Abstract

Both the wheat midge (*Sitodiplosis mosellana* (Géhin) (Diptera: Cecidomyiidae)) and the Hessian fly (*Mayetiola destructor* (Say)) (Diptera: Cecidomyiidae) belong to a group of insects called gall midges (Diptera: Cecidomyiidae), and both are destructive pests of wheat. From Hessian fly larvae, a large number of genes have been identified to encode secreted salivary gland proteins (SSGPs), which are presumably critical for the insect to feed on and manipulate host plants. For comparison, we conducted an analysis on transcripts encoding SSGPs from the first instar larvae of the wheat midge. In total, 3,500 cDNA clones were sequenced, from which 1,301 high-quality sequences were obtained. Approximately 25% of the cDNAs with high-quality sequences encoded SSGPs. The SSGPs were grouped into 97 groups based on sequence homology. Among the SSGP-encoding transcripts, 206 encoded unique proteins with no sequence similarity to any known protein and 29 encoded proteins similar to known proteins including proteases, serpines, thioesterases, ankyrins, and ferritins. Most (~80%) SSGP-encoding genes appear under strong selection for mutations that generate amino acid changes within the coding region. Identification and characterization of SSGPs in wheat midge larvae provide a foundation for future work to reveal molecular mechanisms behind wheat midge–wheat interactions and the role of these putative effector proteins in insect virulence. Availability of the SSGP transcripts will also facilitate comparative analyses of insect effectors from related species.

Key words: orange wheat blossom midge, transcriptome analysis, secreted salivary gland protein, *Sitodiplosis mosellana*, insect effector
Several avirulence effectors have been cloned from the Hessian fly, and all of them were SSGPs (Aggarwal et al. 2014; Zhao et al. 2015, 2016). Many secreted proteins have also been identified in the saliva of several aphid species (Thorpe et al. 2016), and many of these aphid proteins act as effectors either to suppress or trigger plant defense responses (Elzinga et al. 2014). Therefore, identification of SSGPs from insects provides an efficient way to identify putative effectors of insect species.

Whether the interaction between wheat midge and wheat follows a gene-for-gene model remains to be investigated. However, a highly effective resistance gene, named Sm1, to the wheat midge was discovered in winter wheat genotype in 1996 (Barker and McKenzie 1996), and wheat cultivars with Sm1 can significantly limit kernel damage and yield loss (Blake et al. 2014, Smith et al. 2014). The existence of a major resistance gene in wheat suggests that a gene-for-gene relationship is possible in the wheat midge–wheat interaction. Two groups of SSGPs in wheat midge larvae have been reported previously (Chen et al. 2010). However, large-scale identification of SSGP-encoding genes in the wheat midge has not been conducted. The objective of this study is to conduct a more extensive analysis of SSGPs from insects provides an efficient way to identify putative effectors of insect species.

**Materials and Methods**

**Insects and Salivary Gland Preparation**

The insect population used in this research was derived from a colony consisting approximately 20,000 individuals collected from Divide County in North Dakota in 2013. The colony has been maintained in a greenhouse at North Dakota State University, Fargo, ND, since then.

Seeds of Roblin hard red spring wheat, an early maturing Canadian variety that is susceptible to wheat midge, were planted in a greenhouse at North Dakota State University to rear the wheat midge. Wheat plants were maintained at 20°C with a photoperiod of 18:6 (L:D) h cycles. Meanwhile, dormant pupae at 4°C have been placed at room temperature to break down the dormancy stage for adult's emergence. When wheat plants were at Zadok's growth stages 55–59 (the inflorescence is half or more emerged from the sheath), two or more gravid females were placed into a glass cylinder covering an individual wheat head. After 24 h of exposure to the wheat mid females, the glass cylinder was removed and the head was covered with a glassine pollination bag to help protect the eggs from desiccating as they develop. Egg hatch and larval migration to their larval feeding sites on the surface of the developing seed occurs 3 d after oviposition. For RNA analysis, 3- to 4-d-old wheat midge larvae were collected from wheat heads with the aid of a 20× dissecting microscope. Salivary glands were obtained by dissecting first instar larvae in saline buffer. Dissection was achieved by pulling away the anterior tip of a larva with a pair of forceps while holding the posterior end of the larva steady with another pair of forceps. The salivary glands of the larva move out of the cascade during this process along with other mouthpart tissues. Clean salivary glands were then obtained by removing unwanted mouthpart tissues. For RNA analysis and cDNA library construction, the dissected glands were transferred into TRI reagent (Molecular Research, Inc., Cincinnati, OH) and frozen in liquid nitrogen as soon as they were obtained.

