of thess 41,706 or 20.6% matched spectra corresponding to 270 independent proteins, mostly peptides. The protein and peptide FDRs were 0.4-0.5% and 0.0% for the gel-derived and AcCN treated samples, respectively. The low FDRs suggested our positive identifications have high reliability. Since 157 proteins were detected in both samples, we identified a total of 898 hemolymph proteins, substantially more than 55 found in the previous study (Furusawa et al. 2008). The successful identification is partly caused by increases in *M. sexta* sequences: based on the sequences deposited at GenBank over the years, we identified 126 hemolymph proteins in the gel slices; including the fat body and hemocyte transcriptome data added 349 new IDs; using the genome sequence resulted in another 292 new ones (Fig.4). Similarly, we found 54, 161, and 205 proteins/peptides in AcCN-treated samples using GenBank, transcriptome, and genome sequence data, respectively. Since there is no major increase in *M. sexta* sequences at GenBank since the proteomic study was published, our identification of 126 proteins (instead of 55 in the early study) is a result of method/instrumental improvements. The selection of day 1, 5^{th} instar larvae also seems critical: the insects were large enough to provide adequate hemolymph yet not too many storage proteins had accumulated at high levels in the plasma of late 5^{th} instar larvae. These abundant proteins tend to mask signals from other proteins. Besides, the gradient gel better separated proteins based on sizes and slicing gel on the basis of staining pattern allowed us to

load appropriate amount of proteins to the mass spectrometer individually according to the relative abundances.

Quantitative analysis using spectral count

Spectral counting has been widely adopted as a robust label free method for quantitative analysis in proteomic studies. Spectral counts (SCs), defined as the number of observations of certain proteins in an MS/MS experiment, were used to estimate protein abundances in this study. The SCs for individual proteins were first normalized based on the total number of spectral counts in each sample. The pairwise Pearson correlation coefficients (Table 11) indicate the data consistency between/within all the induced hemolymph and control hemolymph samples, as well as gel-extracted and AcCN-treated samples. We detected strong correlation between samples after the same biological treatment (C or I). The correlation coefficients between control or those between induced were 0.971~0.990 in the "protein" group, whereas those for the "peptide" samples were 0.936~0.965. On the other hand, the Pearson correlation coefficients between C and I were relatively lower. For the "peptide" samples, the correlation coefficients were 0.554~ 0.654, indicating drastic differences in peptide levels after the bacterial injection. For the "protein" samples, the correlation coefficients were 0.933~ 0.950, suggesting that a smaller proportion of proteins underwent less dramatic changes after immune challenge.

To compare the methods of sample preparation, we analyzed the 157 proteins that were identified in both "protein" and "peptide" samples and noticed some interesting differences (Table 12). Fifty-one proteins showed significant changes and 46 proteins had insignificant changes after the immune challenge in both samples, as supported by the t-test results. This showed the consistency of t-test results for most proteins identified in both sources. However, 60 other proteins showed significant changes in one sample but not in the other. This may be related to differences in sample handling: while gel-extracted samples went through gel electrophoresis, staining and excision, and in-gel trypsin digestion, AcCN treatment is simple and does not

introduce much operational error. On the other hand, dependent on their properties and concentrations, AcCN might introduce inequality to proteins/peptides, especially the high molecular ones. Consequently, when different quantitative results were obtained from these two sources, we relied more on data from AcCN-treated samples for small peptides and from gel-extracted samples for large proteins. We checked the percentage of proteins showing significant changes after immune challenge. The percentages for total proteins from the gel-extracted and AcCN-treated samples were 25.6 and 57.8%, respectively. Comparing with the numbers for overlapping proteins (gel: 65/157, AcCN: 97/157), we could see the difference for gel-extracted samples. The reason appears to be that most of the overlapping proteins were small peptides whose expression levels were highly induced after the immune process. In fact, Pearson pairwise correlation data (Table11) showed drastic changes of small peptides after immune challenge.

Up-regulated proteins

After t-test and ratio calculation, we found 94 and 48 up-regulated proteins from the "protein" and "peptide" samples, respectively. As we anticipated, a majority of the proteins are associated with immune responses (Table 13). These include 8 PRRs, such as hemolin, PGRPs, immulectins and hemicentin. Their corresponding mRNA levels were also up-regulated after the immune challenge, suggesting a reinforcement of pathogen detection. We also found 24 signaling proteins including hemolymph proteases, protease inhibitors, and an intracellular signaling protein called Dorsal. The most drastic up-regulation was observed in the category of AMPs, such as attacins, cecropins, lebocins, gloverin, gallerimycin, and psychimicin. Most of their levels increased more than 10 folds after immune challenge. Some attacins showed 500-fold changes. These data are consistent with the fact that AMPs mRNA levels elevated greatly after the immune challenge (See above). Besides we also found some proteins in the up-regulated protein list, which are not known to be immunity-related. For instance, a cysteine-rich peptide (contig04199) was highly up-regulated and similar in sequence to a salivary protein. We suspect it could be a

new antimicrobial protein whose structure is stabilized by disulfide bonds. Similarly, another hypothetical protein (lrc512) containing eight Cys residues is also induced after the immune challenge. Several homologs of heat shock protein 25.4 (contig04865, contig04960, contig05548, contig05861, contig08771) are involved in stress responses, in this case, presence of microbes and their surface molecules injected into the hemocoel. Another up-regulated protein named "cold-related protein" (CUFF.24081.1) may also be a stress response protein. Additionally, we noticed coherent increase of mRNA and protein levels of several lipases and esterases (CUFF17912.1, CUFF.17913.2, CUFF.19298.1, CUFF.19800.1, CUFF.22549.1, CUFF.25705.2, CUFF25705.3, c707). Perhaps, due to increased energy need for fighting infection, lipid mobilization and metabolism are enhanced by these hydrolytic enzymes.

Down-regulated proteins

We, based on the I/C values and statistical analysis, identified 103 and 62 proteins from "protein" and "peptide" samples respectively, whose levels significantly reduced after the immune challenge (Table 14). Several immunity-related proteins were found in this list including signaling proteins and proteins involved in cellular response. Some proteinases and serine protease inhibitors were down-regulated to modulate immunity signals to appropriate levels. The precursor of plasmatocyte-spreading peptide (PSP) and its binding protein were both found to be down-regulated after the bacterial injection. Interestingly, some ribosomal proteins account for a substantial portion of the down-regulated protein lists.

Immunity-related proteins

Among all the identified proteins in both samples, 220 may participate in immune responses of *M. sexta* including PRRs, signaling proteins, AMPs and others (Table 15). PRRs recognize surface components of invading microbes. In addition to ones described above as up-regulated, we also identified different PRRs including immulectins, microbe binding protein (MBP), β-1,3-

glucan recognition proteins (βGRPs), leureptins, nimrod, and draper. Most of these PRRs did not show significant change after immune challenge. Nearly half of the identified immunity-related proteins were extracellular proteins involved in signal transduction and regulation, such as hemolymph protease and protease inhibitors. Some of the proteases may function as components of an enzyme network that respond to pathogen recognition by PRRs through specific proteolytic activation of zymogens in a cascade mode. Protease inhibitors form inactive complexes with the pathway members that diffuse away from the site of infection. While some of the proteases and inhibitors are up-regulated (Table 13), others did not change much or even decreased. Such orchestrated changes in mRNA and protein levels may finally result in a more potent yet balanced immune reaction against secondary infection. Interestingly, in the plasma proteomes, we identified four intracellular proteins suggesting that our analysis was sensitive enough to detect small amount of intracellular proteins released from ruptured cells.

Of all the defense proteins, AMPs are induced to high ratios after the immune challenge (Fig. 5). As the major effectors of the immune system, almost all the AMPs were up-regulated and some attacins showed ~500-fold increase after bacterial injection. In our dataset, we found attacins, cecropins, lebocins, gloverins, lysozymes and tranferrins that attack bacteria. We also identified homologs of antifungal peptides such as gallerimycin, diapausin and psychimicin. Activity assays are needed to confirm the predicted antimicrobial functions of highly induced small proteins with multiple Cys residues in their sequences (see above). While almost all the currently known AMPs are identified in the *M. sexta* plasma proteome, we did not find moricin, a Lys-rich peptide that may not yield peptides at appropriate sizes after trypsinolysis.

Protein VS mRNA

In this proteomic study, we used the fat body and hemocyte transcriptome data to construct the protein database for mass data analyses (Fig.2). Consequently, we found mass spectra of the gel-extracted and AcCN-treated samples matched to 336 “protein” and 135 “peptide” sequences

from their cDNAs. Of these, 296 “proteins” and 134 “peptides” and their respective contigs were identified for retrieving the read numbers for CF, IF, CH, and IH. Using these and corresponding spectra counts, we attempted to first correlate the mRNA and proteins levels after the immune challenge. As shown in Fig. 5 (A, B), there was a positive correlation between CF and C (0.66, gel) and between IF and I (0.51, gel). The positive correlations remained when “peptide” data were analyzed, but the correlation coefficients reduced to 0.38 (CF vs. C) and 0.33 (IF vs. I) (Fig.5, C and D). When we plotted the IH and CH read numbers with the “protein” and “peptide” data in the “C” and “I” groups (Fig. 5, E-H), the coefficients (-0.14~+0.08) did not indicate either positive or negative correlation. Perhaps, due to its sheer volume and contribution to hemolymph factors, a moderate positive correlation exists between fat body mRNA and plasma protein levels. Hemocytes do play critical roles in cellular immune responses but, due to lower cell number and protein synthesis, their contribution to the plasma protein pool seems limited. As we understand, since transcript abundance is only one of the factors that govern final protein levels, the positive correlation coefficients (0.33~0.66) between fat body mRNA and plasma protein levels seem reasonable.

We then tested if there is a stronger correlation between the mRNA and protein level changes after the immune challenge. To make sure the comparisons are of statistical significance, we selected “proteins” or “peptides” whose levels significantly changed (> 5-fold change or >2-fold change with t-test significance) after the bacterial injection. The logarithm scatter plot (Fig. 6) clearly demonstrated a positive correlation: most proteins showed the same tendency of change with their mRNAs and only a small number of proteins showed mRNA level increases but protein level decreases or vice versa. Interestingly, most of the inconsistent ones participate in immune signal transduction and regulation. Perhaps the dynamics of these mRNAs and proteins are unusual as compared with other defense gene products. The better correlation from the “peptide” samples (including most AMPs) suggests that differences in gene expression and functions do impact overview of the system.

CHAPTER V

SUMMARY AND CONCLUSIONS

We used pyrosequencing technology and mass spectrometry to quantify transcriptomic and proteomic changes in the 5th instar larvae of *M. sexta* after a bacterial injection. We obtained 19,020 fat body and hemocyte cDNA contigs and identified 898 proteins from the plasma samples. With the read numbers and spectral counts available, our quantitative analyses helped us identify hundreds of differentially expressed cDNAs and proteins from *M. sexta* after immune challenge. These are the first quantitative transcriptome and proteome studies of this biochemical model insect.

The transcriptomic study expanded our knowledge on *M. sexta* mRNAs sequences as well as their expression levels in different tissues and physiological states. It proved the value of NGS technologies in quantitative transcriptomic studies. Since pyrosequencing offers the longest fragment sequences among different NGS technologies, it has particular advantages in studies on organisms with unknown genome sequences. Compared with microarray technology, NGS provides digital data on the expression levels which are interfered with much lower, if any, background noise. NGS-based transcriptomic studies do not necessarily rely on genome sequences. With constantly decreasing cost of sequencing, NGS technologies have been regarded as cost-effective methods in biological research.

As the second study on *M. sexta* plasma proteome, this work made greater progress, when compared with the first one published in 2008. We identified a lot more proteins along with information on their relative abundances. The improvements are based on the following factors: advanced instrument, improved protein database, and optimized sample preparation. The normalized spectral counts served well for the quantitative analysis and this work proved the robustness of spectral counting as a semi-quantitative analysis method. Biological and technical replicates were utilized to decrease the random error. Pairwise Pearson correlation tests confirmed superb reproducibility of our work. However, due to the treatment to the plasma samples (gel electrophoresis or AcCN precipitation) for more protein IDs, we sacrificed the quantitative accuracy to some extent. Moreover, although the transcriptome and draft genome sequences helped us identify more proteins, the redundant protein database affected the quality of our lists and caused the "grouping ambiguity" problem. It is necessary to redo protein identification when the official *M. sexta* protein dataset is released.

The transcriptomic and proteomic data showed hundreds of immunity-related genes. Different PRRs were identified and some were up-regulated at mRNA- and protein-levels after the immune challenge, such as hemolins, PGRPs, and some immulectins. Signaling proteins constituted the largest category among all the identified proteins including hemolymph proteases and protease inhibitors. Some of them were up- regulated, others down-regulated after the bacterial injection, but most did not show drastic changes. Antimicrobial peptides, as major defense effectors, were highly induced at both mRNA- and protein-levels. Based on this feature, we also found some highly up-regulated small protease inhibitors and Cys-rich proteins and hypothesize they have antimicrobial activity. The two "omics" projects provided long lists of DNAs/proteins, some of which are key molecules in the insect immune system. They are valuable in terms of what systems biology could offer. Nevertheless, we still need to focus on

characterizing selected candidate molecules by traditional biochemical techniques to elucidate their specific functions.

Table 1. Summary statistics for pyrosequencing analysis of *M. sexta* ESTs

	06 [a]	CF	CH	IF	IH	CIFH [b]	06CIFH [c]
Total number of reads	95,458 (*95,358*)	227,302	647,587	405,739	541,024	1,821,652	1,917,110
Average reads length (bp)	185 (*185*)	296	287	293	287	289	284
Total number of contigs	1,471 (*7,231*)	2,118	11,540	4,063	10,600	19,020	19,504
Contigs size (avg./longest in bp)	391/3,552 (*300/3,909*)	770/12,740	827/11,667	764/8,482	832/10,591	923/23,095	911/23,097
Total assemebled reads	64,874 (*69,429*)	191,156	561,054	349,028	465,561	1,677,738	1,757,333
Singlet reads	28,518 (*25,929*)	32,518	68,861	49,444	61,108	108,587	120,670
Singlet length (avg. in bp)	179	244	245	235	254	209	200
Total BLASTable sequences	29,989	34,636	80,401	53,507	71,708	127,607	140,174
Orphan sequences (no BLAST match, #/%)	19,963/67	17,982/52	51,968/65	28,649/54	46,521/65	73,915/58	89,948/64
Contigs and reads with functional assignment	10,026	16,654	28,433	24,858	25,187	53,692	50226

[a] Results from reanalysis of the 2006 sequence data. The numbers in parentheses (adopted from Zou et al., 2008) are listed for comparing with the new results. [b] Analysis of the 2009 EST sequences of control fat body (CF), control hemocytes (CH), induced fat body (IF), and induced hemocytes (IH) from *M. sexta* larvae. [c] Analysis of the combined reads of 2006 (raw flow signals interpreted with the up-graded software) and 2009 (CF, CH, IF, and IH).

Table 2. A list of 19 UP CIFH contigs with similarity to pattern recognition receptors*

CIFH contig #	Original read # CF	CH	IF	IH	Total	RA or ARN IF/CF	IH/CH	BLAST results
00131	11	41	137	61	250	6.4	1.8	gi|198430641|ref|XP_002123478.1| ~ hemicentin 1, Ig domains [Ciona intestinalis]
00575	3	0	259	5	267	44.0	5.9	gi|154240658|dbj|BAF74637.1| peptidoglycan recognition protein-D [Samia cynthiaricini]
01326	1	9	19	70	99	9.7	9.2	gi|52782739|sp|Q8ISB6.1|BGBP2_MANSE Beta-1,3-glucan recognitionprotein 2
03442	1	12	1468	40	1521	748.5	4.0	gi|511297|gb|AAC46916.1| hemolin [Manduca sexta]
04775	1	0	89	0	90	45.4	0.0	gi|237869126|gb|AAF91316.3|AF242202_1 immulectin-2 [Manduca sexta]
04808	0	0	426	2	428	217.2	2.4	gi|237861314|gb|AAV41237.2| immulectin-4 [Manduca sexta]
06630	2	40	44	77	163	11.2	2.3	gi|55139125|gb|AAV41236.1| immulectin-3 [Manduca sexta]
08247	27	2	122	18	169	2.3	10.7	gi|208972535|gb|ACI32828.1| beta-1,3-glucan recognition protein 3 [Helicoverpa armigera]
08467	0	0	113	0	113	57.6	0.0	gi|112983866|ref|NP_001036858.1| peptidoglycan recognition protein-6 [Bombyx mori]
11458	0	0	55	0	55	28.0	0.0	gi|148298818|ref|NP_001091784.1| multi-binding protein [Bombyx mori]
11845	0	2	9	17	28	4.6	10.1	gi|18202160|sp|O76537.1|PGRP_TRINI peptidoglycan recognitionprotein;
13190	15	0	117	9	141	4.0	10.7	gi|27733423|gb|AAO21509.1|AF413068_1 peptidoglycan recognition protein 1A [Manduca sexta]
14104	14	0	173	13	200	6.3	15.4	gi|27733423|gb|AAO21509.1|AF413068_1 peptidoglycan recognition protein 1A [Manduca sexta]
14278	1	34	17	179	231	8.7	6.3	gi|83583693|gb|ABC24706.1| hemicentin-like protein 1, Ig domains [Spodoptera frugiperda]
14515	2	0	34	0	36	8.7	0.0	gi|148298818|ref|NP_001091784.1| multi-binding protein [Bombyx mori]
14700	0	0	183	2	185	93.3	2.4	gi|260765453|gb|ACX49764.1| peptidoglycan recognition protein 2 [Manduca sexta]
14752	0	0	118	2	120	60.2	2.4	gi|260765453|gb|ACX49764.1| peptidoglycan recognition protein 2 [Manduca sexta]
15639	10	0	109	0	119	5.6	0.0	gi|148298818|ref|NP_001091784.1| multi-binding protein [Bombyx mori]
15857	0	1	0	9	10	0.0	10.7	gi|27733411|gb|AAO21503.1|AF413062_1 leureptin, LPS-binding [Manduca sexta]

* RA and ARN are calculated using original read numbers as described in Section 2.3. Listed here are contigs with $RA_{IF/CF}$ >5, $RA_{IH/CH}$ >8, ARN_{IF} >10 when RN_{CF} =0, or ARN_{IH} >10 when RN_{CH} =0. $RA_{IF/CF}$ and $RA_{IH/CH}$ values are shown in red if they are greater than 5 and 8, respectively. ARN_{IF} and ARN_{IH} values are shown in blue if they are higher than 10. In the columns of RA or ARN, cells shaded yellow and blue represent fat body- and hemocyte-specific gene expression, respectively.

Table 3. A list of 40 UP CIFH contigs with similarity to extracellular signal modulators*

CIFH contig #	Original read #					RA or ARN		BLAST results
	CF	CH	IF	IH	Total	IF/CF	IH/CH	
00915	0	21	21	26	68	10.7	1.5	gi|91084647|ref|XP_966816.1| ~ AGAP002414-PA, Zn protease [Tribolium castaneum]
00940	0	0	209	7	216	106.6	8.3	gi|1352212|sp|P48861.1|DDC_MANSE dopa decarboxylase (DDC)
02023	1	0	33	7	41	16.8	8.3	gi|148611442|gb|ABQ95973.1| tyrosine hydroxylase isoform A [Manduca sexta]
01667	0	7	98	33	138	50.0	5.6	gi|26006435|gb|AAL76085.1| prophenoloxidase-activating proteinase-2 [Manduca sexta]
01818	0	26	45	66	137	22.9	3.0	gi|60299972|gb|AAX18637.1| prophenoloxidase-activating proteinase-3 [Manduca sexta]
02361	7	4	70	1	82	5.1	0.3	gi|56418425|gb|AAV91020.1| hemolymph proteinase 22 [Manduca sexta]
02382	0	2	109	69	180	55.6	41.0	gi|4090964|gb|AAD09279.1| immune-related Hdd1 [Hyphantria cunea]
02693	21	7	310	19	357	7.5	3.2	gi|27733415|gb|AAO21505.1|AF413064_1 serpin 3a [Manduca sexta]
02813	108	9	313	72	502	1.5	9.5	gi|242351233|gb|ACS92763.1| serine proteinase-like protein 1b [Manduca sexta]
02985	3	0	158	0	161	26.9	0.0	gi|56418466|gb|AAV91027.1| serine proteinase-like protein 4 [Manduca sexta]
03018	0	54	22	79	155	11.2	1.7	gi|56418395|gb|AAV91005.1| hemolymph proteinase 7 [Manduca sexta]
03778	0	11	192	28	231	97.9	3.0	gi|74813957|sp|Q86RS3.1|DFP_MANSE putative defense protein Hdd11-like, precursor
03989	0	1	8	24	33	4.1	28.5	gi|56418399|gb|AAV91007.1| hemolymph proteinase 9 [Manduca sexta]
05186	0	0	8	13	21	4.1	15.4	gi|56418413|gb|AAV91014.1| hemolymph proteinase 17 [Manduca sexta]
05606	1	0	19	4	24	9.7	4.7	gi|4090968|gb|AAD09281.1| immune-related Hdd13 [Hyphantria cunea]
05831	3	8	97	25	133	16.5	3.7	gi|45594232|gb|AAS68507.1| serpin-5A [Manduca sexta]
06149	21	22	686	32	761	16.7	1.7	gi|27733421|gb|AAO21508.1|AF413067_1 serine protease-like protein [Manducasexta]
06215	29	1	108	8	146	1.9	9.5	gi|112983872|ref|NP_001036857.1| Serpin-like protein (SEP-LP) or serpin-12 [Bombyx mori]
06581	0	0	13	10	23	6.6	11.9	gi|4090970|gb|AAD09282.1| immune-related Hdd23 [Hyphantria cunea]
07639	651	0	1237	14	1902	1.0	16.6	gi|134436|sp|P14754.1|SERA_MANSE serpin-1
08231	0	1	0	34	35	0.0	40.4	gi|56418417|gb|AAV91016.1| hemolymph proteinase 18 [Manduca sexta]
10791	1	0	1081	1	1083	551.2	1.2	gi|4262357|gb|AAD14591.1| scolexin A [Manduca sexta]
10792	0	0	333	0	333	169.8	0.0	gi|4262357|gb|AAD14591.1| scolexin A [Manduca sexta]
13453	5	4	58	7	74	5.9	2.1	gi|45594232|gb|AAS68507.1| serpin-5A [Manduca sexta]
13454	0	1	17	10	28	8.7	11.9	gi|45594232|gb|AAS68507.1| serpin-5A [Manduca sexta]
14093	1	0	14	0	15	7.1	0.0	gi|56418419|gb|AAV91017.1| hemolymph proteinase 19 [Manduca sexta]
14248	0	6	0	196	202	0.0	38.8	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta]
14393	2	4	132	11	149	33.7	3.3	gi|27733421|gb|AAO21508.1|AF413067_1 serine protease-like protein [Manduca sexta]
14456	0	0	1	52	53	0.5	61.7	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta]
15055	1	1	16	0	18	8.2	0.0	gi|112983896|ref|NP_001037394.1| paralytic peptide binding protein 1 [Bombyxmori]
15111	1	48	8	800	857	4.1	19.8	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta]
16520	1	0	664	1	666	338.6	1.2	gi|4262357|gb|AAD14591.1| scolexin A [Manduca sexta]
16917	0	40	2	519	561	1.0	15.4	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta]
17048	0	1	0	95	96	0.0	112.8	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta]
17058	0	32	4	545	581	2.0	20.2	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta]
17751	0	24	1	269	294	0.5	13.3	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta]
18441	0	1	0	65	66	0.0	77.2	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta]
18669	0	0	285	0	285	145.3	0.0	gi|4262357|gb|AAD14591.1| scolexin A [Manduca sexta]
18670	0	0	139	0	139	70.9	0.0	gi|4262357|gb|AAD14591.1| scolexin A [Manduca sexta]
18963	0	0	254	0	254	129.5	0.0	gi|4262357|gb|AAD14591.1| scolexin A [Manduca sexta]

* See Table 2.

