Available online at www.sciencedirect.com
ScienceDirect
Journal of Insect Physiology 54 (2008) 1–16
Journal of Insect Physiology
www.elsevier.com/locate/jinsphys

Analysis of transcripts and proteins expressed in the salivary glands of Hessian fly (*Mayetiola destructor*) larvae ☆

Ming-Shun Chen[a,*], Hui-Xian Zhao[b], Yu Cheng Zhu[c], Brian Scheffler[c], Xuming Liu[d], Xiang Liu[d], Scot Hulbert[e], Jeffrey J. Stuart[f]

[a]*USDA-ARS Plant Science and Entomology Research Unit, 4008 Throckmorton Hall, 1515 College Avenue, Manhattan, KS 66506, USA*
[b]*College of Life Sciences, Northwest A & F University, Yangling, Shaanxi, China*
[c]*USDA-ARS-JWDSRC, 141 Exp Stn Rd, Stoneville, MS 38776, USA*
[d]*Department of Entomology, Kansas State University, Manhattan, KS 66506, USA*
[e]*Department of Plant Pathology, Washington State University, Pullman, WA 99164, USA*
[f]*Department of Entomology, Purdue University, West Lafayette, IN 47907, USA*

Received 8 June 2007; received in revised form 20 July 2007; accepted 24 July 2007

Abstract

Hessian fly (*Mayetiola destructor*) larvae are thought to manipulate host growth and metabolism through salivary secretions. However, the transcriptome and proteome of Hessian fly salivary glands have not been systematically analyzed. In this research, we analyzed Expressed-Sequence-Tags (EST) representing 6106 cDNA clones randomly selected from four libraries made from dissected salivary glands. We also analyzed the protein composition of dissected salivary glands using one- and two-dimensional gel electrophoresis as well as LC-MS/MS analysis. Transcriptomic analysis revealed that approximately 60% of the total cDNA clones and 40% of assembled clusters encoded secretory proteins (SP). The SP-encoding cDNAs were grouped into superfamilies and families according to sequence similarities. In addition to the high percentage of SP-encoding transcripts, there was also a high percentage of transcripts encoding proteins that were either involved directly in protein synthesis or in house-keeping functions that provide conditions necessary for protein synthesis. Proteomic analysis also revealed a high percentage of proteins involved in protein synthesis either directly or indirectly. The high percentage of SP-encoding transcripts and high percentage of proteins related to protein synthesis suggested that the salivary glands of Hessian fly larvae are indeed specialized tissues for synthesis of proteins for host injection. However, LC-MS/MS analysis of 64 proteins did not identify any SPs corresponding to the cDNA sequences. The lack of accumulation of SPs in the salivary glands indicated the SPs were likely secreted as soon as they were synthesized.

Keywords: Hessian fly; *Mayetiola destructor*; Secretory proteins; Salivary glands; Transcriptome; Proteome

1. Introduction

The Hessian fly, *Mayetiola destructor*, belongs to a group of insects called gall midges (Cecidomyiidae, Diptera) (Harris et al., 2003). The insect has been one of the most destructive pests of wheat, *Triticum aestivum*, ever since it was introduced into the USA during the Revolutionary War in 1770s (Hatchett et al., 1987; Pauly, 2002). The insect has four different developmental stages: adult, larva, pupa, and egg (Gagne and Hatchett, 1989). The larval stage is responsible for economic damage to crop plants. Soon after hatching, a neonate larva moves down the leaf sheath to the base of the plant and begins to attack wheat seedlings. A Hessian fly larva has three instars, of which the first larval instar, which lasts approximately 5 days at 20 °C, is the most critical for survival on a specific host plant. On susceptible plants, such as Karl92 or Newton,

☆The GenBank accessions for ESTs are EV465518–EV467318. The GenBank accessions for full cDNA sequences are AY318789–AY318798, AY542316–AY542323, AY596471–AY596479, AY665658–AY669864, AY828552–AY828563, DQ196068–DQ196076, DQ232690–DQ232718, and DQ458295–DQ458315.

*Corresponding author. Tel.: +1 785 532 4719; fax: +1 785 532 6232.
E-mail address: mchen@ksu.edu (M.-S Chen).

0022-1910/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jinsphys.2007.07.007

first instar larvae establish a permanent feeding site, and develop into immobile second instar larvae (Byers and Gallun, 1971). Third instar larvae exist in a non-feeding, pre-pupa stage. On resistant plants, such as Iris or Molly, first instar larvae survive for 3–5 days before dying.

Host plant resistance is the most effective and cost efficient way to control Hessian fly (Ratcliffe and Hatchett, 1997). Numerous Hessian fly resistance genes have been identified and are being used in wheat cultivars (Sardesai et al., 2005). However, the major challenge to use of resistant cultivars is the ability of the fly to constantly develop new biotypes that overcome the resistance of specific genes after they are deployed (Ratcliffe et al., 1994, 2000). For continued success of the host resistance strategy, we need to understand the molecular mechanism of biotype differentiation.

Insect biotypes are defined as populations with different host specificities, a similar concept to the races or pathotypes of plant pathogens. The differentiation of unique races or pathotypes of a pathogen is typically caused by a change(s) in the proteins that are secreted into host plants by the pathogen (Grant et al., 2006). These secretory proteins (SP), also referred to as effector proteins, can elicit various physiological changes within host plants to create a favorable environment for the pathogen to parasitize (Tang et al., 1996). Like pathogens, insects also inject substances including proteins into host plants during feeding. The salivary glands of insects are the primary tissues to synthesize proteins for host injection (Lehane, 1991; Miles, 1999). Therefore, an analysis of the genes expressed in insect salivary glands and the identification of the proteins that are injected into host plants could lead to the elucidation of the molecular mechanism(s) resulting in insect biotype differentiation.

The salivary glands of Hessian fly larvae consist of a basal region and a filament region (Stuart and Hatchett, 1987). The basal region is likely the place that proteins for host injection are synthesized since it is connected directly to mandibles specialized for saliva secretion and injection (Hatchett et al., 1990). The basal region exists only during the first instar, when larvae actively establish a permanent feeding site and inhibit plant growth (Haseman, 1930; Byers and Gallun, 1971). The basal region decays at the beginning of the second instar, leaving the salivary glands with only a pair of filaments for the remainder of the larval stage (Stuart and Hatchett, 1987).

Previously, we identified several genes that encode SPs from Hessian fly salivary glands (Chen et al., 2004, 2006; Liu et al., 2004). Here we provide a systematic analysis of the transcripts obtained from several cDNA libraries derived from dissected salivary glands of first instar larvae. In addition, we analyzed the protein composition of the salivary glands via proteomic profiling. These genomic and proteomic analyses provide a foundation for future research on the role of the salivary glands in Hessian fly virulence and biotype differentiation.

2. Materials and methods

2.1. Insects and salivary gland preparation

Hessian fly biotype GP was from a laboratory colony collected in Ellis County, Kansas in 1988 (Gagne and Hatchett, 1989). Biotype L was provided by Sue Cambron, USDA-ARS at West Lafayette, Indiana. A Syrian population was provided by Mustapha El-Bouhssini at the International Center for Agricultural Research in the Dry Areas, Aleppo, Syria. All insects were maintained on susceptible wheat seedlings ('Newton' or 'Karl 92') in environmental chambers at 20 °C and 12:12 (L:D) photoperiod.

Salivary glands were obtained by dissecting 3-day old, first instar larvae in DEPC-treated distilled water. For RNA analysis and cDNA library construction, the dissected glands were transferred into TRI reagent™ (Molecular Research, Inc., Cincinnati, OH) directly. For proteomic analysis, the glands were frozen in liquid nitrogen as soon as they were obtained.

2.2. cDNA library construction and sequencing

Total RNA was isolated from 200 pairs of salivary glands for each library using TRI reagent™ following the protocol provided by the manufacturer. The cDNA libraries were constructed using a 'SMART™' library construction kit from Clontech (Palo Alto, CA). To identify as many unique transcripts as possible, two types of cDNA libraries were constructed. A phage library was constructed to recover larger transcripts by using cDNAs larger than 1.5 kb as inserts. The phage library was made from biotype GP following the protocol provided by the manufacturer (Clontech). Since plasmid library is less time consuming, three plasmid libraries were made from biotypes GP, biotype L, and the Syrian population, respectively. The plasmid libraries were made following the same procedure as the phage library with one modification: cDNA inserts were ligated into the pPCR-XL-TOPO plasmid using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA) instead to a phage vector.

For sequencing, phage libraries were plated and individual plagues were picked. The plagues were converted into plasmid clones *in vivo* before DNA isolation. Plasmid DNA was then isolated using a Qiagen BioRobot 3000 and sequenced in an ABI 3700 DNA analyzer. For each cDNA clone from the phage library, one EST was produced using the 5′-specific primer since the insert was directionally ligated into the phage vector (Clontech). For cDNAs from the plasmid libraries, the clones were first sequenced with the M13 forward primer. If the sequence was derived from the 3′-end of a clone, this clone was then sequenced again using the M13 reverse primer. Only those sequences generated from the 5′-ends of the cDNAs were counted as ESTs in this paper, even though all sequences were included in our own databases for analysis.

2.3. Sequence analysis

Cluster analysis of ESTs was conducted using the CAP3 Sequence Assembly Program (http://pbil.univ-lyon1.fr/cap3.php) (Huang and Madan, 1999). ORF and sequence similarity analysis were performed using ORF finder. Sequence alignment and similarity analysis were performed using various BLAST programs (http://www.ncbi.nlm.nih.gov/) of the National Center for Biotechnology Information (Bethesda, MD). Initial database search were conducted with BLASTN and BLASTX. Sequence alignments with *E*-values greater than 10^{-2} were considered no sequence similarity. Sequence alignments with *E*-values smaller than 10^{-10} were considered similar sequences. Sequence alignments with *E*-values between 10^{-2} and 10^{-10} were further examined individually regarding the length, gaps, and residues potentially involved in domain structures within the alignments. Alignments of more than 50 residues with multiple gaps and no conserved residues involved in potential domain structures are considered no sequence similarity. For those sequences that did not show sequence similarity with any known sequences, the nucleotide sequences were then translated into proteins of all possible ORFs. The putative proteins were then analyzed using BLASTP, PSI-BLAST, PHI-BLAST, RPSBLAST from the same website, and other searching methods including PFAM and SMART for short matches or domain conservation.

Analysis for secretion signal peptides was carried out using the SignalP v. 1.1 (Center for Biological Sequence Analysis, Technical University of Denmark, http://www.cbs.dtu.dk/services/SignalP/) and PSORT II analysis (Prediction of Protein Sorting Signals and Localization Sites in Amino Acid Sequences, http://psort.nibb.ac.jp/).

2.4. Real-time PCR analysis

Total RNA was isolated from dissected tissues as well as from whole larvae as described above. The total RNA was reverse-transcribed into cDNA using superscript reverse transcriptase following the manufacturer's procedure (BD Biosciences, San Jose, CA). After removing the RNA in the reaction by DNase-free RNase, the cDNAs were used as template for real-time PCR (RT-PCR) analysis.