**cDNA Library Construction and Sequencing**

Total RNA was isolated from 300 pairs of salivary glands using TRI reagent following the protocol provided by the manufacturer. RNA quality and integrity were assessed using a Bioanalyzer (Agilent Technologies, Santa Clara, CA). cDNA libraries were constructed using a “SMART” library construction kit from Clontech (Palo Alto, CA) as described by Chen et al. (2004). Briefly, cDNA inserts were ligated into the pCRX-TOPO plasmid contained in a TOPO TA cloning kit (Invitrogen, Carlsbad, CA) instead of a phage vector. Individual clones were picked up for plasmid DNA isolation, which were sequenced with the M13 forward and reverse primers following the Sanger DNA sequencing method via a commercial contract (GENEWIZ, South Plainfield, NJ).

**Sequence Analysis**

Vector sequences were trimmed from raw reads after cDNA clones were sequenced. Sequences from sense and antisense directions were aligned to examine if a clone was sequenced fully from both directions. If no overlap was found between the sense and antisense reads, new primers were synthesized for further sequencing.

Cluster analyses of cDNAs were conducted using BlastStation-Local 64 program. Open reading frames (ORF) were identified using the ORF finder. Sequence alignment and similarity analysis were performed using various BLAST programs (http://www.ncbi.nlm.nih.gov/). Initial database search was conducted with BLASTN and BLASTX. Sequence alignments with E-values greater than 10^3 were considered to have no meaningful sequence similarity between the two sequences. Sequence alignments with E-values smaller than 10^-10 were considered that two sequences share significant similarity. Sequence alignments with E-values between 10^-3 and 10^-10 were further examined individually to determine if two sequences share similarity based on the length and gaps of the alignments. Analysis for secretion signal peptides was carried out using the SignalP v4.1 (Center for Biological Sequence Analysis, Technical University of Denmark; http://www.cbs.dtu.dk/services/SignalP/).

**Calculation of Synonymous and Nonsynonymous Mutation Rates**

The percentages of synonymous and nonsynonymous mutations were calculated based on sequence alignments of members within a group. For example, the percentages of nonsynonymous mutations were derived by dividing the number of nonsynonymous mutations by the number of total mutations among group members. If there are multiple members that share the same mutation at the same position, then the mutation is counted only once. However, if different members have two or more different mutations at the same position, then the mutation was counted as two or more.

**Results**

**Composition of Transcripts Obtained From Dissected Salivary Glands**

In total, 3,523 cDNA clones were sequenced. After removing clones with small inserts and bad quality sequences, 1,301 cDNA sequences were retained. Among these cDNAs, 330 (25.3%) encode SSGPs and the remaining 971 (74.7%) encode proteins without a typical secretion signal peptide. Among the SSGP-encoding cDNAs, 235 encoded unique proteins with no sequence similarity to any known sequences in GenBank, whereas 33 encoded proteins with sequence similarity to known proteins such as carboxypeptidases, peptidases, lysosomal thioesterases, serpins, ankyrins, and ferritins (Supp Fig. S1 and Supp Table S1 [online only]).

Among the 971 non-SSGP transcripts, 295 (30.4%) encode proteins with no meaningful (E-values greater than 10^3) sequence similarity to any proteins in GenBank, 321 (33.1%) encode proteins
with sequence similarity to proteins with unknown function, and the remaining 353 (36.4%) encode proteins with sequence similarity to proteins with various functions. For the transcripts encoding known proteins, 153 (43.1%) are proteins with functions in protein synthesis and the remaining 205 (57.7%) with other house-keeping functions, including energy-metabolic enzymes, structural proteins, transporters, and others (Supp Table S2 [online only]).

