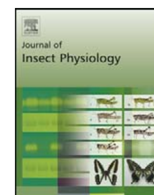
Journal of Insect Physiology 56 (2010) 1198–1206

Contents lists available at ScienceDirect

Journal of Insect Physiology

journal homepage: www.elsevier.com/locate/jinsphys



The gut transcriptome of a gall midge, *Mayetiola destructor*☆

Shize Zhang [a], Richard Shukle [b], Omprakash Mittapalli [c], Yu Cheng Zhu [d], John C. Reese [e], Haiyan Wang [f], Bao-Zhen Hua [a], Ming-Shun Chen [g,*]

[a] Department of Entomology, College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] USDA-ARS, Department of Entomology, Purdue University, West Lafayette, IN 47907, United States
[c] Department of Entomology, The Ohio State University, Ohio Agricultural Research and Development Center, Wooster, OH 44691, United States
[d] USDA-ARS-JWDSRC, PO Box 346, Stoneville, MS 38776, United States
[e] Department of Entomology, Kansas State University, Manhattan, KS 66506, United States
[f] Department of Statistics, Kansas State University, Manhattan, KS 66506, United States
[g] USDA-ARS and Department Entomology, Kansas State University, Manhattan, KS 66506, United States

ARTICLE INFO

Article history:
Received 11 February 2010
Received in revised form 15 March 2010
Accepted 16 March 2010

Keywords:
Hessian fly
Mayetiola destructor
Gut
Secretory proteins
Salivary glands
Transcriptome

ABSTRACT

The Hessian fly, *Mayetiola destructor*, is a serious pest of wheat and an experimental organism for the study of gall midge–plant interactions. In addition to food digestion and detoxification, the gut of Hessian fly larvae is also an important interface for insect–host interactions. Analysis of the genes expressed in the Hessian fly larval gut will enhance our understanding of the overall gut physiology and may also lead to the identification of critical molecules for Hessian fly–host plant interactions. Over 10,000 Expressed Sequence Tags (ESTs) were generated and assembled into 2007 clusters. The most striking feature of the Hessian fly larval transcriptome is the existence of a large number of transcripts coding for so-called small secretory proteins (SSP) with amino acids less than 250. Eleven of the 30 largest clusters were SSP transcripts with the largest cluster containing 11.3% of total ESTs. Transcripts coding for diverse digestive enzymes and detoxification proteins were also identified. Putative digestive enzymes included trypsins, chymotrypsins, cysteine proteases, aspartic protease, endo-oligopeptidase, aminopeptidases, carboxypeptidases, and α-amylases. Putative detoxification proteins included cytochrome P450s, glutathione S-transferases, peroxidases, ferritins, a catalase, peroxiredoxins, and others. This study represents the first global analysis of gut transcripts from a gall midge. The identification of a large number of transcripts coding for SSPs, digestive enzymes, detoxification proteins in the Hessian fly larval gut provides a foundation for future studies on the functions of these genes.

Published by Elsevier Ltd.

1. Introduction

The insect gut is involved in various physiological and biological processes including food digestion, detoxification, interactions with hosts and/or symbiotic microbes, and developmental regulations (Nation, 2002). Different insect species live in different ecological environments and ingest different types of food. Each species, therefore, has evolved a unique set of genes expressed in the gut to meet specific challenges. Analysis of specific gut transcriptomes will contribute to knowledge of the molecular components in the gut of individual insect species and may also identify molecules that have the potential for practical applications (Hughes and Vogler, 2006).

Studies on genes expressed in the insect gut were initially focused on characterization of individual genes, particularly those involved in digestion and detoxification. Recently, more global approaches were adopted to characterize the entire set of genes expressed in the insect gut. Gut transcriptomes of numerous insect species using these approaches were analyzed including blood-sucking disease vectors *Lutzomyia longipalpis* (Dillon et al., 2006; Jochim et al., 2008) and *Phlebotomus papatasi* (Ramalho-Ortigao et al., 2007), and plant-feeding insects such as the cowpea weevil *Callosbruchus maculates* (Pedra et al., 2003; Chi et al., 2009), European corn borer *Ostrinia nubilalis* (Goates et al., 2008; Khajuria et al., 2009), light brown apple moth *Epiphyas postvittana* (Simpson et al., 2007), green peach aphid *Myzus persicae* (Ramsey et al., 2007), pea aphid *Acyrthosiphon pisum* (Sabater-Munoz et al., 2006), and Japanese rotten-wood termite *Hodotermopsis sjostedti* (Yuki et al., 2008). In addition, genomes of several insect species have been sequenced, and the genes expressed in the guts of these insects have been analyzed with microarrays (Li et al., 2008; Oviedo et al., 2008).

So far, no gut transcriptome has been analyzed from any gall midges (Cecidomyiidae: Diptera). Unlike mobile insects, gall

☆ GenBank accession numbers for singleton ESTs are GR305974–GR307142. GenBank accession numbers for contig ESTs are GR557681–GR564524. GenBank accession numbers for contigs are EZ406257–EZ407128. NCBI deposition numbers (GEO accession) for microarray data are GSE18412, GSE18413, and GSE18414.

* Corresponding author. Tel.: +1 785 532 4719; fax: +1 785 532 6232.
E-mail addresses: ming-shun.chen@ars.usda.gov, mchen@ksu.edu (M.-S. Chen).

0022-1910/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.jinsphys.2010.03.021

midges live within plant tissues at a fixed feeding site, creating abnormal plant growths called galls (Rohfritsch, 1992, 2005). A few examples include willow tree gall midge *Dasineura marginemtorquens* (Hoglund et al., 2005), Asian rice gall midge *Orseolia oryzae* (Bentur and Kalode, 1996), orange wheat blossom midge *Sitodiplosis mosellana*, and Hessian fly, *Mayetiola destructor* (Hatchett et al., 1987). Analysis of gall midge gut transcriptomes will not only enhance our understanding of unique features of gut physiology and biochemistry in these types of insects, but may also identify specific targets that have the potential for pest management.

Many gall midges are important agricultural pests and some possess exceptional physiological traits. Among galling insects, Hessian fly is rapidly becoming an experimental organism to study insect–plant interactions because of its intriguing behavior, ease of maintenance in culture and relatively well-characterized genetics for a plant-feeding dipteran (Harris et al., 2003; Stuart et al., 2008). On the economic front, it is one of the most destructive pests of wheat worldwide (Hatchett et al., 1987). Hessian fly larvae live within wheat plants as a parasite, killing infested seedlings or causing serious yield reduction in mature plants. The most effective measure for controlling this insect pest is the development and deployment of resistant wheat cultivars (Ratcliffe and Hatchett, 1997). All resistance genes so far identified and deployed in wheat are dominant major genes. The challenge for this major-gene approach is the development of virulent biotypes that can overcome resistance in wheat over a short period of time (6–8 years). Since the gut is one of the critical interfaces for interactions between insects and plants (Terra and Ferreira, 1994; Herrero et al., 2001), studies on genes expressed in the gut may help to understand the toxicity mechanisms of resistant wheat on Hessian fly larvae, which may eventually lead to improved durability of host plant resistance. In addition, characterization of gut genes may also identify targets for alternative approaches, such as transgenes with RNA interference that targets critical gut genes (Baum et al., 2007; Mao et al., 2007).

Several genes expressed in the Hessian fly larval gut have been characterized, including trypsins, chymotrypsins, carboxypetidases, cytochrome P450s, glutathione S-transferases (GST), protease inhibitors, and several small secretory gut proteins (Zhu et al., 2005; Chen et al., 2006; Liu et al., 2006; Maddur et al., 2006; Mittapalli et al., 2007). A global analysis of the gut transcriptome of Hessian fly larvae should provide comprehensive information on the physiological and biochemical processes in this intriguing insect species. In this study, we have generated more than 10,000 Expressed Sequence Tags (ESTs) from two cDNA libraries made from the gut tissue of Hessian fly larvae. A total of 2007 clusters (contigs and singletons) were obtained from these ESTs. Microarray were applied for the analysis of the abundance of representative transcripts in different tissues.

2. Materials and methods

2.1. Insects and gut preparation

Biotype L was isolated from a greenhouse population of biotype N by both progeny and single egg selection (Sosa, 1978). Biotype GP was derived from a field colony collected from Kansas (Harris and Rose, 1989). The insect populations have been maintained on susceptible wheat seedlings ('Newton' or 'Karl 92′) in environmental chambers at 20 °C and 12:12 (L:D) photoperiod.