Table 4. A list of 18 UP CIFH contigs with similarity to intracellular signal transducers*

CIFH contig #	Original read #					RA or ARN		BLAST results
	CF	CH	IF	IH	Total	IF/CF	IH/CH	
00461	1	48	14	108	171	7.1	2.7	gi|47217104|emb|CAG02605.1| unnamed protein product, integrin β6 precursor [Tetraodon nigroviridis]
00537	1	32	10	32	75	5.1	1.2	gi|270009406|gb|EFA05854.1| TcasGA2_TC008649 Tyr protein kinase [Triboliumcastaneum]
00671	1	12	10	46	69	5.1	4.6	gi|189235637|ref|XP_967498.2| ral guanine nucleotide exchange factor, putative [Tribolium castaneum]
01020	42	4	63	27	136	0.8	8.0	gi|91082721|ref|XP_972476.1| ~ eiger CG12919-PA, JNK [Triboliumcastaneum]
01044	9	70	163	105	347	9.2	1.8	gi|289629214|ref|NP_001166191.1| cactus [Bombyx mori]
01313	2	52	33	35	122	8.4	0.8	gi|242009174|ref|XP_002425367.1| Ser-Thr protein kinase, plant-type, putative [P. humanus corporis]
01390	1	31	14	19	65	7.1	0.7	gi|46403173|gb|AAS92609.1| vrille transcription factor [Antheraea pernyi]
01970	1	16	12	29	58	6.1	2.2	gi|157118595|ref|XP_001659169.1| guanine nucleotide exchange factor [Aedes aegypti]
04802	2	42	25	66	135	6.4	1.9	gi|157412326|ref|NP_001098704.1| BmRelish2 [Bombyx mori]
05836	2	1	0	7	10	0.0	8.3	gi|189235110|ref|XP_971078.2| receptor tyrosine phosphatase type r2a [Tribolium castaneum]
06304	1	0	11	1	13	5.6	1.2	gi|170038257|ref|XP_001846968.1| dipeptidyl peptidase 4, apoptosis, immunity [Culex quinquefasciatus]
06868	0	1	1	11	13	0.5	13.1	gi|193713771|ref|XP_001946690.1| ankyrin repeat domain 54 [Acyrthosiphon pisum]
06893	0	1	1	20	22	0.5	23.7	gi|126635756|gb|ABO21763.1| Toll receptor [Manduca sexta]
11311	0	1	3	9	13	1.5	10.7	gi|189237512|ref|XP_972880.2| protein phosphatase type 2c [Tribolium castaneum]
11356	0	1	4	7	12	2.0	8.3	gi|156551808|ref|XP_001603899.1| arf6 guanine nucleotide exchange factor [Nasonia vitripennis]
13966	0	1	0	9	10	0.0	10.7	gi|190570736|ref|YP_001975094.1| Ankyrin repeat domain protein [Wolbachia of C. quinquefasciatus Pel]
15532	1	19	12	9	41	6.1	0.6	gi|157412326|ref|NP_001098704.1| BmRelish2 [Bombyx mori]
18001	0	1	0	7	8	0.0	8.3	gi|126635756|gb|ABO21763.1| toll receptor [Manduca sexta]

* See Table 2.

Table 5. A list of 65 UP CIFH contigs with similarity to antimicrobial proteins*

CIFH contig #	Original read # CF	CH	IF	IH	Total	RA or ARN IF/CF	IH/CH	BLAST results
02067	1	0	280	82	363	142.8	97.4	gi|110649240|emb|CAL25129.1| gloverin [Manduca sexta]
02145	0	15	20	95	130	10.2	7.5	gi|157134051|ref|XP_001663123.1| transferrin [Aedes aegypti]
03142	1	7	420	121	549	214.2	20.5	gi|33860163|sp|P82176.2|IMPI_GALME Inducible metalloproteinase inhibitor protein; IMPIα precursor
03674	1	0	5	21	27	2.5	24.9	gi|110347837|gb|ABG72720.1| protease inhibitor-like protein [Antherae amylitta]
03746	0	7	55	389	451	28.0	66.0	gi|148298709|ref|NP_001091749.1| possible antimicrobial peptide [Bombyx mori]
04175	0	7	40	45	92	20.4	7.6	gi|114052803|ref|NP_001040277.1| salivary cysteine-rich peptide [Bombyx mori]
04903	0	0	279	6	285	142.3	7.1	gi|187281722|ref|NP_001119732.1| lebocin 3 precursor [Bombyx mori]
05197	0	0	20	1	21	10.2	1.2	gi|115392217|gb|ABI96910.1| brasiliensin precursor, thrombin inhibitor [Triatoma brasiliensis]
06782	0	0	102	17	119	52.0	20.2	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
07116	1	4	902	3	910	459.9	0.9	gi|171262319|gb|ACB45566.1| lebocin-like protein [Antheraea pernyi]
07203	2	3	312	22	339	79.5	8.7	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
08286	0	0	139	23	162	70.9	27.3	gi|56462340|gb|AAV91453.1| protease inhibitor 6 [Lonomia obliqua]
08421	4	2	28	99	133	3.6	58.8	gi|7327646|gb|AAB31190.2| lysozyme [Manduca sexta]
08902	0	0	164	14	178	83.6	16.6	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
09484	1	0	134	56	191	68.3	66.5	gi|29469961|gb|AAO74637.1| antimicrobial peptide moricin [Manduca sexta]
10234	0	1	249	7	257	127.0	8.3	gi|169264911|dbj|BAG12297.1| gallerimycin [Samia cynthia ricini]
10722	9	3	102	3	117	5.8	1.2	gi|110347833|gb|ABG72718.1| protease inhibitor-like protein [Antherae amylitta]
10853	0	0	113	1	114	57.6	1.2	gi|171262319|gb|ACB45566.1| lebocin-like protein [Antheraea pernyi]
11027	59	0	694	0	753	6.0	0.0	gi|136206|sp|P22297.1|TRF_MANSE transferrin precursor
11040	0	4	51	249	304	26.0	73.9	gi|29469969|gb|AAO74640.1| antimicrobial protein attacin 2 [Manduca sexta]
11711	0	7	85	1317	1409	43.3	223.4	gi|29469969|gb|AAO74640.1| antimicrobial protein attacin 2 [Manduca sexta]
12151	0	0	153	0	153	78.0	0.0	gi|116084|sp|P14665.1|CEC5_MANSE Bactericidin B-5P; Cecropin-like;
13563	0	0	657	0	657	335.0	0.0	gi|110347786|gb|ABG72695.1| attacin-like protein [Antheraea mylitta]
13894	0	0	48	29	77	24.5	34.4	gi|112984238|ref|NP_001037460.1| cecropin B precursor [Bombyx mori]
13916	1	0	741	0	742	377.8	0.0	gi|219958086|gb|ACL68097.1| lebocin-related protein precursor [Manduca sexta]
13936	0	0	25	0	25	12.7	0.0	gi|123725|sp|P26227.1|HTIB_MANSE hemolymph trypsin inhibitor B, BPI-type
14343	0	0	186	7	193	94.8	8.3	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
14380	0	0	106	0	106	54.0	0.0	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
14568	0	0	2	68	70	1.0	80.7	gi|148298709|ref|NP_001091749.1| possible antimicrobial peptide [Bombyx mori]
14641	0	0	157	0	157	80.1	0.0	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
14937	13	0	164	0	177	6.4	0.0	gi|136206|sp|P22297.1|TRF_MANSE: transferrin precursor
14997	0	0	34	10	44	17.3	11.9	gi|29469965|gb|AAO74638.1| antimicrobial peptide cecropin 6 [Manduca sexta]
15041	0	0	36	0	36	18.4	0.0	gi|116084|sp|P14665.1|CEC5_MANSE Bactericidin B-5P; Cecropin-like;
15159	0	0	0	15	15	0.0	17.8	gi|15963410|dbj|BAB69462.1| attacin [Samia cynthia ricini]
15732	0	1	253	43	297	129.0	51.1	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
15744	0	0	0	35	35	0.0	41.6	gi|29469969|gb|AAO74640.1| antimicrobial protein attacin 2 [Manduca sexta]
15931	40	37	1504	364	1945	19.2	11.7	gi|7327646|gb|AAB31190.2| lysozyme [Manduca sexta]
15953	1	0	43	6	50	21.9	7.1	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
15997	0	0	142	4	146	72.4	4.7	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
15998	0	0	1	10	11	0.5	11.9	gi|73921456|gb|AAZ94260.1| immune related protein X-tox [Spodoptera frugiperda]
16018	0	0	40	12	52	20.4	14.2	gi|116833115|gb|ABK29470.1| immune reactive putative protease inhibitor [Helicoverpa armigera]
16129	1	0	212	35	248	108.1	41.6	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
16133	47	57	1719	440	2263	18.6	9.2	gi|233964|gb|AAB19535.1| lysozyme (peptide partial, 120 aa]
16150	0	1	145	3	149	73.9	3.6	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
16292	0	0	1	34	35	0.5	40.4	gi|148298709|ref|NP_001091749.1| possible antimicrobial peptide [Bombyx mori]
16576	0	0	0	18	18	0.0	21.4	gi|74767320|sp|Q5MGE6.1|DFP3_LONON Defense protein 3 precursor, attacin E
16606	8	0	164	0	172	10.5	0.0	gi|136206|sp|P22297.1|TRF_MANSE transferrin precursor
17135	0	9	103	1157	1269	52.5	152.6	gi|110649242|emb|CAL25130.1| attacin II [Manduca sexta]
17184	0	11	76	449	536	38.8	48.5	gi|73921456|gb|AAZ94260.1| immune related protein, X-tox [Spodoptera frugiperda]
17206	3	0	136	0	139	23.1	0.0	gi|136206|sp|P22297.1|TRF_MANSE transferrin precursor
17301	1	0	272	0	273	138.7	0.0	gi|219958086|gb|ACL68097.1| lebocin-related protein precursor [Manduca sexta]
17304	0	1	412	13	426	210.1	15.4	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
17350	0	0	205	0	205	104.5	0.0	gi|29469969|gb|AAO74640.1| antimicrobial protein attacin 2 [Manduca sexta]
17434	1	0	314	0	315	160.1	0.0	gi|219958086|gb|ACL68097.1| lebocin-related protein precursor [Manduca sexta]
17439	0	0	98	31	129	50.0	36.8	gi|110649236|emb|CAL25127.1| like moricin [Manduca sexta]
17632	0	0	83	6	89	42.3	7.1	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
17705	0	0	36	0	36	18.4	0.0	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
18150	0	0	0	18	18	0.0	21.4	gi|148298709|ref|NP_001091749.1| possible antimicrobial peptide [Bombyx mori]
18239	3	0	67	0	70	11.4	0.0	gi|136206|sp|P22297.1|TRF_MANSE transferrin precursor
18308	15	0	169	0	184	5.7	0.0	gi|136206|sp|P22297.1|TRF_MANSE transferrin precursor
18324	0	0	25	0	25	12.7	0.0	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
18699	0	1	26	114	141	13.3	135.4	gi|148298709|ref|NP_001091749.1| possible antimicrobial peptide [Bombyx mori]
18814	0	0	235	29	264	119.8	34.4	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]
18819	0	5	59	405	469	30.1	96.2	gi|73921456|gb|AAZ94260.1| immune related protein X-tox [Spodoptera frugiperda]
18977	0	1	20	2	23	10.2	2.4	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta]

See Table 2.

Table 6. A list of 22 UP CIFH contigs without BLAST hit*

CIFH contig #	Original read number CF	CH	IF	IH	total	RA or ARN IF/CF	IH/CH	Length (aa)	domain	Protein sequence
00327	2	9	154	31	196	39.3	4.1	60		MSHSHTFILVSIALTIVEGVGYRSFYANRSSIKSLFAKILKKNHTFFFHIYNILNKYNIE*
01714	3	2	90	1	96	15.3	0.6	84		MIAFTRLYLLVILASLAIIQSDANEDSVQLNEIADIYSNTVGGRIEVSYHDSNGNIVDQEVPVEGNHYVMWPSPQSNERDSRRY*
02467	1	37	45	46	129	22.9	1.5	160		MAKSVAITLAFLLAVFEIGSCLRCYQCNSQEDPACADPYKSAKAPVDCASQDSINYNQLYLRNMLPPEVFGTVAGAPRYCHKIVTQTGTTVRTCLDANPADINHTCRLIENSSKMASVESKKIKHCSVCDKDNCNGSGTVSLSAPLAALALVASYLYYKQ*
02669[a]	2	0	158	0	160	40.3	0.0	>135 &179	hsp25.4	…QNMQRNMPKSGASIEITNNHYILQVYLNGYEDKDIEVKSKPGWVMIKAIHRDQNGKDKNYLEMLSLPDNVDPAGEWTYSQGVLKIDFKIKNSFENSNVIWHSVISVDNNNIYGHNVGQAEVQHDNNDQMYINVKH* and MFFVVLCSVLLGTAAAHTDSYSDEQNQYNAPSLNRNFDKFWDKFNRSMDKFDRGIAKLQREIEKLWRKMPHTRKNIEVTDNRYTLKMYFNVKFRDIEVKSKPGTVMIKAIYKGKDGRDKNYSDVLSLPDNVDPAGKWTYSQGVLKIDFKKKNGPENPVVISDSVVDVDNDLNEYDVEYE*
03381	1	32	34	44	111	17.3	1.6	322	5 ankyrin	MSAKKGYETKIVEEENMDSGIVSGELESYEISGEVDSGVIDCDKKYEGVPSEVLELTDKFKSVNVREKSCPDVPPLADLFHPDNDGDTQLHIASVHGCEKSVSTIIRVCPDKEWLDLPNDYGHTPLHLAVMSGNAVVTRMLVIAGASLAIRDFMGETPLHKATAARNQECLKALLAPVPEQPNRKLSSILDQRNYNGQCCVHLAASIGSVETLQTLVYYGADINARENLAGWTALHIAARRGDVRVVQFLRSRCAGAATRPRDYAGRTPRRLARRTKAAAAFDDKDDSDSDSDSDDDDMYDSDSETLFEKLRESLSTSINVA*
04720	26	1	1025	0	1052	20.1	0.0	66		MYSWKAAVLRFRVGQLYWVFLLSISPESGIYPRYGDRLAPHHIMGCNTHGEKGMQRCASAYSFKDM*
05532[b]	0	0	110	0	110	56.1	0.0	100	hsp25.4	MFIVLASLVSLAAAAPRLAPNTRFSQSDIEQYANIYPNMDDFGVTKNQLYLQFNSAPWTVTTKDLQYVMSMPMPGYNKEDIEVLALNKGITVRAITKKRK*
06987[b]	0	0	74	0	74	37.7	0	152	hsp25.4	MFIVLASLVTFAAAAPRVAPYTHFSQSDIEQYSNIYPNLDDLGVTKNQFYLQFNSAPWTVSTKDLQYVMSMPMPGYNKEDIEVLALNKGITVRAIQKEGDDIVKSQIVVTLLPAYVNPLGRWTYDGVLRIAFPIKWFSDDGSSYAIPVIIDA*
07536	7	0	383	2	392	27.9	2.4	199		MPTIKPYLLFILLGVACATAQSYYGVSVHDNNVQGSVEINLSDAKLQTYNSRRQNPSGQGNNVPPQNVPQGSYDQQRNFNSGPSQQGPYDGDSNTQTIRLNGPGQADVRLFQEIGEDGSTRRAINYGDSRRQDYPSYRQDNSERNPQSQPITHKVTVEDFESTTKRYFTRSTDNSNYGWDAVPVTYNGRKKVCYCPKRS*
08371	0	0	79	0	79	40.3	0.0	125		MFLKTSVFCACIAMIAAQDFNLDVSNKTLAPKPEKFKNIEGCYIPEKDTVIPLNARVAWKDKCLEYRCYSQSYEIAECSTQVPDFKNNPKCFMHRDYEKPYPECCPKIACYIRSISGVNAFDNIF*
08751[b]	1	0	72	0	73	36.7	0.0	163	hsp25.4	MFLLLASLVTLAAAAPHTLPEAHFTKSAFKQYFQNIFPDKNSLGVTMNTFNILYDSAPWTITSTGTEYTLSMPMSGYQKEDIAVMAHTGILAMRAIHKERGVIKKSETSLNFLPLLVNPAGWWTYHDGVLKVTFPINGRNTDAGATRDVNASDVEQIVLDGNN*
13238[a]	3	1	524	0	528	89.1	0.0	156 &179	hsp25.4	MTSIRVWMNFIAAWTSYNAAYQNMQRNMPKSGASIEITNNHYILQVYLNGYEDKDIEVKSKPGWVMIKAIHRDQNGKDRNYLEMLSLPDNVDPAGEWTYSQGVLKIDFKIKNSFENSNVVWHSVISVDNNNIYGHNVGQAEVQHDNNDQMYINVKH* and MFFVVLCSVLLGTAAAHTDSYSDEQNQYNAPSLNRNFDKFWDKFNRSMDKFDRGIAKLQREIEKLWRKMPHTRKNIEVTDNRYTLKMYFNVKFRDIEVKSKPGTVMIKAIYKGKDGRDKNYSDVLSLPDNVDPAGKWTYSQGVLKIDFKKKNGPENPVVISDSVVDVDNDLNEYDVEYE*
15852[c]	0	0	134	0	134	68.3	0.0	>375	kazal	MQNSYVVILTFAVLGNTATATFWRRHGHGNWGHRPHWGHPHHVGYAYGQHPYHHLHPHNWGFNQGYPNWGYPINTITSPPPHFPSLPNSEIGTNFIPGLPTATVGTIQQTINPPNTIGTIQQTINPPNTIGTIQQTIKPTHAIGTTQQTLLPSDTTKPNHPPSNYTSVHINPPRITPSVTTPVYQPVCGTNYATYENMDQFLSAQKSGKNIRIFLRRPCPLLAVGETKSIKVQCIASCYKSEDNQPICGSDDITYDNPAMLLCVQMCGYDVKVKHISPCRQVNNYTNNNPSISNGNDQTVNPLSPAVQISMNTFMDIIELCMTMNVSSNQIDYEQYCGDNIATLQNLLCTQNDNVNQIIITTPTPSATPTPTPTP…
15910	1	31	4	472	508	2.0	18.1	381	serpin2	MDAAAFSSAVAQFSTKFCNELDNTTNIVCSPLSAENLLALLTLGSTDPARTELLKALGFPDNDDHKSIRSTFGALNGKLKAIKGVTLLVANKIYIKDGGYEVEPELKKDAEDIFDTEFEKINFKDSASAAQLINQWVEHKTKNQIKDLFSSSSFSAFTRLVLVNALYFKGVWKNQFNPKDTIKQVFHLDDKKTVKIPMMFKEQKFNYYASPDLQAQLLEVSYAGEETSMVFILPDDIVGLNAVMQNLADGHDLMSEIKKMTPTKVKATLPKFKVETEIDLTKLLPQLGIKAIFNKDDSGLSELLSPAQEVYVTEAIQKVYIEVNETGREGGDGSGIDIRPTSYMADAVTPQSAYFRAVHPFLYLLMGPDNTILFIGAYRGN*
16754[d]	2	0	147	0	149	37.5	0.0	>85		MFFVILCLFLLSTTAAYTWQQTAYPEQQYQYNTPSLNLNFDQYWNDFNRGMNEFHRGMDELQRSIRKHARETCRRVEHPLKITKQ…
16782[d]	1	0	228	0	229	116.3	0.0	>85		MFFVILCLFLLGTTAAYTWQQTAYPEQQYQYNAPSLNLNFDQYWNDFNKGMDEFHRGMDELQRGIRKHARETCRRVEHPLKITKQ…
17202	0	0	216	1	217	110.1	1.2	185	SPH	MFSSKQSVVLAVAAVLFGCACAAPNPGANDIQLNQKLSIEAKGAKQPIDTRAVNERYPYAVRSFGGFCGGTIISPTWILTAGHCSILYAGSGLPAGTNITEVSSLYRFPKRLVIHPLFSIGPVWLNATEFNFKQVAARWDFLLIELEEPLPLDGKILAAAKLDDQPDLPAGLDVGYPSYRPTLRG*
17316[c]	0	0	98	0	98	50.0	0.0	496	Kazal	MQNSYVVILTFAVLGNTATATFWRRHGHGNWGHRPHWGHPHHVGYAYGQHPYHHLHPHNWGFNQGYPNWGYPINTITSPPPHFPSLPNSEIGTNFIPGLPTATVGTIQQTINPPNTIGTIQQTIKPTHAIGTTQQTLLPSDTTKPNHPPSNYTSVHINPPRITPSVTTPVYQPVCGTNYATYENMDQFLSAQKSGKNIRIFLRRPCPLLAVGETKSIKVQCIASCYKSEDNQPICGSDDITYDNPAMLLCVQMCGYDVKVKHISPCRQVNNYTNNNPSISNGNDQTVNPLSPAVQISMNTFMDIIELCMTMNVSSNQIDYEQYCGDNIATLQNLLCTQNDNVNQIIITTPTPSATPTPTPTPTTFWQPELVACVKVCPHTPEYNPVCGTNGITFENLSILRCVQLCGVSVNIHRASACTPAVVTQQETNDKNENKPLPNNNSSETIPSEVPLDNFGGVTKTTTEADETIDIDPRILQIANNKT*
17537[e]	0	0	1	320	321	0.5	380.0	>209	DM9	MYSVSPDAVFPPGTATSLNVRVSFLSAFPSNTMASPIRWIAATSDDASYLAPIAVVGGEDNYCNDKQALWVIRAKYECDLIPGELNSQRHTAYVPSDGVAHAVKDIEVCCAPRDKIQWITAGNGEVPP…RIGDAIVFEGNADKNDTRTFSEVAVPGGKTASGETLYIGRAKEHKSLIPGKIQPSLGHLYVTFKGKEIAKKYYEVLCTVN*
17568[e]	0	0	0	400	400	0.0	474.9	>118	DM9	…NCRAVASLNVRVSFLSAFPSNTMASPIRWIAATSDDASYLAPIAVVGGEDNYCNDKQALWVIRAKYECDLIPGELNSQRHTAYVPSDGVAHAVKDIEVCCAPRDKIQWITAGNGEVPP…
17610	9	2	82	159	252	4.6	94.4	>484	esterase	…FGVPYAKVDEGNPFGNSLNQPAFKKPFIANDSTILCPQATVFVGGILQCLTLNIYVPNQAGPSNTKAVFVWFYGGGFFFGYAGQYGGQYLVQQDIVVITVNYRLGPYGFLCLNDPKVPGNQGLKDQIAALRWIKANIGAFGGDPTKITIAGESYGGGSVDFHLYSMYEKLFDKTIVQSGSIFTPYVFGKGDPNAAIELAGKMGHESMNNDDAIKYLARADPVEVMKFARNLTNILRPCKEKPFKGVSSFMTTDPYHFQNSAKIRNAKILIGYNSKETFDTFVDKDDAFYNGLQNAFSERLGQIFTIKKEELETLAKIVRDFYLGGKPIAKESRLELSDFLSDFMVNHAAERSVNNYVKFGAEKVYKYIFSYIGGSPYKSLSGVGAFHTEELQYLFQMTQNLTSDEQIMMRNRMTEMWANFAKYGDPTPQVTSLLPVRWTPVTNTEARPYMNIDVNMEVKNYVNQQRMAFWDLIWNQYWKTSPAI*
18018	0	5	63	30	98	32.1	7.1	346	estarase	MIQLLVLLVMAAASDAHRQHGAAPSTDGPLTTSPSGTFRGSWMTTRKGQRIEAYRGVRYAEPPVGELRFKPPRLITQYKDVVDATQEGPACPQPVQNDYPVDEDCLRLNVYTHGHKGKLLPVIMYMHAGGFYSVSGRSDVAGPDYLLDRDVVLVTINYRLGSLGFLSTGDEQAPGNNGFKDQVMAMRWIQRNIAAFGGDPNLVTITGYSAGSFSVFLHMVSPTSLKVLFHRAISMSGSPVSQIMIPEHQRHLAERQARLLGVSTDSSKAIVDGLKTKTAKELGDSLLGMFEFDYDPVLLWVPVVEKDFGQERFLTMQPLDAIRQGKMHQVPYIISQTKDEFFWKAF*