SSGP Classification

SSGP-encoding transcripts were sorted into 97 groups according to the sequence similarity among the cDNAs and derived proteins (Supp Table S1 [online only]). Among the 97 groups, 66 have either a single clone or multiple clones that encode the same protein. The remaining 31 groups have multiple clones that encode at least two different proteins. Proteins within a group share at least 30% amino acid identity and have a highly conserved secretion signal peptide. Proteins between different groups share no meaningful \((E > 10^4)\) sequence similarity and have a completely different secretion signal peptide (Supp Fig. S1 [online only]). Figure 1 shows amino acid sequence alignments of two representative groups. Both groups have a highly conserved signal peptide and a more diversified mature protein. The overall conservation among group members particularly in the signal peptide region suggests that the transcripts within a group may have been derived from genes that share the same evolutionary origin and, therefore, can be considered the same gene family (Fig. 1). Some sequence variation may have also resulted from different alleles of the same gene. Amino acid sequence alignments of all groups with multiple members are shown in Supp Fig. S2 (online only).

Sequence Variations Among Group Members

Group members among those with significant sequence variations were divided into mature protein (MP)-coding region, signal peptide (SP)-coding region, and noncoding regions, and percentages of nucleotides with sequence variation in each region were analyzed (Table 1). Among these three regions, sequence variation in the MP-coding region was the lowest except group 24, probably due to the functional constraint of the secretion role of signal peptides. Variation rates in MP-coding region and noncoding regions were much higher (Supp Table S3 and Supp Fig. S2 [online only]).

To examine if group members were under selection pressure for diversification, the percentages of nonsynonymous and synonymous mutations in the MP-coding region were also analyzed. Over 70% of nucleotide substitutions were nonsynonymous (Table 1).

Discussion

Many insects inject effectors into host tissues to manipulate plants including suppressing host defense, inhibiting plant growth, and reprogramming plant metabolism (Stuart 2015). Some insects also inject effectors into host tissues for predigesting food before ingestion and for various other functions (Miles 1999, Harris et al. 2015). The salivary glands of insects are the main tissue to produce effector proteins for host injection. Therefore, analyzing transcripts in the salivary glands and identifying those proteins with a secretion signal peptide is an efficient way to identify putative effector proteins (Chen et al. 2004, 2008). In this study, we analyzed the composition of transcripts in salivary glands of the first instars of the wheat midge through a traditional Sanger sequencing approach. There are two reasons to follow a traditional sequencing approach in this study. First, previous studies have shown that effector genes from gall midges are conserved unconventionally (Chen et al. 2010), which would cause problems in correctly assembling short sequence reads from high-throughput sequencing. Second, the wheat midge is an understudied species genomically and may be difficult to annotate small transcript fragments.

Our analysis resulted in the identification of 97 groups of transcripts encoding SS-GPs. Among these groups, 64% (62 groups) are singletons, indicating that our analysis is very preliminary and further sequencing more clones is likely to identify much more unique SSGP transcripts. The most abundant group is group 1, which has 48 unique transcripts (99 including redundant sequences). SSGP proteins encoded by group 1 transcripts share no sequence similarity with any known sequences in GenBank, and therefore, the functions of this group of genes remain to be determined. The fact that members among this group have been under strong positive selection indicates that this group of genes are likely to play important roles in the wheat midge–wheat interaction. Other abundant transcript groups include group 2, group 3, group 4, group 12, group 13, group 24, group 29, group 40, group 45, and group 67 (Supp Table S1 [online only]).

There are commonalities and differences between the putative SSGPs from wheat midge larvae and those from Hessian fly larvae, a species that has been studied more extensively for SSGP-encoding genes (Chen et al. 2004, 2008, 2010; Zhao et al. 2015, 2016). A commonality is that most of the SSGPs are small peptides (50–150 amino acid residues), and those small SSGPs share no sequence similarity with any known proteins in available databases. In addition, SSGP-encoding genes from both the wheat midge and Hessian fly appear to be under strong diversifying selection pressure. Evidence for this is the fact that over 70% of point mutations among group members are nonsynonymous (Table 1). A similar phenomenon was also found in Hessian fly SSGP-encoding genes, where over 80% of point mutations among group members were nonsynonymous (Chen et al. 2004). The fast-evolving nature of SSGP-encoding genes in both insect species is another indicator that these genes are involved in interactions with their host plants (Thompson 1998). There is no sequence similarity between SSGPs from wheat midge larvae and those from Hessian fly larvae, suggesting that SSGPs from these two insect species perform different biochemical functions and have different mechanisms to manipulate host plants. In addition, many SSGP-encoding genes from Hessian fly exhibit an unconventional conservation pattern, in which the 5′- and 3′-noncoding regions and introns are highly conserved, whereas the regions encoding mature proteins are highly diversified (Chen et al. 2010, Zhao et al. 2015). No such unconventional conservation pattern was found among group members of SSGP-encoding genes from the wheat midge.