Gut tissue was obtained by dissecting 3-day old, 1st instar larvae in DEPC-treated distilled water. Salivary glands and Malphigian tubules were carefully removed. The gut was then punctured to let out content in the alimentary canal. The gut tissue was washed in a large volume of DEPC water to further remove alimentary content. The clean tissue was then transferred into TRI reagent™ (Molecular Research, Inc., Cincinnati, OH) for RNA isolation.

2.2. RNA isolation and cDNA library construction

Total RNA was isolated from 200 guts for each library with TRI reagent™ following the protocol provided by the manufacturer. The cDNA libraries were constructed with a 'SMART™' library construction kit from Clontech (Palo Alto, CA). Plasmid libraries were made following the procedure provided by the manufacturer with one modification: cDNA inserts were ligated into the pPCR-XL-TOPO plasmid included in the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) instead of the provided phage vector. For sequencing, Plasmid DNA was then isolated using a Qiagen BioRobot 3000 and sequenced in an ABI 3700 DNA analyzer. The clones were sequenced using either the M13 forward or reverse primer, but not both. For a small percentage of clones, the cDNA samples were sequenced a second time using a different primer either from the same end (T3 or T7 primers) or from the opposite end if the sequences from the first round were not good enough for comparison with database sequences. For those clones sequenced twice, only one sequence for each clone was used for assembly.

2.3. Sequence analyses

DNA sequences were preprocessed using EGassembler (Masoundi-Nejad et al., 2006) using default parameters. Briefly, the Sequence Cleaning Process was followed to trim the vector/adaptor and mitochondrial sequences from the ESTs with default settings ($\geq$96% identity with $\geq$11 bp alignment) against the EGvec vector library and the NCBI metazoan mitochondrial database. ORFs of the assembled clusters were identified using the ORF predictor software (http://www.bioinformatics.org/sms/orf_find.html). Gene Ontology (GO) annotation was derived using Blast2GO software (http://www.blast2go.de/). For blast searching, a minimum of 9.0e−3 *E*-value was used as a significant similarity.

Analysis of secretion signal peptides was carried out with 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/).

ClustalW (gap weight penalty = 3, gap length weight penalty = 0.2) of the MegAlign module (DNAStar, Ver. 8.02) was used to conduct multiple sequence alignments. Phylogenetic analysis was performed using PAUP* 4.0b10 and the maximum parsimony method was applied in the analysis (Swofford, 1999). Bootstrapping (1000 replicates) was performed to generate a phylogenetic tree. The Heuristic tree search method was performed in maximum parsimony analysis. In Heuristic search, all characters were set as unordered and equally weighted.

2.4. Microarray analysis

For tissue-specific expression, three biological replicates were included for gut and salivary glands, and two replicates were included for Malpighian tubules and carcass. For each biological replicate, 200 dissected tissues from 3-day-old larvae were used for preparation of RNA samples. Since 10–12 probe sequences were included for each probe set in the microarray, these individual probes can be taken as technical duplicates. Therefore, no additional duplicates were carried out. The quantity of RNA was measured using a nanodrop spectrophotometer while the quality was checked using a Bioanalyzer (Agilent, Foster City, CA).

A customized oligo-based microarray was designed and manufactured by Affymetrix (Santa Clara, CA). The microarray

contains two parts of probes: probes (~95%) that are targeted to transcripts isolated from the salivary glands of Hessian fly larvae and probes (the remaining ~5%) that are targeted to gut transcripts. Microarray processing and data analysis were carried out in the Integrated Gene Expression Facility at Kansas State University following the same procedure as described previously (Liu et al., 2007).

For hybridization, 50 ng of total RNA was converted to antisense cDNA using an Ovation RNA Amplification System V2 kit (NuGEN technologies, San Carlos, CA). The single-stranded cDNA was then purified using a Minelute PCR purification kit (Qiagen, Valencia, CA) and quantified using the nanodrop spectrophotometer. The purified cDNA (3.75 μg) was fragmented and labeled using a FL-Ovation cDNA Biotin module V2 kit (NuGEN technologies, Inc.). The labeled probe was checked by running the fragmented cDNA through a RNA nano-chip with the Agilent Bioanalyzer. Hybridization mixture was prepared according to the protocol included in the FL-Ovation cDNA Biotin module V2 kit. The hybridization mixture was then injected in the Hessian fly arrays. After 18 h incubation in a genchip oven, arrays were washed following standard protocol (www.affymetrix.com/support/technical/manual/expression_manual.affx) and stained with streptavidin phycoerythrin (SAPE) in a genechip fluidic station 450. Hybridization quality was verified by scaling factor, overall hybridization rate, and signal strength of several bacterial spike controls. The spike controls were hybridized with labeled targets in different concentrations resulting particular ratios between different spikes.

The arrays were then scanned with the genechip scanner 3000-7G. The Genechip operating software (GCOS) version 1.4 was used for generating initial image (dat) and scaled image (cel) files. The presence/absence detection call for each probe set was performed through Wilcoxon signed rank-based test using R function *mas5calls* from an *Affy* package (Liu et al., 2001, 2002). Signal intensity for each probe set was modeled as weighted combinations of individual probes with MAS5 expression summary algorithm implemented in the R *affy* package (Pepper et al., 2007). This algorithm combines the signals from the multiple Perfect-Match (PM) and Mismatch (MM) probes to produce a value of the hybridization signal for each probe set. Parameters for models were determined through model-fitting techniques taking consideration of all available chip data.

For tissue analysis, relative abundance of a transcript in a specific tissue was calculated by taking the ratio (percentage) of the signal intensity in that tissue over the sum of signal intensity from all tissues (gut, salivary glands, Malphigian tubules, and the leftover referred as carcass). Standard error was calculated as square root of $p(1-p)/T$, where T is the sum of signal intensity. Goodman's test for equality was conducted for pairwise comparison and grouping (Goodman, 1965). The overall family-wise type I error was controlled at 0.01 level. Due to multiple probe sets with multiple tissues, significance of each test was adjusted by Bonferroni correction, namely, two values were considered different when the P-value from Goodman's test was less than $0.01/(6 \times 48) = 3.47 \times 10^{-5}$, where 6 is the number of pairwise comparisons for each probe set and 48 is the number of probe sets considered in the analysis.

3. Results

Two cDNA libraries were constructed, one from biotype L, the most virulent population isolated so far (Sosa, 1978), and the other from biotype GP, the most avirulent population (Harris and Rose, 1989). These two biotypes are widely used for studies in the Hessian fly community and therefore were selected for cDNA library construction. Since both biotypes consisted of mixed genotypes, differences in cDNAs could be due to different genes and/or different alleles. Gut tissue from first instar Hessian fly larvae were dissected for library construction because the first instar represents the most active feeding stage and is critical in establishing a sustained feeding site that determines growth and development on the host plant.

A total of 10,051 clones were sequenced (Table 1). Excluding sequences with poor quality, 8281 high quality ESTs were obtained. The ESTs were assembled into 2007 unique clusters, with 873 contigs and 1134 singletons. The average length of the clusters is 800 bp. Open reading frame (ORF) analysis revealed only a small fraction (4.9%) of the clusters had no ORF. The sizes of clusters that lacked an ORF were under 500 bp, and therefore most likely represented sequences from the 3′-noncoding regions of incomplete transcripts. The majority of the clusters (77.4%) had ORFs fewer than 250 amino acids. The remainder of the clusters (17.7%) had ORFs equal to or more than 250 amino acids.

Among the clusters with similarity to GenBank sequences, 64.6% had similarity to functionally known proteins, and the remaining 35.4% shared similarity to unknown or hypothetical proteins. The clusters with similarity to known proteins are grouped into nine categories according to their first hits. They are transcripts encoding digestive enzymes, small secretory proteins (SSPs) with amino acids less than 250, detoxification proteins, proteins involved in protein synthesis and folding, metabolic enzymes, structural proteins, regulators, and others (Fig. 1). Of these clusters, 4.2% coded for digestive enzymes, 10.7% for SSPs, 3.6% for detoxification enzymes/proteins, 24.4% for proteins involved in protein synthesis and folding, 21.2% for proteins involved in metabolism, 5.9% for structural proteins, 15.1% for regulators, 7.0% for transporters, and 7.8% for proteins with other functions. In terms of ESTs contained in these clusters, 3.7% coded for digestive enzymes, 3.6% for detoxification enzymes/proteins, 5.0% for regulators, 44.0% for SSPs, 16.4% for proteins involved in protein synthesis and folding, 18.6% for proteins involved in metabolism, 2.2% for structural proteins, 2.1% for transporters, and 4.5% for proteins with other functions.