* RA and ARN are calculated using original read numbers as described in Section 2.3. Listed here are contigs with $RA_{IF/CF}$ >15, $RA_{IH/CH}$ >15, ARN_{IF} >30 when RN_{CF} =0, or ARN_{IH} >30 when RN_{CH} =0. Contigs with total read numbers lower than 70 or GC content lower than 35% are not listed. Some of the contig sequences have been extended using sequences in dataset "06" (Table 1, Zou et al., 2008) and in the *M. sexta* gut EST dataset (Pauchet et al., 2009). $RA_{IF/CF}$ and $RA_{IH/CH}$ values are shown in red if they are greater than 15, while ARN_{IF} and ARN_{IH} values are shown in blue if they are higher than 30. In the two columns of RA or ARN, cells shaded yellow and blue represent fat body- and hemocyte-specific gene expression, respectively. The contigs labeled with the same letter (a to e) in superscript indicate high sequence similarity between them, as highlighted with different colors at certain key sites of the protein sequences.

Table 7. A list of DN CIFH contigs with BLAST hits*

CIFH contig #	Original read # CF	CH	IF	IH	Total	RA or ARN CF/IF	CH/IH	BLAST results
00010	29	464	3	286	782	19.0	1.4	gi|242005387|ref|XP_002423550.1| cAMP-dependent protein kinase subunit [Pediculus humanus corporis]
00015 &	112	4705	17	3918	8752	12.9	1.0	gi|6164595|gb|AAF04457.1|AF078161_1 lacunin [Manduca sexta] [00015, 02717]
00248	7	200	1	155	363	13.7	1.1	gi|157113908|ref|XP_001657920.1| n-acetyllactosaminidebeta-1,3-NAG transferase [Aedes aegypti]
00379	10	308	1	184	503	19.6	1.4	gi|170037242|ref|XP_001846468.1| Leu-rich repeat-containing protein 1[Culex quinquefasciatus]
00623	12	527	1	443	983	23.5	1.0	gi|157132531|ref|XP_001656056.1| odd Oz protein [Aedes aegypti]
00628	7	38	1	39	85	13.7	0.8	gi|170030982|ref|XP_001843366.1| rho/rac/cdc GTPase-activating protein [Culex quinquefasciatus]
00773	49	12	93	1	155	1.0	10.1	gi|157103945|ref|XP_001648193.1| dihydropyrimidine dehydrogenase [Aedes aegypti]
00851	6	42	1	26	75	11.8	1.4	gi|158300087|ref|XP_320080.3| AGAP009284-PA [Anopheles gambiae]
01289	7	45	1	31	84	13.7	1.2	gi|187281809|ref|NP_001119723.1| kinesin-like protein Ncd [Bombyx mori]
02637	5	12	9	1	27	1.1	10.1	gi|116789445|gb|ABK25249.1| unknown [Picea sitchensis]
02730	8	15	8	1	32	2.0	12.6	gi|2970687|gb|AAC06038.1| beta-glucosidase precursor [Spodoptera frugiperda]
03286 etc.	62	1976	7	586	2631	17.4	2.8	gi|254746344|emb|CAX16637.1| C1A Cys protease precursor [Manduca sexta] [03286, 05560, 15201, 17978]
03654	21	686	2	647	1356	20.6	0.9	gi|157134123|ref|XP_001663157.1| atlastin [Aedes aegypti]
03792	7	20	1	5	33	13.7	3.4	gi|91090218|ref|XP_968156.1| PREDICTED: similar to E1a binding protein P400 [Tribolium castaneum]
03996	6	6	1	6	19	11.8	0.8	gi|170052039|ref|XP_001862040.1| small GTP-binding protein [Culex quinquefasciatus]
05824	8	0	1	4	13	15.7	0.0	gi|116326818|ref|YP_803355.1| hypothetical protein TNAV2c gp132 [Trichoplusiani ascovirus 2c]
06497 etc.	225	10451	12	4266	14954	36.8	2.1	gi|217262|dbj|BAA03124.1| lectin [Bombyx mori] [06497, 15047, 15764, 16677, 16801, 16877, 16886, 17700]
06713	0	12	0	1	13	0.0	10.1	gi|193613364|ref|XP_001943860.1| limkain b1 [Acyrthosiphon pisum]
06902	12	3	2	0	17	11.8	2.5	gi|114050917|ref|NP_001040414.1| 3-hydroxyacyl-CoA dehydrogenase [Bombyx mori]
07139	21	767	2	262	1052	20.6	2.5	gi|110649216|emb|CAL25117.1| dVA-AP3 [Manduca sexta]
07515	7	1	1	0	9	13.7	0.8	gi|158295141|ref|XP_316035.4| AGAP005993-PA [Anopheles gambiae]
07642	9	601	1	153	764	17.7	3.3	gi|55139125|gb|AAV41236.1| immulectin-3 [Manduca sexta]
07754	0	12	1	1	14	0.0	10.1	gi|71895231|ref|NP_001026433.1| coiled-coil domain containing 93 [Gallus gallus]
08686 &	21	854	3	680	1558	13.7	1.1	gi|82880638|gb|ABB92836.1| scavenger receptor C-like protein [Spodoptera frugiperda] [08686, 15116]
08705	8	10	1	5	24	15.7	1.7	gi|224084416|ref|XP_002192181.1| selenium binding protein 1[Taeniopygia guttata]
08707	6	9	1	13	29	11.8	0.6	gi|24585081|ref|NP_609923.2| CG10639 [Drosophila melanogaster]
08801	1	14	1	1	17	2.0	11.8	gi|91081401|ref|XP_972667.1| exosome component 8[Tribolium castaneum]
09847	0	13	0	1	14	0.0	10.9	gi|194745608|ref|XP_001955279.1| GF16313 [Drosophila ananassae]
10124 etc.	115	4638	8	2848	7609	28.2	1.4	gi|114050871|ref|NP_001040411.1| carboxylesterase [Bombyx mori] [10124, 16922, 17330, 18860]
10316	0	13	1	1	15	0.0	10.9	gi|157106599|ref|XP_001649397.1| hypothetical protein AaeL_AAEL004554 [Aedes aegypti]
10439	12	0	1	0	13	23.5	0.0	gi|183979241|dbj|BAG30782.1| cuticular protein CPR41B [Papilio xuthus]
11030	13	0	2	0	15	12.7	0.0	gi|3121953|sp|Q25504.1|CU16_MANSE larval cuticle protein 16/17 precursor
11098	40	0	3	0	43	26.1	0.0	gi|159526|gb|AAA29320.1| methionine-rich storage protein 1 [Manduca sexta]
11161	0	12	1	1	14	0.0	10.1	gi|125808686|ref|XP_001360831.1| GA18253 [Drosophila pseudoobscurapseudoobscura]
11280 etc.	143	7866	11	2589	10609	25.5	2.6	gi|91090548|ref|XP_971239.1| hemolectin CG7002-PA [Tribolium castaneum] [11280, 15506, 15594, 18551]
12095	10	0	1	0	11	19.6	0.0	gi|194741936|ref|XP_001953465.1| GF17208 [Drosophila ananassae]
12848	0	16	0	1	17	0.0	13.5	gi|2822109|sp|P14730.2|EXPI_RAT extracellular peptidase inhibitor; WDNM1 precursor
13013	7	1	1	0	9	13.7	0.8	gi|189031278|gb|ACD74812.1| cuticle protein 1 [Helicoverpa armigera]
13094	15	10	1	5	31	29.4	1.7	gi|183979298|dbj|BAG30762.1| similar to CG5304-PA [Papilio xuthus]
13813	31	2398	4	848	3281	15.2	2.4	gi|110758905|ref|XP_395067.3| PREDICTED: similar to Hemolectin CG7002-PA [Apis mellifera]
13842	14	677	2	228	921	13.7	2.5	gi|138601|sp|P19616.1|VITM_MANSE microvitellogenin precursor
14129	7	0	1	0	8	13.7	0.0	gi|91078692|ref|XP_971204.1| phospholipase A2, grp VI (cytosolic, Ca-independent) [Tribolium castaneum]
14570 etc.	559	28386	29	10677	39651	37.8	2.2	gi|162462371|ref|NP_001104817.1| lectin [Bombyx mori] [14570, 15250, 15380, 15792, 16289, 16291, 16594, 16842, 17159, 17421, 17471, 17732, 17769, 18032, 18067, 18097, 18286, 18326, 18719, 18721, 18794, 18997]
14760 etc.	57	3372	3	1184	4616	37.3	2.4	gi|156545430|ref|XP_001606650.1| CG7002-PA [Nasonia vitripennis] [14760, 18045]
14781	28	0	3	1	32	18.3	0.0	gi|114052677|ref|NP_001040269.1| phosphoserine aminotransferase 1 [Bombyx mori]
15132	9	0	1	0	10	17.7	0.0	gi|112984526|ref|NP_001037199.1| promoting protein [Bombyx mori]
15465	6	0	1	1	8	11.8	0.0	gi|170574840|ref|XP_001892989.1| hypothetical protein Bm1_07595 [Brugia malayi]
16105	10	23	1	42	76	19.6	0.5	gi|91087179|ref|XP_975411.1| CG9471-PB [Tribolium castaneum]
16288 etc.	63	3044	4	1126	4237	30.9	2.3	gi|2738863|gb|AAB94557.1| hemocyte protease-1 [Manduca sexta] [16288, 16719, 17102]
17085 etc.	236	11035	27	7455	18753	17.1	1.2	gi|74763772|sp|O44249.3|MANSE proPO-P1 [17085, 17315, 17420, 17612, 17629, 18065, 18463, 18887]
17958 etc.	130	5309	19	3669	9127	16.3	1.2	gi|75038472|sp|Q25519.3|MANSE proPO-p2 [17958, 18004, 18516]
18482	11	0	0	0	11	21.6	0.0	gi|114240|sp|P14296.1|ARYA_MANSE Arylphorin subunit alpha precursor
18611	0	12	4	1	17	0.0	10.1	gi|12585261|sp|Q9U639.1|HSP7D_MANSE heat shock 70 kDa protein cognate 4, Hsp 70-4

* RA and ARN are calculated using original read numbers as described in Section 2.3. Listed here are contigs with $RA_{CF/IF}$ >10, $RA_{CH/IH}$ >10, ARN_{CF} >20 when RN_{IF} =0, or ARN_{CH} >20 when RN_{IH} =0. $RA_{CF/IF}$ and $RA_{CH/IH}$ values are shown in red if they are greater than 10, whereas ARN_{CF} and ARN_{CH} values are shown in blue if they are higher than 20. In the two columns of RA or ARN, cells shaded yellow and blue represent fat body- and hemocyte-specific gene expression, respectively. Contigs with identical BLAST results are combined, with their average RAs or ARNs calculated based on the sums of original reads in CF, CH, IF, and IH for each group.

Table 8. A list of HC CIFH contigs with BLAST hits*

CIFH contig #	Original read # CF	CH	IF	IH	Total	RA or ARN CH/CF	IH/IF	BLAST results
00010	29	464	3	286	782	3.3	46.0	gi|242005387|ref|XP_002423550.1| cAMP-dependent protein kinase catalytic subunit, [Pediculus humanus corporis]
00015 etc.	119	5073	20	4227	9439	8.8	102.0	gi|6164595|gb|AAF04457.1|AF078161_1 lacunin [Manduca sexta] (00015, 02717, 15269)
00028	13	958	4	754	1729	15.3	91.0	gi|91081003|ref|XP_975140.1| ~ odd Oz protein [Triboliumcastaneum]
00248	7	200	1	155	363	5.9	74.8	gi|157113908|ref|XP_001657920.1| n-acetyllactosaminidebeta-1,3-n-acetylglucosaminyltransferase [Aedes aegypti]
00379	10	308	1	184	503	6.4	88.8	gi|170037242|ref|XP_001846468.1| leucine-rich repeat-containing protein 1[Culex quinquefasciatus]
00541	14	567	7	760	1348	8.4	52.4	gi|170029717|ref|XP_001842738.1| Leu-rich repeat-containing G-protein coupled receptor 4 [Culex quinquefasciatus]
00569	4	182	1	176	363	9.4	85.0	gi|283135216|ref|NP_001164363.1| homeobox protein prospero [Nasoniavitripennis]
00623	12	527	1	443	983	9.1	213.8	gi|157132531|ref|XP_001656056.1| odd Oz protein [Aedes aegypti]
00752	0	38	1	164	203	7.9	79.2	gi|194859640|ref|XP_001969420.1| GG23966 [Drosophila erecta]
00802	3	203	3	253	462	14.0	40.7	gi|260840271|ref|XP_002613791.1| hypothetical protein BRAFLDRAFT_85332[Branchiostoma floridae]
00839	3	340	1	226	570	23.5	109.1	gi|242021897|ref|XP_002431379.1| conserved hypothetical protein [Pediculushumanus corporis]
00882	7	268	0	261	536	7.9	126.0	gi|112983326|ref|NP_001037620.1| ras-related GTP-binding protein Rab3 [Bombyxmori]
01064	5	134	1	116	256	5.6	56.0	gi|48095930|ref|XP_394560.1| Jagged-1 precursor (Jagged1)(hJ1) (CD339 antigen) [Apis mellifera]
01429 &	27	924	4	827	1782	7.1	99.8	gi|157134123|ref|XP_001663157.1| atlastin [Aedes aegypti] (01429, 03654)
01609	1	71	1	144	217	14.7	69.5	gi|134001247|gb|ABO45233.1| reverse transcriptase [Ostrinia nubilalis]
02159	3	101	1	144	249	7.0	69.5	gi|114052056|ref|NP_001040346.1| septin [Bombyx mori]
02473	10	255	2	382	649	5.3	92.2	gi|281362668|ref|NP_651533.2| eater [Drosophila melanogaster]
02852	23	1128	7	885	2043	10.2	61.0	gi|66391199|ref|YP_239364.1| hypothetical protein [Microplitis demolitorbracovirus]
03225	1	25	1	143	170	5.2	69.0	gi|195445668|ref|XP_002070431.1| GK11035 [Drosophila willistoni]
03246 &	4	182	2	245	433	9.4	59.1	gi|83583697|gb|ABC24708.1| G protein-coupled receptor [Spodoptera frugiperda] (03246, 06319)
03287	7	493	0	237	737	14.6	114.4	gi|114052174|ref|NP_001040228.1| aminoacylase [Bombyx mori]
04085	0	34	3	268	305	7.0	43.1	gi|206725499|ref|NP_001128673.1| cathepsin L like protein [Bombyx mori]
04278	3	141	1	154	299	9.7	74.3	gi|270001550|gb|EEZ97997.1| hypothetical protein TcasGA2_TC000395 [Triboliumcastaneum]
04746 etc.	0	0	16	1939	1955	0.0	58.5	gi|195486646|ref|XP_002091593.1| GE13745 [Drosophila yakuba] (04746, 13353, 14100)
05560	24	965	4	440	1433	8.3	53.1	gi|254746344|emb|CAX16637.1| putative C1A cysteine protease precursor [Manducasexta]
05577	4	157	22	1895	2078	8.1	41.6	gi|254746342|emb|CAX16636.1| putative C1A cysteine protease precursor [Manducasexta]
05933 etc.	39	1862	8	1395	3304	9.9	84.2	gi|82880638|gb|ABB92836.1| SR-C-like protein [Spodopterafrugiperda] (05933, 08686, 13271, 15116, 15350, 15564)
06497 etc.	237	11297	15	4531	16080	9.9	145.8	gi|217262|dbj|BAA03124.1| lectin [Bombyx mori] (06497, 15047, 15764, 15986, 16677, 16801, 16877, 16886, 17700)
07139	21	767	2	262	1052	7.6	63.2	gi|110649216|emb|CAL25117.1| dVA-AP3 [Manduca sexta]
07199	2	73	1	102	178	7.6	49.2	gi|110649250|emb|CAL25134.1| immulectin III [Manduca sexta]
07480	3	248	2	193	446	17.1	46.6	gi|91086517|ref|XP_971701.1| ~ NtR CG6698-PA [Triboliumcastaneum]
07642 etc.	17	1246	3	562	1828	15.2	90.4	gi|55139125|gb|AAV41236.1| immulectin-3 [Manduca sexta] (07642, 13452, 14991)
07883	0	0	3	792	795	0.0	127.4	gi|157128533|ref|XP_001661472.1| hypothetical protein AaeL_AAEL011180 [Aedesaegypti]
08524 etc.	74	3481	7	1984	5546	9.8	136.8	gi|2738863|gb|AAB94557.1| hemocyte protease-1 [Manduca sexta] (08524, 12527, 16288, 16719, 17102)
10124 etc.	162	6970	18	4204	11354	8.9	112.7	gi|114050871|ref|NP_001040411.1| carboxylesterase [Bombyx mori] (10124, 15112, 16627, 16922, 17330, 18860)
11280 etc.	143	7866	11	2589	10609	11.4	113.6	gi|91090548|ref|XP_971239.1| Hemolectin CG7002-PA[Tribolium castaneum] (11280, 15506, 15594, 18551)
13813	31	2398	4	848	3281	16.0	102.3	gi|110758905|ref|XP_395067.3| ~ hemolectin CG7002-PA [Apismellifera]
13842	14	677	2	228	921	10.0	55.0	gi|138601|sp|P19616.1|VITM_MANSE Microvitellogenin precursor
14248 etc.	1	150	15	2329	2495	31.1	74.9	gi|2149091|gb|AAB58491.1| serpin-2 [Manduca sexta] (14248, 15111, 16917, 17058, 17751)
14570 etc.	562	29402	26	11144	41134	10.8	206.9	gi|162462371|ref|NP_001104817.1| lectin [B. mori] (14570, 15250, 15380, 15792, 16278, 16289, 16291, 16594, 16842, 17159, 17421, 17471, 17732, 17769, 18032, 18067, 18073, 18089, 18097, 18286, 18326, 18719, 18721, 18794)
14760 &	57	3372	3	1184	4616	12.3	190.5	gi|156545430|ref|XP_001606650.1| ~CG7002-PA [Nasoniavitripennis] (14760, 18045)
14811	5	136	1	121	263	5.6	58.4	gi|221055473|ref|XP_002258875.1| hypothetical protein, conserved in Plasmodium [Plasmodium knowlesi strain H]
15584	3	241	1	202	447	16.7	97.5	gi|66535330|ref|XP_623280.1| ~atlastin CG6668-PA, isoformA [Apis mellifera]
16815 etc.	208	9161	39	6243	15651	9.1	77.3	gi|75038472|sp|Q25519.3|PRP2_MANSE proPO-2 (16815, 17417, 17958, 18004, 18516, 18811)
17085 etc.	261	12058	33	8286	20638	9.6	121.2	gi|74763772|sp|O44249.3|PRP1_MANSE proPO-1 (17085, 17315, 17420, 17612, 17629, 17562, 18065, 18463, 18887)

* RA and ARN are calculated using original read numbers as described in Section 2.3. Listed here are contigs with $RA_{IH/IF}$ >40, $RA_{CH/CF}$ >40, ARN_{IH} >80 when RN_{IF} =0, or ARN_{CH} >80 when RN_{CF} =0. $RA_{IH/IF}$ and $RA_{CH/CF}$ values are shown in red if they are greater than 40, whereas ARN_{IH} and ARN_{CH} values are shown in blue if they are higher than 80. In the columns of RA or ARN, cells shaded green and orange represent down- and up-regulated gene expression, respectively. Contigs with identical BLAST results are combined, with their average RAs or ARNs calculated based on the sums of original reads in CF, CH, IF, and IH for each group.