RT-PCR was performed with iQ SYBR Green Supermix on a iCycler real-time detection system (Bio-Rad, Hercules, CA). Each reaction was carried out with 2 μl of a 1/40 (v/v) dilution of the first cDNA strand, 0.5 μM of each primer in a total volume of 25 μl. The cycling conditions were: 95 °C for 5 min followed by 45 cycles of denaturation at 95 °C for 20 s, annealing and extension at 62–64.5 °C, depending on the primer set, for 45 s. At the end of the cycles, PCR amplification specificity was verified by obtaining a dissociation curve, derived by cooling the denatured samples to 55 °C and raising the temperature 0.5 °C for 10 s for each cycle, for a total of 80 cycles until reaching 95 °C. The PCR products were analyzed on 1.5% agarose gels, and subsequently purified and sequenced to confirm faithful amplification. Primers were designed using the Beacon Designer (v2.0) software from Biosoft International (Palo Alto, CA) and the primer sequences were listed in Table 1. Plasmid DNA containing the corresponding insert was used to generate a calibration standard curve, where cycle threshold (Ct) values are plotted to serve as standard concentrations. The transcript concentration for each sample was calculated based on the standard concentrations. A negative control without template was always included for each primer set. Template concentrations of different tissue samples were normalized by the ribosomal protein P1 (RP-P1) gene. For each sample analyzed, results represent the mean of values obtained from at least two independent PCR reactions, and from at least three independent biological replicates.

Table 1
Primer sets for real-time PCR

Primer name	Primer sequence
RP-P1_F	AAGAATTGGCTTGCGTTTACTCC
RP-P1_R	CGTTGGCGGCTTTCAAAATGG
RP-P2_F	GGTGACCTGGCTGGCAAAAG
RP-P2_R	ACGGCACCACCTCCAGATG
NDK_F	AGTTGTTGCTATGGTATGGGAAGG
NDK_R	TGAGTCTGCTGGATTGGTTGC
Hsc70_F	AATTGACAACGAATCCAGAGAACC
Hsc70_R	GCGTCGGCTTGAACATTTGG
PGDH_F	TTGGTATGCGAGTTATTGGTTACG
PGDH _R	GTGGAATGAGTGGTGTGTGAAC
1A2_F	GCCAACAGCCCAATCAAACC
1A2_R	GGCAGGTCCATTAGAATTTCCG
1C2_F	CTTGACGAAGATGAGCACTGG
1C2_R	TGCTTCATTAGGGCCATTATCATC
4C1_F	AGTCTAGCAACCCCATCAATATCC
4C1_R	CAGATATACCCCTCCACCAATGC
6A1_F	CGAAAGCCGAGTCTTCTAATGC
6A1_R	CCTCTCCAAGTCTGCCGAAAG
7C1_F	GGTGGTAAGGGTAAGGATGATGG
7C1_R	TCAGAGTCTCCGAGAACTTTTGG
10A1_F	AGCGCTGCAGGGTTGCC
10A1_R	CAATTTCTCCTGTTCCGTAGGTAG
12A3_F	AACCGACGCCCCAAACAAG
12A3_R	TCCGCACCATTATCCGAATCTG

RP-P1 and *RP-P2* are ribosomal protein P1 and P2. *NDK*, *Hsc70*, and *PGDH* represent nucleotide diphosphate kinase, heat shock protein cognate 70 kDa, and phosphoglycerate dehydrogenase, respectively. *1A2*, *1C2*, *4C1*, *6A1*, *7C1*, *10A1*, and *12A3* are seven representative SP-encoding genes.

2.5. One-dimensional (1D) gel electrophoresis

Approximately 300 pairs of salivary glands, dissected from 3-day old, first instar larvae, were homogenized and dried completely using a Speed Vacuum. The dried samples were mixed with 30 μl Lammeli buffer (Bio-Rad, Hercules, CA) with 50 mM DTT. The sample was then incubated for 30 min at 65 °C, and was then loaded onto a 4–20% gradient sodium dodecyl sulfate/polyacrylamide gel

electrophoresis gel (SDS/PAGE). The gel was run for 1 h at 200 V in TGS Buffer (Bio-Rad), fixed for 30 min in 49:50:1 ratio of water, methanol, and acetic acid, and washed twice with water for 5 min each. The gel was then stained for 1 h in Bio-Safe Coomassie, washed, and stored in water at 4 °C until imaging.

2.6. Two-dimensional (2D) gel electrophoresis

Approximately 300 pairs of glands were homogenized and dried as described above. The dried samples were mixed with 185 µl of rehydration buffer (7 M urea, 2 M thiourea, 2% Chaps, traces of bromophenol blue, 0.5% IPG 3–10 buffer). After incubation at room temperature for 30 min, the samples were centrifuged at 13,000 rpm for 5 min. The supernatant was then loaded onto an iso-electric focusing (IEF) tray, which was overlaid with an 11 cm, 3–10 IEF strip. The IEF gel strip was run in a pre-programmed device until the voltage reached 35,000 V. The strip was removed from the device and incubated for 15 min in an equilibration solution (6 M urea, 30% glycerol, 2% SDS, 50 mM Tris pH 8.8) with 2% DTT, followed by 15 min incubation in the equilibration solution with 2.5% iodoacetamide. The strip was then loaded onto the top of an 8–16% gradient SDS/PAGE gel, sealed with 0.5% molten agarose, and run in TGS buffer for 1 h at 200 V. The gel was fixed, washed, stained and stored as described above. To improve separation of small proteins, another 2D gel was run using a 10–20% gradient SDS/PAGE gel.

2.7. Gel imaging and quantification of specific protein bands

Gels were imaged in a Fujifilm Intelligent Dark Box II with a Fujifilm LAS-1000 Luminescent Image Analyzer (CCD camera). Image intensities of specific protein spots were analyzed quantitatively using Bio-Rad's PDQuest 2D gel image analysis software following the standard procedure provided by the manufacturer. For background subtraction, the non-spot method was used. For normalization, the total volume multiplied by total area method was used.

2.8. LC-MS/MS protein identification

LC-MS/MS protein identification and database searching was carried out by Cognition Therapeutics (Rockville, MD) through a commercial contract as follows. Protein bands of interest were removed from the gel using a clean polypropylene pipette with the end modified to make the appropriate size cut. The gel cuts were placed in Eppendorf tubes, and washed with the following solutions sequentially: 200 µl of water/acetonitrile (ACN) (1:1, v/v) twice for 15 min each, 100 µl of 0.1 M NH_4HCO_3 for 5 min, and a solution containing 0.1 M NH_4HCO_3 and acetonitrile at a 1:1 ratio for 15 min. The gel cuts were then dried using a speed vacuum. Proteins in the dried gel cuts were digested with 20 µl digestion buffer (12.5 ng/µl of trypsin in 50 mM NH_4HCO_3) for 45 min on ice. The digested peptides were extracted once with a 40 µl solution of 25 mM NH_4HCO_3 and CAN (1:1, v/v) for 30 min, and twice with a 50 µl solution containing a mixture of 5% formic acid and ACN (1:1, v/v). The extracts were combined and analyzed by LC-MS/MS.

For LC-MS/MS analysis, the peptide solution was injected using the microliter pick-up method on a LC Packings (Sunnyvale, CA) Famos autosampler. A typical reversed-phase gradient was used from low to high organic solvent over about 60 min. Mobile phase A was 0.1% formic acid and B was 100% MeCN with 0.1% formic acid. The system utilized a small split flow resulting in a column flow rate of approximately 0.3 ml/min. The program was run with 95% A and 5% B for 10 min to desalt, followed by a gradient separation step from 95% A and 5% B to 35% A and 65% B within 28 min. The column was then cleaned with 5% A and 95% B for 5 min, and re-equilibrated with 95% A and 5% B for 15 min.

The mass/charge (m/z) ratios of peptides and their fragmented ions were recorded as spectra by the mass spectrometer. LC-MS/MS data were processed using ProteinLynx v2.15 software (Micromass) and searched against the NCBI non-redundant (nr) and a Hessian fly database including ESTs, full cDNAs, and available genomic sequences, using MASCOT (Matrix Science Inc., Boston, MA, USA). Searches were performed using carbamidomethylation of cysteine as the fixed modification and oxidation of methionine as the variable modification, allowing for one missed cleavage during trypsin digestion.

3. Results

3.1. Salivary transcript distribution

A total of 6106 ESTs (one from each cDNA) were obtained by sequencing random clones from cDNA libraries as described in Section 2. Of the total ESTs, 59.7% encoded SPs, while the other 40.3% encoded non-secretory proteins (NSPs). The ESTs were then assembled into 2017 clusters (singlets and contigs of overlapping sequences). Among these clusters, 784 (38.9%) were predicted to encode SPs while the other 1233 (61.1%) encoded NSPs. Among the SP-encoding sequences, 767 encoded unique proteins (with no sequence similarity to any known sequences), while the other 17 encoded known proteins (with sequence similarity to those with defined functions). These 17 sequences are discussed below with the NSP-encoding sequences that also encoded known proteins.

To further characterize the unique SPs, full-length cDNA sequences were obtained from representative clones. The SPs can be sorted into groups according to their secretion signal peptides, protein structures, and cDNA sequence similarities. Proteins within a group have the same or very similar secretion signal peptides, and

conserved non-coding regions in their encoding cDNAs (data not shown). But the mature proteins between group members could be very different from each other (Fig. 1A) due to a higher rate of diversification in coding regions than that in non-coding regions (Chen et al., 2004). Since group members are clustered within short chromosome regions (data not shown) and may have originated from gene duplications (Chen et al., 2006), the SPs within the same group were considered a superfamily. Some of the superfamilies contained several families. A family consisted of proteins with sequence differences including amino acid substitutions, small insertions/deletions, or divergent carboxy termini (Fig. 1B).

Twenty-five superfamilies were identified that represented more than a single SP gene family and these included a total of 622 of the 767 unique SPs. Some of these superfamilies, such as superfamilies 4 and 9, contained more than 30 families with over 100 members in each. Another 123 of the SP belong to gene families that are not members of superfamilies and the remaining 22 SP are not members of gene families. Although it is possible that some of the small families represent single genes with abundant within-biotype allelic variations, it is clear that the large majority of the predicted SP-encoding genes belong to large gene families or superfamilies. The majority of the unique SP is very small, with 50–200 amino acids.