Table 1
Summary of the 10,051 ESTs derived from the Hessian fly gut tissue.

Library	Number of ESTs	Chromatograph quality		Sequence quality		Average length (bp)	Contig[a]	Singletons
		Good	Poor[b]	Good	Poor[c]			
Biotype GP[d]	2727	2193	534	2043	150	643		
Biotype L	7324	6088	1236	6077	11	581		
Total	10051	8281	1770	8120	161	597	873	1134

[a] The numbers of contigs and singletons were based on the analysis of all the ESTs sequenced from the two libraries.
[b] Chromatographs with peak heights varied greater than 3-fold were defined as poor quality.
[c] Sequences with less than 100 bp were defined as poor quality. Poor quality sequences were not included in the analysis.
[d] The library from biotype GP was made at Kansas State University whereas the library from biotype L was made at Purdue University following the same protocol.

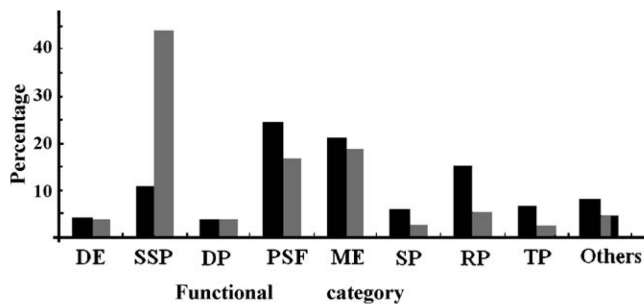


Fig. 1. Putative functions of clusters according to their similarity to proteins in GenBank. A black bar represents the percentage of the number of clusters in a functional category against the total clusters contained in the nine categories. A grey bar represents the percentage of the total number of ESTs in a functional category against the total number of ESTs contained in the nine categories. DE, SSP, DP, PSF, ME, SP, RP, TP represent digestive enzymes, small secretory proteins, detoxification proteins, protein involved in protein synthesis and folding, metabolic enzymes, structural proteins, regulatory proteins, transport proteins, respectively.

3.1. Digestive enzymes

Clusters coding for eight different types of digestive enzymes were identified (Table 2). These transcripts included serine proteinases (trypsins and chymotrypsins), cysteine proteases, aspartic protease, various peptidases (ento-, amino-, and carboxy-peptidases), and α-amylases.

Transcripts coding for major proteinases were trypsins and chymotrypsins. Among them, 19 (including two previously identified) different transcripts coding for trypsins and chymotrypsins were divided into 14 groups: six trypsin groups (MDtryp1 to MDtryp6) and eight chymotrypsin groups (MDchym1 to MDchym8) (Fig. S1A). Transcripts from different groups were likely derived from different genes since amino acid sequence identity among different groups were less than 80%. Transcripts within each group were named A, B, and etc. (for example, MDtryp3A, MDtryp3B, and MDtryp3C are three members in the third trypsin group). Transcripts from the same group coded for similar proteins with sequence identity greater than 90%, and were

Table 2
Clusters coding for putative digestive enzymes.

Cluster ID	No. of ESTs	Length (bp)	First hit	*E*-Value	Organism
Trypsin-like					
Contig46	17	1048	AAT66249	10e−147	*Mayetiola destructor*
Contig49	13	1088	AAT66248	1e−143	*Mayetiola destructor*
Contig465	4	1269	AAN74999	2e−41	*Ochlerotatus epactius*
Contig695	31	1138	AAT81428	4e−145	*Mayetiola destructor*
Contig709	21	1061	AAT66247	6e−93	*Mayetiola destructor*
HFMidgut11B24	1	592	AAT81428	2e−12	*Mayetiola destructor*
HFMidgut11P13	1	785	AAD31269	1e−34	*Rhyzopertha dominica*
HFMidgut17J18	1	441	AAT81427	5e−63	*Mayetiola destructor*
Gg24F7	1	1010	AAG33251	6e−10	*Drosophila. melanogaster*
Gg3H2	1	966	AAG33251	2e−11	*Anopheles gambiae*
Chymotrypsin-like					
Contig87	8	1021	AAT66250	2e−151	*Mayetiola destructor*
Contig329	4	997	AAT66251	1e−75	*Mayetiola destructor*
Contig507	4	932	AAT66251	1e−42	*Mayetiola destructor*
Contig551	21	1063	AAT66244	6e−154	*Mayetiola destructor*
Contig552	2	952	AAT66244	1e−98	*Mayetiola destructor*
Contig726	13	897	AAT66244	2e−64	*Mayetiola destructor*
Gg1H9	1	1168	AAT66243	2e−153	*Mayetiola destructor*
Cysteine protease					
HFMidgut21L13	1	981	XP_001842337	4e−95	*Culex quinquefasciatus*
HFMidgut14G01	1	650	AAY46196	0.001	*Globodera pallida*
Aspartic protease					
Contig333	2	1507	ABA29651	0	*Culex quinquefasciatus*
Endo-Oligopeptidase					
Contig513	2	770	XP_001861876	2e−18	*Culex quinquefasciatus*
Aminopeptidase					
Contig735	4	1833	XP_001656430	5e−137	*Aedes aegypti*
HFMidgut10P07	1	811	XP_001651479	7e−50	*Aedes aegypti*
HFMidgut14O04	1	545	XP_001866008	1e−31	*Culex quinquefasciatus*
Carboxypeptidase					
Contig166	23	1507	ABA29650	5e−130	*Mayetiola destructor*
Gg3E8	1	1539	ABA29654	0.0	*Mayetiola destructor*
Gg5C3	1	1517	ABA29648	0.0	*Mayetiola destructor*
Gg7C8	1	1514	ABA29651	0.0	*Mayetiola destructor*
Gg8D3	1	1503	ABA29649	0.0	*Mayetiola destructor*
Gg11E10	1	1514	ABA29652	0.0	*Mayetiola destructor*
Lg2FD09	1	891	ABA29656	1e−70	*Mayetiola destructor*
Lg4FA06	1	893	ABA29653	5e−103	*Mayetiola destructor*
Midgut08FE11	1	829	ABA29655	3e−108	*Mayetiola destructor*
HFMidgut4B12	1	208	XP_971451	0.003	*Tribolium castaneum*
Amylase (Fig. S3)					
Contig241	2	785	XP_001660907	3e−22	*Aedes aegypti*
HFMidgut18N24	1	528	XP_001847527	3e−5	*Culex quinquefasciatus*

Underline indicates potential novel sequences, which were defined as those with at least 2% differences with any previously characterized Hessian fly sequences at the amino acid level (differences at the 5′- and 3′-unreliable regions of a cluster were excluded).

likely derived from different alleles of the same gene or from similar genes that were duplicated recently. Sequence alignments of predicted full-length trypsins and chymotrypsins revealed that all critical residues and consensus including the active site and serine protease specificity determinant residues are present in the predicted proteins (Fig. S1B), indicating that these proteins are likely active once they are synthesized and secreted into the gut canal.

3.2. Detoxification enzymes

Living within plant tissues with a fixed feeding site, Hessian fly larvae must cope with all types of plant defense molecules. Reactive oxygen species (ROS) and secondary metabolites are common defense molecules produced in plants in response to herbivory (Lamb and Dixon, 1997). A large number of clusters coding for a range of detoxification molecules were identified, including 13 clusters coding for cytochrome P450 (Fig. S2), three for glutathione S-transferases (GST), three for peroxidases, three for ferritins (Fig. S4), two for superoxide dismutases, two for peroxiredoxins, one for catalase, one for adrenodoxin, one for glutaredoxin, and one for glutathione synthetase (Table 3).

3.3. Small secretory proteins (SSPs)

One of the unique characteristics of the Hessian fly larval gut transcriptome in comparison with those from other insects is the existence of a large number of SSPs. One hundred and eleven clusters coded for secretory proteins with amino acids less than 250. Among the 111 clusters, 22 coded for proteins with similarity to functionally known proteins (Table 4). Ten of the 22 clusters coded for protease inhibitor-like proteins (Fig. S5), two for salivary secreted ribonucleases (Rampias et al., 2003), and the remaining 10 for proteins with various functions. Among the remaining 89 clusters, 26 coded for proteins with similarity to unknown or hypothetical proteins from other insects (Fig. S6), and 63 clusters coded for proteins without sequence similarity to any known proteins (Fig. S7).