Table 9. A list of FB CIFH contigs with BLAST hits*

CIFH contig #	Original read # CF	CH	IF	IH	Total	RA or ARN CF/CH	IF/IH	BLAST results
00051	291	1	329	0	621	1403.9	681.6	gi|183979376|dbj|BAG30740.1| muscle myosin heavy chain [Papilio xuthus]
00153 etc.	2069	4	2563	1	4637	2495.3	5309.8	gi|2498144|sp|Q25490.1 apoLp (00153 02405 02406 03748 04510 06831 06834 07770 14087 14589)
00194	37	0	81	1	119	178.5	167.8	gi|48476133|gb|AAT44358.1| calcium-activated potassium channel alpha subunit[Manduca sexta]
00285 &	298	23	921	5	1247	62.5	381.6	gi|73921301|gb|AAG42021.2|AF327882_1 JHE precursor[Manduca sexta] (00285, 00859)
00409	168	0	216	0	384	810.5	447.5	gi|110750043|ref|XP_394261.3| plexin A CG11081-PA, isoform A [Apis mellifera]
00414	58	1	50	0	109	279.8	103.6	gi|195382713|ref|XP_002050074.1| GJ21937 [Drosophila virilis]
00423	149	0	220	0	369	718.8	455.8	gi|158295580|ref|XP_316291.4| AGAP006225-PA [Anopheles gambiae str. PEST]
00465	134	1	230	0	365	646.4	476.5	gi|149755131|ref|XP_001491560.1| hemicentin 1 [Equuscaballus]
00535	67	1	100	0	168	323.2	207.2	gi|242015135|ref|XP_002428229.1| Thrombospondin-3 precursor [Pediculus humanus corporis]
00575	3	0	259	5	267	14.5	107.3	gi|154240658|dbj|BAF74637.1| peptidoglycan recognition protein-D [Samia cynthiaricini]
00609	324	0	762	0	1086	1563.1	1578.6	gi|225542786|gb|ACN91276.1| dentin sialophosphoprotein precursor [Bos taurus]
00737	2	4	131	2	139	2.4	135.7	gi|198466442|ref|XP_002135189.1| GA23919 [Drosophila pseudoobscurapseudoobscura]
00748	131	4	118	2	255	158.0	122.2	gi|29346557|ref|NP_810060.1| glycine dehydrogenase [Bacteroidesthetaiotaomicron VPI-5482]
00766	45	0	74	1	120	217.1	153.3	gi|158293377|ref|XP_314728.3| AGAP008632-PA [Anopheles gambiae str. PEST]
00773	49	12	93	1	155	19.7	192.7	gi|157103945|ref|XP_001648193.1| dihydropyrimidine dehydrogenase [Aedesaegypti]
00785	120	2	139	2	263	289.5	144.0	gi|193795848|gb|ACF21977.1| paramyosin [Bombyx mandarina]
00884	39	1	23	0	63	188.1	47.6	gi|156553304|ref|XP_001599652.1| GA21752-PA [Nasoniavitripennis]
00960	52	2	99	1	154	125.4	205.1	gi|157107996|ref|XP_001650030.1| sarcosine dehydrogenase [Aedes aegypti]
01095	64	0	99	1	164	308.8	205.1	gi|169639235|gb|ACA60733.1| venom acid phosphatase [Pteromalus puparum]
01097	134	2	436	5	577	323.2	180.7	gi|55139125|gb|AAV41236.1| immulectin-3 [Manduca sexta]
01127	41	1	52	1	95	197.8	107.7	gi|189491898|gb|ACE00761.1| adipokinetic hormone receptor [Manduca sexta]
01454	599	3	1337	3	1942	963.2	923.3	gi|91082539|ref|XP_973726.1| inter-α (globulin) inhibitor H4 (Kallikrein-sensitive) [T. castaneum]
01480	211	0	729	0	940	1017.9	1510.3	gi|183979392|dbj|BAG30748.1| hypothetical protein [Papilio xuthus]
01601	60	1	79	0	140	289.5	163.7	gi|270005801|gb|EFA02249.1| hypothetical protein TcasGA2_TC007912 [Triboliumcastaneum]
01742	65	0	75	0	140	313.6	155.4	gi|283100192|gb|ADB08386.1| sugar transporter 4 [Bombyx mori]
01743	27	0	112	0	139	130.3	232.0	gi|134252572|gb|ABO65045.1| beta-hexosaminidase [Ostrinia furnacalis]
01870	184	0	323	0	507	887.7	669.2	gi|242010783|ref|XP_002426138.1| conserved hypothetical protein [Pediculushumanus corporis]
01892	82	0	108	0	190	395.6	223.7	gi|158289807|ref|XP_311448.4| AGAP010734-PA [Anopheles gambiae str. PEST]
01915	85	2	275	0	362	205.0	569.7	gi|110757936|ref|XP_623940.2| Peroxidase precursor (DmPO)[Apis mellifera]
01956	127	0	99	0	226	612.7	205.1	gi|156551746|ref|XP_001602035.1| ENSANGP00000015052[Nasonia vitripennis]
01972 etc.	383	0	3327	0	3710	1847.7	6892.5	gi|136206|sp|P22297.1|transferrin (01972 10382 11027 14937 17193 17206 17395 16606 18234 18308)
02101	51	0	75	0	126	246.0	155.4	gi|186909546|gb|ACC94296.1| glucose oxidase-like enzyme [Helicoverpa armigera]
02104	59	1	67	1	128	284.6	138.8	gi|91079628|ref|XP_967731.1| PREDICTED: similar to AGAP002355-PA [Triboliumcastaneum]
02137	101	0	24	0	125	487.2	49.7	gi|91084191|ref|XP_967340.1| PREDICTED: similar to AGAP002557-PA [Triboliumcastaneum]
02144	82	0	132	3	217	395.6	91.2	gi|62002223|gb|AAX58711.1| pheromone-degrading enzyme 1 [Antheraea polyphemus]
02166	60	0	57	0	117	289.5	118.1	gi|193876254|gb|ACF24761.1| lipid storage droplet protein 1 [Manduca sexta]
02184	53	2	111	1	167	127.8	230.0	gi|226342886|ref|NP_001139705.1| serpin 13 [Bombyx mori]
02219	454	3	971	3	1431	730.1	670.5	gi|219815604|gb|ACL36977.1| putative ecdysone oxidase [Helicoverpa zea]
02329	143	0	411	0	554	689.9	851.5	gi|112984054|ref|NP_001037422.1| yellow1 [Bombyx mori]
02337 &	107	2	170	7	286	258.1	50.3	gi|91079867|ref|XP_967070.1| AGAP005945-PB [Triboliumcastaneum] (02337, 15796)
02361	7	4	70	1	82	8.4	145.0	gi|56418425|gb|AAV91020.1| hemolymph proteinase 22 [Manduca sexta]
02393	45	0	77	5	127	217.1	31.9	gi|156545523|ref|XP_001607196.1| Dihydroxyacetone kinase2 homolog (yeast) [Nasonia vitripennis]
02394	28	1	23	0	52	135.1	47.6	gi|91077746|ref|XP_966706.1| conserved hypotheticalprotein [Tribolium castaneum]
02409	113	0	187	0	300	545.1	387.4	gi|109502352|gb|ABE01157.2| carboxylesterase [Spodoptera litura]
02482	63	0	85	1	149	303.9	176.1	gi|66519258|ref|XP_625210.1| PREDICTED: similar to CG6188-PA [Apis mellifera]
02609	97	0	146	2	245	468.0	151.2	gi|156968285|gb|ABU98614.1| alpha-amylase [Helicoverpa armigera]
02638 &	241	0	206	0	447	1162.6	426.8	gi|41016826|sp|Q27772.3|C1TC_SPOFR C-1-THF synthase, cytoplasmic (02638, 07658)
02651	24	0	124	0	148	115.8	256.9	gi|5326830|gb|AAD42058.1|AF122899_1 plasmatocyte-spreading peptide precursor [Manduca sexta]
02800	28	0	97	0	125	135.1	201.0	gi|260765449|gb|ACX49762.1| beta-fructofuranosidase 1 [Manduca sexta]
02847	33	0	103	0	136	159.2	213.4	gi|114051702|ref|NP_001040423.1| zinc-containing alcohol dehydrogenase [Bombyxmori]
02931 &	187	0	429	0	616	902.1	888.8	gi|1658003|gb|AAB18243.1| microsomal epoxide hydrolase [Trichoplusia ni] (02931, 04388)
02947	518	21	981	56	1576	119.0	36.3	gi|259493819|gb|ACW82749.1| hemocyte aggregation inhibitor protein precursor [Manduca sexta]
02979	49	0	92	4	145	236.4	47.6	gi|52782757|sp|Q9NJ98.1|BGRP1_MANSE Beta-1,3-glucan recognitionprotein 1; BetaGRP-1
02985	3	0	158	0	161	14.5	327.3	gi|56418466|gb|AAV91027.1| serine proteinase-like protein 4 [Manduca sexta]
03185	106	0	234	10	350	511.4	48.5	gi|157117489|ref|XP_001658792.1| 3-hydroxyacyl-CoA dehyrogenase [Aedesaegypti]
03224	98	0	477	0	575	472.8	988.2	gi|226342906|ref|NP_001139715.1| serpin 22 [Bombyx mori]
03226	222	0	663	0	885	1071.0	1373.5	gi|153791757|ref|NP_001093275.1| myo-inositol oxygenase [Bombyx mori]
03395	22	1	24	0	47	106.1	49.7	gi|157908523|dbj|BAF81491.1| juvenile hormone epoxide hydrolase [Bombyx mori]
03415	190	0	216	1	407	916.6	447.5	gi|2708688|gb|AAB92583.1| acyl-CoA delta-9 desaturase [Trichoplusia ni]
03434	1	0	387	0	388	4.8	801.7	gi|189234566|ref|XP_001815977.1| Kaz1-ORFB CG1220-PE [Tribolium castaneum]
03454	28	0	102	0	130	135.1	211.3	gi|6560669|gb|AAF16712.1|AF117590_1 unknown [Manduca sexta]
03483	280	0	374	0	654	1350.8	774.8	gi|283558277|gb|ADB27116.1| aliphatic nitrilase [Bombyx mori]
03712	49	2	157	5	213	118.2	65.1	gi|170779021|gb|ACB36909.1| glutathione S-transferase theta [Antheraea pernyi]
03737	167	1	197	0	365	805.6	408.1	gi|56462300|gb|AAV91433.1| putative serine protease-like protein 2 [Lonomiaobliqua]
03776 etc.	204	8	960	51	1223	123.0	39.0	gi|112983872|ref|NP_001036857.1| Serpin-like protein [Bombyx mori] (03776, 06215, 06531, 17814)
04012 &	167	3	727	11	908	268.5	136.9	gi|27733411|gb|AAO21503.1|AF413062_1 leureptin [Manduca sexta] (04012, 08453)
04413	69	1	133	1	204	332.9	275.5	gi|194743582|ref|XP_001954279.1| GF18195 [Drosophila ananassae]
04424	72	0	64	0	136	347.3	132.6	gi|114052020|ref|NP_001040445.1| tropomyosin 1 [Bombyx mori]
04430	74	0	68	0	142	357.0	140.9	gi|114052573|ref|NP_001040481.1| phosphoribosyl pyrophosphate synthetase[Bombyx mori]
04498	46	0	115	0	161	221.9	238.2	gi|90025232|gb|ABD85119.1| juvenile hormone epoxide hydrolase [Spodopteraexigua]
04504	53	0	135	0	188	255.7	279.7	gi|7239259|gb|AAF43151.1|AF226857_1 hemolymph JHBP precursor [Manduca sexta]
04722 &	578	0	861	0	1439	2788.4	1783.7	gi|116791778|gb|ABK26104.1| unknown [Picea sitchensis] (04722, 04994)
04781	56	0	237	0	293	270.2	491.0	gi|118359591|ref|XP_001013035.1| PHD-finger family protein [Tetrahymenathermophila]
04786	61	0	62	0	123	294.3	128.4	gi|219686082|emb|CAW30924.1| putative aldo-ketose reductase 1 [Papiliodardanus]

04791	144	0	200	0	344	694.7	414.3	gi|116788175|gb|ABK24783.1| unknown [Picea sitchensis]
04806	518	1	372	0	891	2499.0	770.7	gi|157122933|ref|XP_001659963.1| actin [Aedes aegypti]
04808	0	0	426	2	428	0.0	441.3	gi|237861314|gb|AAV41237.2| immulectin-4 [Manduca sexta]
04830 etc.	59	2	755	6	822	142.3	260.7	gi|169646838|ref|NP_001112375.1| heat shock protein 25.4 [Bombyx mori] (04830, 04887, 05717)
05038 &	101	0	175	1	277	487.2	362.5	gi|110759694|ref|XP_394781.3| rTS beta protein [Apismellifera] (05038, 05832)
05136	1074	11	1041	37	2163	471.0	58.3	gi|114051966|ref|NP_001040198.1| mitochondrial aldehyde dehydrogenase [Bombyxmori]
05324	68	0	88	0	156	328.0	182.3	gi|225346695|gb|AcCN86370.1| troponin I transcript variant C [Bombyx mandarina]
05348	50	0	67	0	117	241.2	138.8	gi|189234391|ref|XP_974849.2| GA16498-PA [Triboliumcastaneum]
05417 etc.	273	0	917	0	1190	1317.0	1899.7	gi|260907784|gb|ACX53694.1| alcohol DH [Heliothis virescens] (05417, 05461, 07389, 07432)
05984	89	0	97	0	186	429.4	201.0	gi|56462260|gb|AAV91413.1| myosin 3 light chain [Lonomia obliqua]
06175	11	0	52	1	64	53.1	107.7	gi|170070451|ref|XP_001869584.1| conserved hypothetical protein [Culexquinquefasciatus]
06227	251	1	715	0	967	1210.9	1481.3	gi|124527|sp|Q00630.1|ICYB_MANSE insecticyanin-B (INS-b), blue biliprotein
06251	66	2	57	7	132	159.2	16.9	gi|158289206|ref|XP_310956.4| AGAP000179-PA [Anopheles gambiae str. PEST]
06394	51	0	228	0	279	246.0	472.3	gi|110611262|gb|ABG77980.1| alanine-glyoxylate transaminase 1 [Glossinamorsitans morsitans]
06588	60	0	75	0	135	289.5	155.4	gi|56462256|gb|AAV91411.1| myosin 1 light chain [Lonomia obliqua]
06597	60	0	200	0	260	289.5	414.3	gi|56462320|gb|AAV91443.1| putative secreted peptide 30 [Lonomia obliqua]
06732	115	1	244	0	360	554.8	505.5	gi|25090512|sp|Q25513.1|HGLY_MANSE 27 kDa hemolymph glycoprotein;
06789 &	159	0	460	0	619	767.1	953.0	gi|156968291|gb|ABU98617.1| unknown [Helicoverpa armigera] (06789, 06876)
06975 &	106	2	212	2	322	255.7	219.6	gi|189237651|ref|XP_001813448.1| N-acetyl neuraminatelyase [Tribolium castaneum] (06975, 14637)
07116 &	1	4	1015	4	1024	1.2	525.7	gi|171262319|gb|ACB45566.1| lebocin-like protein [Antheraea pernyi] (07116, 10853)
07565	24	1	14	0	39	115.8	29.0	gi|7862150|gb|AAF70499.1|AF255341_1 3-dehydroecdysone 3alpha-reductase[Spodoptera littoralis]
07608 etc.	353	3	3931	0	4287	567.7	8143.9	gi|159526|gb|AAA29320.1| methionine-rich storage protein 1 (07608, 07975, 08141, 14688)
07629	65	0	82	0	147	313.6	169.9	gi|77415676|emb|CAJ01507.1| hypothetical protein [Manduca sexta]
07639 &	811	0	1616	18	2445	3912.5	186.0	gi|134436|sp|P14754.1|Alaserpin; serpin-1 (07639, 15891)
07671	227	3	450	3	683	365.0	310.8	gi|195164814|ref|XP_002023241.1| GL21066 [Drosophila persimilis]
08076 &	47	3	115	2	167	75.6	119.1	gi|226342878|ref|NP_001139701.1| serpin 7 [Bombyx mori] (08076, 14528)
08224 etc.	7528	8	10093	0	17629	4539.6	20909.2	gi|1168527|sp|P14297.2|arylphorin β subunit (08224, 16474, 16501, 16664, 16715, 16764, 18695)
08467	0	0	113	0	113	0.0	234.1	gi|112983866|ref|NP_001036858.1| T7 lysozyme-like protein 1(BTL-LP1) [Bombyx mori]
08500	138	0	407	0	545	665.7	843.2	gi|156406857|ref|XP_001641261.1| predicted protein [Nematostella vectensis]
08821	246	0	436	2	684	1186.8	451.6	gi|112983550|ref|NP_001036879.1| fibrillin-like protein [Bombyx mori]
08845	27	0	130	0	157	130.3	269.3	gi|195029763|ref|XP_001987741.1| GH19797 [Drosophila grimshawi]
08854 &	302	5	5234	0	5541	291.4	10843.3	gi|5869985|emb|CAB55603.1| moderately Met-rich storage protein [Spodoptera litura] (08854, 15324)
09928	30	0	106	0	136	144.7	219.6	gi|242090851|ref|XP_002441258.1| hypothetical protein SORBIDRAFT_09g023310[Sorghum bicolor]
10071 &	15	1	1243	0	1259	72.4	2575.1	gi|228382|prf||1803340A Met-rich storage protein SP1A (10071, 17516)
10326	284	4	299	11	598	342.5	56.3	gi|56462160|gb|AAV91363.1| hypothetical protein 10 [Lonomia obliqua]
10791 etc.	2	0	2756	2	2760	9.6	2854.8	gi|4262357|gb|AAD14591.1| scolexin A [Manduca sexta] (10791, 10792, 16520, 18669, 18670, 18963)
11039 etc.	13962	11	19836	0	33809	6123.3	41094.2	gi|114240|sp|P14296.1|arylphorin α subunit (11039 16171 16537 16814 17492 18240 18257 18556)
11830	26	1	33	0	60	125.4	68.4	gi|260780799|ref|XP_002585527.1| hypothetical protein BRAFLDRAFT_89257 [B. floridae]
11922 &	901	12	1052	3	1968	362.2	726.5	gi|114058|sp|P13276.1|ApoLp-III; (11922, 13093)
12005	154	0	2177	0	2331	742.9	4510.1	gi|2625150|gb|AAB86646.1| moderately methionine rich hexamerin precursor[Hyalophora cecropia]
12151	0	0	153	0	153	0.0	317.0	gi|116084|sp|P14665.1|Bactericidin B-5P; Cecropin-like peptide B-5; precursor
12749	135	0	1462	0	1597	651.3	3028.8	gi|159530|gb|AAA29322.1| methionine-rich storage protein 3 [Manduca sexta]
13563	0	0	657	0	657	0.0	1361.1	gi|110347786|gb|ABG72695.1| attacin-like protein [Antheraea mylitta]
13916 etc.	3	0	1327	0	1330	14.5	2749.1	gi|219958086|gb|ACL68097.1| lebocin-related protein precursor [M. sexta] (13916, 17301, 17434)
13994	57	0	62	0	119	275.0	128.4	gi|112983654|ref|NP_001036872.1| Bombyrin [Bombyx mori]
14173	45	0	32	0	77	217.1	66.3	gi|153792114|ref|NP_001093267.1| phosphatidylethanolamine binding protein[Bombyx mori]
14375 etc.	400	0	681	0	1081	1929.7	1410.8	gi|400673|sp|P31420|OMBP Ommochrome-binding protein precursor (14375, 14659, 17494, 17813)
14380 etc.	0	1	408	3	412	0.0	281.8	gi|67906420|gb|AAY82587.1| attacin-1 [Manduca sexta] (14380, 14641, 16150)
14700 &	0	0	301	4	305	0.0	155.9	gi|260765453|gb|ACX49764.1| peptidoglycan recognition protein 2 [Manduca sexta] (14700, 14752)
15089	271	0	194	0	465	1307.4	401.9	gi|158293921|ref|XP_315269.4| AGAP011516-PA [Anopheles gambiae str. PEST]
15639	10	0	109	0	119	48.2	225.8	gi|148298818|ref|NP_001091784.1| multi-binding protein [Bombyx mori]
16000	61	0	138	0	199	294.3	285.9	gi|109458629|ref|XP_001073545.1| hypothetical protein [Rattusnorvegicus]
16223	22	1	47	2	72	106.1	48.7	gi|242003442|ref|XP_002422733.1| bifunctional purine biosynthesis protein [Pediculus corporis]
16281 &	358	0	541	0	899	1727.1	1120.8	gi|134103857|gb|ABO60878.1| cationic peptide CP8 precursor [Manduca sexta] (16281, 17312)
16849	134	0	541	0	675	646.4	1120.8	gi|114051738|ref|NP_001040426.1| alcohol dehydrogenase [Bombyx mori]
17199	42	2	33	4	81	101.3	17.1	gi|3108073|gb|AAC15763.1| putative multifunctional protein ADE2 [Manducasexta]
17350	0	0	205	0	205	0.0	424.7	gi|29469969|gb|AAO74640.1| antimicrobial protein attacin 2 [Manduca sexta]
18797	9	0	549	0	558	43.4	1137.4	gi|39843367|gb|AAR32136.1| VHDL receptor [Helicoverpa zea]

* RA and ARN are calculated using original read numbers as described in Section 2.3. Listed here are contigs with $RA_{IF/IH}$ >100, $RA_{CF/CH}$ >100, ARN_{IF} >200 when RN_{IH} =0, or ARN_{CF} >200 when RN_{CH} =0. $RA_{IF/IH}$ and $RA_{CF/CH}$ values are shown in red if they are greater than 100, whereas ARN_{IF} and ARN_{CF} values are shown in blue if they are higher than 200. In the columns of RA or ARN, cells shaded green and orange represent down- and up-regulated gene expression, respectively. Contigs with identical BLAST results are combined, with their average RAs or ARNs calculated based on the sums of original reads in CF, CH, IF, and IH for each group.