A

```
L1C12     MFNSQKLIICCLLFAAVWVQSLKAALTPGLSIYLDEDKDRDLLVPPGSQYEKYLISSANR 60
G9B3      MFNSQKLIICCLLFAAVWVQSLKAALTPGLSIYLDEDKDRDLLVPPGSQYEKYLISSANR 60
SSGP-4A1  MFNSQKLIIFCLLFAAVWVQFSKAAPKR-----RPEDVDLTLSLATPS-------ISDEQ 48
G15B9     MFNSQKLIICCLLFAVVLLQSLEAGSNG-------KEIDLNLSIAQPGSPNTHALLSPEN 53
S8A3      MLNSQKLIICCLLFAAVLLQSLKATPKR-----KGKEIDLNLTMEQPG--------TSKR 47

L1C12     YSKIFHRGEHEIKSIRLIEYYLLFEIHNYKNAAEKYEFDDTKFWSWWGNEIGFDKKSELG 120
G9B3      YSKIFHRGEHEIKSIRLIEYYLLVNFYNYKDTAKDFGFDDTKFWSWWGDQIGFDKKSELG 120
SSGP-4A1  KSQIMFLGQHLDAAVLAYIDLLGYNVDNYVH---------------WWRGISDDVKAKIA 93
G15B9     VESVELLEQGIMCVAMKYWERMKLDRSQFED---------------WWLKLDSKTKLKLV 98
S8A3      LIKAFKAALGAVRSDNLDQKALLA----------------------YWKTLSIEKQLEFA 85

L1C12     AAFELKAMEIRRDGRQLVETLKKRFSK------CARETLDSYISKMEEENNIVTPFQKEW 174
G9B3      AAFELKAIKIPKIQLRTDENLDKQFVR------FARETLDSYISKINKENNIVTPFQKEW 174
SSGP-4A1  SKFHSVAKDKKTVGSSNNVGREKELDK------LAFETVRSFIKDG------TIPQQECW 141
G15B9     QRHQACLKRIATDIMRVADMRLSKVCF------CAHEVLDAVFRPTPRR-VKIEEITEQW 151
S8A3      TKYEECVERKTERAASFAEVPDSETERRAVSDACVIEIIDEHFGLDNPE---LKPCSSCR 142

L1C12     YKFRKAVEA-------ADPVLQHNIPT--- 194
G9B3      YKFQTAVNAKFQINNPTSSEHRFNIGMQNK 204
SSGP-4A1  F---------------------------- 142
G15B9     DA---------------------------- 153
S8A3      IKNK-------------------------- 146
```

B

```
G2C8      -----MLNSKSWIICSFLFVAVLAMGTDNSAENDKMAKWVKSLDGSISYLRSKNFKNKEF
L4H12     -----MLNSKSWIICSFLFAAVLAVGTDNSAENEKMAKWVKSLDGSISYLRSKNFKNKEF
G20G4     MNCFKMLNSKSWIICSFLFAAVLAMGTDNSAENKKMAKWVKSLDGSISYLRSKNFKNKEF
L3A9      -----MLNSKSWIICSFLFVAVLAMGTDNSAENDEMAKWVKSLDGSISYLRSKNFKNKEF
G13E6           MLNSKSWIICSFLFAAVLAVGTDNSEENDEMAKWVKSLDGSISYLKSKNFKNTEF

G2C8      TDLNSYEWTDWEKMPYKLKASAAYTYKTEMLKKNYKTNRNADQFSATCAYKAFKELLRKC
L4H12     TDLNLYEWSDWEKIPYRLKAAASYTYKTEMLKKNYKTNRNADQFSATRAYKAFKEFLRKR
G20G4     TDLNSYEWTDWEKMPYKLKSSAAYTYKTEMLKKNYKTNRNADQFSATCAYKAFKELLRKC
L3A9      TDLNLYEWSDWEKMPYKLKAAAAYTYKTEMLKKNYKTNRNANQFSATCAYKAFKELLRKC
G13E6     TDLNLYVWSDWEKMPYRLKAFASYTYKTEMLKKKYKPNRNADQFTATCAYKAFKDFLKKQ

G2C8      DDIIYFQLTFHKDIRAKYNEIFNDKANYLKFLIDFVAGLEDKDIFVPETSKIPRLAE
L4H12     DEFMYFHLTFHTDLKMKYDEIFNDKVKFLKLIIDFEAVLEEKDASFMDILYIPFVAD
G20G4     DDIIYFQLIFQKDIRAKYNEIFNDKANYLKFLIDFVAGLKDKDIFVPETSKIPRLAE
L3A9      DEIIYLRLTFNKDLKAKYDEIFNDKANYLKFLINFVAGLEDKDISVSETIIIPLAAE
G13E6     DELIYFHLTFHKALKAKYDEIFNDKANYLKFLIDFEAELEDENGC

G2C8      AFRNYSKSKASGAGGSSSTS
L4H12     VFRDYLKGKASESGASSSTS
G20G4     AFRNYSKSKASGAGGSSSTS
L3A9      SFRDYFNSKASGAGGSSSTS
```

Fig. 1. Sequence alignment among superfamily and family members: (A) Sequence comparison among members of a representative superfamily. Predicted secretion signal peptides are underlined and italicized. Residues identical among superfamily members are in bold. (B) Sequence comparison among members of a representative family that belongs to the superfamily shown in (A). Predicted secretion peptides are underlined and italicized. Residues different among family members are in bold.

Among the 1233 NSP-encoding clusters, 773 (62.7%) were predicted to encode unique proteins due to their lack of significant homology to protein databases. Another 190 (15.4%) had homology to predicted proteins with no known function, and the remaining 270 (21.9%) were homologous to characterized proteins with predicted functions. For the sequences encoding known proteins, 40.1% encoded proteins with functions in protein synthesis or folding, 40.8% encoded house-keeping proteins, and the other 19.1% encoded proteins with various other functions (Table 2). The proteins involved in protein synthesis and folding included 66 ribosomal proteins, 21 translation

Table 2
A list of transcripts encoding proteins with similarity to known proteins

Cluster	No.	First hit	Accession	*E*-value
(A) **Protein synthesis and folding**				
1. Ribosomal proteins				
460	1	60S ribosomal protein L4 (L1) [*Xenopus laevis*]	P09180	1.0E−110
438	1	Mitochondrial ribosomal protein L4 [*Homo sapiens*]	NP_057040	5.0E−59
490	1	60S ribosomal protein L5 [*Anopheles gambiae*]	O44248	1.0E−112
493	1	Ribosomal protein L6 [*Rattus norvegicus*]	NP_446423	2.0E−25
416	3	Ribosomal protein L7 [*Ictalurus punctatus*]	AAK95131	4.0E−72
425	1	60S ribosomal protein L7A [*D. melanogaster*]	P46223	1.0E−76
446	1	60S ribosomal protein L7/L12 [*Spodoptera frugiperda*]	AAK72378	3.0E−20
464	1	Ribosomal protein L8 [*Spodoptera frugiperda*]	AAL26575	1.0E−125
11	1	60S ribosomal protein L8 [*Aedes albopictus*]	P41569	3.0E−23
499	1	Ribosomal protein L10 [*Bombyx mori*]	AAV34820	9.0E−67
308	1	Ribosomal protein L10a [*Rattus norvegicus*]	NP_112327	4.0E−67
100	1	Ribosomal protein L11 [*D. melanogaster*]	P46222	1.0E−56
405	1	60S ribosomal protein L13 [*Ictalurus punctatus*]	Q90YV5	5.0E−57
171	1	Mitochondrial ribosomal protein L13 [*Mus musculus*]	NP_081035	8.0E−13
422	1	Ribosomal protein L16a [*Emericella nidulans*]	AAD54383	3.0E−49
379	2	ribosomal protein L17 [*Mus musculus*]	XP_128846	1.0E−63
20	4	Ribosomal protein L18 [*Branchiostoma lanceolatum*]	AAN73381	2.0E−74
380	1	Ribosomal protein L18A [*D. melanogaster*]	NP_523774	1.0E−78
504	1	Ribosomal protein L18a [*Aedes albopictus*]	AAV90708	1.0E−29
507	1	Ribosomal protein L18a [*Lysiphlebus testaceipes*]	AAX62415	3.0E−49
387	1	60S ribosomal protein L21 [*Homo sapiens*]	NP_000973	8.0E−64
454	1	Mitochondrial ribosomal protein L21 [*Homo sapiens*]	NP_852615	3.0E−22
41	2	Ribosomal protein L22 [*Caenorhabditis elegans*]	NP_740781	1.0E−36
111	2	60S ribosomal protein L23 [*D. melanogaster*]	P48159	1.0E−21
484	1	Ribosomal protein L23a [*D. melanogaster*]	AAD19340	3.0E−44
388	1	Ribosomal protein L24 [*Danio rerio*]	genbank:NP_775342NP_775342	5.0E−35
429	2	60S ribosomal protein L26 [*Homo sapiens*]	NP_000978	6.0E−41
427	7	Ribosomal protein L27a [*D. melanogaster*]	JC2392	1.0E−66
80	1	Ribosomal protein L29 [*Homo sapiens*]	NP_000983	2.0E−14
12	5	60S ribosomal protein L30 [*Homo sapiens*]	NP_000980.1	1.0E−47
133	3	60S ribosomal protein L34 [*Ochlerotatus triseriatus*]	Q9NB34	5.0E−44
305	2	Ribosomal protein L35 [*Gallus gallus*]	NP_989604	1.0E−28
381	1	Ribosomal protein L35A [*D. melanogaster*]	NP_649539	1.0E−39
87	3	60S ribosomal protein L37 [*Emericella nidulans*]	Q9C0T1	6.0E−22
47	10	60S ribosomal protein L37A [*Cryptochiton stelleri*]	O61462	1.0E−31
7	4	60S ribosomal protein L38 [*Homo sapiens*]	NP_000990	6.0E−23
88	4	60S ribosomal protein L39 [*Ictalurus punctatus*]	Q90YS9	7.0E−18
428	1	60S ribosomal protein L44 [*Ochlerotatus triseriatus*]	Q9NB33	3.0E−43
423	1	Acidic ribosomal protein P0 [*Sarcophaga crassipalpis*]	AAF31449	1.0E−136
393	1	Acidic ribosomal protein P2 [*Spodoptera frugiperda*]	AAL62467	7.0E−19
93	1	Ribosomal protein, acidic [*Caenorhabditis elegans*]	NP_740801	2.0E−25
461	1	Ribosomal protein S2, cytosolic [*D. melanogaster*]	S30395	1.0E−102
6	11	40S ribosomal protein S3a [*Ictalurus punctatus*]	AAK95185.1	9.0E−26
10	15	40S ribosomal protein S4 [*D. melanogaster*]	P41042	7.0E−83
390	1	40S ribosomal protein S5 [*Homo sapiens*]	NP_001000	2.0E−94
431	1	Ribosomal protein S6 [*Anopheles stephensi*]	AAO88054	4.0E−86
508	1	Ribosomal protein S7 [*Culex pipiens*]	Q9NB21	6.0E−84
247	2	40S ribosomal protein S8 [*Platichthys flesus*]	CAC27398	7.0E−19
477	1	Ribosomal protein S9 [*Plutella xylostella*]	BAD26701	1.0E−87
105	1	40S ribosomal protein S10 [*Homo sapiens*]	NP_001005	1.0E−43
304	1	40S ribosomal protein S11 [*Homo sapiens*]	NP_001006	2.0E−19
299	2	40S ribosomal protein S13 [*Homo sapiens*]	NP_001008	4.0E−71
108	1	40S ribosomal protein S15 [*Homo sapiens*]	NP_001009	2.0E−54

Table 2 (*continued*)