The tissue-specific expression of SSPs was examined using a customized Hessian fly microarray developed through a commercial contract with Affymetrix (Santa Clara, CA). The microarray was designed to study expression and mutation of genes expressed in Hessian fly larval salivary glands, however, 37 clusters of gut transcripts were also included in this microarray. Tissue-specific distribution of the transcripts corresponding to the 37 clusters is

Table 3
Clusters coding for proteins potentially involved in detoxification.

Cluster ID	No. of ESTs	Length (bp)	First hit	*E*-Value	Organism
Cytochrome p450					
Contig281	3	1696	XP_001649311	2e−138	*Aedes aegypti*
Contig319	6	1566	XP_001869138	5e−103	*Culex quinquefasciatus*
Contig512	6	1129	XP_001869138	9e−50	*Culex quinquefasciatus*
Contig638	10	1720	AAX35340	1e−131	*Mayetiola destructor*
Contig683	5	1577	AAX35341	1e−113	*Mayetiola destructor*
Contig765	15	1158	XP_001649108	2e−66	*Aedes aegypti*
Contig859	5	1610	XP_001652217	5e−88	*Aedes aegypti*
HFMidgut10C01	1	559	XP_563963	1e−33	*Aedes aegypti*
HFMidgut11P15	1	340	XP_001855204	7e−20	*Culex quinquefasciatus*
HFMidgut12A06	1	381	XP_001652218	5e−21	*Aedes aegypti*
HFMidgut19C19	1	995	AAX35340	2e−54	*Mayetiola destructor*
HFMidgut2P17	1	945	XP_001652224	8e−36	*Aedes aegypti*
HFMidgut4L06	1	484	XP_001867632	2e−22	*Aedes aegypti*
Glutathione S-Transferase					
Contig783	9	1156	ABG56084	1e−119	*Mayetiola destructor*
Contig797	4	1141	ABG56083	8e−102	*Mayetiola destructor*
HFMidgut8G21	1	748	XP_001658060	7e−33	*Aedes aegypti*
Peroxidase					
Contig234	5	1019	ABD83336	2e−92	*Mayetiola destructor*
Contig350	3	759	ABD83337	6e−85	*Mayetiola destructor*
HFMidgut6L10	1	255	XP_001843438	0.007	*Culex quinquefasciatus*
Ferritin					
Contig235	53	1457	AAP57194	2e−61	*Drosophila ananassae*
Contig294	35	1174	ABV44741	2e−53	*Phlebotomus papatasi*
HFMidgut10E11	1	1044	NP_733361	3e−32	*Drosophila melanogaster*
Catalase					
Contig833	24	2008	ABL09376	1e−180	*Anopheles gambiae*
Superoxide dismutase					
HFMidgut3H18	1	857	ABE28533	9e−86	*Mayetiola destructor*
HFMidgut6F01	1	604	XP_001866335	9e−60	*Culex quinquefasciatus*
Peroxiredoxin					
Contig796	2	924	XP_001663718	5e−85	*Aedes aegypti*
HFMidgut13C03	1	806	XP_001658149	3e−69	*Phlebotomus papatasi*
Adrenodoxin					
Contig222	2	687	XP_001659837	6.00e−61	*Aedes aegypti*
Glutaredoxin					
HFMidgut21C21	1	629	XP_309539	1e−31	*Anopheles gambiae*
Glutathione synthetase					
HFMidgut6N02	1	970	XP_001653706	1e−114	*Aedes aegypti*

Underline indicates potential novel sequences, which were defined as those with at least 2% differences with any previously characterized Hessian fly sequences at the amino acid level (differences at the 5′- and 3′-unreliable regions of a cluster were excluded).

Table 4
Clusters coding for SSPs with similarity to Genbank sequences.

Cluster ID	No. of ESTs	Size (bp)	First hit	*E*-Value	Putative function [Organism]
Contig43	5	496	ABB70541	4e−56	Protease inhibitor Lg2F7 [*Mayetiola destructor*]
Contig51	25	501	ABB70525	2e−58	Protease inhibitor G14A4 [*Mayetiola destructor*]
Contig55	14	487	ABB70534	2e−51	Protease inhibitor Lg2A3 [*Mayetiola destructor*]
Contig637	7	478	ABB70519	6e−57	Protease inhibitor L5H2 [*Mayetiola destructor*]
Contig857	11	591	ABV60319	2e−08	Serine protease inhibitor [*Lutzomyia longipalpis*]
midgut5RE07	1	921	ABC25079	5e−33	Serine protease inhibitor [*Glossina morsitans*]
Contig345	2	952	XP_001841753	8e−26	Serpin-4 [*Drosophila willistoni*]
HFMidgut12M20	1	435	XP_001865070	2e−15	Serpin-4 [*Culex quinquefasciatus*]
Contig674	3	1104	NP_001106745	2e−13	Carboxypeptidase inhibitor [*Bombyx mori*]
Contig693	3	895	NP_001106745	2e−13	Carboxypeptidase inhibitor [*Bombyx mori*]
Contig432	2	992	XP_001841753	3e−37	Salivary secreted ribonuclease [*C. quinquefasciatus*]
HFMidgut2O22	1	885	XP_001841753	1e−34	Salivary secreted ribonuclease [*C. quinquefasciatus*]
Contig511	10	416	AAY82237	1e−45	Defensin I [*Mayetiola destructor*]
Contig832	2	587	XP_001983268	2e−11	Peptidase m23b [*Drosophila grimshawi*]
HFMidgut19C18	1	753	XP_001851278	3e−54	Cornichon protein [*C. quinquefasciatus*]
Contig279	2	443	ABG21230	9e−51	Diptericin [*Mayetiola destructor*]
Contig325	2	869	XP_002061003	7e−64	Mesoderm development candidate 2 [*D. willistoni*]
Contig57	7	989	XP_001867883	8e−30	Odorant-binding protein 99a [*C. quinquefasciatus*]
Contig510	8	651	XP_001656586	2e−22	Pupal cuticle protein 78E, putative [*Aedes aegypti*]
Contig269	4	482	XP_001660648	2e−16	Pupal cuticle protein, putative [*Aedes aegypti*]
Contig808	5	571	XP_001660648	9e−16	Pupal cuticle protein, putative [*Aedes aegypti*]
Contig472	6	825	XP_973909	1e−23	Cuticular protein Ld-CP3-like [*Tribolium castaneum*]

Underline indicates potential novel sequences, which were defined as those with at least 2% differences with any previously characterized Hessian fly sequences at the amino acid level (differences at the 5′- and 3′-unreliable regions of a cluster were excluded).

Table 5
Tissue-specific expression of selected clusters.