Table 10. Summary statistics for mass spectra and protein IDs

	gel 1[a]	gel 2	gel 3	gel 4	gel 5	gel 6	gel 7	gel 8	gel 9	gel 1-9	AcCN[b]
apparent M_r range (kDa)	201-350	121-200	81-120	71-80	51-70	46-50	25-45	20-24	10-19	10-350	n.c.
total spectra	144,558	165,475	188,699	184,148	210,706	177,400	231,982	119,325	172,220	1,594,513	202,077
matching spectra [c]	20,004	33,885	41,226	31,484	56,536	31,476	50,820	14,478	27,265	314,866	41,706
# of protein IDs	59	121	171	111	271	167	319	123	255	785	270
Protein FDR[d]	1.7%	0.8%	0.6%	0.9%	0.4%	0.6%	0.0%	0.0%	0.4%	0.5%	0.4%
Peptide FDR	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%

[a] Gel 1~9: protein samples from gel slices 1~9; [b] AcCN: acetonitrile-treated plasma samples; [c] matching spectra: number of spectra that match certain protein in the database; [d] FDR: false discovery rate.

Table 11. Pearson pairwise correlation among biological replicates

		AcCN-treated or "peptide" samples						
		CH1	CH2	CH3	IH1	IH2	IH3	
Gel-extracted or "protein" samples	CH1	1	0.9490	0.9381	0.5990	0.5463	0.5536	AcCN-treated or "peptide" samples
	CH2	0.9760	1	0.9647	0.6425	0.5766	0.5918	
	CH3	0.9884	0.9714	1	0.6541	0.5807	0.5907	
	IH1	0.9334	0.9448	0.9360	1	0.9362	0.9629	
	IH2	0.9405	0.9415	0.9504	0.9901	1	0.9416	
	IH3	0.9385	0.9453	0.9454	0.9881	0.9896	1	
Gel extracted or "protein" samples								

Table 12. T-test results of 157 overlapping proteins*

	"protein" samples			
"peptide" samples		significant	insignificant	total
	significant	51	46	97
	insignificant	14	46	60
	total	65	92	157

Table 13. A list of 115 up-regulated proteins after immune challenge*

ID	Function	p-value	IH/CH	Source
gi|29469965	antimicrobial peptide cecropin 6	0.0008	17.31	AcCN, Gel
gi|29469969	antimicrobial protein attacin 2	0.0000	106.97	AcCN, Gel
CUFF.24771.1	apyrase	0.0170	2.70	Gel
CUFF.19825.1	attacin II	0.0000	98.84	Gel
c25131	attacin-1	0.0000	40.86	AcCN, Gel
c2978	attacin-1	0.0001	130.79	AcCN, Gel
CUFF.19826.2	attacin-1	0.0003	72.41	AcCN
CUFF.19826.3	attacin-1	0.0000	13.05	Gel
CUFF.19826.4	attacin-1	0.0000	508.58	AcCN, Gel
CUFF.19826.5	attacin-1	0.0000	176.15	AcCN
CUFF.19828.1	attacin-1	0.0000	162.75	AcCN, Gel
CUFF.19828.3	attacin-1	0.1200	98.55	AcCN
gi|67906420	attacin-1	0.0000	143.86	AcCN, Gel
lrc6486	attacin-1	0.3700	29.21	AcCN, Gel
contig12105	bactericidin	0.0005	25.83	AcCN
CUFF.10471.3	Cadherin-23 precursor, putative	0.2500	5.52	Gel
contig01648	CALNUC	0.0450	10.60	Gel
CUFF.17912.1	carboxylesterase CarE-7	0.0140	4.05	Gel
gi|115654	Casocidin-I	0.4700	5.12	Gel
contig13987	cecropin 3	0.0029	2.18	AcCN
CUFF.24081.1	cold-related protein	0.3000	8.15	Gel
CUFF.21538.1	cuticle protein 1	0.0180	6.53	Gel
contig11699	cuticle protein 4	0.1800	5.19	Gel
lrc23	dorsal	0.0069	2.08	Gel
CUFF.25705.2	esterase	0.0022	15.67	AcCN, Gel
CUFF.25705.3	esterase	0.0013	2.50	Gel
contig03139	FK506-binding protein precursor	0.0370	2.86	Gel
contig10194	gallerimycin	0.0000	21.00	AcCN, Gel
contig02032	gloverin	0.0014	3.53	Gel
gi|110649240	gloverin	0.0002	3.60	AcCN
contig04865	heat shock protein 25.4	0.0120	10.87	Gel
contig04960	heat shock protein 25.4	0.0012	3.75	Gel
contig05548	heat shock protein 25.4	0.0790	12.90	AcCN, Gel
contig05861	heat shock protein 25.4	0.1200	27.94	Gel
contig08771	heat shock protein 25.4	0.1200	5.77	Gel
CUFF.23729.1	hemolin	0.0001	12.54	Gel
gi|511297	hemolin	0.0000	5.44	AcCN
CUFF.14527.1	hemolymph proteinase 17	0.0720	12.93	Gel
contig04271	hemolymph proteinase 20	0.0069	4.16	AcCN
CUFF.25442.1	hemolymph proteinase 20	0.0004	2.75	Gel
contig02033	hemolymph proteinase 22, partial	0.0490	15.83	Gel
CUFF.31629.1	hemolymph proteinase 5	0.0370	5.76	Gel
lrc512	hypothetical protein CHLNCDRAFT_55808	0.0250	9.28	AcCN, Gel
contig04364	hypothetical protein KGM_01134	0.0490	4.96	Gel
contig08449	hypothetical protein KGM_06199	0.0092	18.32	AcCN, Gel
CUFF.31784.1	hypothetical protein KGM_16225	0.0006	9.17	Gel
CUFF.1806.1	hypothetical protein KGM_20797	0.0170	7.39	Gel
CUFF.19650.2	hypothetical protein KGM_21511	0.0090	2.15	Gel
contig01508	IML1	0.0170	4.95	Gel
contig04357	immulectin-4	0.0890	7.98	Gel
gi|27733419	immune-induced protein 1	0.0100	23.09	AcCN, Gel
contig02384	immune-related Hdd1	0.0000	10.85	AcCN, Gel
CUFF.19800.1	integument esterase 2	0.0570	9.32	Gel
c707	integument esterase 2 precursor	0.0000	4.71	AcCN, Gel
CUFF.17913.2	integument esterase 2 precursor	0.0000	106.93	AcCN, Gel
c2567	juvenile hormone binding protein	0.0110	7.30	Gel
contig05216	Kazal-type inhibitor	0.0001	10.42	AcCN, Gel
CUFF.7975.1	Kazal-type serine proteinase inhibitor	0.0037	8.79	Gel
contig06672	lacunin	0.0260	5.11	Gel
gi|291603839	lebocin-like protein 1	0.0005	9.28	AcCN
gi|291603841	lebocin-like protein 2	0.0000	47.87	AcCN
contig10931	lebocin-like protein B	0.0000	48.61	AcCN
contig04845	lebocin-like protein C	0.0000	29.87	AcCN

c2002	legumaturain	0.2600	11.49	Gel
CUFF.18786.1	legumaturain	0.2100	24.58	Gel
CUFF.18788.1	legumaturain	0.0006	18.43	Gel
contig00798	MBF2	0.0460	8.07	Gel
contig00953	methylenetetrahydrofolate dehydrogenase	0.3700	19.67	Gel
contig02555	mitochondrial aldehyde dehydrogenase	0.0170	4.51	Gel
contig02897	molting fluid carboxypeptidase A precursor	0.0011	9.41	Gel
contig09790	Multiple coagulation factor deficiency protein 2	0.0280	2.66	Gel
CUFF.22549.1	neutral lipase	0.0430	2.52	AcCN
contig06828	New attacin	0.0000	94.66	AcCN, Gel
CUFF.32773.2	New attacin II	0.0017	109.37	AcCN
CUFF.32773.5	New attacin II	0.0000	356.15	Gel
contig14322	New attacin-1	0.0004	93.02	AcCN, Gel
CUFF.3622.1	New lebocin-like protein B	0.0000	40.08	AcCN
gi|28070937	non-muscle actin	0.3700	28.04	Gel
CUFF.18570.1	ORF	0.0041	33.27	Gel
gi|260765453	peptidoglycan recognition protein 2	0.0500	18.41	Gel
contig00589	peptidoglycan recognition protein-D	0.0055	4.72	Gel
CUFF.23390.3	peptidoglycan recognition protein-like protein	0.0002	3.68	AcCN, Gel
contig01410	peptidyl-glycine alpha-amidating monooxygenase 1	0.0022	3.11	Gel
contig00080	phosphoribosylformylglycinamidine synthase-like	0.1500	5.49	Gel
CUFF.18564.1	phytosulfokine receptor kinase	0.0280	12.82	AcCN
CUFF.21811.1	neuroendocrine convertase 1-like	0.0380	6.41	Gel
CUFF.19272.1	similar to ubiquitin-activating enzyme E1	0.0470	8.26	Gel
gi|219958088	pro-lebocin	0.0026	2.84	AcCN, Gel
contig01664	prophenoloxidase-activating proteinase-2	0.0059	39.33	Gel
gi|26006435	prophenoloxidase-activating proteinase-2	0.0330	15.82	AcCN
CUFF.24908.2	prophenoloxidase-activating proteinase-3	0.0024	9.36	Gel
gi|35277829	prophenoloxidase-activating proteinase-3 precursor	0.0001	6.50	AcCN
contig07412	protease inhibitor 6	0.0000	33.03	AcCN, Gel
contig03465	protease inhibitor-like protein	0.0009	18.51	AcCN, Gel
CUFF.35951.1	psychimicin	0.0170	13.91	AcCN, Gel
contig00143	putative hemicentin 1	0.0140	5.57	Gel
CUFF.19083.1	putative nidogen	0.0100	3.02	Gel
contig00965	putative odd Oz protein	0.0150	13.17	Gel
contig01992	putative Zn carboxypeptidase family protein	0.1300	9.79	Gel
CUFF.24674.1	Pv-fam-d protein	0.0200	5.55	Gel
contig04199	salivary cysteine-rich peptide precursor	0.0330	41.11	Gel
lrc474	scolexin A	0.0002	54.40	AcCN, Gel
CUFF.28292.2	serine protease inhibitor 11 precursor	0.0035	8.68	Gel
CUFF.16810.10	serine protease inhibitor 28	0.0450	28.84	AcCN
CUFF.16810.8	serine protease inhibitor 28	0.0000	89.43	AcCN, Gel
gi|27733421	serine protease-like protein	0.3700	13.51	Gel
contig03020	serine proteinase-like protein 4	0.0027	2.78	Gel
gi|27733415	serpin 3a	0.0430	13.78	AcCN, Gel
gi|2149091	serpin-2	0.0080	7.08	Gel
gi|45594232	serpin-5A	0.0000	37.67	Gel
CUFF.19298.1	triacylglycerol lipase	0.0006	24.38	Gel
contig01723	Vanin-like protein 1	0.0001	6.12	Gel
CUFF.30901.2	Vanin-like protein 1	0.0004	6.05	Gel
CUFF.29762.1	Xaa-Pro dipeptidase	0.4500	6.28	Gel
CUFF.9332.4	yellow-d	0.0054	3.16	Gel

*T-test was conducted using normalized spectral counts of proteins from induced samples and control samples. Those with significant changes after immune challenge were marked red (p-value < 0.05). IH/CH was calculated using the average normalized spectral counts of induced and control samples. If the number for control sample was zero, we used one for the calculation.

Table 14. A list of 155 down-regulated proteins after immune challenge*

ID	Function	P-Value	IH/CH	Source
contig01651	26S proteasome non-ATPase regulatory subunit 1	0.2000	0.00	Gel
CUFF.14802.2	3-dehydroecdysone 3beta-reductase	0.2100	0.04	Gel
gi|1350990	40S ribosomal protein S3	0.0260	0.25	Gel
gi|1351005	40S ribosomal protein S7	0.2800	0.10	Gel
gi|28207648	ADP/ATP translocase	0.0083	0.12	Gel
CUFF.20827.1	alcohol dehydrogenase	0.3300	0.13	Gel
contig03267	aldo-keto reductase	0.0080	0.18	AcCN
CUFF.22393.1	aldo-keto reductase	0.0240	0.00	AcCN
CUFF.12719.1	aldo-keto reductase 2E	0.0009	0.37	AcCN
contig03322	aldose-1-epimerase	0.0005	0.06	AcCN
CUFF.696.1	alpha-amylase 3	0.0110	0.42	AcCN
CUFF.909.1	alpha-esterase 45	0.2100	0.00	Gel
CUFF.23067.1	apolipophorin-III	0.0069	0.48	AcCN
gi|159491	arylphorin beta subunit precursor	0.1600	0.13	AcCN
CUFF.30159.1	Arylsulfatase B	0.0056	0.06	AcCN
CUFF.4487.2	basement membrane-specific heparan sulfate proteoglycan core pr.	0.0370	0.14	Gel
contig05808	BCP inhibitor precursor	0.0087	0.31	Gel
contig05808	BCP inhibitor precursor	0.0400	0.50	AcCN
contig04133	β-1,3-glucanase	0.0100	0.32	AcCN
contig00514	β-N-acetylglucosaminidase 1	0.0045	0.36	AcCN
CUFF.18887.2	Bombyrin	0.4500	0.15	AcCN
contig06322	Bombyrin precursor	0.0027	0.48	AcCN
CUFF.13061.1	calcium-dependent protein 2	0.0240	0.00	Gel
contig02851	Calmodulin	0.1900	0.00	AcCN
gi|62738877	Insecticyanin	0.0390	0.42	Gel
contig02044	chitin deacetylase 1	0.0370	0.18	Gel
F5KO9GJ01A5X4X	collagen, type IV	0.1600	0.00	Gel
contig11118	cuticle protein 4	0.0001	0.04	AcCN
CUFF.21137.1	cuticular protein analogous to peritrophins 3-A1	0.0097	0.33	AcCN
CUFF.657.1	cuticular protein RR-1 motif 3 precursor	0.0043	0.41	Gel
contig03131	cytosolic malate dehydrogenase	0.1100	0.08	Gel
CUFF.27792.1	deoxyribonuclease I	0.0920	0.19	AcCN
CUFF.32682.1	diapausin precursor	0.0010	0.00	AcCN
gi|159499	diazepam binding inhibitor-like peptide	0.0009	0.34	AcCN
contig03386	DNA supercoiling factor	0.0160	0.16	Gel
contig05586	elongation factor 1-beta'	0.0004	0.18	Gel
CUFF.11497.2	endoprotease FURIN	0.0260	0.21	AcCN
contig01115	eukaryotic initiation factor 5A	0.0750	0.00	AcCN
contig02840	eukaryotic translation initiation factor 4A	0.0690	0.17	Gel
contig00923	FK506-binding protein FKBP59 homologue	0.0140	0.13	Gel
CUFF.15602.1	flavin-dependent monooxygenase FMO2 precursor	0.0050	0.00	AcCN
gi|113608	fructose-bisphosphate aldolase A	0.3700	0.00	Gel
contig09929	gamma-glutamyl cyclotransferase-like venom protein isoform 1	0.0015	0.46	Gel
contig03063	glutaminyl-peptide cyclotransferase-like	0.0050	0.28	AcCN
contig08432	glutaredoxin	0.2600	0.09	AcCN
gi|121746	glutathione S-transferase P	0.3700	0.00	Gel
CUFF.9633.1	glutathione S-transferase sigma 1	0.0240	0.39	Gel
contig03213	glycerophosphoryl diester phosphodiesterase	0.0130	0.37	AcCN
CUFF.15250.1	glycosyl hydrolase family 31 protein	0.0550	0.00	Gel
contig12629	Hdd1-like protein	0.0430	0.39	AcCN
lrc245	heat shock protein 25.4	0.1800	0.00	Gel
contig00548	hemolymph proteinase 6	0.0340	0.00	AcCN
CUFF.31628.6	hemolymph proteinase 8	0.0007	0.49	AcCN
CUFF.6831.1	hemolymph trypsin inhibitor A;	0.0000	0.48	Gel
gi|123725	hemolymph trypsin inhibitor B	0.0170	0.44	Gel
gi|505621	high affinity nuclear JH binding protein	0.1900	0.00	Gel
contig06072	histone H2A-like protein 2	0.0180	0.33	Gel
contig02045	hydroxypyruvate isomerase	0.0034	0.36	AcCN
contig15813	hypothetical protein KGM_01763	0.0130	0.08	AcCN
contig13712	hypothetical protein KGM_06638	0.0033	0.50	AcCN
contig03958	hypothetical protein KGM_08730	0.3100	0.00	Gel

contig04595	hypothetical protein KGM_10974	0.0560	0.10	AcCN
contig05943	hypothetical protein KGM_16230	0.0370	0.28	AcCN
CUFF.22336.1	hypothetical protein KGM_21980	0.1300	0.00	Gel
gi|9716	insecticyanin a form	0.0440	0.43	Gel
CUFF.17913.1	integument esterase 2 precursor	0.0190	0.23	Gel
gi|73921301	juvenile hormone esterase precursor	0.0027	0.11	Gel
contig00254	juvenile hormone esterase precursor	0.0260	0.00	Gel
CUFF.27967.12	kynurenine formamidase	0.0610	0.12	Gel
gi|48428995	lysozyme C	0.3700	0.00	Gel
contig06389	mating type protein MAT1-1-1	0.0260	0.00	Gel
CUFF.23390.1	microvitellogenin	0.0500	0.16	AcCN
lrc75	moderately methionine rich storage protein	0.1200	0.00	Gel
CUFF.13545.2	multiplexin, isoform M	0.0820	0.08	Gel
contig00808	N-acetylglucosaminidase	0.0770	0.10	Gel
contig05820	nascent polypeptide associated complex α subunit	0.0340	0.32	Gel
gi|1708635	neuroglian	0.2300	0.00	AcCN
CUFF.9504.1	new antennal binding protein 7	0.0026	0.16	AcCN
CUFF.17954.6	new sensory appendage protein 1	0.0038	0.50	AcCN
CUFF.17957.2	new sensory appendage protein 1	0.0097	0.10	AcCN
CUFF.5287.1	nonclathrin coat protein gamma2-COP	0.3700	0.00	Gel
contig13241	odorant binding protein	0.3700	0.00	AcCN, Gel
lrc1777	ommochrome-binding protein	0.0011	0.38	AcCN
CUFF.8839.1	p270	0.2300	0.00	Gel
CUFF.18772.10	paralytic peptide binding protein 1	0.3700	0.00	Gel
CUFF.20588.18	peroxisomal N1-acetyl-spermine oxidase	0.0008	0.19	AcCN
CUFF.20588.11	peroxisomal N1-acetyl-spermine oxidase	0.0014	0.20	AcCN
contig14653	phosphoglucomutase	0.0460	0.49	Gel
CUFF.38214.1	phosphoglucomutase	0.0870	0.16	Gel
CUFF.32110.3	plasmatocyte-spreading peptide precursor	0.0330	0.45	Gel
CUFF.12385.1	plasminogen activator inhibitor 1 RNA-binding pr.	0.3700	0.00	Gel
CUFF.31619.1	apolipophorins-like, partial	0.0730	0.07	Gel
CUFF.17519.2	collagen alpha-1(IV) chain-like	0.0016	0.37	Gel
contig00668	lamin Dm0-like isoform 1	0.3700	0.00	Gel
contig07975	proteasome subunit β type-2-like	0.3700	0.00	Gel
CUFF.21412.1	similar to GA14337-PA	0.1500	0.00	AcCN
CUFF.32240.1	protein disulfide isomerase	0.0026	0.09	Gel
gi|58864722	putative annexin IX-B	0.0360	0.27	Gel
gi|254746340	putative C1A cysteine protease precursor	0.0065	0.39	AcCN
gi|254746338	putative C1A cysteine protease precursor	0.0450	0.41	AcCN
CUFF.17517.1	putative collagen alpha-2IV chain protein	0.0220	0.42	Gel
CUFF.21813.1	putative collagen and Ca^{2+}-binding EGF domains 1	0.0003	0.49	AcCN
CUFF.21813.1	putative collagen and Ca^{2+}-binding EGF domains 1	0.0003	0.26	Gel
contig02362	putative enolase protein	0.4500	0.15	AcCN
CUFF.10253.3	inosine-uridine preferring nucleoside hydrolase	0.0005	0.00	AcCN
CUFF.10254.1	inosine-uridine preferring nucleoside hydrolase	0.1600	0.00	AcCN
CUFF.19054.1	putative mannosidase, βA, lysosomal	0.0048	0.04	AcCN
CUFF.7410.1	putative protease inhibitor 4	0.0200	0.43	AcCN
CUFF.16862.1	putative rab GDP-dissociation inhibitor	0.0350	0.18	Gel
contig04965	putative ribophorin II	0.3700	0.00	Gel
contig01389	ras-related GTP-binding protein Rab11	0.0004	0.31	Gel
CUFF.31934.2	regulator of chromosome condensation	0.3700	0.00	Gel
CUFF.24819.1	reticulon/nogo receptor	0.0300	0.08	Gel
gi|268306444	ribosomal protein L10	0.3300	0.15	Gel
gi|268306376	ribosomal protein L11	0.1300	0.03	Gel
gi|268306382	ribosomal protein L12	0.2500	0.11	Gel
gi|268306480	ribosomal protein L13	0.0680	0.05	Gel
gi|268306366	ribosomal protein L13A	0.0170	0.18	Gel
CUFF.2904.1	ribosomal protein L14	0.0040	0.48	Gel
gi|268306486	ribosomal protein L15	0.0200	0.00	Gel
gi|268306468	ribosomal protein L18	0.1400	0.05	Gel
contig08946	ribosomal protein L18A	0.2700	0.00	Gel
gi|268306370	ribosomal protein L26	0.1100	0.10	Gel
gi|268306352	ribosomal protein L3	0.1800	0.19	Gel
gi|268306418	ribosomal protein L31	0.0400	0.36	Gel
gi|268306428	ribosomal protein L35A	0.0046	0.11	Gel
gi|268306434	ribosomal protein L5	0.0001	0.04	Gel

gi|268306462	ribosomal protein L7	0.0370	0.33	Gel
gi|268306466	ribosomal protein L7A	0.0046	0.14	Gel
gi|268306420	ribosomal protein L8	0.1900	0.00	Gel
gi|268306396	ribosomal protein L9	0.0660	0.14	Gel
contig11395	ribosomal protein S12	0.0560	0.06	Gel
gi|268306426	ribosomal protein S13	0.0610	0.00	Gel
contig13755	ribosomal protein S15A	0.0042	0.14	Gel
gi|268306384	ribosomal protein S18	0.0180	0.27	Gel
gi|268306498	ribosomal protein S2	0.2100	0.13	Gel
gi|268306374	ribosomal protein S20	0.0340	0.32	Gel
gi|268306364	ribosomal protein S27	0.0032	0.00	Gel
contig06573	ribosomal protein S3A	0.0017	0.09	Gel
gi|268306408	ribosomal protein S4	0.0450	0.48	Gel
CUFF.1918.1	ribosomal protein S6	0.0012	0.00	Gel
gi|268306412	ribosomal protein S8	0.2700	0.00	Gel
gi|268306404	ribosomal protein S9	0.0056	0.23	Gel
contig01488	S-adenosyl-L-homocysteine hydrolase	0.0300	0.32	Gel
CUFF.10507.1	secreted protein acidic and rich in cysteine	0.0017	0.30	Gel
CUFF.2529.2	seminal fluid protein CSSFP028	0.2700	0.00	Gel
gi|1378132	serpin 1	0.0019	0.49	AcCN
CUFF.22827.1	serpin 7	0.0710	0.05	Gel
contig02334	small GTP-binding protein	0.0440	0.30	Gel
CUFF.10507.1	sparc	0.0006	0.10	AcCN
contig05971	superoxide dismutase	0.0480	0.15	AcCN
gi|6560635	thioredoxin-like protein	0.0180	0.13	AcCN
CUFF.7756.1	Tolloid-like protein 2	0.0043	0.36	AcCN
gi|136429	trypsin	0.0069	0.40	AcCN
CUFF.10235.1	twelve cysteine protein 1	0.0180	0.00	AcCN
CUFF.19786.1	unknown	0.0087	0.43	AcCN
CUFF.19786.1	unknown	0.0099	0.40	Gel
CUFF.5575.1	unknown	0.0200	0.24	AcCN
contig01080	venom acid phosphatase	0.0024	0.20	AcCN
CUFF.10876.1	venom acid phosphatase	0.0040	0.50	Gel
contig02578	vitellogenic carboxypeptidase	0.0440	0.07	Gel