Cluster	No.	First hit	Accession	*E*-value
37	2	40S ribosomal protein S15a [*Homo sapiens*]	NP_001010	2.0E−63
58	3	40S ribosomal protein S16 [*Homo sapiens*]	NP_001011	5.0E−35
434	2	Ribosomal protein S18 [*Mus musculus*]	NP_035426	1.0E−55
501	1	Ribosomal protein S19 [*D. melanogaster*]	CAA51677	1.0E−59
417	1	40S ribosomal protein S20 [*Xenopus laevis*]	P23403	3.0E−41
99	3	40S ribosomal protein S21 [*Homo sapiens*]	NP_001015	4.0E−27
66	1	28S ribosomal protein S21 [*Homo sapiens*]	NP_114107	1.0E−14
426	2	40S ribosomal protein S24 [*Homo sapiens*]	NP_001017	3.0E−38
503	1	Ribosomal protein S25 [*Spodoptera frugiperda*]	AAK92193	1.0E−40
424	1	40S ribosomal protein S26-2 [*Ictalurus punctatus*]	AAK95209	9.0E−45
443	1	Ribosomal protein S27a [*Aedes aegypti*]	AAS79344	4.0E−50
159	1	40S ribosomal protein S28 [*Ictalurus punctatus*]	Q90YP3	1.0E−15
4	8	40S ribosomal protein S30 [*Ictalurus punctatus*]	AAK95215.1	6.0E−26
2. Translation factors, tRNA/amino acid synthases				
517	1	Translation initiation factor [*Anopheles gambiae*]	AAA18901	2.0E−53
204	1	Translation initiation factor 2 gamma [*Clytus arietis*]	CAB98197	7.0E−25
445	1	Translation initiation factor 3, subunit 6 [*Danio rerio*]	NP_957133	6.0E−56
415	1	Translation initiation factor eIF3 p40 [*Homo sapiens*]	AAC84044	2.0E−71
404	1	Translation initiation factor 4γ-3 [*Homo sapiens*]	NP_003751	2.0E−27
506	1	Translation initiation factor 4A [*D. melanogaster*]	CAA48790	7.0E−84
433	1	Translation initiation factor 5A [*Spodoptera exigua*]	Q9TVJ8	3.0E−83
398	1	Elongation factor 1-beta [*Bombyx mori*]	P29522	2.0E−76
507	1	Translation elongation factor-1ε-2 [*Xenopus laevis*]	S57631	2.0E−33
463	1	Translation elongation factor-2 [*D. melanogaster*]	S05988	5.0E−91
382	1	Elongation factor SIII p15 subunit [*Aedes aegypti*]	EAT43917	1.0E−55
421	5	Elongation factor-1alpha F2 [*Apis mellifera*]	AAC38959	1.0E−126
467	1	Elongation factor tu [*Schizosaccharomyces pombe*]	NP_595746	6.0E−72
432	1	Ribosomal protein L9 [*D. melanogaster*]	NP_477161	3.0E−87
317	1	Tyrosyl-tRNA synthetase [*Drosophila virilis*]	AAD03560	6.0E−45
327	1	Asparaginyl-tRNA synthetase [*Brugia malayi*]	P10723	1.0E−07
389	1	Seryl-tRNA synthetase 2 [*Bos taurus*]	NP_776882	5.0E−42
482	1	Phenylalanyl tRNA synthetase [*D. melanogaster*]	AAB65750	2.0E−60
520	1	tRNA-nucleotidyltransferase 1 [*Homo sapiens*]	Q96Q11	3.0E−52
455	1	Glutamine synthetase [*D. melanogaster*]	CAA36970	5.0E−62
441	1	Proline synthetase co-transcribed [*Homo sapiens*]	NP_009129	1.0E−42
3. Chaperones, protein folding and secretion				
525	1	Prefoldin subunit 1 [*Homo sapiens*]	NP_002613	1.0E−22
332	1	Endoplasmic translocon component [*Homo sapiens*]	NP_009145	4.0E−28
527	1	Ribophorin I [*Sus scrofa*]	NP_999498	6.0E−52
458	1	Polypeptide associated protein-α [*D. melanogaster*]	CAA70166	8.0E−56
450	1	Calreticulin [*D. melanogaster*]	BAA85379	8.0E−81
403	1	GrpE-like protein cochaperone [*Homo sapiens*]	NP_079472	2.0E−40
532*	1	Dnajb11 protein [*Danio rerio*]	AAH66411	4.0E−45
452	1	Heat-shock protein 60 [*D. melanogaster*]	CAA67720	4.0E−99
242	1	heat shock protein 25 [*Caenorhabditis elegans*]	NP_509009	4.0E−13
448*	1	Heat-shock cognate 70 kDa [*Pimephales promelas*]	AAS46619	1.0E−96
145*	1	Heat-shock protein 70 cognate 3 [*D. melanogaster*]	P29844	9.0E−54
357*	1	Heat shock 70 kDa proteins [*Ceratitis capitata*]	AAC23392	7.0E−52
406*	1	Heat-shock 70 kDa protein 1A [*Bos taurus*]	AAN78092	6.0E−16
474*	1	Hsp70 binding protein [*Homo sapiens*]	NP_036399	6.0E−16
412	1	Clathrin interactor 1 [*Homo sapiens*]	NP_055481	1.0E−62
413*	1	Heat-shock protein 82 [*Drosophila pseudoobscura*]	AAC07919	5.0E−84
63	1	Surf1 [*Gallus gallus*]	BAC65170	1.0E−26
352	1	Vps28-prov protein [*Xenopus laevis*]	AAH56011	2.0E−78
26	1	Protein disulfide isomerase P5 [*Danio rerio*]	AAK71636.1	8.0E−32
53	1	Protein disulfide isomerase [*D. melanogaster*]	S68280	4.0E−24
479	1	Protein disulfide isomerase [*Bombyx mori*]	AAG45936	3.0E−40
4. RNA synthesis and processing				
222	1	DNA-directed RNA polymerase [*Mus musculus*]	Q63871	3.0E−17
129	2	TAF9 RNA polymerase II [*Homo sapiens*]	NP_057367	2.0E−39
350	1	RNA polymerase II 14.4 kDa [*Homo sapiens*]	NP_068809	6.0E−30
375	1	RNA polymerase II 18 kD subunit [*D. melanogaster*]	NP_524910	2.0E−34

Table 2 (*continued*)

Cluster	No.	First hit	Accession	*E*-value
372	1	RNA polymerase II rpb4 [*D. melanogaster*]	AAN88031	1.0E−52
345	1	RNA polymerase III subunit RPC8 [*Homo sapiens*]	NP_612211	2.0E−40
481	1	Splicing factor Prp8 [*Homo sapiens*]	AAC61776	1.0E−116
(B) House keeping				
1. Energy metabolism				
2	3	Monooxygenase [*Nicotiana tabacum*]	T02995	1.0E−41
385	1	Flavin-containing monooxygenase [*D. melanogaster*]	AAK94940	4.0E−51
27	1	Cytochrome oxidase homolog 1 [*Homo sapiens*]	NP_004580.1	3.0E−21
439	1	Cytochrome b [*Acryllium vulturinum*]	AAN10094	4.0E−41
466	1	Cytochrome b5 [*Anopheles gambiae*]	AAO24766	5.0E−42
486	1	Cytochrome c [*Apis mellifera*]	P00038	1.0E−50
411	1	Cytochrome c oxidase subunit 2 [*Apis mellifera*]	AAB96800	2.0E−15
435	1	Cytochrome c oxidase IV [*Drosophila simulans*]	AAP88303	2.0E−39
82	2	Cytochrome c oxidase VIa [*Oncorhynchus mykiss*]	O13085	6.0E−22
376	2	Cytochrome c oxidase VIa [*Cyprinus carpio*]	U83907	3.0E−18
196	1	Cytochrome c oxidase VIIc [*Mus musculus*]	NP_031775	7.0E−10
223	1	Cytochrome c oxidase VIII [*Drosophila mauritiana*]	AAP88325	0.005
377	1	Cytochrome c oxidase VIIIb [*Mus musculus*]	AAH27531	0.004
378	1	Cytochrome c oxidase assembly [*T. castaneum*]	XP_967529	9.0E−11
418	1	ATP synthase F0 subunit 6 [*Locusta migratoria*]	NP_007294	4.0E−29
112	2	ATP synthase, mitochondrial F0, S-E [*Bos taurus*]	NP_788812	4.0E−06
59	1	F1-atpase [*Rattus norvegicus*]	1MAB_A	4.0E−19
370	2	Coilin-interacting nulcear ATPase [*Aedes aegypti*]	EAT33025	3.0E−52
485	1	ATP synthase, gamma subunit [*D. melanogaster*]	CAA73233	8.0E−51
24	1	ATP synthase coupling factor 6 [*Homo sapi*ens]	NP_001676.2	2.0E−13
472	1	Vacuolar ATPase B subunit [*Aedes aegypti*]	AAD27666	1.0E−116
430	1	Vacuolar ATP synthase subunit C [*Aedes aegypti*]	O16110	2.0E−59
369	1	H+ transporting ATP synthase S-E [*Bombyx mori*]	ABF51335	1.0E−12
420	1	Vacuolar ATP synthase subunit F [*Manduca sexta*]	P31478	6.0E−50
383	1	Mitochondrial ATP synthase F chain [*Aedes aegypti*]	ABF18130	1.0E−35
392	1	Vacuolar ATP synthase subunit G [*Manduca sexta*]	Q25532	2.0E−22
234	1	ATP synthase g chain, mitochondrial [*Bos taurus*]	Q28852	3.0E−14
399	1	Mitochondrial F0 complex, subunit c [*Bos taurus*]	NP_788786	2.0E−15
203	1	ATP synthase regulatory factor B [*Mus musculus*]	AAH27442	6.0E−05
151	1	Reticulum-type calcium ATPase [*Heliothis virescens*]	AAD09820	3.0E−45
447	1	ATP synthase lipid-binding protein [*Manduca sexta*]	Q9U505	2.0E−32
471	1	ATP synthase γ chain, mitochondrial [*Bos taurus*]	P05631	3.0E−52
320	1	NADH dehydrogenase 1-α S5 [*Bos taurus*]	NP_787023	9.0E−12
449	1	NADH dehydrogenase S1 [*Ostrinia nubilalis*]	NP_563595	3.0E−25
30	4	NADH dehydrogenase S4 [*Bactrocera oleae*]	NP_957750	2.0E−40
397	1	NADH dehydrogenase B22 subunit [*Homo sapiens*]	NP_004996	5.0E−29
31	1	NADH thyroid oxidase [*Caenorhabditis elegans*]	NP_490684	1.0E−42
391	1	Enolase [*D. melanogaster*]	P15007	6.0E−98
410	2	Triosephosphate isomerase [*Culex tarsalis*]	P30741	1.0E−105
451	1	G3P dehydrogenase [*D. melanogaster*]	P07486	1.0E−91
400	1	Phosphoglycerate dehydrogenase [*Rattus norvegicus*]	NP_113808	1.0E−83
475	1	Phosphoglycerate kinase [*Aedes aegypti*]	AAL58081	1.0E−81
468	1	Pyruvate dehydrogenase E1 Sα [*Ascaris suum*]	P26268	1.0E−65
473	1	Transaldolase [*Homo sapiens*]	A49985	4.0E−58
478	1	Isocitrate dehydrogenase [*Caenorhabditis elegans*]	NP_497927	6.0E−37
543	1	Citrate synthase [*Mus musculus*]	NP_080720	6.0E−67
407	1	ATP citrate lyase [*Mus musculus*]	NP_598798	7.0E−80
492	1	Succinate-ubiquinone oxidoreductase [*Homo sapiens*]	AAA35708	1.0E−48
161	1	Acyl-CoA-binding protein [*Sus scrofa*]	P12026	3.0E−06
310	1	Ornithine decarboxylase antizyme [*D. melanogaster*]	P54361	1.0E−07
528	1	M-semialdehyde dehydrogenase [*D. melanogaster*]	T13418	2.0E−80
529	1	acyl-CoA desaturase [*Helicoverpa assu*lta]	AAM28480	9.0E−38
2. Structural proteins				
9	1	Laminin A chain [*D. melanogaster*]	AAA28661.1	5.0E−14
291	2	Histone 1 [*Drosophila sechellia*]	BAD02447	2.0E−21
18	1	Tropomyosin isoform 9D [*D. melanogaster*]	AAA28969	3.0E−14
79	1	Cytoplasmic light-chain dynein [*Sus scrofa*]	AAL30831	0.007
70	1	Cardiac myosin light chain 2 [*Danio rerio*]	AAF00097	4.0E−14