In addition to small SSGPs, there are a few transcripts that encode secreted proteins with sequence similarity to known proteins, which include proteases, protease inhibitors, lysosomal thioesterases, ankyrins, and ferritins. Whether these proteins are injected into host plants or secreted into body fluid of the insect remains to be determined. Proteases and protease inhibitors have also been found in saliva from other insect species (Miles 1999, Chen et al. 2008, Liu et al. 2016). Proteases could act as digestive enzymes for preoral digestion of food before ingestion, whereas protease inhibitors could neutralize defense proteases from host plants (Pechan et al. 2002). Lysosomal thioesterases, ankyrins, and ferritins play house-keeping functions inside insects. However, some proteins with house-keeping functions in insects can also be injected into host plants and play effector roles in insect–plant interactions (Miles 1999).

In summary, we have conducted a global analysis on genes expressed in the salivary glands of first instars of the wheat midge.
for the first time and identified numerous genes encoding SSGPs. The availability of the putative effector genes provides a foundation for further research to characterize the roles of these genes in wheat midge and wheat interactions. For example, the cDNAs could be used to produce recombinant proteins for various biochemical assays or for antibody production to analyze tissue distribution within both the insect bodies and the host tissues if they are injected into plants during feeding. The availability of these genes is also

Fig. 1. Amino acid alignments of two representative groups. The boundary between predicted signal peptide and mature proteins is indicated by an arrow. Only partial alignment for the second group is shown in the figure.
useful for comparative analysis of salivary proteins from different insect species.

**Supplementary Data**

Supplementary data are available at *Journal of Insect Science* online.

**Acknowledgments**

The research is joint effort between USDA–ARS and Kansas State University. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer. This study has been conducted in the Department of Entomology, Kansas State University, Manhattan, KS.

Z.A., K.M.A., and O.M. conducted research and analyzed data. M.O.H. and R.J.W. contributed funds and reagents, analyzed data, and revised manuscript. Z.A. and M.S.C. designed experiments and wrote the article.

**References Cited**

Aggarwal, R., S. Subramanyam, C. Zhao, M. S. Chen, M. O. Harris, and J. J. Stuart. 2014. Avrulence effector discovery in a plant galling and plant parasitic arthropod, the Hessian fly (*Mayetiola destructor*). PLoS ONE. 9: e100958.

Barker, P. S., and R. I. H. McKenzie. 1996. Possible sources of resistance to the wheat midge in wheat. Can. J. Plant Sci. 76: 689–695.

Berzonsky, W. A., H. Ding, S. D. Haley, M. O. Harris, R. J. Lamb, R. I. H. McKenzie, H. W. Ohn, E. L. Patterson, P. Pearis, D. R. Porter, et al. 2003. Breeding wheat for resistance to insects. Plant Breed. Rev. 22: 221–296.

Blake, N. K., R. N. Stougaard, B. Bohannon, D. K. Weaver, H. Y. Heo, P. F. Lamb, D. Nash, D. M. Wichman, K. D. Kephart, J. H. Miller, et al. 2014. Registration of ‘Egan’ wheat with resistance to orange wheat blossom midge. J. Plant Regen. 8: 298–302.

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

Chen, M. S., H. X. Zhao, Y. C. Zhu, B. Scheffler, X. Liu, X. Liu, S. Hulbert, and J. J. Stuart. 2008. Analysis of transcripts and proteins expressed in the salivary glands of Hessian fly (*Mayetiola destructor*) larvae. J. Insect Physiol. 54: 1–16.

Chen, M. S., X. Liu, Z. Yang, H. Zhao, R. H. Shukle, J. J. Stuart, and S. Hulbert. 2010. Unusual conservation among genes encoding small secreted salivary gland proteins from a gall midge. BMC Evol. Biol. 10: 296.