*See Table 13

Table 15. A list of 211 immunity-related proteins*

ID	Function	P-Value	IH/CH	Source	Type
contig03013	β-1,3-glucan recognition protein	0.0100	0.83	Gel	1
gi|52782757	β -1,3-glucan recognition protein 1	0.8400	0.90	AcCN, Gel	1
contig01306	β-1,3-glucan recognition protein 2	0.0660	1.82	Gel	1
contig14217	β-1,3-glucan recognition protein 3	0.3200	0.76	Gel	1
contig13657	β-1,3-glucan recognition protein 3	0.5600	0.34	Gel	1
CUFF.37621.1	β-1,3-glucan recognition protein 3	0.6600	1.12	Gel	1
CUFF.32840.1	β-1,3-glucan recognition protein 3	0.9500	1.01	Gel	1
contig04133	β-1,3-glucanase	0.0100	0.32	AcCN	1
contig04133	β-1,3-glucanase	0.2700	0.72	Gel	1
contig01513	β-galactosidase	0.9600	0.99	AcCN, Gel	1
CUFF.3641.4	C-type lectin 10 precursor	0.0530	0.24	Gel	1
CUFF.3641.1	C-type lectin 10 precursor	0.2800	0.50	Gel	1
CUFF.3645.2	C-type lectin 10 precursor	0.3100	0.25	AcCN, Gel	1
CUFF.1879.2	draper	0.3800	1.43	Gel	1
CUFF.29563.2	galectin-4	0.4500	0.71	Gel	1
GNBP	GNBP	0.6100	1.15	AcCN, Gel	1
CUFF.15846.1	GNBP-like protein	0.2600	1.16	Gel	1
contig15094	hemicentin-like protein	0.2500	1.39	Gel	1
CUFF.4030.1	hemicentin-like protein 1	0.5700	1.17	Gel	1
contig04393	hemicentin-like protein 2	0.2600	1.29	Gel	1
gi|259493819	hemocyte aggregation inhibitor protein precursor	0.0250	0.75	AcCN	1
CUFF.8487.3	hemocyte aggregation inhibitor protein precursor	0.0300	0.86	Gel	1
CUFF.27533.1	hemocytin	0.1100	0.62	Gel	1
CUFF.27538.1	hemocytin	0.3300	0.82	Gel	1
gi|511297	hemolin	0.0000	5.44	AcCN	1
CUFF.23729.1	hemolin	0.0001	12.54	Gel	1
contig01508	Immulectin-1	0.0170	4.95	Gel	1
gi|237869126	immulectin-2	0.6800	0.76	AcCN, Gel	1
CUFF.5601.4	immulectin-3	0.1000	0.74	AcCN, Gel	1
contig04357	immulectin-4	0.0890	7.98	Gel	1
CUFF.32498.1	lectin	0.9100	1.07	Gel	1
CUFF.11764.1	leureptin	0.0710	0.92	Gel	1
CUFF.11766.1	leureptin	0.0940	0.66	Gel	1
gi|27733411	leureptin	0.2400	0.78	Gel	1
CUFF.3007.14	New immulectin	0.0390	1.70	Gel	1
CUFF.3007.22	New immulectin	0.7700	1.07	Gel	1
CUFF.5601.1	New immulectin-3	0.0014	0.64	Gel	1
CUFF.5595.1	New immulectin-3	0.5000	4.24	AcCN, Gel	1
CUFF.18040.1	nimrod B precursor	0.0280	0.69	AcCN, Gel	1
c2114	nimrod B precursor	0.2000	0.79	Gel	1
gi|27733423	peptidoglycan recognition protein 1A	0.9800	1.00	Gel	1
gi|27733409	peptidoglycan recognition protein 1B	0.4500	0.93	AcCN	1
gi|260765453	peptidoglycan recognition protein 2	0.0500	18.41	Gel	1
contig00589	peptidoglycan recognition protein-D	0.0055	4.72	Gel	1
CUFF.23390.3	peptidoglycan recognition protein-like protein	0.0002	3.68	AcCN, Gel	1
CUFF.28117.2	hemocytin-like	0.0700	0.72	Gel	1
contig00143	putative hemicentin 1	0.0140	5.57	Gel	1
contig00476	putative hemicentin 1	0.0150	0.64	AcCN, Gel	1
CUFF.237.1	putative hemicentin 1	0.1600	1.32	Gel	1
CUFF.239.1	putative hemicentin 1	0.3500	2.06	AcCN, Gel	1
contig08927	putative hemicentin 1	0.4700	1.24	Gel	1
contig00583	putative hemicentin 1	0.5200	1.27	Gel	1
CUFF.19863.5	putative hemicentin-1	0.0740	0.44	Gel	1
contig15247	putative hemicentin-1	0.8200	1.11	Gel	1
contig05808	BCP inhibitor precursor	0.0400	0.50	AcCN, Gel	2
CUFF.26093.1	cysteine proteinase inhibitor precursor	0.0170	0.76	AcCN, Gel	2
CUFF.26093.2	cysteine proteinase inhibitor precursor	0.3100	1.23	Gel	2
CUFF.26093.3	cysteine proteinase inhibitor precursor	0.0720	0.77	AcCN, Gel	2
gi|260234113	cysteine proteinase inhibitor precursor	0.3100	0.86	AcCN, Gel	2
contig00585	cysteine-rich/pacifastin venom protein 2	0.7200	1.20	AcCN, Gel	2
CUFF.18357.1	dorsal	0.6300	1.37	AcCN, Gel	2
lrc23	dorsal	0.0230	1.77	AcCN, Gel	2
contig12629	Hdd1-like protein	0.0430	0.39	AcCN, Gel	2

CUFF.10416.1	hemocyte protease-1	0.0074	0.51	AcCN, Gel	2
contig06406	hemocyte protease-2	0.8500	1.09	Gel	2
gi|56418409	hemolymph proteinase 15	0.2700	1.82	Gel	2
CUFF.29694.1	hemolymph proteinase 16	0.2000	1.66	Gel	2
CUFF.14527.1	hemolymph proteinase 17	0.0720	12.93	Gel	2
gi|56418419	hemolymph proteinase 19	0.1100	1.56	Gel	2
contig04271	hemolymph proteinase 20	0.0069	4.16	AcCN	2
CUFF.25442.1	hemolymph proteinase 20	0.0004	2.75	Gel	2
contig02445	hemolymph proteinase 21	0.0580	0.48	AcCN, Gel	2
contig02033	hemolymph proteinase 22, partial	0.0490	15.83	Gel	2
CUFF.31629.1	hemolymph proteinase 5	0.0370	5.76	Gel	2
contig00548	hemolymph proteinase 6	0.0340	0.00	AcCN, Gel	2
CUFF.31628.4	hemolymph proteinase 8	0.7000	0.93	Gel	2
CUFF.31628.6	hemolymph proteinase 8	0.0007	0.49	AcCN	2
gi|56418399	hemolymph proteinase 9	0.2300	1.77	Gel	2
CUFF.6831.1	hemolymph trypsin inhibitor A	0.0017	0.53	AcCN, Gel	2
gi|123725	hemolymph trypsin inhibitor B	0.0170	0.44	Gel	2
contig05216	Kazal-type inhibitor	0.0001	10.42	AcCN, Gel	2
gi|6560641	Kazal-type proteinase inhibitor	0.0067	0.66	AcCN, Gel	2
CUFF.7975.1	Kazal-type serine proteinase inhibitor	0.0037	8.79	Gel	2
contig11836	Kunitz-type protease inhibitor precursor	0.0250	0.58	AcCN	2
lrc464	Kunitz-type protease inhibitor precursor	0.0026	0.68	Gel	2
contig03315	New hemolymph proteinase	0.1800	1.32	Gel	2
CUFF.25432.2	New hemolymph proteinase 20	0.2600	2.55	Gel	2
CUFF.29541.2	pattern recognition serine proteinase precursor	0.0600	1.62	Gel	2
gi|39655053	pattern recognition serine proteinase precursor	0.4700	1.17	Gel	2
CUFF.23027.1	prophenoloxidase subunit 1	0.0930	3.40	AcCN, Gel	2
CUFF.26961.2	prophenoloxidase	0.0720	0.82	Gel	2
contig03105	prophenoloxidase-activating proteinase-1	0.0320	0.84	AcCN, Gel	2
contig01664	prophenoloxidase-activating proteinase-2	0.0059	39.33	Gel	2
gi|26006435	prophenoloxidase-activating proteinase-2	0.0330	15.82	AcCN	2
CUFF.24908.2	prophenoloxidase-activating proteinase-3	0.0024	9.36	Gel	2
gi|35277829	prophenoloxidase-activating proteinase-3	0.0001	6.50	AcCN	2
CUFF.7977.1	protease inhibitor 1	0.5700	1.10	AcCN	2
contig07412	protease inhibitor 6	0.0000	33.03	AcCN, Gel	2
contig03465	protease inhibitor-like protein	0.0009	18.51	AcCN, Gel	2
contig02101	putative C1A Cys protease precursor	0.6200	0.74	Gel	2
CUFF.14914.4	putative C1A Cys protease precursor	0.8900	1.15	Gel	2
gi|254746338	putative C1A Cys protease precursor	0.0450	0.41	AcCN, Gel	2
gi|254746340	putative C1A Cys protease precursor	0.0065	0.39	AcCN, Gel	2
contig00474	putative f-spondin	0.0022	0.62	AcCN, Gel	2
CUFF.7410.1	putative protease inhibitor 4	0.0200	0.43	AcCN, Gel	2
contig03750	putative serine protease-like protein 2	0.0054	0.58	AcCN	2
CUFF.23176.4	putative serine protease-like protein 2	0.0100	0.66	Gel	2
gi|136429	trypsin precursor	0.0069	0.40	AcCN, Gel	2
CUFF.21587.1	REPAT31	0.0052	0.65	AcCN, Gel	2
contig00315	REPAT32	0.2300	0.75	AcCN, Gel	2
lrc474	scolexin A	0.0002	54.40	AcCN, Gel	2
contig04027	scolexin B	0.0680	0.74	AcCN, Gel	2
lrc45	serine protease 17	0.2900	0.64	Gel	2
lrc477	serine protease 17	0.4400	1.89	Gel	2
contig06430	serpin 11 precursor	0.0830	4.15	Gel	2
CUFF.28292.2	serpin 11 precursor	0.0035	8.68	Gel	2
CUFF.16311.1	serpin 13 precursor	0.1200	0.79	Gel	2
CUFF.25074.1	serpin 23 precursor	0.1300	0.66	Gel	2
CUFF.16810.1	serpin 28	0.0230	0.50	AcCN, Gel	2
CUFF.16810.10	serpin 28	0.0450	28.84	AcCN	2
CUFF.16810.8	seripn 28	0.0000	89.43	AcCN, Gel	2
CUFF.25446.2	serine protease-like protein	0.2900	2.22	Gel	2
CUFF.25446.3	serine protease-like protein	0.1600	1.34	Gel	2
CUFF.25446.4	serine protease-like protein	0.0390	1.92	Gel	2
gi|27733421	serine protease-like protein	0.3700	2.23	AcCN, Gel	2
contig02838	serine proteinase-like protein 1b	0.0004	0.54	AcCN, Gel	2
CUFF.33286.1	serine proteinase-like protein 1b	0.5700	0.90	Gel	2
gi|21630233	serine proteinase-like protein 2	0.3400	0.76	AcCN, Gel	2
contig03020	serine proteinase-like protein 4	0.0027	2.78	Gel	2

CUFF.30294.14	serpin 1	0.0120	0.73	Gel	2
CUFF.30294.20	serpin 1	0.0720	0.85	Gel	2
CUFF.30294.6	serpin 1	0.0039	0.82	Gel	2
gi|1378127	serpin 1	0.0044	0.74	Gel	2
gi|1378132	serpin 1	0.0019	0.49	AcCN	2
gi|27733415	serpin 3a	0.0430	13.78	AcCN, Gel	2
contig05880	serpin 2	0.3700	1.82	Gel	2
contig15945	serpin 2	0.3700	2.06	Gel	2
CUFF.16959.12	serpin 2	0.8100	1.31	Gel	2
gi|2149091	serpin 2	0.0080	7.08	Gel	2
gi|45594224	serpin 4A	0.2700	1.74	Gel	2
gi|45594226	serpin 4B	0.0180	1.35	Gel	2
gi|45594232	serpin 5A	0.0000	37.67	Gel	2
CUFF.16998.1	serpin 6	0.0570	1.29	Gel	2
CUFF.22827.1	serpin 7	0.0710	0.05	Gel	2
serpin8	serpin 8	0.0380	1.89	Gel	2
contig13837	silk protease inhibitor 1 precursor	0.0400	0.72	AcCN, Gel	2
contig14017	silk protease inhibitor 1 precursor	0.2500	0.68	AcCN	2
CUFF.6839.1	silk protease inhibitor 1 precursor	1.0000	1.00	AcCN, Gel	2
CUFF.10507.1	sparc	0.0006	0.10	AcCN, Gel	2
SPH6	SPH6	0.5600	3.05	Gel	2
contig05893	Spz1B	0.1400	0.63	AcCN, Gel	2
contig07592	Spz1B	0.1700	0.37	AcCN, Gel	2
contig15228	antileukoproteinase precursor	0.1200	4.07	AcCN	3
gi|29469965	antimicrobial peptide cecropin 6	0.0008	17.31	AcCN, Gel	3
CUFF.21534.1	antimicrobial peptide MGD2b precursor	0.3800	0.61	AcCN, Gel	3
CUFF.26341.2	antimicrobial protein 6Tox	0.2000	4.70	AcCN	3
gi|29469969	antimicrobial protein attacin 2	0.0000	106.97	AcCN, Gel	3
CUFF.19826.5	attacin-1	0.0000	176.15	AcCN	3
CUFF.19828.1	attacin-1	0.0000	162.75	AcCN, Gel	3
CUFF.19826.3	attacin-1	0.0000	13.05	Gel	3
c25131	attacin-1	0.0000	40.86	AcCN, Gel	3
gi|67906420	attacin-1	0.0000	143.86	AcCN, Gel	3
CUFF.19826.4	attacin-1	0.0000	508.58	AcCN, Gel	3
c2978	attacin-1	0.0001	130.79	AcCN, Gel	3
CUFF.19826.2	attacin-1	0.0003	72.41	AcCN	3
CUFF.19828.3	attacin-1	0.1200	98.55	AcCN	3
lrc6486	attacin-1	0.3700	29.21	AcCN, Gel	3
contig12105	bactericidin	0.0005	25.83	AcCN	3
contig13987	cecropin 3	0.0029	2.18	AcCN	3
CUFF.32682.1	diapausin precursor	0.0010	0.00	AcCN	3
contig15549	diapausin precursor	0.0220	1.75	AcCN, Gel	3
contig10194	gallerimycin	0.0000	21.00	AcCN, Gel	3
gi|110649240	gloverin	0.0002	3.60	AcCN	3
contig02032	gloverin	0.0014	3.53	Gel	3
lrc512	hypothetical protein CHLNCDRAFT_55808	0.0250	9.28	AcCN, Gel	3
contig08449	hypothetical protein KGM_06199	0.0092	18.32	AcCN, Gel	3
gi|291603839	lebocin-like protein 1	0.0005	9.28	AcCN	3
gi|291603841	lebocin-like protein 2	0.0000	47.87	AcCN	3
contig10931	lebocin-like protein B	0.0000	48.61	AcCN	3
contig04845	lebocin-like protein C	0.0000	29.87	AcCN	3
CUFF.12293.1	lysozyme	0.0001	1.81	AcCN, Gel	3
gi|48428995	lysozyme C	0.3700	0.00	Gel	3
gi|260765455	lysozyme-like protein 1	0.2100	0.60	AcCN, Gel	3
contig06828	new attacin	0.0000	94.66	AcCN, Gel	3
CUFF.32773.5	new attacin II	0.0000	356.15	Gel	3
CUFF.32773.2	new attacin II	0.0017	109.37	AcCN	3
contig14322	new attacin-1	0.0004	93.02	AcCN, Gel	3
CUFF.3622.1	new lebocin-like protein B	0.0000	40.08	AcCN	3
gi|219958088	pro-lebocin	0.0026	2.84	AcCN, Gel	3
CUFF.35951.1	psychimicin	0.0170	13.91	AcCN, Gel	3
CUFF.13075.1	WAP four-disulfide core domain 2 isoform 1	0.9700	1.02	AcCN	3
contig04199	salivary cysteine-rich peptide precursor	0.6000	0.87	AcCN, Gel	3
gi|136206	transferrin	0.9100	1.02	AcCN	3
CUFF.7001.1	transferrin	0.0069	1.35	Gel	3
c4856	transferrin	0.6800	0.87	Gel	3

c6909	transferrin	0.6900	0.94	Gel	3
contig12389	transferrin	0.8800	1.26	AcCN, Gel	3
contig06672	lacunin	0.0260	5.11	Gel	4
gi|6164595	lacunin	0.2000	1.39	Gel	4
CUFF.17122.1	laminin β-2 chain	0.7500	0.93	Gel	4
gi|1708635	neuroglian	0.2300	0.00	AcCN, Gel	4
CUFF.18772.10	paralytic peptide binding protein 1	0.3700	0.00	Gel	4
CUFF.18772.6	paralytic peptide binding protein 2	0.3800	0.41	Gel	4
CUFF.32110.3	plasmatocyte-spreading peptide precursor	0.0013	0.57	AcCN, Gel	4
contig00030	putative laminin A chain	0.2500	1.11	Gel	4
contig00004	putative laminin A chain	0.7600	1.12	Gel	4
contig01397	45 kDa immunophilin FKBP45	0.4300	0.50	Gel	5
gi|27733419	immune-induced protein 1	0.0100	23.09	AcCN, Gel	5
contig02384	immune-related Hdd1	0.0000	10.85	AcCN, Gel	5
contig04966	peroxiredoxin	0.4500	0.54	Gel	5
CUFF.4124.1	thioredoxin peroxidase	0.4300	0.76	AcCN	5

*T-test was conducted using normalized spectral counts of proteins from induced samples and control samples. Those with significant changes after immune challenge were marked red (p-value < 0.05). IH/CH was calculated using the average normalized spectral counts of induced and control samples. If the number for control sample was zero, we used one for the calculation. Type means different types of immunity related proteins (1: PRR; 2: signaling proteins; 3: AMPs; 4: proteins involved in hemocyte adhesion; 5: Others).

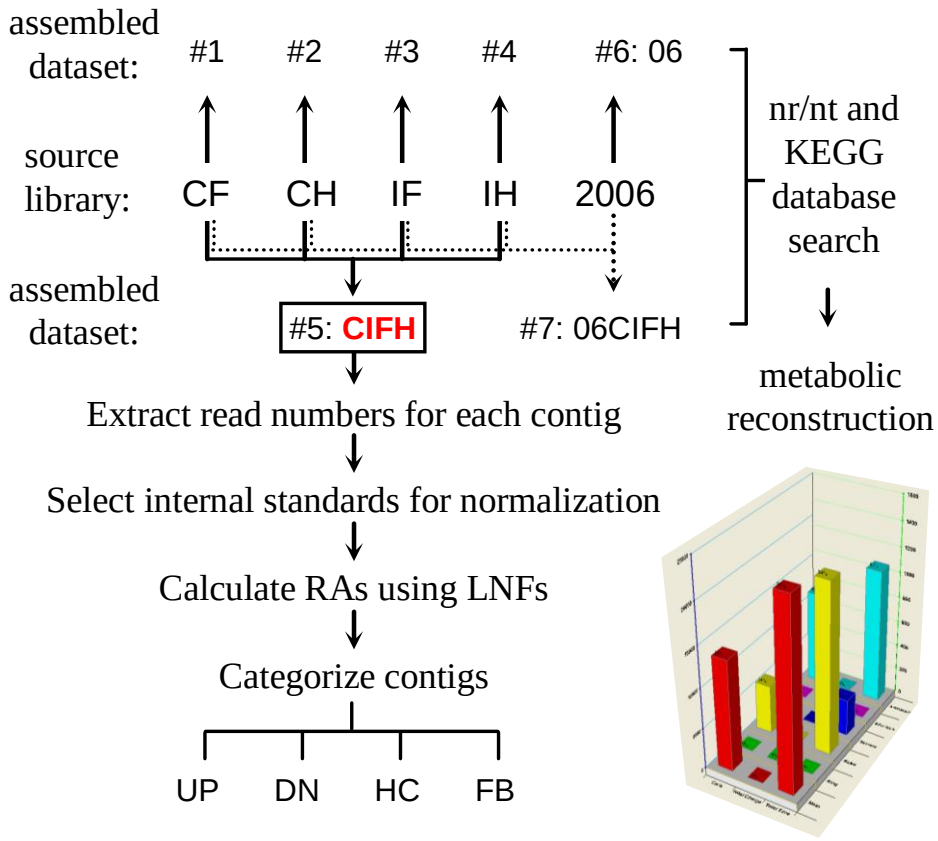


Fig. 1. Scheme of library sequencing, dataset assembling, read normalization, contig categorization, and function prediction. Five cDNA libraries (CF, CH, IF, IH, and 2006) were assembled into seven datasets, one of which (#5: CIFH) was further analyzed by extracting numbers of CF, CH, IF and IH reads assembled into each contig. As described in Section 2.3, read numbers were calibrated using library normalization factors (LNFs) for the calculation of relative abundances (RAs) or adjusted read numbers (ARNs). Based on thresholds set arbitrarily, contigs were categorized into four groups: UP and DN for up- and down-regulated; HC and FB for hemocyte- or fat body-specific.