Table 2 (*continued*)

Cluster	No.	First hit	Accession	*E*-value
444	1	Myosin light (16.0 kD) [*Caenorhabditis elegans*]	NP_499813	6.0E−39
263	1	Actin CA15 [Styela clava]	A61043	7.0E−07
343	1	Actin, muscle [*Manduca sexta*]	P49871	7.0E−74
45	2	Tubulin alpha chain [*Bombyx mori*]	P52273	6.0E−11
248	2	Tubulin alpha chain [*Bombyx mori*]	P52273	7.0E−04
118	2	Nucleolar protein family A, P2 [*Homo sapiens*]	NP_060308	7.0E−24
107	1	snRNP-associated protein; SmB [*Danio rerio*]	AAF72188	5.0E−22
177	1	U6 snRNA-associated protein [*Homo sapiens*]	NP_036454	4.0E−36
255	1	Small nuclear ribonucleoprotein G [*Homo sapiens*]	NP_003087	9.0E−28
364	1	Muscular protein 20 [*Cicindela puritana*]	CAD55779	2.0E−39
394	1	Collagen IV alpha 1 chain [*Anopheles gambiae*]	AF313909	9.0E−166
414	1	Nucleolar protein [*Drosophila subobscura*]	CAB92783	2.0E−69
436	1	Nucleolar protein Nop56 [*Homo sapiens*]	O00567	7.0E−54
497	1	Tropomyosin [*D. melanogaster*]	AAF55163	0.015
3. Protein processing and degradation				
22	1	Cytosolic non-specific dipeptidase [*Mus musculus*]	NP_075638.2	3.0E−25
156*	2	Cysteine proteinase 2 [*Homarus americanus*]	P25782	4.0E−07
359*	2	Cathepsin l [*Aedes aegypti*]	EAT45919	3.0E−14
361*	1	Serine protease K15/F2R3 [*Chrysomya bezziana*]	AAG30233	4.0E−15
384*	1	Zinc carboxypeptidase [*Aedes aegypti*]	EAT44906	1.0E−130
385*	3	Trypsin-related protease [*Anopheles gambiae*]	CAA80514	5.0E−12
396	1	Mitochondrial processing protease [*R. norvegicus*]	BAA03007	3.0E−48
8	4	Microsomal signal peptidase 12 kDa [*Homo sapiens*]	Q9Y6A9	8.0E−13
127	1	26S proteinase regulatory complex [*D. melanogaster*]	S66528	3.0E−23
157	1	Proteasome 26S regulatory S6 [*Danio rerio*]	NP_956585	2.0E−45
235	1	Microsomal signal peptidase 21 kDa [*Canis familiaris*]	P13679	3.0E−31
237	1	Proteosome subunit RPN2 [*D. melanogaster*]	AAG33625	6.0E−36
360	3	Signal peptidase subunit [*Aedes aegypti*]	ABF18229	3.0E−28
401	1	Microsomal signal peptidase [*Canis familiaris*]	Q28250	1.0E−39
408	1	Proteasome chain 1 [*D. melanogaster*]	JQ0681	1.0E−106
470	1	Signal peptidase precursor [*Caenorhabditis elegans*]	NP_498755	1.0E−34
402	1	Ubiquitin-activating enzyme E1 [*Homo sapiens*]	NP_003325	3.0E−96
395	1	Ubiquitin Conjugating enzyme [*C. elegans*]	NP_502065	2.0E−81
286	1	Ubiquinone-binding protein QP-C [*Bos taurus*]	UYBO	3.0E−13
155	1	Ubiquitin thiolesterase [*Homo sapiens*]	Q96FW1	2.0E−23
530	1	Arih1 protein [*Mus musculus*]	AAH57680	1.0E−129
4. Transportation				
14	1	ADP/ATP translocase [*D. melanogaster*]	AAB23114	5.0E−41
97	2	Phosphate carrier protein [*Caenorhabditis elegans*]	NP_502087	1.0E−10
456	1	Phosphate transporter [*D. melanogaster*]	AAD24490	1.0E−114
109	1	Zinc transporter [*Mus musculus*]	Q31125	8.0E−32
265	1	Protein translocation complex β [*Homo sapiens*]	NP_006799	1.0E−18
374	1	Mitochondrial import receptor S20 [*Aedes aegypti*]	EAT40912	1.0E−24
276	1	Solute carrier family 2 [*Ovis aries*]	P47843	2.0E−06
371	2	Copii-coated vesicle protein P24 [*Aedes aegypti*]	EAT48799	4.0E−60
349	1	Coated vesicle membrane protein [*Rattus norvegicus*]	NP_113910	5.0E−52
322	2	Signal recognition particle 9kDa [*Homo sapiens*]	NP_003124	2.0E−10
150	1	Tip associating protein [*D. melanogaster*]	CAB75849	2.0E−04
419	1	DCAPL1 [*D. melanogaster*]	BAB62017	5.0E−34
446	1	UDP-galactose transporter [*Cricetulus griseus*]	AAG53653	2.0E−24
457	1	Secretory carrier protein [*D. melanogaster*]	AAF35367	2.0E−32
459	1	Mitochondrial porin [*D. melanogaster*]	CAA63143	1.0E−102
469	1	Amino acid transporter Ae_AAT1 [*Aedes aegypti*]	AAN40410	3.0E−75
483	1	Aquaporin 5 [*Rattus norvegicus*]	NP_036911	4.0E−28
488	1	Vacuole membrane protein 1 [*Rattus norvegicus*]	NP_620194	8.0E−37
495	1	Solute carrier family 17, member 5 [*Homo sapiens*]	NP_036566	3.0E−19
500	1	Acyl carrier protein [*D. melanogaster*]	Q94519	1.0E−35
505	1	Translocon-associated protein-γ [*Anopheles gambiae*]	EAA10212	7.0E−11
510	1	Innexin Inx2 [*Schistocerca Americana*]	Q9XYN1	1.0E−65
526	1	2-oxoglutarate carrier protein [*Homo sapiens*]	A56650	4.0E−62
534	1	Urea transporter S-A5 [*Anopheles gambiae*]	EAA14036	2.0E−06

Table 2 (*continued*)

Cluster	No.	First hit	Accession	*E*-value
(C) Other categories				
1. Detoxification				
78*	3	Glutathione S-transferase I [*Anopheles dirus*]	AAB41104	3.0E−05
440*	1	Glutathione S-transferase S1-2 [*Anopheles gambiae*]	AAM53611	1.0E−79
409*	1	Glutathione transferase zeta 1 [*Mus musculus*]	NP_034493	7.0E−70
510*	1	Glutathione transferase [*Anopheles gambiae*]	AAP37003	5.0E−09
275	1	Thioredoxin-like protein [*Manduca sexta*]	AAF16695	2.0E−08
290	1	Thioredoxin peroxidase [*Ostertagia ostertagi*]	CAD20737	1.0E−28
367	1	Metallothionein-I activator [*Mus musculus*]	XP_127976	2.0E−17
386	2	Glutaredoxin-related protein [*Leptospira interrogans*]	AAS70397	4.0E−27
476	1	CYP9 cytochrome P450 [*Drosophila mettleri*]	AAC33299	7.0E−31
489*	1	Glutathione peroxidase [*Schistosoma mansoni*]	Q00277	1.0E−43
487	1	Ferredoxin precursor [*Digitalis lanata*]	CAD79348	4.0E−29
496	1	Cytochrome P450 [*Culex pipiens quinquefasciatus*]	BAA92152	7.0E−44
2. Regulation				
524	1	Protein tyrosine phosphatase [*Gallus gallus*]	NP_989645	8.0E−12
465	1	Acid phosphatase [*D. melanogaster*]	S64682	4.0E−43
511	1	Alkaline phosphatase 5 [*Mus musculus*]	NP_031459	5.0E−12
44	2	Protein phosphatase 1 S8 [Bos taurus]	NP_777007	1.0E−52
437	1	Protein kinase kinase kinase kinase [*Homo sapiens*]	NP_663720	1.0E−102
251	1	Mitogen-activated protein-binding [*Homo sapiens*]	NP_054736	7.0E−24
116	1	A-kinase anchor protein [*D. melanogaster*]	T03094	3.0E−05
316	2	Protein kinase raf 1 [*Seriola quinqueradiata*]	BAB18860	2.0E−05
362	1	Protein kinase C inhibitor [*Aedes aegypti*]	EAT39737	7.0E−32
480	1	Casein kinase 1 homolog [*Encephalitozoon cuniculi*]	NP_597600	2.0E−06
35	1	Pin1 protein homolog dodo [*D. melanogaster*]	T08426	2.0E−43
531	1	GTP-binding nuclear protein [*Chironomus tentans*]	CAE55862	1.0E−101
509	1	RAW [*D. melanogaster*]	AAF28462	2.0E−35
3. Others				
265	1	Nucleoside diphosphate kinase [*D. melanogaster*]	P08879	8.0E−70
297	1	Calcium-activated nucleotidase [*Homo sapiens*]	AAM94564	3.0E−35
491	1	RNA-binding protein S1 [*Homo sapiens*]	NP_006702	8.0E−27
442	1	RNA helicase [*D. melanogaster*]	S11485	5.0E−67
368	1	ATP-dependent RNA helicase [*Homo sapiens*]	NP_003162	6.0E−55
53	1	Ribonuclease Z, mitochondrial [*D. melanogaster*]	Q8MKW7	1.0E−04
515	1	Spermidine synthase [*Dictyostelium discoideum*]	Q9XY92	4.0E−68
373	1	Pyrroline-5-carboxylate reductase [*Aedes aegypti*]	EAT41304	3.0E−93
260	1	Trp monooxygenase activatioprotein [*R. norvegicus*]	NP_113791	7.0E−18
92	4	cAMP generating peptide [*Sarcophaga bullata*]	AAB36093	5.0E−08
95	1	Arch protein [*Mus musculus*]	AAH26442	1.0E−22
193	1	Galactosyltransferase 3 beta 1, 4 [*Mus musculus*]	NP_766470	6.0E−39
363	1	Stromal cell-derived factor 2 [*Rattus norvegicus*]	XP_213377	2.0E−44
256	3	Retinoic acid binding protein [*Manduca sexta*]	AAC24317	1.0E−32
312	1	Golgi compartment lectin [*Ixodes scapularis*]	AAQ86834	8.0E−48
326	2	Transposase [*Drosophila simulans*]	AAC16614	2.0E−13
334	1	Cuticle protein [*Bombyx mori*]	BAB32485	3.0E−14
342	1	Tumor protein [*Mus musculus*]	NP_033455	5.0E−15
514	1	lipase 1 [*D. melanogaster*]	O46108	4.0E−16
515	1	Esterase lipase [*Caenorhabditis elegans*]	NP_493077	3.0E−07
292	1	Liprin-α 2 [*Homo sapiens*]	NP_003616	6.0E−26
141	1	Beta-NAC-like protein [*Reticulitermes flavipes*]	AAP33157	3.0E−11
36	2	Ecdysone-induced protein [*D. melanogaster*]	A24254	2.0E−28
187	1	Suppressor of var1 3-like 1 [*Caenorhabditis elegans*]	NP_502088	2.0E−43
512	1	Nuclear pore complex protein [*Homo sapiens*]	NP_065134	9.0E−44
516	1	Chromatin assembly factor 1 [*Gallus gallus*]	NP_990183	1.0E−129
519	1	Peptidyl-prolyl isomerase [*Blattella germanica*]	P54985	3.0E−79
521	1	Nucleoplasmin-like protein-short [*D. melanogaster*]	AAC47294	4.0E−13
522	1	Ribonucleotide reductase M2 [*Homo sapiens*]	NP_001025	3.0Ee−59
533	1	Odorant-binding protein [*Anopheles gambiae*]	AAO12096	1.0E−16