Cluster	Probe set	Relative abundance (%) of transcript among tissues				Function
		Gut	S. Glands	M. Tubules	Carcass	
Contig695	AY669864_at	$\mathbf{77} \pm \mathbf{1.2}^{a}$	2.6 ± 0.4^{c}	17.8 ± 1.1^{b}	2.7 ± 0.5^{c}	Trypsin (MDtryp5A)
Contig511	DQ017267_at	$\mathbf{71.9} \pm \mathbf{2.2}^{a}$	6.6 ± 1.2^{b}	15 ± 1.8^{b}	6.5 ± 1.2^{c}	Defensin
Contig471	DQ017266_at	$\mathbf{71.3} \pm \mathbf{1.1}^{a}$	10 ± 0.8^{c}	16.2 ± 0.9^{b}	2.5 ± 0.4^{d}	Defensin
Contig46	AY596477_s_at	$\mathbf{69.9} \pm \mathbf{1.1}^{a}$	8.8 ± 0.7^{c}	17.9 ± 0.9^{b}	3.3 ± 0.4^{d}	Trypsin (MDtryp4A)
Contig58	Lg3A6_s_at	$\mathbf{66.8} \pm \mathbf{0.9}^{a}$	9.1 ± 0.5^{c}	19.6 ± 0.7^{b}	4.5 ± 0.4^{d}	Unknown SSP
Contig857	Lg1F10_at	$\mathbf{64.3} \pm \mathbf{0.9}^{a}$	4.6 ± 0.4^{c}	25.7 ± 0.8^{b}	5.3 ± 0.4^{c}	Protease inhibitor
Contig558	Sg9G7_s_at	$\mathbf{62.4} \pm \mathbf{1}^{a}$	5.2 ± 0.5^{c}	28.5 ± 1^{b}	3.8 ± 0.4^{c}	Unknown SSP
Contig143	Gg6D8_s_at	$\mathbf{60.5} \pm \mathbf{0.7}^{a}$	6.2 ± 0.4^{c}	26.2 ± 0.6^{b}	7.1 ± 0.4^{c}	Protease inhibitor
Contig49	AY596476_s_at	$\mathbf{57.3} \pm \mathbf{0.7}^{a}$	5.4 ± 0.3^{d}	29.5 ± 0.6^{b}	7.8 ± 0.4^{c}	Trypsin (MDtryp3A)
Contig194	Gg5C9_s_at	$\mathbf{57.1} \pm \mathbf{0.8}^{a}$	9.3 ± 0.5^{c}	26.9 ± 0.7^{b}	6.7 ± 0.4^{c}	Unknown SSP
Contig448	Gg9D3_s_at	$\mathbf{56.9} \pm \mathbf{0.7}^{a}$	5 ± 0.3^{d}	30 ± 0.6^{b}	8 ± 0.4^{c}	Unknown SSP
Contig718	Lg2A10_x_at	$\mathbf{55.8} \pm \mathbf{0.8}^{a}$	5.9 ± 0.4^{c}	34.6 ± 0.8^{b}	3.7 ± 0.3^{d}	Unknown SSP
G2F1	MDEST1117_at	$\mathbf{55.5} \pm \mathbf{1.6}^{a}$	16 ± 1.2^{b}	23.3 ± 1.3^{b}	5.3 ± 0.7^{c}	Unknown SSP
Contig467	Lg1A12_x_at	$\mathbf{54.8} \pm \mathbf{0.7}^{a}$	7.9 ± 0.4^{c}	29 ± 0.7^{b}	8.3 ± 0.4^{c}	Unknown SSP
Contig464	Gg7G6_x_at	$\mathbf{54.6} \pm \mathbf{0.8}^{a}$	8.8 ± 0.5^{c}	30.8 ± 0.7^{b}	5.8 ± 0.4^{d}	Unknown SSP
Contig87	AY596478_s_at	$\mathbf{51.2} \pm \mathbf{0.7}^{a}$	8.2 ± 0.4^{c}	32.4 ± 0.6^{b}	8.3 ± 0.4^{c}	Chymotrypsin (MDchym2A)
Contig328	Lg3E3_s_at	$\mathbf{46.5} \pm \mathbf{0.}^{a}$	7 ± 0.3^{d}	31.9 ± 0.5^{b}	14.7 ± 0.4^{c}	Unknown SSP
Contig666	Lg2C4_x_at	$\mathbf{46.2} \pm \mathbf{0.6}^{a}$	9.6 ± 0.4^{d}	32.1 ± 0.6^{b}	12 ± 0.4^{c}	Unknown SSP
Contig255	Lg2G7_x_at	$\mathbf{44.8} \pm \mathbf{0.5}^{a}$	8.9 ± 0.3^{d}	32.1 ± 0.5^{b}	14.3 ± 0.4^{c}	Unknown SSP
Contig563	Lg4A1_x_at	$\mathbf{42.3} \pm \mathbf{0.6}^{a}$	12.2 ± 0.4^{c}	32 ± 0.6^{b}	13.6 ± 0.4^{c}	Unknown SSP
Contig551	AY596471_s_at	$\mathbf{41.9} \pm \mathbf{0.5}^{a}$	10.6 ± 0.3^{d}	32.4 ± 0.5^{b}	15.1 ± 0.4^{c}	Chymotrypsin (MDchym1B)
Contig431	Lg2A10_s_at	$\mathbf{40.1} \pm \mathbf{0.5}^{a}$	9.3 ± 0.3^{d}	32.6 ± 0.5^{b}	18 ± 0.4^{c}	Unknown SSP
Contig354	Sg8E8_at	$\mathbf{34.1} \pm \mathbf{0.4}^{a}$	10 ± 0.3^{d}	30.7 ± 0.4^{b}	25.2 ± 0.4^{c}	Unknown SSP
Contig866	Lg3B12_at	22.9 ± 0.5^{b}	5 ± 0.3^{d}	17.4 ± 0.5^{c}	$\mathbf{54.7} \pm \mathbf{0.6}^{a}$	Unknown SSP
Contig781	SM2E4_at	22.1 ± 0.4^{b}	8.9 ± 0.3^{c}	$\mathbf{36.1} \pm \mathbf{0.5}^{a}$	$\mathbf{32.9} \pm \mathbf{0.5}^{a}$	Unknown SSP
Contig118	Gg5C1_s_at	19.8 ± 0.8^{b}	7.2 ± 0.5^{c}	$\mathbf{51.2} \pm \mathbf{1.1}^{a}$	21.9 ± 0.9^{b}	Unknown SSP
Contig689	Lg3F8_s_at	19.5 ± 0.5^{b}	7.4 ± 0.3^{c}	$\mathbf{55.5} \pm \mathbf{0.7}^{a}$	17.6 ± 0.5^{b}	Unknown SSP
Contig44	Lg2F10_x_at	19.4 ± 0.6^{b}	12.4 ± 0.5^{c}	$\mathbf{34} \pm \mathbf{0.8}^{a}$	$\mathbf{34.2} \pm \mathbf{0.8}^{a}$	Unknown SSP
Contig57	Lg3A4_s_at	19.4 ± 0.5^{b}	8.1 ± 0.3^{c}	$\mathbf{37.8} \pm \mathbf{0.6}^{a}$	$\mathbf{34.7} \pm \mathbf{0.5}^{a}$	Odorant-binding protein
Contig55	Lg2A3_s_at	18.6 ± 0.4^{c}	9.1 ± 0.3^{d}	$\mathbf{46.6} \pm \mathbf{0.6}^{a}$	25.7 ± 0.5^{b}	Protease inhibitor
Contig51	Lg1A1_x_at	18.4 ± 0.8^{c}	9.7 ± 0.6^{d}	$\mathbf{48} \pm \mathbf{1}^{a}$	23.9 ± 0.8^{b}	Protease inhibitor
Contig43	Lg2F7_s_at	17.7 ± 0.5^{b}	8.7 ± 0.4^{c}	$\mathbf{55.4} \pm \mathbf{0.6}^{a}$	18.2 ± 0.5^{b}	Protease inhibitor
Contig275	Lg1E7_at	17.2 ± 6.7^{b}	5.3 ± 4^{b}	$17.2 \pm 6.7^{a,b}$	$60.4 \pm \mathbf{8.7}^{a}$	Unknown SSP
Contig598	Lg2A8_at	16.8 ± 0.7^{b}	7.5 ± 0.5^{d}	12.5 ± 0.6^{c}	$\mathbf{63.1} \pm \mathbf{0.9}^{a}$	Unknown SSP
Contig472	Sg9D11_s_at	15.4 ± 0.3^{b}	4.2 ± 0.2^{d}	12.5 ± 0.3^{c}	$\mathbf{67.9} \pm \mathbf{0.4}^{a}$	Cuticle protein
Lg1RF07	Lg1F7_at	15.2 ± 1^{b}	$\mathbf{72.1} \pm \mathbf{1.2}^{a}$	8 ± 0.7^{c}	4.7 ± 0.6^{c}	Unknown SSP
Contig15	Lg1E5_at	14.8 ± 1.1^{b}	$\mathbf{59.7} \pm \mathbf{1.5}^{a}$	15.3 ± 1.1^{b}	10.2 ± 0.9^{b}	Unknown SSP

S. Glands and M. Tubules represent salivary glands and Malpighian tubules, respectively. Carcass contains the remaining tissue after removing salivary glands, gut, and Malpighian tubules. Underline (bold) represents tissue with the highest transcript level. The superscripts "a, b, c, and d" indicate groups with significant differences at family error rate of 0.01 based on Goodman's test (see Section 2).

given in Table 5. The majority (62.1%) of the 37 clusters exhibited the highest levels of transcripts in the gut. Among those clusters with highest expression in the gut were all trypsins and chymotrypsins, consistent with their role in digestion. In addition to proteases, several transcripts coding for SSP were also at high levels in the gut.