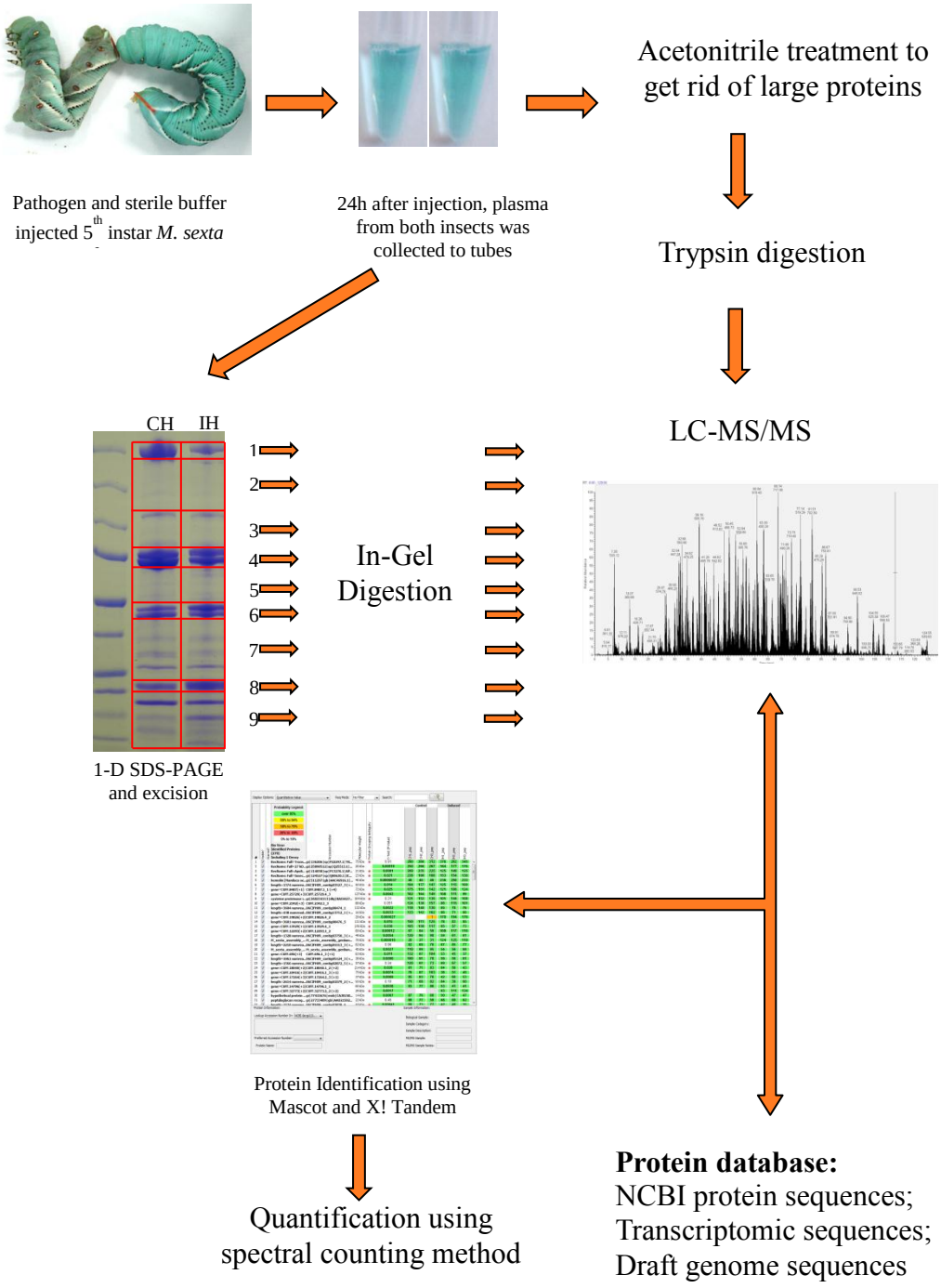


Fig. 2. Experimental workflow for the proteome analysis. Plasma collected from pathogen- and buffer-injected insects was treated using two methods – gel electrophoresis and acetonitrile treatment. Both protein samples were digested by trypsin and loaded on LTQ-Orbitrap mass spectrometer. Proteins were identified using Mascot and X! Tandem. Spectral counts for each protein were used for quantitative analysis.

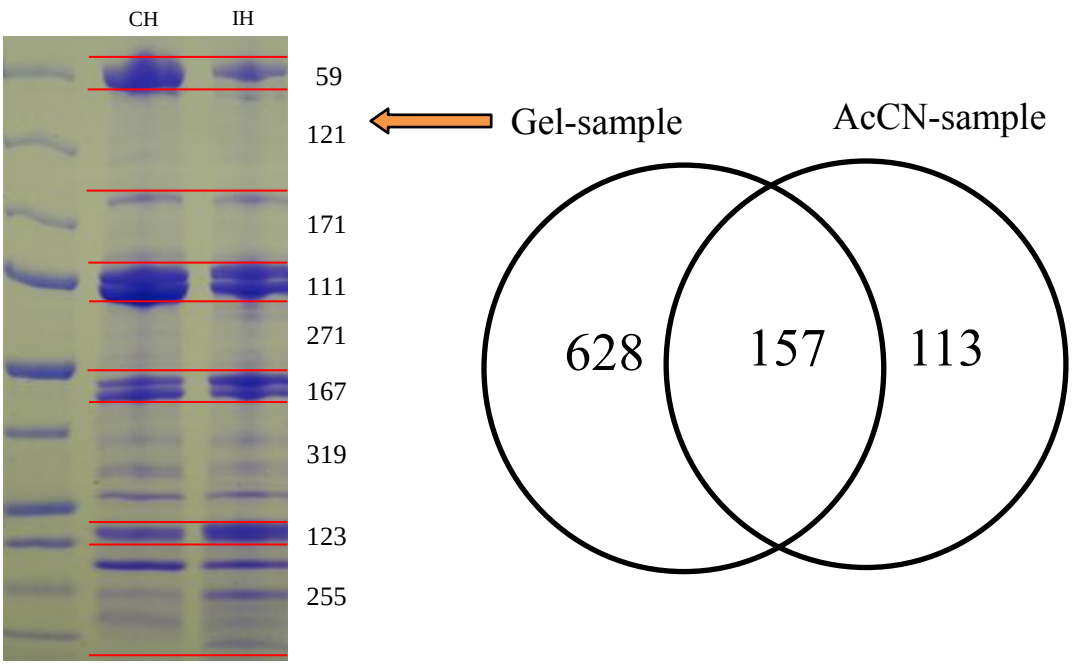
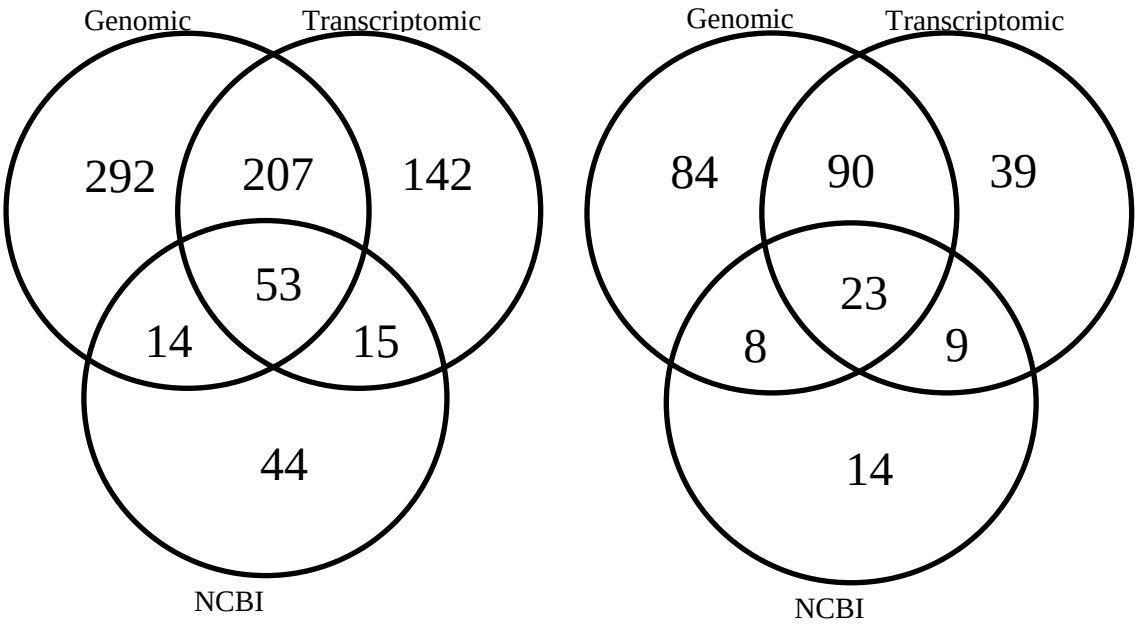


Fig. 3. Numbers of proteins identified using both methods. 785 proteins were identified from the gel-extracted samples and 270 proteins were identified from the ACN treated samples. 157 proteins were identified from both samples.



A. Gel-extracted samples **B. AcCN-treated samples**

Fig. 4. Numbers of proteins identified using different sequence sources.

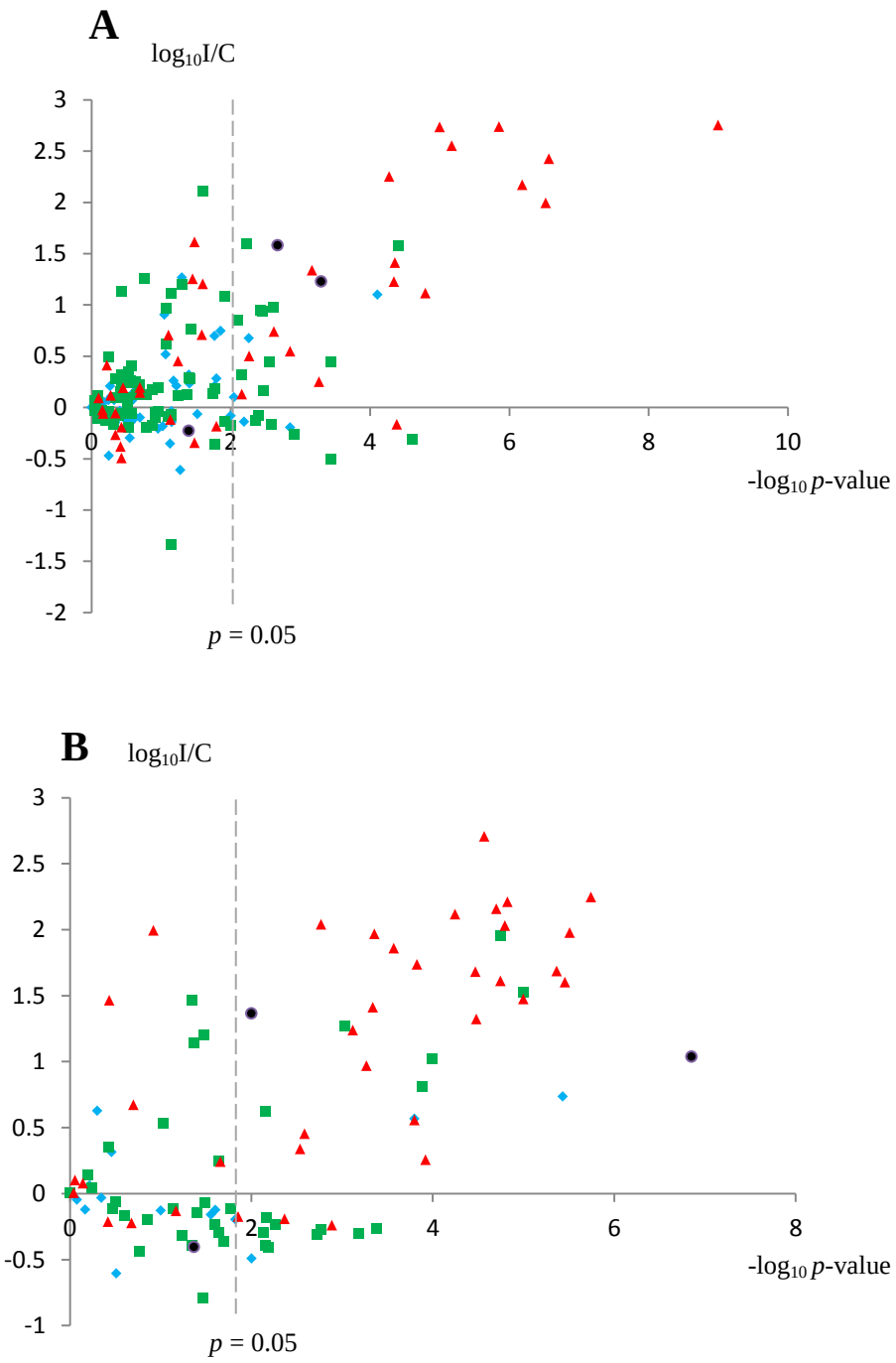


Fig. 5. Differential expression of immunity-related proteins after immune challenge. Proteins from gel slices (**A**) or AcCN-treated samples (**B**). Pattern recognition proteins (♦), signal mediators/modulators (■), effectors (▲), and others (●).

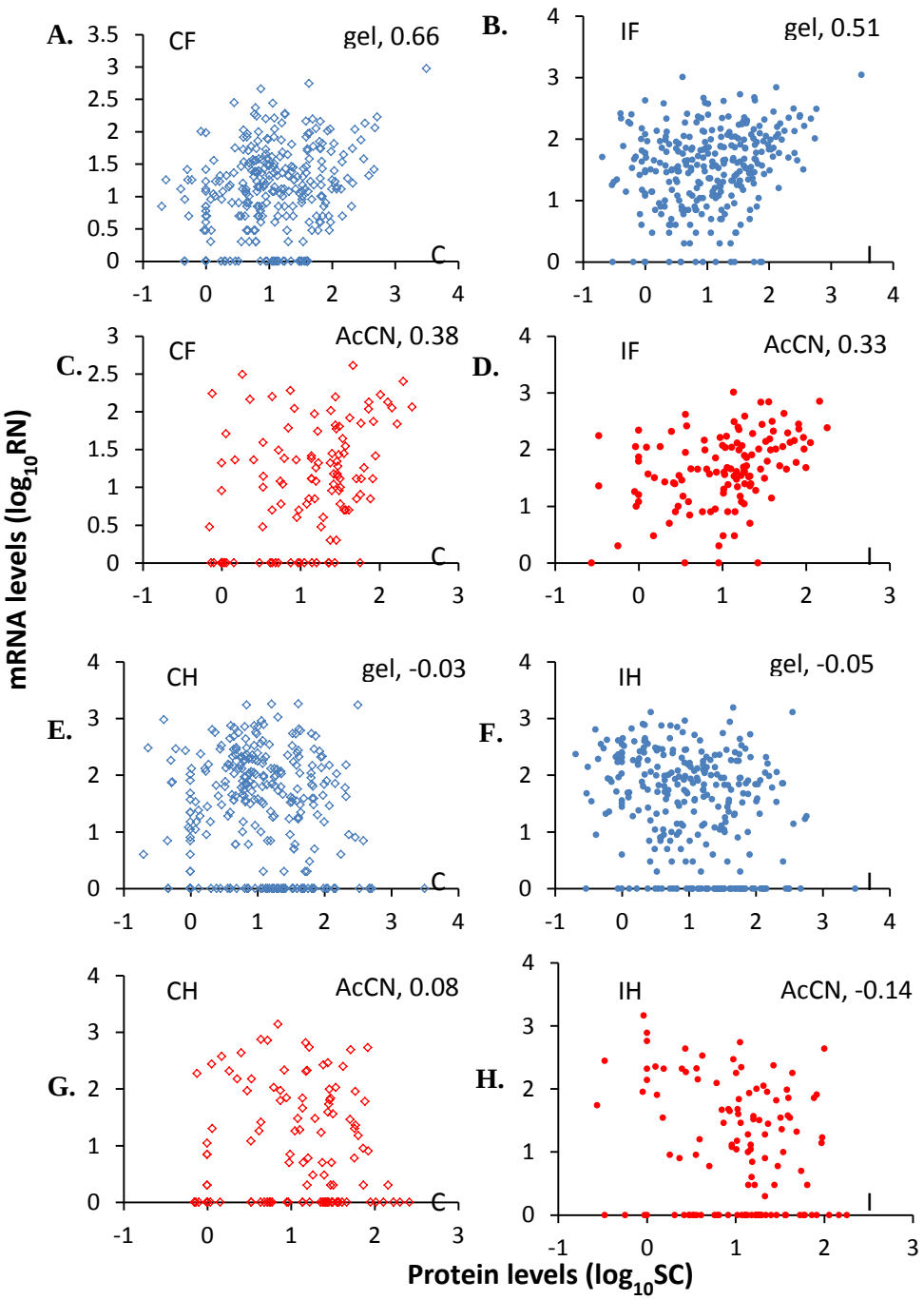


Fig. 6. Comparison of the abundances of proteins and mRNAs. Figure A-H showed the correlation of abundances of plasma proteins (SCs) and fat body /hemocyte mRNAs (RNs) from both induced and control *M. sexta*. Proteins from gel-extracted samples were marked blue and those from AcCN-treated samples were marked red.

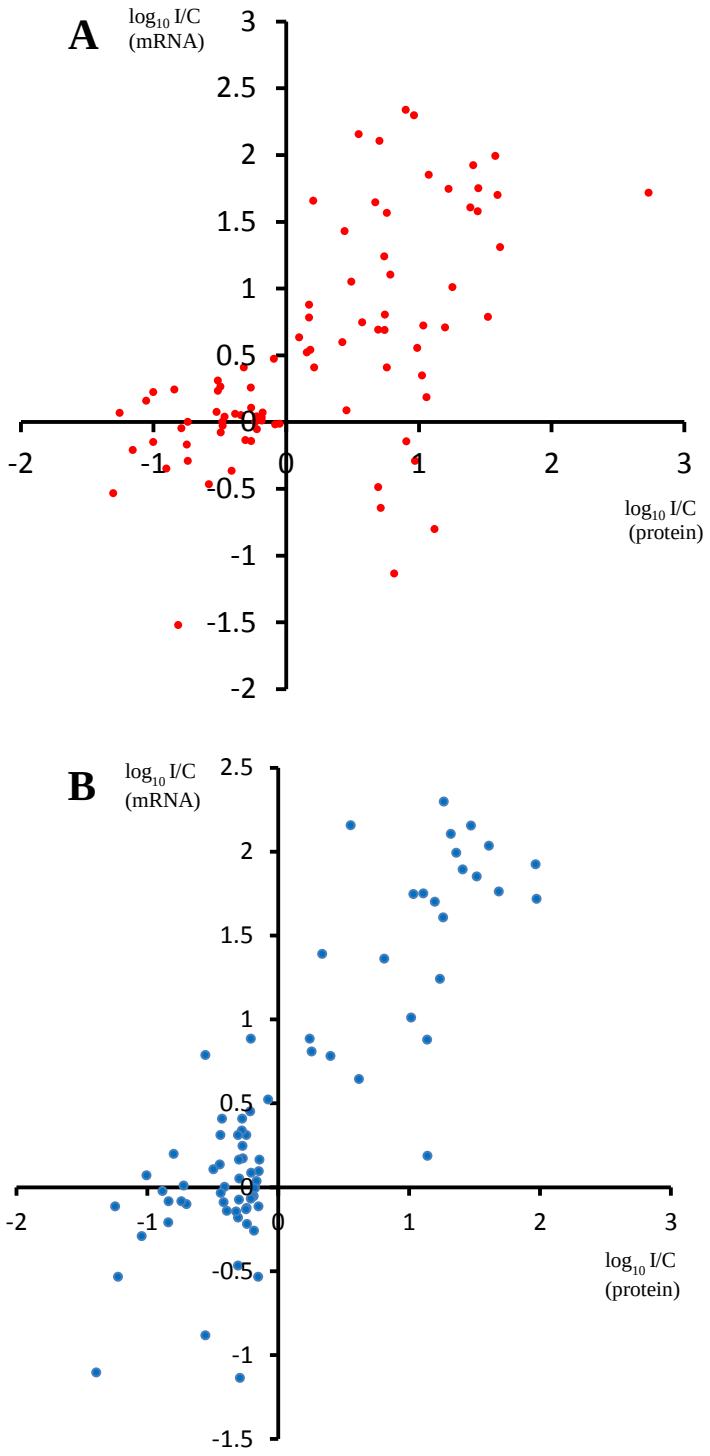


Fig. 7. Comparison of protein- and mRNA-level changes. Figure A and B showed the relative protein and fat body mRNA level changes after immune challenge for proteins with significant differences between C and I (p <0.05), I/C >5, or C/I >5.

REFERENCES

Abdel-latief M, Hilker M. 2008. Innate immunity: eggs of Manduca sexta are able to respond to parasitism by Trichogramma evanescens. *Insect Biochem Mol Biol* 38: 136-145.