"No." represents the number of cDNA clones in each assembled contig. The symbol '*' indicates a SP-encoding transcript.

initiation/elongation factors or tRNA/amino acid synthases, 21 chaperones such as heat-shock proteins or other proteins involved in protein folding, and seven RNA polymerase/splicing factors. The proteins involved in housekeeping functions included 52 energy-metabolic enzymes such as oxidases, ATP synthases, and dehydrogenases;

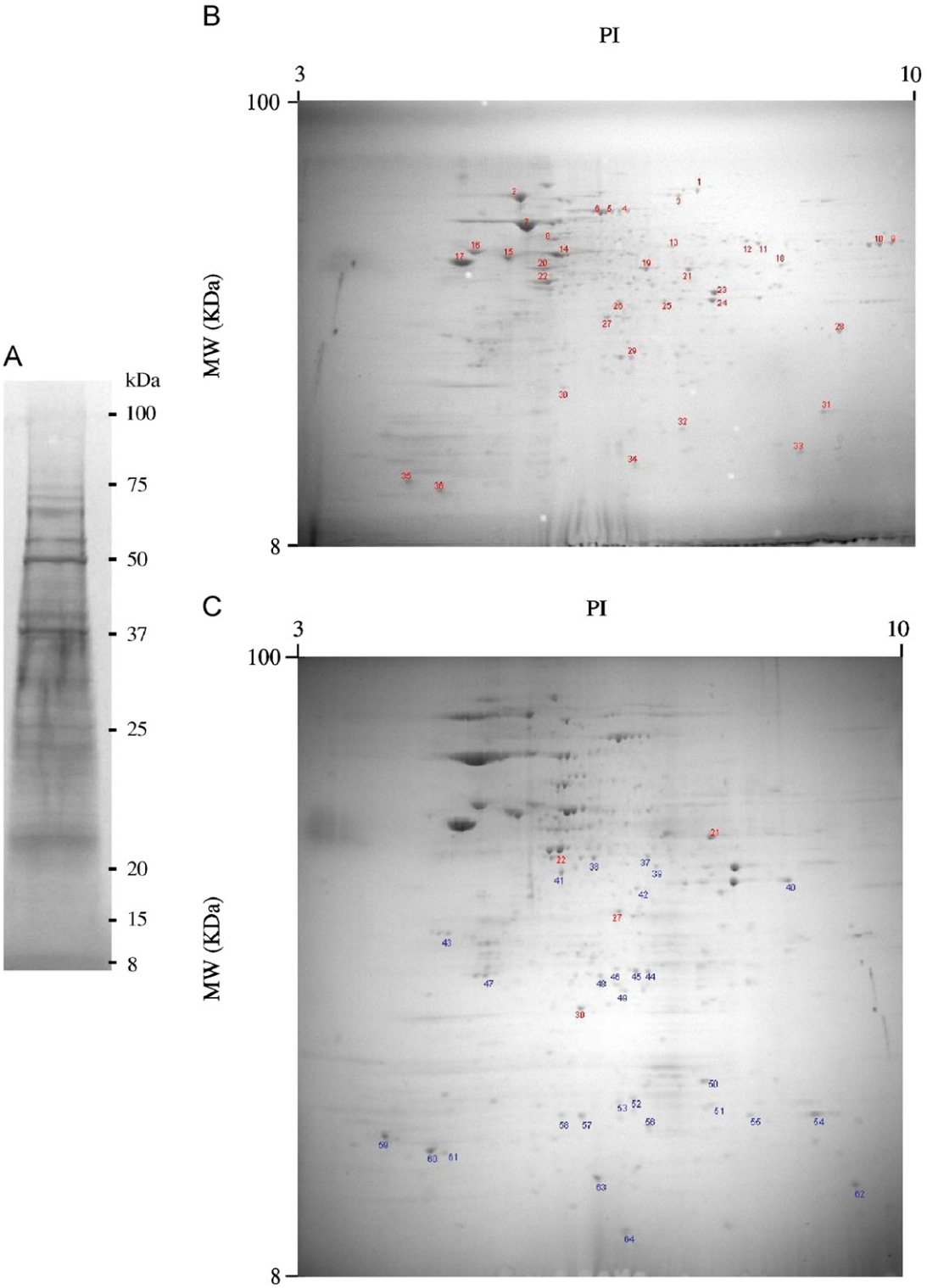

Fig. 2. Analysis of proteins in the salivary glands of Hessian fly larvae by gel electrophoresis: (A) One-dimensional SDS/PAGE analysis. Molecular size markers were indicated on the right side of the gel. The proteins were analyzed on a 4–20% gradient SDS/PAGE gel. (B) Two-dimensional gel analysis with isoelectrofocusing pH 3–10 and SDS/PAGE gradient 8–16%. The protein bands subjected to LC-MS/MS analysis are numbered. (C) Two-dimensional gel analysis with isoelectrofocusing pH 3–10 and SDS/PAGE gradient 10–20%. Bands subjected to LC-MS/MS analysis are numbered. Blue numbers indicate new bands; red numbers indicate bands reanalyzed to serve references between panels B and C.

19 structural proteins including actins, tubulins, and nucleolar proteins; 22 proteases or processing proteasome components; and 24 proteins involved in transport. The proteins with other functions included 12 detoxification enzymes such as glutathione S-transferases, thioredoxins, and a cytochrome P450; 13 proteins involved in regulation (protein phosphatases and protein kinases); and 30 proteins with various other functions.

3.2. Salivary protein composition

The majority of proteins observed on the 1D, 10–20% gradient gel from the salivary glands of 3-day old larvae were between 21 and 75 kDa (Fig. 2A). There were five major bands between 50 and 75 kDa, two major bands between 37 and 40 kDa, several faint bands at approximately 25 kDa, and one major, diffuse band at approximately 21 kDa. On the 2D, 8–16% gradient gel, the majority of proteins were located in the area with pIs from 4.5 to 8 and molecular weight from 25 to 90 kDa (Fig. 2B). The most abundant proteins were located in the area with pIs from 4.5 to 6, and molecular weight between 40 and 90 kDa. On the 2D, 10–20% gradient gel, the separation of lower molecular proteins was much better as expected, but the patterns of the protein distribution were the same as the previous 2D gel (Fig. 2C). Many proteins identified on the 2D gels had the same molecular weight but different pIs. These proteins were likely isozymes (see below).

3.3. Salivary protein identification

To identify proteins, a total of 64 protein bands, 36 from the lower percentage gel (as identified in Fig. 2B) and 28 from the higher percentage gel (as identified in Fig. 2C), were selected for LC-MS/MS analysis. Among these 64 protein bands, 46 were reliably identified (Table 3). Twenty-four of the 46 proteins were related to protein synthesis and folding. These included 13 ribosomal proteins (proteins 27, 35, 36, 50, 52, 53, 56, 57, 58, 59, 60, 61, and 64), four chaperones (proteins 2, 7, 8, 17, and 43), four protein initiation or elongation factors (proteins 3, 21, 39, and 47), and two protein disulfide isomerases (proteins 16 and 20). Eight of the identified proteins were related to energy metabolism, including three ATP synthases (proteins 9, 10, and 14) and five metabolic enzymes (proteins 18, 19, 23, 24, and 38). Five of the identified proteins were related to amino acid metabolism, including three homocysteine methyltransferase isozymes (proteins 4, 5, and 6), one glutamate dehydrogenase (protein 12), and one glutamine synthetase (protein 37). The remaining nine were β-tubulin (protein 15), actin (protein 22), porin (protein 28), glutathione S-transferases (protein 30), nucleotide-diphosphate kinases (proteins 33 and 54), signal sequence receptor-β (protein 44), glutathione peroxidase (protein 46), and 25 kDa proteasome subunit (protein 49). The three most abundant proteins (numbered 2, 7, 17, Fig. 2A) were hsp108, hsc70, and calreticulin, which are heat-shock proteins.

3.4. Quantification of representative transcripts and proteins

Analysis of ESTs derived from Hessian fly salivary glands showed a high percentage of transcripts encoding novel SPs. Yet LC-MS/MS analysis of the 64 most abundant proteins showed that none of these proteins were predicted to be these SPs. The inconsistency between transcript and protein data prompted us to determine the copy numbers of the transcripts encoded by several representative genes using RT-PCR and the abundance of several known proteins. Among the analyzed genes were ribosomal protein P1 (*RP-P1*) and P2 (*RP-P2*), nucleotide-diphosphate kinase (*NDK*), heat-shock protein 70 kD cognate (*Hsc70*), and phosphoglycerate dehydrogenase (*PGDH*). *RP-P1*, *RP-P2*, and *NDK* showed high levels of transcripts, ~23 million copies, as well as abundant proteins (Fig. 3). *Hsc70* and *PGDH*, however, showed low levels of transcripts, but high levels of proteins. Specifically, Hsc70 was the most abundant protein with protein level above 35, yet it contained only 3.6 million transcripts. PGDH had a protein level of ~9, but it contained only 700,000 transcripts.