4. Discussion

BLASTx analysis of the 2007 clusters revealed that 56.7% shared similarity with GenBank sequences. Similar research on Europen corn borer (*O. nubilalis*), a plant-feeding lepidopteran, revealed that 62.7% of larval gut clusters shared similarity with GenBank sequences (Khajuria et al., 2009). The slightly lower percentage of Hessian fly clusters with similarity to known sequences might be due to the fact that more unique genes were expressed in Hessian fly larval gut. Alternatively, it might simply reflect the fact that fewer genes expressed in the gut of galling dipterans have been studied even though the genomes of several non-galling dipterans have been sequenced, whereas gut transcriptomes of several lepidopterans including the keratin-feeding clothes moths (Nation, 2002), wild silkmoth (*Antheraea mylitta*) (Gandhe et al., 2006), and the European corn borer itself (Goates et al., 2008) have been partially characterized before.

Transcripts coding for different types of digestive enzymes were identified in the Hessian fly larval gut. The existence of various types of digestive enzymes indicated that Hessian fly larvae use a wide range of food sources as nutrition. Hessian fly larvae ingest cell content after destruction of cellular and subcellular structures (Harris et al., 2006). Cell content is rich in proteins as well as other substances. Among the transcripts coding for digestive enzymes, those coding for trypsins and chymotrypsins (Fig. S1) were the most abundant, and therefore could be useful targets for pest management such as engineered wheat with high content of protease inhibitors. In resistant wheat seedlings, elevation of protease inhibitors was observed following an Hessian fly attack (Wu et al., 2008).

The Hessian fly larval gut is also rich in transcripts coding for various detoxification enzymes including cytochrome P450s and GSTs, which can convert toxic chemicals such as plant secondary metabolites into less toxic or nontoxic chemicals. The presence of a large number of transcripts coding for different P450 enzymes in the Hessian fly larval gut may suggest a highly complex mode of detoxification evolved to counter-defend the host plant chemical warfare. This is consistent with previous findings that genes involved in the synthesis of secondary metabolites are up-regulated in resistant wheat upon Hessian fly attack (Liu et al., 2007), and that ROS is part of plant defense against Hessian fly attack (Liu et al., 2010). Insect P450s have been long suggested a vital role in detoxification of plant secondary metabolites produced by host plants (Feyereisen, 1999). However, few P450 transcripts have been reported from the characterized insect-gut transcirptomes including plant-feeding insects the cowpea weevil (Pedra et al., 2003), aphids (Sabater-Munoz et al., 2006; Ramsey et al., 2007), and European corn borer (Goates et al., 2008; Khajuria et al., 2009), as well as the blood feeding sand fly (Ramalho-Ortigao et al., 2007; Jochim et al., 2008). This might be explained that the scientists for those studies have other interests and therefore did not focus on P450 transcripts even though they might be part of their EST collections. Alternatively, P450 genes were not expressed or expressed at low levels in the guts of those insects. Transcripts coding for other detoxification enzymes including GSTs, catalases, peroxiredoxins, and ferritins were reported in the sand fly gut transcriptome (Jochim et al., 2008).

In addition to P450 and GST, transcripts coding for enzymes for removal of ROS were also identified, including those coding for peroxidases, ferritins, catalase, peroxiredoxins, and several other enzymes. These detoxification enzymes could form the molecular basis for Hessian fly larvae to overcome basal and induced host-plant defenses. The cluster groupings for ferritin represented the most ESTs. Ferritin is a protein complex that chelates free Fe^{3+} ions, which are toxic to cells because they act as catalyst in the formation of free radicals from ROS via the Fenton Reaction (Orino et al., 2001). Among the three ferritin clusters, contig235 coded for a full-length heavy chain protein, whereas contig294 and singleton HFMidgut6L10 coded for a full-length light chain protein. HFMidgut6L10 was separated from contig294 by two small insertions in the noncoding region (data not shown), but the coding region was the same. The protein encoded by contig235 shared 48% identity with a ferritin heavy chain protein from *Drosophila ananassae* (Fig. S4A). The protein encoded by contig294 shared 57% identity with a ferritin light chain protein from *P. papatasi* (Fig. S4B). The existence of abundant transcripts coding for ferritins and the upregulation of genes coding for other anti-ROS proteins such as gluthionine peroxidase, catalase, and superoxide dismutase in resistant host plants following a Hessian fly attack (Mittapalli et al., 2007) suggests that ROS is also an important type of defense chemicals from host plants (Liu et al., 2010). The abundant and complex detoxification network in the Hessian fly gut is likely the molecular basis for this insect's ability to overcome basal and induced host-plant defenses.

The most striking feature of the gut transcriptome of Hessian fly larvae is the presence of a large number of transcripts coding for SSPs. Little sequence similarity was shared among the SSPs, indicating their likely participation in different biological functions. Further research is needed to elucidate the specific functions of individual SSP-encoding genes. These SSP-encoding genes could also be targets for novel transgenic approaches such as silencing transgenes via RNAi for Hessian fly management (Baum et al., 2007; Mao et al., 2007). The Hessian fly belongs to the order Diptera. Whole genomes of many different dipterans have been sequenced. The fact that these SSPs did not match any GenBank sequences suggests that they are likely unique to the Hessian fly or related gall midges and, therefore, perform unique functions characteristic of this insect or related species. The exact functions of these SSPs remain to be delineated. One possibility is that some of these proteins may play a role in host plant–insect interactions. Gall midges live within plant tissues and have the ability to manipulate plant growth (Rohfritsch, 1992, 2005). Hessian fly larvae can inhibit plant growth (Anderson and Harris, 2006), reprogram gene transcription of infested plants (Liu et al., 2007), and induce nutritive tissues (Harris et al., 2006). The primary source for effector proteins is the salivary glands (Miles, 1999). However, effectors from oral secretions have been also reported from various insects (Kessler and Baldwin, 2002; Schmetz et al., 2006). Alternatively, these SSPs may be secreted into the haemocoel as feedback regulators for physiological processes in the gut, or secreted directly into the alimentary canal and protect gut tissue from damaging microorganisms, or act as inhibitors of toxic enzymes such as proteases ingested from host plants (Pechan et al., 2000).

Abundant transcripts of genes coding for digestive enzymes and detoxification proteins are characteristic of many insect-gut transcriptomes (Pedra et al., 2003; Dillon et al., 2006; Hughes and Vogler, 2006; Ramalho-Ortigao et al., 2007; Goates et al., 2008; Jochim et al., 2008; Chi et al., 2009; Khajuria et al., 2009). However, the high proportion of unique transcripts coding for SSPs has not been reported in gut transcriptomes of other insects so far. The fact that SSP transcripts represent only 10.7% of clusters, but 44.0% of ESTs indicates that the transcripts coding for SSPs were the most abundant in gut of Hessian fly larvae. This can be further seen from the fact that 11 out of the 30 largest contigs coded for SSPs

Table 6
The 30 contigs with the largest numbers of ESTs.

Contig	EST No.	% of total	Length (bp)	First hit	Putative function	*E*-Value	Organism
718[a]	915	11.3	860	ABE26919	Small secreted gut protein – Lg3H4	2e−81	MD
460[a]	265	3.3	801	No hit	Small secretory protein		
535	233	2.9	720	ABQ96857	Unknown protein	0.015	HQ
431[a]	195	2.4	839	ABE26927	Small secreted gut protein – Pg7A3	2e−81	MD
161	190	2.3	2102	CAI11090	Cytochrome oxidase subunit I	7e−142	CM
258	82	1.0	1354	ABI52743	10 kDa putative secreted protein	1e−23	AM
591[a]	82	1.0	752	No hit	Small secretory protein		
497[a]	80	1.0	882	XP_001868961	14.5 kDa salivary protein	5e−21	CQ
725	76	1.0	1681	CAB63100	Serine protease inhibitor-serpin-5	4e−47	DM
354[a]	73	0.9	787	No hit	Small secretory protein		
328[a]	67	0.8	1346	No hit	Small secretory protein		
30	62	0.8	1844	XP_001864493	40S ribosomal protein S3	1e−110	CQ
60	54	0.7	658	YP_002261331	Cytochrome c oxidase subunit III	1e−32	MM
235	53	0.7	1457	AAP57194	Ferritin subunit 1	2e−61	LD
643[a]	53	0.7	511	AAY82237	Defensin I	0.011	MD
75	52	0.6	2388	XP_001647991	Hypothetical protein	1e−22	AA
563[a]	52	0.6	850	ABE26913	Small secreted gut protein – Lg4A1	9e−72	MD
658	47	0.6	1650	XP_001844836	Aldehyde reductase 1	3e−77	CQ
63	44	0.5	1005	AAC34860	Hypothetical protein	1e−12	DM
98	44	0.5	866	YP_973149	NADH dehydrogenase subunit 4	7e−49	CD
467[a]	43	0.5	562	No hit	Small secretory protein		
478[a]	41	0.5	929	No hit	Small secretory protein		
616	37	0.5	1843	XP_002118266	Senescence-associated protein	5e−50	TA
295	35	0.4	1004	XP_001649329	Preprotein translocase secy subunit	6e−109	AA
744	34	0.4	1069	EEB20322	Trehalose-6-phosphate synthase 1	2e−54	PHC
453	33	0.4	662	NP_649560	Ribosomal protein l13a	2e−66	DM
117	32	0.4	411	ABQ96857	Unknown	0.006	HQ
225	32	0.4	983	AAV65760	Cysteine-rich protein	1e−17	CLF
382	31	0.4	598	No hit			
695	31	0.4	1138	AAT81428	Trypsin precursor	4e−145	MD