Agaisse H, Petersen UM, Boutros M, Mathey-Prevot B, Perrimon N. 2003. Signaling role of hemocytes in Drosophila JAK/STAT-dependent response to septic injury. *Developmental cell* 5: 441-450.

Altincicek B, Vilcinskas A. 2008. Identification of a lepidopteran matrix metalloproteinase with dual roles in metamorphosis and innate immunity. *Dev Comp Immunol* 32: 400-409.

Alwine JC, Kemp DJ, Stark GR. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* 74: 5350-5354.

Ao JQ, Ling E, Yu XQ. 2008. A Toll receptor from Manduca sexta is in response to Escherichia coli infection. *Molecular Immunology* 45: 543-552.

Armstrong PB. 2006. Proteases and protease inhibitors: a balance of activities in host-pathogen interaction. *Immunobiology* 211: 263-281.

Arrese EL, Soulages JL. 2010. Insect Fat Body: Energy, Metabolism, and Regulation. *Annual Review of Entomology* 55: 207-225.

Axen A, Carlsson A, Engstrom A, Bennich H. 1997. Gloverin, an antibacterial protein from the immune hemolymph of Hyalophora pupae. *European journal of biochemistry / FEBS* 247: 614-619.

Baggerman G, Verleyen P, Clynen E, Huybrechts J, De Loof A, Schoofs L. 2004. Peptidomics. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* 803: 3-16.

Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. 2007. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 389: 1017-1031.

Becker-Andre M, Hahlbrock K. 1989. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic acids research* 17: 9437-9446.

Boguski MS, Tolstoshev CM, Bassett DE, Jr. 1994. Gene discovery in dbEST. *Science* 265: 1993-1994.

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M et al. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature biotechnology* 18: 630-634.

Brown SE, Howard A, Kasprzak AB, Gordon KH, East PD. 2009. A peptidomics study reveals the impressive antimicrobial peptide arsenal of the wax moth Galleria mellonella. *Insect Biochem Mol Biol* 39: 792-800.

Bulet P, Stocklin R, Menin L. 2004. Anti-microbial peptides: from invertebrates to vertebrates. *Immunol Rev* 198: 169-184.

Dai H, Rayaprolu S, Gong Y, Huang R, Prakash O, Jiang H. 2008. Solution structure, antibacterial activity, and expression profile of Manduca sexta moricin. *J Pept Sci* 14: 855-863.

Durell SR, Raghunathan G, Guy HR. 1992. Modeling the ion channel structure of cecropin. *Biophysical journal* 63: 1623-1631.

Elmore JM, Liu J, Smith B, Phinney B, Coaker G. 2012. Quantitative proteomics reveals dynamic changes in the plasma membrane during Arabidopsis immune signaling. *Mol Cell Proteomics* 11: M111 014555.

Elrod-Erickson M, Mishra S, Schneider D. 2000. Interactions between the cellular and humoral immune responses in Drosophila. *Curr Biol* 10: 781-784.

Feldhaar H, Gross R. 2008. Immune reactions of insects on bacterial pathogens and mutualists. *Microbes and infection / Institut Pasteur* 10: 1082-1088.

Feng X, Liu X, Luo Q, Liu BF. 2008. Mass spectrometry in systems biology: an overview. *Mass spectrometry reviews* 27: 635-660.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. 1989. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* 246: 64-71.

Finnerty CM, Karplus PA, Granados RR. 1999. The insect immune protein scolexin is a novel serine proteinase homolog. *Protein Sci* 8: 242-248.

Furukawa S, Tanaka H, Ishibashi J, Imanishi S, Yamakawa M. 2009. Functional Characterization of a Cactus Homolog from the Silkworm Bombyx mori. *Biosci Biotech Bioch* 73: 2665-2670.

Furusawa T, Rakwal R, Nam HW, Hirano M, Shibato J, Kim YS, Ogawa Y, Yoshida Y, Kramer KJ, Kouzuma Y et al. 2008. Systematic investigation of the hemolymph proteome of Manduca sexta at the fifth instar larvae stage using one- and two-dimensional proteomics platforms. *J Proteome Res* 7: 938-959.

Gan H, Wang Y, Jiang HB, Mita K, Kanost MR. 2001. A bacteria-induced, intracellular serpin in granular hemocytes of Manduca sexta. *Insect Biochem Molec* 31: 887-898.

Gerhard DS Wagner L Feingold EA Shenmen CM Grouse LH Schuler G Klein SL Old S Rasooly R Good P et al. 2004. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res* 14: 2121-2127.

Girard PA, Boublik Y, Wheat CW, Volkoff AN, Cousserans F, Brehlin M, Escoubas JM. 2008.

X-tox: An atypical defensin derived family of immune-related proteins specific to Lepidoptera. *Developmental and Comparative Immunology* 32: 575-584.

Gorman MJ, An CJ, Kanost MR. 2007. Characterization of tyrosine hydroxylase from Manduca sexta. *Insect Biochem Molec* 37: 1327-1337.

Gupta S, Wang Y, Jiang HB. 2005. Manduca sexta prophenoloxidase (proPO) activation requires proPO-activating proteinase (PAP) and serine proteinase homologs (SPHs) simultaneously. *Insect Biochem Molec* 35: 241-248.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* 17: 994-999.

Hara S, Yamakawa M. 1995. Moricin, a novel type of antibacterial peptide isolated from the silkworm, Bombyx mori. *J Biol Chem* 270: 29923-29927.

Hernandez P, Muller M, Appel RD. 2006. Automated protein identification by tandem mass spectrometry: issues and strategies. *Mass spectrometry reviews* 25: 235-254.

Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. 2005. The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry : JMS* 40: 430-443.

Hu S, Yang X. 2000. dFADD, a novel death domain-containing adapter protein for the Drosophila caspase DREDD. *J Biol Chem* 275: 30761-30764.

Hultmark D, Engstrom A, Andersson K, Steiner H, Bennich H, Boman HG. 1983. Insect immunity. Attacins, a family of antibacterial proteins from Hyalophora cecropia. *EMBO J* 2: 571-576.

International Human Genome Sequencing C. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945.

Janeway CA, Jr., Medzhitov R. 2002. Innate immune recognition. *Annu Rev Immunol* 20: 197-216.

Jiang H, Wang Y, Gu Y, Guo X, Zou Z, Scholz F, Trenczek TE, Kanost MR. 2005. Molecular

identification of a bevy of serine proteinases in Manduca sexta hemolymph. *Insect Biochem Mol Biol* 35: 931-943.

Jiang H, Wang Y, Kanost MR. 1999. Four serine proteinases expressed in Manduca sexta haemocytes. *Insect Mol Biol* 8: 39-53.

Jiang H, Wang Y, Ma C, Kanost MR. 1997. Subunit composition of pro-phenol oxidase from Manduca sexta: molecular cloning of subunit ProPO-P1. *Insect Biochem Mol Biol* 27: 835-850.

Jiang H, Wang Y, Yu XQ, Kanost MR. 2003a. Prophenoloxidase-activating proteinase-2 from hemolymph of Manduca sexta. A bacteria-inducible serine proteinase containing two clip domains. *J Biol Chem* 278: 3552-3561.

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

Jiang HB. 2008. The biochemical basis of antimicrobial responses in Manduca sexta. *Insect Sci* 15: 53-66.

Jiang HB, Kanost MR. 2000. The clip-domain family of serine proteinases in arthropods. *Insect Biochem Molec* 30: 95-105.

Jiravanichpaisal P, Lee BL, Soderhall K. 2006. Cell-mediated immunity in arthropods: hematopoiesis, coagulation, melanization and opsonization. *Immunobiology* 211: 213-236.

Kanamori Y, Hayakawa Y, Matsumoto H, Yasukochi Y, Shimura S, Nakahara Y, Kiuchi M, Kamimura M. 2010. A eukaryotic (insect) tricistronic mRNA encodes three proteins selected by context-dependent scanning. *J Biol Chem* 285: 36933-36944.

Kanost MR. 1999. Serine proteinase inhibitors in arthropod immunity. *Dev Comp Immunol* 23: 291-301.

Kanost MR, Zepp MK, Ladendorff NE, Andersson LA. 1994. Isolation and characterization of a

hemocyte aggregation inhibitor from hemolymph of Manduca sexta larvae. *Arch Insect Biochem Physiol* 27: 123-136.

Karas M, Hillenkamp F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* 60: 2299-2301.

Kauppila S, Maaty WS, Chen P, Tomar RS, Eby MT, Chapo J, Chew S, Rathore N, Zachariah S, Sinha SK et al. 2003. Eiger and its receptor, Wengen, comprise a TNF-like system in Drosophila. *Oncogene* 22: 4860-4867.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical chemistry* 74: 5383-5392.

Kocks C, Cho JH, Nehme N, Ulvila J, Pearson AM, Meister M, Strom C, Conto SL, Hetru C, Stuart LM et al. 2005. Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in Drosophila. *Cell* 123: 335-346.

Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D, Imamura K, Kai C, Harbers M et al. 2006. CAGE: cap analysis of gene expression. *Nat Methods* 3: 211-222.

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

Ladendorff NE, Kanost MR. 1991. Bacteria-induced protein P4 (hemolin) from Manduca sexta: a member of the immunoglobulin superfamily which can inhibit hemocyte aggregation. *Arch Insect Biochem Physiol* 18: 285-300.

Lau KW, Jones AR, Swainston N, Siepen JA, Hubbard SJ. 2007. Capture and analysis of quantitative proteomic data. *Proteomics* 7: 2787-2799.

Lavine MD, Strand MR. 2002. Insect hemocytes and their role in immunity. *Insect Biochem Mol Biol* 32: 1295-1309.

Leclerc V, Reichhart JM. 2004. The immune response of Drosophila melanogaster. *Immunol Rev* 198: 59-71.

Lesch C, Goto A, Lindgren M, Bidla G, Dushay MS, Theopold U. 2007. A role for Hemolectin in coagulation and immunity in Drosophila melanogaster. *Dev Comp Immunol* 31: 1255-1263.

Ling E, Rao XJ, Ao JQ, Yu XQ. 2009. Purification and characterization of a small cationic protein from the tobacco hornworm Manduca sexta. *Insect Biochem Mol Biol* 39: 263-271.

Liu X, Yao Q, Wang Y, Chen K. 2010. Proteomic analysis of nucleopolyhedrovirus infection resistance in the silkworm, Bombyx mori (Lepidoptera: Bombycidae). *J Invertebr Pathol* 105: 84-90.

Lundgren DH, Hwang SI, Wu L, Han DK. 2010. Role of spectral counting in quantitative proteomics. *Expert review of proteomics* 7: 39-53.

Ma C, Kanost MR. 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.

Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. *Trends in genetics : TIG* 24: 133-141.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.

Matsumoto Y, Oda Y, Uryu M, Hayakawa Y. 2003. Insect cytokine growth-blocking peptide triggers a termination system of cellular immunity by inducing its binding protein. *J Biol Chem* 278: 38579-38585.

Meister M, Lemaitre B, Hoffmann JA. 1997. Antimicrobial peptide defense in Drosophila. *BioEssays : news and reviews in molecular, cellular and developmental biology* 19:

1019-1026.

Merrell K, Southwick K, Graves SW, Esplin MS, Lewis NE, Thulin CD. 2004. Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry. *Journal of biomolecular techniques : JBT* 15: 238-248.

Metzker ML. 2010. Sequencing technologies - the next generation. *Nat Rev Genet* 11: 31-46.

Morozova O, Hirst M, Marra MA. 2009. Applications of new sequencing technologies for transcriptome analysis. *Annual review of genomics and human genetics* 10: 135-151.

Mulnix AB, Dunn PE. 1994. Structure and induction of a lysozyme gene from the tobacco hornworm, Manduca sexta. *Insect Biochem Mol Biol* 24: 271-281.

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

Nardi JB, Martos R, Walden KK, Lampe DJ, Robertson HM. 1999. Expression of lacunin, a large multidomain extracellular matrix protein, accompanies morphogenesis of epithelial monolayers in Manduca sexta. *Insect Biochem Mol Biol* 29: 883-897.

Nardi JB, Zhuang S, Pilas B, Bee CM, Kanost MR. 2005. Clustering of adhesion receptors following exposure of insect blood cells to foreign surfaces. *J Insect Physiol* 51: 555-564.

Nesvizhskii AI, Keller A, Kolker E, Aebersold R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Analytical chemistry* 75: 4646-4658.

Nichol H, Law JH, Winzerling JJ. 2002. Iron metabolism in insects. *Annu Rev Entomol* 47: 535-559.

Noguchi H, Tsuzuki S, Tanaka K, Matsumoto H, Hiruma K, Hayakawa Y. 2003. Isolation and characterization of a dopa decarboxylase cDNA and the induction of its expression by an insect cytokine, growth-blocking peptide in Pseudaletia separata. *Insect Biochem Molec* 33: 209-217.

Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA,

Ahn NG. 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 4: 1487-1502.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1: 376-386.

Pappin DJ, Hojrup P, Bleasby AJ. 1993. Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* 3: 327-332.

Pauchet Y, Wilkinson P, Vogel H, Nelson DR, Reynolds SE, Heckel DG, Ffrench-Constant RH. 2010. Pyrosequencing the Manduca sexta larval midgut transcriptome: messages for digestion, detoxification and defence. *Insect Molecular Biology* 19: 61-75.

Pinheiro VB, Ellar DJ. 2006. How to kill a mocking bug? *Cellular microbiology* 8: 545-557.

Pozhitkov AE, Tautz D, Noble PA. 2007. Oligonucleotide microarrays: widely applied--poorly understood. *Briefings in functional genomics & proteomics* 6: 141-148.

Rabilloud T. 2002. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* 2: 3-10.

Ragan, E.J., An, C., Jiang, H., Kanost, M.R., 2009. Roles of hemolymph proteins in antimicrobial defenses of Manduca sexta. In: Reynolds, S., Rolff, J. (Eds.), Insect Infection and Immunity. Oxford University Press, pp. 34e48.

Rayaprolu S, Wang Y, Kanost MR, Hartson S, Jiang H. 2010a. Functional analysis of four processing products from multiple precursors encoded by a lebocin-related gene from Manduca sexta. *Dev Comp Immunol* 34: 638-647.

Reddy KV, Yedery RD, Aranha C. 2004. Antimicrobial peptides: premises and promises. *International journal of antimicrobial agents* 24: 536-547.

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

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S,

Dey S, Daniels S et al. 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3: 1154-1169.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467.

Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.

Schulze WX, Usadel B. 2010. Quantitation in mass-spectrometry-based proteomics. *Annual review of plant biology* 61: 491-516.

Shin SW, Park SS, Park DS, Kim MG, Kim SC, Brey PT, Park HY. 1998. Isolation and characterization of immune-related genes from the fall webworm, Hyphantria cunea, using PCR-based differential display and subtractive cloning. *Insect Biochem Molec* 28: 827-837.

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

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.

Tanaka H, Ishibashi J, Fujita K, Nakajima Y, Sagisaka A, Tomimoto K, Suzuki N, Yoshiyama M, Kaneko Y, Iwasaki T et al. 2008. A genome-wide analysis of genes and gene families involved in innate immunity of Bombyx mori. *Insect Biochem Molec* 38: 1087-1110.

Tanaka H, Matsuki H, Furukawa S, Sagisaka A, Kotani E, Mori H, Yamakawa M. 2007. Identification and functional analysis of Relish homologs in the silkworm, Bombyx mori. *Bba-Gene Struct Expr* 1769: 559-568.

Tzou P, De Gregorio E, Lemaitre B. 2002. How Drosophila combats microbial infection: a model

to study innate immunity and host-pathogen interactions. *Current opinion in microbiology* 5: 102-110.

Tzou P, Ohresser S, Ferrandon D, Capovilla M, Reichhart JM, Lemaitre B, Hoffmann JA, Imler JL. 2000. Tissue-specific inducible expression of antimicrobial peptide genes in Drosophila surface epithelia. *Immunity* 13: 737-748.

Unlu M, Morgan ME, Minden JS. 1997. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18: 2071-2077.

VanGuilder HD, Vrana KE, Freeman WM. 2008. Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* 44: 619-626.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. *Science* 270: 484-487.

Vogel BE, Hedgecock EM. 2001. Hemicentin, a conserved extracellular member of the immunoglobulin superfamily, organizes epithelial and other cell attachments into oriented line-shaped junctions. *Development* 128: 883-894.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57-63.

Wolters DA, Washburn MP, Yates JR, 3rd. 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Analytical chemistry* 73: 5683-5690.

Yu XQ, Jiang HB, Wang Y, Kanost MR. 2003. Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, Manduca sexta. *Insect Biochem Molec* 33: 197-208.

Yu XQ, Tracy ME, Ling EJ, Scholz FR, Trenczek T. 2005. A novel C-type immulectin-3 from Manduca sexta is translocated from hemolymph into the cytoplasm of hemocytes. *Insect Biochem Molec* 35: 285-295.

Zang XX, Maizels RM. 2001. Serine proteinase inhibitors from nematodes and the arms race between host and pathogen. *Trends Biochem Sci* 26: 191-197.

Zhu Y, Johnson TJ, Myers AA, Kanost MR. 2003. Identification by subtractive suppression hybridization of bacteria-induced genes expressed in Manduca sexta fat body. *Insect Biochem Molec* 33: 541-559.

Zhu YF, Ragan EJ, Kanost MR. 2010. Leureptin: A soluble, extracellular leucine-rich repeat protein from Manduca sexta that binds lipopolysaccharide. *Insect Biochem Molec* 40: 713-722.

Ziganshin R, Arapidi G, Azarkin I, Zaryadieva E, Alexeev D, Govorun V, Ivanov V. 2011. New method for peptide desorption from abundant blood proteins for plasma/serum peptidome analyses by mass spectrometry. *Journal of proteomics* 74: 595-606.

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

Zou Z, Picheng Z, Weng H, Mita K, Jiang H. 2009. A comparative analysis of serpin genes in the silkworm genome. *Genomics* 93: 367-375.

APPENDIX

ABBREVIATION LIST

AMP: antimicrobial peptide

NGS: next generation sequencing

MS: mass spectrometry/mass spectrometer

PRR: pattern recognition receptor

PAMP: pathogen-associated molecular pattern

LPS: lipopolysaccharide

LTA: lipoteichoic acid

βGRP: β-1, 3-glucan recognition protein

GNBP: Gram-negative bacteria binding protein

PGRP: peptidoglycan recognition protein

CTL: C-type lectin

HP: hemolymph protease

proPO/PPO: pro-phenoloxidase

PO: phenoloxidase

SPH: serine protease homologue

serpin: serine protease inhibitor

DIF: dorsal-related immune factor

EST: expressed sequence tag

MALDI: matrix assisted laser desorption/ionization

ESI: electrospray ionization

QIT: quadrupole ion trap

LIT: linear ion trap

TOF: time of flight

FTICR: Fourier transform ion cyclotron resonance

LC: liquid chromatography

MudPIT: multidimensional protein identification

DiGE: difference gel electrophoresis

SILAC: stable isotope-labeling with amino acids in cell culture

ICAT: isotope coded affinity tag

iTRAQ: isobaric tagging technology

CF, IF, CH, IH: fat body (F) or hemocyte (H) mRNA samples/sequences from immune induced (I) or control (C) insects

CIFH: contigs generated using reads from CF, CH, IF, and IH samples

06CIFH: contigs generated using reads from CF, CH, IF, IH samples and cDNA sequences obtained in 2006

LNF: library normalization factor

RA: relative abundance

ARN: adjusted read number

UP: up-regulated genes

DN: down-regulated genes

HC: genes preferentially expressed in hemocyte

FB: genes preferentially expressed in fat body

FDR: false discovery rate

PAP: pro-phenol oxidase activating proteinase

HAIP: hemocyte aggregation inhibitor protein

SC: spectral count

PSP: plasmatocyte spreading peptide

VITA

Shuguang Zhang

Candidate for the Degree of

Master of Science

Thesis: QUANTITATIVE TRANSCRIPTOME AND PROTEOME ANALYSES OF IMMUNITY-RELATED MOLECULES IN *MANDUCA SEXTA* LARVAL HEMOLYMPH

Major Field: Entomology

Biographical:

Education:

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

Completed the requirements for the Bachelor of Science in biological sciences at University of Science and Technology of China, Hefei, China in July, 2009.

Experience:

Graduate Research Assistant at Oklahoma State University from 2009-2012

Name: Shuguang Zhang Date of Degree: July, 2012

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: QUANTITATIVE TRANSCRIPTOME AND PROTEOME ANALYSES OF IMMUNITY-RELATED MOLECULES IN *MANDUCA SEXTA* LARVAL HEMOLYMPH

Pages in Study: 86 Candidate for the Degree of Master of Science

Major Field: Entomology

Scope and Method of Study:

This study explores immunity-related molecules in *Manduca sexta* larvae by studying the quantitative change of mRNAs and proteins after immune challenge. To investigate the mRNA level changes, we isolated mRNA samples from fat body and hemocytes of both pathogen- and buffer-injected insects. The cDNA samples were separately sequenced on 454 pyrosequencer. Short reads from different samples were assembled to contigs and read numbers from each sample were retrieved for all the contigs. These numbers were used to calculate relative abundances of the corresponding genes in each sample. Thus we compared the read numbers from different samples and found immune induced/suppressed genes and tissue preferentially expressed genes. In the proteome study, we collected plasma from both pathogen- and buffer-injected insects. After gel electrophoresis or acetonitrile precipitation, plasma samples were digested by trypsin and subsequently analyzed on mass analyzer. A manually constructed protein database was used for protein identification and spectral counts for each protein were used for semi-quantitative analysis.

Findings and Conclusions:

In the transcriptome study, we obtained two millions of reads which were assembled to 19,020 contigs. By comparing read numbers from different samples we found 528 up-regulated genes including different immune factors. We also found hundreds of down-regulated genes and fat body/hemocyte preferentially expressed genes. The set of genes enlarged our knowledge about immune factors in *M. sexta*. They also contributed to the *M. sexta* genome annotation. In the proteome study, we identified 785 proteins from the gel-extracted samples and 270 samples from the acetonitrile treated samples. Of them, 115 and 155 proteins were found to be increased and decreased after immune challenge, respectively. A total of 211 immunity-related proteins were discovered in this study. The transcriptome study showed the efficiency and accuracy of pyrosequencing in quantitative transcriptomic analyses. The proteome study proved the robustness of spectral counting method. These two studies showed that many immune factors did not show drastic change except that antimicrobial peptides were highly induce at both mRNA and protein levels.

ADVISER'S APPROVAL: Dr. Haobo Jiang