The other seven analyzed genes encoded unique SPs. These seven genes were selected for transcript analysis because antibodies generated against the corresponding recombinant proteins or synthetic peptides did not detect any proteins either in the insect or in infested plants (data not shown). Because their proteins were not detected, only the copy numbers of their transcripts were determined. The copy number for *10A1* was 2.2 million, the lowest among the SP-encoding genes, but three times as high as that of *PGDH*. The copy numbers for *6A1, 7C1, 12A3,* and *4C1* were 3.4, 3.8, 4.9, and 5.8 million, respectively. These numbers were in the same range as the copy number for *Hsc70*. The copy number for *1C2* was 18.4 million, slightly less than the copy numbers for *RP-P1*, *RP-P2*, or *NDK*. The copy number of *1A2* was 448.8 million, almost 20 times greater than those for *RP-P1*, *RP-P2*, or *NDK*.

4. Discussion

There has been a great deal of interest in insect saliva because it plays a crucial role in host–insect interactions (Miles, 1999). The biggest difficulty to study insect saliva is the lack of ways to obtain sufficient amounts of saliva for various biological and biochemical analyses, since most insects inject saliva only into host tissues, and not into artificial media for collection. As a result, many studies on insect saliva have been carried out indirectly using dissected salivary glands. With rapid advances in genomic and proteomic technologies, the transcriptomes and proteomes of salivary glands from many insects have been analyzed. EST analysis of these tissues combined with protein identification have proven effective in identifying novel

Table 3
List of proteins identified by LC-MS/MS analysis

Spot	HF EST	First Hit Accession	First Hit Protein Name [species]	EST Score/First Hit Score	Peptides matched/ % coverage	Database MW/pI	Observed MW/pI
From the lower percentage gel							
1	No hit	No hit					
2	S17F3	AAK69350	Hsp108 [*Gallus gallus*]	563/49	5/8	91.3/4.86	94.8/5.35
3	S17B3	AAL85604	Elongation factor 2 [*Aedes aegypti*]	1033/58	7/12	94.3/6.18	94.7/7.11
4	No hit	P93263	Homocysteine methyltransferase [*M. crystallinum*][a]	NA/70	7/13	84.8/5.90	84.9/6.72
5	No hit	P93263	Homocysteine methyltransferase [*M. crystallinum*]	NA/78	8/17	84.8/5.90	84.9/6.51
6	No hit	P93263	Homocysteine methyltransferase [*M. crystallinum*]	NA/82	10/23	84.8/5.90	84.9/6.43
7	G15A12	AAA28626	Hsc70 [*Drosophila melanogaster*]	1545/80	7/25	72.3/5.22	71.8/5.47
8	Gg6B12	CAA67720	Hsp60 [*D. melanogaster*]	1379/138	14/27	60.8/5.38	61.2/5.78
9	Gg8E2	CAA69202	ATP synthase [*D. melanogaster*]	894/145	16/32	59.4/9.09	60.0/9.59
10	Gg8E2	AAB59266	Alpha ATP synthase [*Bos taurus*]	788/128	11/19	59.7/9.16	59.8/9.40
11	No hit	No hit					
12	No hit	P54385	Glutamate dehydrogenase [*D. melanogaster*]	NA/60	8/15	62.5/8.44	62.1/8.15
13	No hit	No hit					
14	Gg9A6	AAD27666	Vacuolar ATPase B [*A. aegypti*]	560/79	9/21	55.2/5.38	56.3/5.85
15	Lg2E4	BAB86853	Beta-tubulin [*Bombyx mori*]	495/76	8/15	50.2/4.75	54.2/5.18
16	S14F3	AAG45936	Protein disulfide isomerase [*Bombyx mori*]	1264/113	12/23	55.6/4.60	55.4/4.83
17	G27C11	AAN73309	Calreticulin [*Cotesia rubecula*]	1062/94	11/18	46.6/4.40	50.5/4.81
18	Gg9D2	CAA78404	Phosphoglycerate kinase [*D. melanogaster*]	893/108	13/20	43.9/7.60	44.6/8.46
19	S10G4	AAB87891	Enolase [*D. subobscura*]	863/98	9/16	47.0/5.49	44.2/6.89
20	G8E2	CAI20763	Protein disulfide isomerase [*Danio rerio*]	1659/84	10/13	48.0/5.19	45.1/5.66
21	G6G4	P12261	Elongation factor 1-gamma [*Artemia salina*]	1140/76	8/14	49.3/5.85	43.3/7.27
22	No Hit	AF017427	Actin [*Mayetiola destructor*]	NA/2346	29/51	41.8/5.23	40.0/5.76
23	Lg3D10	AAV84238	Phosphoglycerate dehydrogenase [*C. sonorensis*]	735/100	13/21	36.5/6.67	37.2/7.55
24	Lg3D10	AAV84238	Phosphoglycerate dehydrogenase [*C. sonorensis*]	680/76	9/17	36.5/6.67	34.9/7.51
25	No hit	No hit					
26	No hit	No hit					
27	G27D6	AAF31449	Ribosomal protein P0 [*Sarcophaga crassipalpis*]	1294/175	18/33	33.9/6.87	34.6/6.48
28	S10A3	AAL89811	Porin [*Anopheles gambiae*]	1074/87	7/18	30.6/8.64	29.3/9.03
29	No hit	No hit					
30	G14F1	AAB94639	Glutathione S transferase-1 [*C. variipennis*]	530/97	9/19	24.8/6.30	22.1/5.84
31	No hit	No hit					
32	No hit	No hit					
33	Lg1B5	AAC78437	Nucleoside diphosphate kinase [*Columba livia*]	120/62	5/16	16.9/8.02	15.2/8.66
34	No hit	No hit					
35	G11H5	AAL62466	Ribosomal protein P1 [*Spodoptera frugiperda*]	883/79	7/21	11.4/4.21	11.9/4.15
36	S22F2	AAV84269	Ribosomal protein P2 [*Culicoides sonorensis*]	1058/108	9/18	13.8/4.66	10.2/4.54

Table 3 (*continued*)

Spot	HF EST	First Hit Accession	First Hit Protein Name [species]	EST Score/First Hit Score	Peptides matched/ % coverage	Database MW/pI	Observed MW/pI
From the higher percentage gel							
37	G28A9	CAA36970	Glutamine synthetase [*D. melanogaster*]	670/66	5/9	40.9/5.29	40.6/6.95
38	Gg9H1	AAD55084	Isocitrate dehydrogenase [*Strongyloides stercoralis*]	837/69	7/13	43.1/7.15	41.1/6.47
39	Lg1A4	XP_418401	Translation initiation factor 3 [*Gallus gallus*]	280/54	4/10	35.7/7.73	36.2/7.14
40	No hit	No hit					
41	No hit	No hit					
42	No hit	No hit					
43	Gg6E11	CAA70166	Peptide-associated protein alpha [*D. melanogaster*]	570/75	8/15	23.0/4.62	27.2/4.68
44	G5C7	AAY66839	Signal sequence receptor beta [*Ixodes scapularis*]	694/82	8/21	21.4/6.90	20.1/6.98
45	No hit	No hit					
46	S21F9	BAE07196	Glutathione peroxidase [*Bombyx mori*]	539/63	12/14	22.4/8.72	20.5/6.85
47	Lg3F7	AAS79338	Elongation factor 1 beta [*Aedes aegypti*]	420/60	9/26	24.6/4.57	20.8/6.68
48	No hit	No hit					
49	G5D6	CAA49783	Proteasome, 25 kDa subunit [*D. melanogaster*]	380/55	3/12	25.9/6.21	19.5/6.91
50	G8G11	AAV34876	Ribosomal protein S18 [*B. mori*]	1183/79	11/23	17.8/10.39	16.4/7.74
51	No hit	No hit					
52	Sg6D7	CAA51677	ribosomal protein S19 [*D. melanogaster*]	947/114	14/31	17.3/10.12	15.2/6.98
53	Sg6D7	CAA51677	Ribosomal protein S19 [*D. melanogaster*]	836/96	10/18	17.3/10.12	14.8/6.79
54	Lg1B5	AAC78437	Nucleoside diphosphate kinase [*Columba livia*]	835/68	6/19	16.9/8.02	14.5/9.03
55	G11C6	No hit		450/NA			
56	L1D9	AAL85622	Ribosomal protein L17A [*A. aegypti*]	731/59	4/11	14.9/10.83	14.3/7.05
57	S12A6	CAA61806	Ribosomal protein S12 [*D. melanogaster*]	518/68	5/13	16.6/7.09	14.7/6.31
58	S12A6	CAA61806	Ribosomal protein S12 [*D. melanogaster*]	384/53	8/10	16.6/7.09	14.7/6.13
59	G11H5	AAL62466	Ribosomal protein P1 [*Spodoptera frugiperda*]	1218/158	9/29	11.4/4.21	10.9/4.15
60	S22F2	AAV84269	Ribosomal protein P2 [*C. sonorensis*]	1518/183	12/33	13.8/4.66	10.2/4.53
61	S22F2	AAV84269	Ribosomal protein P2 [*C. sonorensis*]	1160/147	11/28	13.8/4.66	10.2/4.65
62	No hit	No hit					
63	No hit	No hit					
64	G3F5	CAJ01877	Ribosomal protein S27Ae [*Cicindela campestris*]	848/104	10/21	17.8/9.77	9.1/7.01

"HF EST"—first hit of the ESTs in the Hessian fly database. "First Hit Accession"—accession number of the first hit in the non-redundant protein database.

[a]Full name: 5-methyltetrahydropteroyltriglutamate—homocysteine methyltransferase [*Mesembryanthemum crystallinum*].

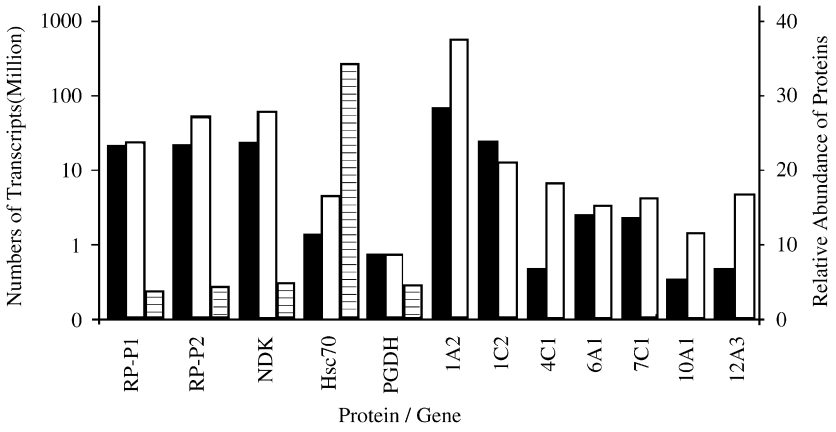

Fig. 3. Copy numbers of transcripts and relative abundance of proteins encoded by specific genes. Black bars represent the copy numbers of a transcript in 1-day old larvae. White bars represent the copy numbers of a transcript in 3-day old larvae. Striped bars represent the relative abundance of the corresponding proteins. RP-P1 and RP-P2 are ribosomal protein P1 and P2. NDK, Hsc70, and PGDH represent nucleotide-diphosphate kinase, heat-shock protein cognate 70 kDa, and phosphoglycerate dehydrogenase, respectively. 1A2, 1C2, 4C1, 6A1, 7C1, 10A1, and 12A3 are seven representative SP-encoding genes.