Abbreviations: MD: *Mayetiola destructor*; HQ: *Haemaphysalis qinghaiensis*; CM: *Chamaesphecia masariformis*; AM: *Argas monolakensis*; CQ: *Culex quinquefasciatus*; DM: *Drosophila melanogaster*; MM: *Myrmecophilus manni*; LD: *Leptinotarsa decemlineata*; AA: *Aedes aegypti*; CD: *Cydistomyia duplonotata*; TA: *Trichoplax adhaerens*; PHC: *Pediculus humanus corporis*; CLF: *Canis lupus familiaris*.
[a] Clusters encoding SSPs.

(Table 6). The high abundance of SSP transcripts indicates that Hessian fly larval gut possesses some unique functions in addition to digestion and detoxification. The exact functions of these SSPs remain to be resolved. Some SSPs were predominantly expressed in the gut while others were mainly expressed in other tissues. Considering the diversity and specific distribution in different tissues, these SSPs were likely to perform various functions. Some of these SSPs may play roles in Hessian fly interaction with other organisms such as host plants and symbiotic bacteria. The ones predominantly expressed in Malpighian tubules might be important regulators of development. The largest cluster (contig718), which coded for an SSP with unknown function (Chen et al., 2006), consisted of 915 ESTs or 11.3% of total ESTs. Further research on this gene and its encoding protein may help to understand unique gut physiology of Hessian fly larvae. Whether the high proportion of abundant transcripts coding for SSPs is unique to Hessian fly larvae or a common feature for galling insects remains to be determined.

Acknowledgments

The authors thank Drs. Kun Yan Zhu and Marcelo Ramalho-Ortigao for valuable suggestions and comments on the manuscript. The manuscript is a contribution no. 09_265-J from the Kansas Agricultural Experiment Station, Kansas State University, Manhattan, Kansas, USA. Mention of commercial or proprietary product does not constitute endorsement by the USDA.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jinsphys.2010.03.021.

References

Anderson, K.G., Harris, M.O., 2006. Does R gene resistance allow wheat to prevent plant growth effects associated with Hessian fly (Diptera: Cecidomyiidae) attack? Journal of Economic Entomology 99, 1842–1853.

Baum, J.A., Bogaert, T., Clinton, W., Heck, G.R., Feldmann, P., Ilagan, O., Johnson, S., Plaetinck, G., Munyikwa, T., Pleau, M., Vaughn, T., Roberts, J., 2007. Control of coleopteran insect pests through RNA interference. Nature Biotechnology 25, 1322–1326.

Bentur, J.S., Kalode, M.B., 1996. Hypersensitive reaction and induced resistance in rice against the Asian rice gall midge *Orseolia oryzae*. Entomologia Experimentalis Applicata 78, 77–81.

Chen, M.S., Liu, X., Zhu, Y.C., Reese, J.C., Wilde, G.E., 2006. Genes encoding a group of related small secreted proteins from the gut of Hessian fly larvae [*Mayetiola destructor* (Say)]. Insect Science 13, 339–348.

Chi, Y.H., Salzman, R.A., Balfe, S., Ahn, J.E., Sun, W., Moon, J., Yun, D.J., Lee, S.Y., Higgins, T.J.V., Pittendrigh, B., Murdock, L.L., Zhu-Salzman, K., 2009. Cowpea bruchid midgut transcriptome response to a soybean cystatin – costs and benefits of counter-defence. Insect Molecular Biology 18 (1), 97–110.

Dillon, R.J., Ivens, A.C., Churcher, C., Holroyd, N., Quail, M.A., Rogers, M.E., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Lehane, M.J., Bates, P.A., 2006. Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect–parasite relationship. Genomics 88 (6), 831–840.

Feyereisen, R., 1999. Insect P450 enzymes. Annual Review of Entomology 44, 507–533.

Gandhe, A.S., Arunkumar, K.P., John, S.H., Nagaraju, J., 2006. Analysis of bacteria-challenged wild silkmoth. *Antheraea mylitta* (Lepidoptera) transcriptome reveals potential immune genes. BMC Genomics 7, 148.

Goates, B.S., Sumerford, D.V., Hellmich, R.L., Lewis, L.C., 2008. Mining an *Ostrinia nubilalis* midgut expressed sequence tag (EST) library for candidate genes and single nucleotide polymophisms (SNPs). Insect Molecular Biology 17 (6), 607–620.

Goodman, L.A., 1965. On simultaneous confidence intervals or multinomial proportions. Technometrics 7, 247–254.

Harris, M.O., Freeman, T.P., Rohfritsch, O., Anderson, K.G., Payne, S.A., Moore, J.A., 2006. Virulent Hessian fly (Diptera: Cecidomyiidae) larvae induce a nutritive tissue during compatible interactions with wheat. Annals of the Entomological Society of America 99, 305–316.

Harris, M.O., Rose, S., 1989. Temporal changes in egglaying behavior of the Hessian fly. Entomologia Experimentalis Applicata 53 (1), 17–29.

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.

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, 625–675.

Herrero, S., Oppert, B., Ferre, J., 2001. Different mechanisms of resistance to Bacillus thuringiensis toxins in the indianmeal moth. Applied and Environmental Microbiology 67 (3), 1085–1089.

Hoglund, S., Larrson, S., Wingsle, G., 2005. Both hypersensitive and non-hypersensitive responses are associated with resistance in Salix viminalis against the gall midge *Dasineura marginemtorquens*. Journal of Experimental Botany 56, 3215–3222.

Hughes, J., Vogler, A.P., 2006. Gene expression in the gut of keratin-feeding clothes moths (Tineola) and keratin beetles (Trox) revealed by subtracted cDNA libraries. Insect Biochemistry and Molecular Biology 36 (7), 584–592.

Jochim, R.C., Teixeira, C.R., Laughinghouse, A., Mu, J., Oliveira, F., Gomes, R.B., Elnaiem, D.-E., Valenzuela, J.G., 2008. The midgut transcriptome of *Lutzomyia longipalpis*: comparative analysis of cDNA libraries from sugar-fed, blood-fed, post-digested and Leishmania infantum chagasi-infested sand flies. BMC Genomics 9, 15.

Kessler, A., Baldwin, I.T., 2002. Plant responses to insect herbivory. Annual Review of Plant Biology 53, 299–328.

Khajuria, C., Zhu, Y.-C., Chen, M.-S., Buschman, L.L., Higgins, R.A., Yao, J., Muthukrishnan, S., Zhu, K.Y., 2009. Expressed sequence tags from the European corn borer (*Ostrinia nubilalis*) larval gut: identification of the genes potentially involved in Bacillus thuringiensis toxicity and resistance. BMC Genomics 10, 286.

Lamb, C., Dixon, R.A., 1997. The oxidative burst in plant disease resistance. Annual Review of Plant Physiology and Plant Molecular Biology 48, 251–275.

Li, H.M., Buczhowski, G., Mittapalli, O., Xie, J., Wu, J., Westerman, R., Schemerhorn, B.J., Murdock, L.L., Pittendrigh, B.R., 2008. Transcriptomic profiles of Drosophila melanogaster third instar larval midgut and responses to oxidative stress. Insect Molecular Biology 17 (4), 325–339.