Ding, H., R. J. Lamb, and N. Ames. 2000. Inducible production of phenolic acids in wheat and antibiotic resistance to *Sitodiplosis mosellana*. J. Chem. Ecol. 26: 969–985.

Doane, J. F., and O. Ollert. 2008. Seasonal development of wheat midge, *Sitodiplosis mosellana* (Géhin) (Diptera: Cecidomyiidae), in Saskatchewan, Canada. Crop Prot. 27: 951–958.

Elinga, D. A., M. De Vos, and G. Jander. 2014. Suppression of plant defenses by a *Myzus persicae* (green peach aphid) salivary effector protein. Mol. Plant. Microbe Interact. 27: 747–756.

Harris, M. O., J. J. Stuart, M. Mohan, S. Nair, R. J. Lamb, and O. Rohrfrisch. 2003. Grasses and gall midges: plant defense and insect adaptation. Annu. Rev. Entomol. 48: 549–577.

Harris, M. O., T. L. Friesen, S. S. Xu, M. S. Chen, D. Giron, and J. J. Stuart. 2015. Pivoting from *Arabidopsis* to wheat to understand how agricultural plants integrate responses to biotic stress. j. Exp. Bot. 66: 513–531.

Lamb, R. J., J. R. Tucker, I. L. Wise, and M. A. H. Smith. 2000. Trophic interaction between *Sitodiplosis mosellana* (Diptera: Cecidomyiidae) and spring wheat: implications for yield and seed quality. Can. Entomol. 132: 607–625.

Liu, X., H. Zhou, J. Zhao, H. Hua, and Y. He. 2016. Identification of the secreted watery saliva proteins of the rice brown planthopper, *Nilaparvata lugens* (Stål) by transcriptome and Shotgun LC-MS/MS approach. J. Insect Physiol. 89: 60–69.

Miles, P. W. 1999. Aphid saliva. Biol. Rev. 74: 41–85.

NDSU. 2016. Integrated pest management of the wheat midge in North Dakota. (https://www.ag.ndsu.edu/publications/crops/integrated-pest-management-of-the-wheat-midge-in-north-dakota) (accessed 15 May 2016).

Pechan, T., A. Cohen, W. P. Williams, and D. S. Luthe. 2002. Insect feeding mobilizes a unique plant defense protease that disrupts the peritrophic matrix of caterpillars. Proc. Natl. Acad. Sci. usa. 99: 13319–13323.

Shorthouse, J. D., and D. Rohrfrisch. 1992. Biology of insect-induced galls. Oxford University Press, New York.

Smith, M. A. H., I. L. Wise, S. L. Fox, C. L. Vera, R. M. DePauw, and O. M. Lukow. 2014. Seed damage and sources of yield loss by *Sitodiplosis mosellana* (Diptera: Cecidomyiidae) in resistant wheat varietal blends relative to susceptible wheat cultivars in western Canada. Can. Entomol. 146: 333–346.

Stuart, J. 2015. Insect effectors and gene-for-gene interactions with host plants. Curr. Opin. Insect Sci. 9: 56–61.

Thompson, J. N. 1998. Rapid evolution as an ecological process. Trends Ecol. Evol. 13: 329–332.

Thorpe, P. J. Cock, and J. Bos. 2016. Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. BMC Genomics 17: 172.

Toruno, T. Y., I. Stergiopoulos, and G. Coaker. 2016. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. Annu. Rev. Phytopathol. 54: 419–441.

Zhao, C., L. N. Escalante, H. Chen, T. R. Benatti, J. Qu, S. Chellapilla, R. M. Waterhouse, D. Wheeler, M. N. Andersson, R. Bao, et al. 2015. A massive expansion of effector genes underlies gall formation in the wheat pest *Mayetiola destructor*. Curr. Biol. 25: 613–620.

Zhao, C., R. Shukle, L. Navarro-Escalante, M. Chen, S. Richards, and J. J. Stuart. 2016. Avrulence gene mapping in the Hessian fly (*Mayetiola destructor*) reveals a protein phosphatase 2C effector gene family. J. Insect Physiol. 84: 22–31.