Table 4
Distribution of transcripts encoding known proteins and the 45 identified proteins

Functional category	Transcript (%)	Protein (%)
Protein synthesis	40	62.3
Ribosomal proteins	23.0	26.7
Chaperones	7.3	15.6
Initiation/elongation factors	7.3	8.9
RNA/amino acid synthesis	2.4	11.1
House keeping	40.8	26.6
Energy metabolism	18.10	17.8
Structural proteins	6.60	4.4
Proteasome components	7.70	2.2
Transportation related	8.40	2.2
Other functions	19.2	11.1
Detoxification	4.2	4.4
Regulation	4.5	
Others	10.5	6.7

genes or proteins that are critical in host–insect interactions (Valenzuela et al., 2003, 2004; Ribeiro et al., 2004). Since Hessian fly larvae need live plants to elicit salivary secretions, there is no existing method to collect saliva directly. Accordingly, we also used dissected salivary glands for transcriptomic and proteomic analyses in order to identify potential genes and proteins that are critical for Hessian fly virulence.

Transcriptomic analyses revealed that nearly 60% of total ESTs from our libraries encoded predicted SPs. The actual percentage of SP-encoding cDNA clones might be even higher since some of the cDNAs in our libraries might be truncated and therefore secretion signal peptides could not be identified. Whether the high percentage of SP-encoding ESTs reflects the same situation of SP-encoding transcripts in the salivary glands of Hessian fly larvae remains to be determined since our library construction might have favored over-representation of small transcripts. In any case, the high percentage of SP-encoding transcripts indicated that the salivary gland tissue of Hessian fly first instar larvae is indeed a specialized structure for producing proteins for secretion. This conclusion was also supported by the fact that the majority of the transcripts encoding known proteins were either directly involved in protein synthesis or involved in house-keeping functions that provide the conditions necessary for protein synthesis (Table 4). Proteomic analysis also revealed that the majority of the 45 identified proteins were either involved in protein synthesis directly (62.3%), or involved in house-keeping functions (26.6%, Table 4).

In addition to the high percentage of SP-transcripts, the complexity of the SP-encoding genes was another intriguing feature of the Hessian fly salivary glands. Most of the 767 unique SPs were members of 160 different gene families, most of which were arranged into highly divergent superfamilies. This is by far one of the most complex salivary gland secretomes for a phytophagous insect (Miles, 1999). The complexity of composition suggested a related complexity in function. Since there were few sequence similarities between the predicted SPs and known proteins, the exact functions of specific SPs are not known. However, Hessian fly and other gall midges manipulate various pathways for plant growth and metabolism (Hori, 1992). Once these SPs are injected into host plants, they likely play roles in regulating wheat growth and metabolism to create a favorable environment for the insect to grow and develop (Zhu et al., 2007).

A high percentage of the SP-encoding cDNA clones was correlated with high levels of transcripts as determined by RT-PCR analysis (Fig. 3). However, this correspondence did not result in a high abundance of corresponding proteins. LC-MS/MS analysis of the 64 most abundant proteins from dissected salivary glands revealed that none of these were SPs. An additional effort to identify SPs by

LC-MS/MS analysis of different regions of 1D SDS/PAGE gels also failed to identify SPs (data not shown). Clearly, there was a discrepancy between the level of transcripts present and the level of proteins retained at the salivary glands. There are two possible explanations for this apparent discrepancy. One explanation is that SPs are secreted upon synthesis, do not accumulate in salivary cells, and therefore are not present in dissected salivary glands. Since Hessian fly larval salivary glands connect directly with wheat tissue via the mandibles (Stuart and Hatchett, 1987), it makes physiological sense to make SPs as needed, and inject them into host tissues as soon as they are synthesized. An alternative explanation would be that the SP transcripts are produced, but not translated into proteins unless they are needed. If this is the case, there must be some kind of communication between the insect and host plants to determine when SP-transcripts are translated. We are currently exploring other fractionation and condensation techniques to accumulate enough proteins with small molecular weights for direct N-terminal sequencing.

It is not clear whether the discrepancy between transcript and protein levels is unique to Hessian fly or a common feature among plant-feeding insects. For blood-feeding insects, the situation is different. Abundant SPs can be obtained from dissected salivary glands of various mosquitoes (Valenzuela et al., 2003; Ribeiro et al., 2004) and the sand fly (Valenzuela et al., 2004). Apparently, insects with different feeding behaviors have different ways to synthesize and secret salivary proteins. Future analyses with other plant-feeding arthropods will reveal whether this is a common trend among them.

Acknowledgments

Mention of commercial or proprietary product does not constitute endorsement by the USDA. The authors thank Drs. Michael Smith and Kun Yan Zhu for reviewing an earlier version of the manuscript. Hessian fly voucher specimens (No. 150) are located in the KSU Museum of Entomological and Prairie Arthropod Research, Kansas State University, Manhattan, Kansas. This work was supported by a NRI grant from the US Department of Agriculture (USDA 04-35607-14861).

References

Byers, R.A., Gallun, R.L., 1971. Ability of the Hessian fly to stunt winter wheat. I. Effect of larval feeding on elongation of leaves. Journal of Economic Entomology 65, 955–958.

Chen, M.S., Fellers, J.P., Stuart, J.J., Reese, J.C., Liu, X., 2004. A group of related cDNAs encoding secreted proteins from Hessian fly [*Mayetiola destructor* (Say)] salivary glands. Insect Molecular Biology 13, 101–108.

Chen, M.S., Fellers, J.P., Zhu, Y.C., Stuart, J.J., Hulbert, S., El-Bouhssini, M., Liu, X., 2006. A super-family of genes coding for secreted salivary gland proteins from the Hessian fly, *Mayetiola destructor*. Journal of Insect Science 6, 12.

Gagne, R.J., Hatchett, J.H., 1989. Instars of the Hessian fly (Diptera: Cecidomyiidae). Annals of the Entomological Society of America 82, 73–79.

Grant, S.R., Fisher, E.J., Chang, J.H., Mole, B.M., Dangl, J.L., 2006. Subterfuge and manipulation: Type III effector proteins of phytopathogenic bacteria. Annual Review of Microbiology 60, 425–449.

Harris, M.O., Stuart, J.J., Mohan, M., Nair, S., Lamb, R.J., Rohfritsch, O., 2003. Grasses and gall midges: plant defense and insect adaptation. Annual Review of Entomology 48, 549–577.

Haseman, L., 1930. The Hessian fly larva and its method of taking food. Journal of Economic Entomology 23, 316–319.

Hatchett, J.H., Starks, K.J., Webster, J.A., 1987. Insect and mite pests of wheat. In: Wheat and Wheat Improvement, Agronomy Mongraph No. 13, pp. 625–675.

Hatchett, J.H., Kreitner, G.L., Elzinga, R.J., 1990. Larval mouthparts and feeding mechanism of the Hessian fly (Diptera: Cecidomyiidae). Annals of the Entomological Society of America 83, 1137–1147.

Hori, K., 1992. Insect secretions and their effect on plant growth, with special reference to Hemipterans. In: Shorthouse, J.D., Rohfritsch, D. (Eds.), Biology of Insect-Induced Galls. Oxford University Press, New York, pp. 157–170.

Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Research 9, 868–877.

Lehane, M.J., 1991. Biology of Blood-sucking Insects. Harper Collins Academic, London, pp. 1–75.

Liu, X.M., Fellers, J.P., Wilde, G.E., Stuart, J.J., Chen, M.S., 2004. Characterization of two genes expressed in the salivary glands of the Hessian fly [*Mayetiola destructor* (Say)]. Insect Biochemistry and Molecular Biology 34, 229–237.

Miles, P.W., 1999. Aphid saliva. Biological Review 74, 41–85.

Pauly, P.J., 2002. Fighting the Hessian fly. Environmental History 7, 385–507.

Ratcliffe, R.H., Hatchett, J.H., 1997. Biology and genetics of the Hessian fly and resistance in wheat. In: Bondari, K. (Ed.), New Developments in Entomology. Research Signpost, Scientific Information Guild, Trivandrum, India, pp. 47–56.

Ratcliffe, R.H., Safranski, G.G., Patterson, F.L., Ohm, H.W., Taylor, P.L., 1994. Biotype status of Hessian fly (Diptera: Cecidomyiidae) populations from the eastern United States and their response to 14 Hessian fly resistance genes. Journal of Economic Entomology 87, 1113–1121.

Ratcliffe, R.H., Cambron, S.E., Flanders, K.L., Bosque-Perez, N.A., Clement, S.L., Ohm, H.W., 2000. Biotype composition of Hessian fly (Diptera: Cecidomyiidae) populations from the southeastern, Midwestern, and northwestern United States and virulence to resistance genes in wheat. Journal of Economic Entomology 93, 1319–1328.

Ribeiro, J.M.C., Charlab, R., Pham, V.M., Garfield, M., Valenzuela, J.G., 2004. An insight into the salivary transcriptome and proteome of the adult female mosquito *Culex pipiens quinquefasciatus*. Insect Biochemistry and Molecular Biology 34, 543–563.

Sardesai, N., Nemacheck, J.A., Subramanyam, S., Williams, C.E., 2005. Identification and mapping of *H32*, a new wheat gene conferring resistance to Hessian fly. Theoretical Applied Geneties 111, 1167–1173.

Stuart, J.J., Hatchett, J.H., 1987. Morphogenesis and cytology of the salivary gland of the Hessian fly, *Mayetiola destructor* (Diptera: Cecidomyiidae). Annals of the Entomological Society of America 80, 475–482.

Tang, X., Frederick, R.D., Zhou, J., Halterman, D.A., Jia, Y., Martin, G.B., 1996. Initiation of plant disease resistance by physical interaction of *Avrpto* and *Pto* kinase. Science 274, 2060–2063.

Valenzuela, J.G., Francischetti, I.M.B., Pham, V.M., Garfield, M.K., Ribeiro, J.M.C., 2003. Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. Insect Biochemistry and Molecular Biology 33, 717–732.

Valenzuela, J.G., Garfield, M., Rowton, E.D., Pham, V.M., 2004. Identification of the most abundant secreted proteins from the salivary glands of the sand fly *Lutzomyia longipalpis*, vector of *Leishmania chagasi*. Journal of Experimental Biology 207, 3717–3729.

Zhu, L., Liu, X.M., Liu, X., Jeannotte, R., Reese, J.C., Harris, M., Stuart, J.J., Chen, M.S., 2007. Hessian fly (*Mayetiola destructor*) attack causes dramatic shift in carbon/nitrogen metabolism in wheat. Molecular Plant–Microbe Interactions. Submitted for publication.