Liu, X.L., Bai, J., Huang, L., Zhu, L., Liu, X., Weng, N., Reese, J.C., Harris, M., Stuart, J.J., Chen, M.S., 2007. Gene expression of different wheat genotypes during attack by virulent and avirulent Hessian fly (*Mayetiola destructor*) larvae. Journal of Chemical Ecology 33, 2171–2194.

Liu, X., Fellers, J.P., Zhu, Y.C., Mutti, N.S., El-Bouhssini, M., Chen, M.S., 2006. Cloning and characterization of cDNAs encoding carboxypeptidase-like proteins from the gut of Hessian fly [*Mayetiola destructor* (Say)] larvae. Insect Biochemistry and Molecular Biology 36, 665–673.

Liu, W.M., Mei, R., Bartell, D.M., Di, X., Webster, T.A., Ryder, T., 2001. Rank-based algorithms for analysis of microarrays. Proceedings of SPIE, Microarrays: Optical Technologies and Informatics 4266, 56–67.

Liu, W.M., Mei, R., Di, X., Ryder, T.B., Hubbell, E., Dee, S., Webster, T.A., Harrington, C.A., Ho, M.H., Baid, J., Smeekens, S.P., 2002. Analysis of high density expression microarrays with signed-rank call algorithms. Bioinformatics 18 (12), 1593–1599.

Liu, X.M., Williams, C.E., Nemacheck, J.A., Wang, H.Y., Subramanyam, S., Zheng, C., Chen, M.S., 2010. Reactive oxygen species are involved in plant defense against a gall midge. Plant Physiology 152, 985–999.

Maddur, A.A., Liu, X.M., Zhu, Y.C., Fellers, F.P., Oppert, B., Park, Y.S., Bai, J., Wilde, G.E., Chen, M.S., 2006. Cloning and characterization of protease inhibitor-like cDNAs from the Hessian fly *Mayetiola destructor* (Say). Insect Molecular Biology 15 (4), 485–496.

Mao, Y.-B., Cai, W.-J., Wang, J.-W., Hong, G.-J., Tao, X.-Y., Wang, L.-J., Huang, Y.-P., Chen, X.-Y., 2007. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nature Biotechnology 25, 1307–1313.

Masoundi-Nejad, A., Tonomura, K., Kawashima, S., Moriya, Y., Suzuki, M., Itoh, M., Kanehisa, M., Endo, T., Goto, S., 2006. EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Research 34 (Web Server issue), W459–W462.

Miles, P.W., 1999. Aphid saliva. Biological Review 74 (1), 41–85.

Mittapalli, O., Neal, J.J., Shukle, R.H., 2007. Antioxidant defense response in a galling insect. Proceedings of the National Academy of Sciences of the United States of America 104 (6), 1889–1894.

Nation, J.L., 2002. Digestion. In: Nation, J.L. (Ed.), Insect Physiology and Biochemistry. CRC Press, Boca Raton, FL, pp. 27–64.

Orino, K., Lehman, L., Tsuji, Y., Ayaki, H., Torti, S.V., Torti, F.M., 2001. Ferritin and the response to oxidative stress. Biochemical Journal 357 (1), 241–247.

Oviedo, M.N., VanEkeris, L., Corena-Mcleod, M.D.P., Linser, P.J., 2008. A microarray-based analysis of transcriptional compartmentalization in the alimentary canal of Anopheles gambiae (Diptera: Culicidae) larvae. Insect Molecular Biology 17 (1), 61–72.

Pechan, T., Ye, J., Chang, Y.M., Mitra, A., Lin, L., Davis, F.M., Williams, W.P., Luthe, D.S., 2000. A unique 33-kDa cysteine proteinase accumulates in response to larval feeding in maize genotypes resistant to fall armyworm and other Lepidoptera. Plant Cell 12, 1031–1040.

Pedra, J.H.F., Brandt, A., Westerman, R., Lobo, N., Li, H.M., Romero-Severson, J., Murdock, L.L., Pittendrigh, B.R., 2003. Transcriptome analysis of the cowpea weevil bruchid: identification of putative proteinases and α-amylases associated with food breakdown. Insect Molecular Biology 12 (4), 405–412.

Pepper, S., Saunders, E., Edwards, L., Wilson, C., Miller, C., 2007. The utility of MAS5 expression summary and detection call algorithms. BMC Bioinformatics 8, 273.

Ramalho-Ortigao, M., Jochim, R.C., Anderson, J.M., Lawyer, P.G., Pham, V.M., Kamhawi, S., Valenzuela, J.G., 2007. Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of sugar-fed, blood-fed and Leishmania major-infected sandflies. BMC Genomics 8, 300.

Rampias, T.N., Sideris, D.C., Fragoulis, E.G., 2003. Cc RNase: the Ceratitis capitata ortholog of a novel highly conserved protein family in metazoans. Nucleic Acids Research 31 (12), 3092–3100.

Ramsey, J.S., Wilson, A.C., de Vos, M., Sun, Q., Tamborindeguy, C., Winfield, A., Malloch, G., Smith, D.M., Fenton, B., Gray, S.M., Jander, G., 2007. Genomic resources for *Myzus persicae*: EST sequencing. SNP identification, and microarray design. BMC Genomics 8, 423.

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

Rohfritsch, O., 1992. Patterns in Gall development. In: Shorthouse, J.D., Rohfritsch, O. (Eds.), Biology of Insect-Induced Galls. Oxford University Press, New York, pp. 60–86.

Rohfritsch, O., 2005. Gall making. In: Goodman, R.M. (Ed.), Encyclopedia of Plant and Crop Science. Marcel Dekker, New York, pp. 1021–1022.

Sabater-Munoz, B., Legeai, F., Rispe, C., Bonhomme, J., Dearden, P., Dossat, C., Duclert, A., Gauthier, J.P., Ducray, D.G., Hunter, W., Dang, P., Kambhampati, S., Martinez-Torres, D., Cortes, T., Moya, A., Nakabachi, A., Philippe, C., Prunier-Leterme, N., Rahbe, Y., Simon, J.C., Stern, D.L., Wincker, P., Tagu, D., 2006. Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera). Genome Biology 7 (3), R21.

Schmetz, E.A., Carroll, M.J., LeClere, S., Phipps, S.M., Meredith, J., Chourey, P.S., Alborn, H.T., Teal, P.E.A., 2006. Fragments of ATP synthase mediate plant perception of insect attack. Proceedings of the National Academy of Sciences of the United States of America 103, 8894–8899.

Simpson, R.M., Newcomb, R.D., Gatehouse, H.S., Crowhurst, R.N., Chagne, D., Gatehouse, L.N., Markwick, N.P., Beuning, L.L., Murray, C., Marshall, S.D., Yauk, Y.-K., Nain, B., Wang, Y.-Y., Gleave, A.P., Christeller, J.T., 2007. Expressed sequence tags from the midgut of *Epiphyas postvittana* (Walker) (Lepidoptera: Tortricidae). Insect Molecular Biology 16 (6), 675–690.

Sosa, O., 1978. Biotype L, ninth biotype of the Hessian fly. Journal of Economic Entomology 71, 458–460.

Stuart, J.J., Chen, M.S., Harris, M., 2008. Hessian fly. In: Hunter, W., Kole, C. (Eds.), Genome Mapping and Genomics in Animals, vol. 1: Genome Mapping and Genomics in Arthropods. Springer, Berlin, Heidelberg, New York, pp. 93–100.

Swofford, D.L., 1999. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: properties, compartmentalization and function. Comparative Biochemistry and Physiology 109B (1), 1–62.

Wu, J.X., Liu, X.M., Zhang, S.Z., Zhu, Y.C., Whitworth, R.J., Chen, M.S., 2008. Differential responses of wheat inhibitor-like genes to Hessian fly, *Mayetiola destructor*, attacks during compatible and incompatible interactions. Journal of Chemical Ecology 34, 1005–1012.

Yuki, M., Moriya, S., Inoue, T., Kudo, T., 2008. Transcriptome analysis of the digestive organs of *Hodotermopsis sjostedti*, a lower termite that hosts mutualistic microorganisms in its hindgut. Zoological Science 25 (4), 401–406.

Zhu, Y.C., Liu, X., Maddur, M., Oppert, B., Chen, M.S., 2005. Cloning and characterization of chymotrypsin- and trypsin-like cDNAs from the gut of the Hessian fly [*Mayetiola destructor* (Say)]. Insect Biochemistry and Molecular Biology 35 (1), 23–32.