Journal of Insect Science, (2019) 18(5): 15; 1–9
doi: 10.1093/jisesa/iey094
Research

Conserved and Unique Putative Effectors Expressed in the Salivary Glands of Three Related Gall Midge Species

Zainab Al-Jbory,[1,2] Mustapha El-Bouhssini,[3] and Ming-Shun Chen[1,4,5]

[1]Department of Entomology, Kansas State University, 123 Waters Hall, Manhattan, KS 66506, [2]College of Agriculture, Green University of Al Qasim, Iraq, [3]International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat BP 6299, Morocco, [4]Hard Winter Wheat Genetics Research Unit, USDA-ARS and Department of Entomology, Kansas State University, Manhattan, KS 66506, and [5]Corresponding author, e-mail: mchen@ksu.edu

Subject Editor: Joanna Chiu

Received 10 May 2018; Editorial decision 7 September 2018

Abstract

Species in the stem gall midge genus *Mayetiola* (Diptera: Cecidomyiidae) cause serious damage to small grain crops. Among *Mayetiola* species are Hessian fly (*Mayetiola destructor Say*), barley midge (*Mayetiola hordei Keiffer*), and oat midge (*Mayetiola avenae Marchal*). Larvae of these species inject saliva into host tissues to manipulate plants. To identify putative effectors, transcriptomic analyses were conducted on transcripts encoding secreted salivary gland proteins (SSGPs) from first instar larvae of the barley and oat midges, since SSGPs are the most likely source for effector proteins delivered into host tissues. From barley midge, 178 SSGP-encoding unigenes were identified, which were sorted into 51 groups. From oat midge, 194 were obtained and sorted into 50 groups. Predicted proteins within a group had a highly conserved secretion signal peptide and shared at least 30% amino acid identity. Among the identified unigenes from both barley and oat midges, ~68% are conserved either among the three species or between two of them. Conserved SSGPs included members belonging to SSGP-1, SSGP-4, SSGP-11, and SSGP-71 families. Unconventional conservation patterns exist among family members within a species and among different gall midges, indicating that these genes are under high selection pressure, a characteristic of effector genes. SSGPs that are unique to each species were also identified. Those conserved SSGPs may be responsible for host manipulation since the three gall midges produce identical phenotypic symptoms to host plants, whereas the SSGPs unique to each species may be responsible for different host specificity.

Key words: barley midge, oat midge, transcriptome analysis, secreted salivary gland protein, insect effector

Cereal crops are attacked by various herbivorous insects (Gagné 1989). Among these insect pests are stem gall midge species from the genus *Mayetiola*, including the Hessian fly (*Mayetiola destructor* Say), the barley stem gall midge (barley midge, *M. hordei* Keiffer), and the oat stem gall midge (oat midge, *M. avenae* Marchal). These species are major insects causing severe damage yearly to their host plants in specific regions or worldwide (Gagné 2010).

Gall midges have four life stages: eggs, larvae, pupae, and adults. The Hessian fly, barley midge, and oat midge cause damage to host plants at the first two instar larval stages (also called maggots), whereas third instar larvae, pupae, and adults do not feed. First instar larvae attack plants by injecting salivary secretions into plant tissues to establish a permanent feeding site. Second instar larvae feed vigorously on plant tissues at the feeding site established by the first instar. The third instar becomes a nonfeeding puparium stage. The first instar is the most critical stage since it paves the way for the second instar larvae to feed. Larvae of the three gall midges are sessile and they feed by sucking up liquid from nutritive tissues (the gall) (Stuart et al. 2012). The adult stage of the life cycle is short (surviving for 1–3 d), and is specialized only for reproduction.

Morphologically, the three gall midge species are very similar at all four developmental stages. However, each species has its own preferred host plants. The preferred host for Hessian fly is wheat, although it can also survive on certain cultivars of barley and oat if no wheat plants are available (Gagné et al. 1991). Previous reports suggested that the preferred host for barley midge is barley, but it has also been recorded on wheat, oat, and rye (Gagné et al. 1991). The preferred host of oat midge is oat, but it can also occur on the wild plant *Avena fatua*, and wheat (Barnes 1956). However, in the field, the barley midge survives and develops well only on barley (Lhaloui 1995). Similarly, the oat midge can feed and survive well on oat, but not on wheat or barley (Barnes 1956).

Although each gall midge has its own preferred host plants, the three species cause almost identical symptoms to infested plants. Infested susceptible plants are stunted due to irreversible inhibition of plant growth, the color of leaves become dark green due to

increased chlorophylls, and the attacked plant will eventually die if no new tillers grow out. The molecular mechanisms required for the three midges to parasitize different host species, which cause identical symptoms to each host, remain to be delineated.

Interactions between the Hessian fly and wheat have been studied relatively extensively due to its important pest status in the United States (Stuart 2015). Previous research efforts have revealed that there are hundreds of secreted salivary gland proteins (SSGPs) produced in the salivary glands of Hessian fly larvae (Chen et al. 2004, 2008; Liu et al. 2004). Systematic transcriptomic analyses have revealed a high proportion (about 60%) of transcripts in salivary glands of first instar larvae that encode SSGPs (Chen et al. 2008). Genome sequencing revealed more than 7% of predicted genes in the Hessian fly genome encode effector-like proteins (Zhao et al. 2015). The great diversity in Hessian fly SSGPs has been indicated in many studies (Chen et al. 2008, 2010; Johnson et al. 2009). A large number of Hessian fly SSGPs contain consensus characteristics of parasite effectors (Zhao et al. 2015). In addition, several avirulence effectors have been cloned from the Hessian fly and all of them are SSGPs (Aggarwal et al. 2014; Zhao et al. 2015, 2016).

Based on Hessian fly data and studies on other sucking insects, SSGPs secreted into host tissues during feeding play critical roles in attacking plants, and are likely the main source for effector proteins (Hatchett et al. 1990, Harris et al. 2003, Chen et al. 2008). The Hessian fly, barley midge, and oat midge share the same feeding mechanism and result in almost identical symptoms on infested host plants, yet each species attacks only certain host plants and cannot survive on the others. Based on these biological observations, we hypothesize that these three midges inject conserved effectors into host tissues to manipulate host plants, resulting in similar symptoms such as inhibition of plant growth, suppression of host defense, and inducing nutritive tissue formation at the feeding site. We also hypothesize that the three midges inject some unique effectors into host tissues to allow each species to live on different host plants. The objectives of this study are to identify SSGP-encoding transcripts from the first instar larvae of barley midge and oat midge, and to compare these SSGPs among the Hessian fly, barley midge, and oat midge for identification of SSGPs conserved among the three species and SSGPs unique to each of them.

Materials and Methods

Insects and Salivary Gland Dissection

Salivary gland preparation was conducted in the Entomological Research Laboratory at the International Center for Agricultural Research in the Dry Area (ICARDA) in Rabat, Morocco. Larvae for dissection were obtained directly from plants in the fields that were infested naturally either by barley midge or oat midge. Three-day-old larvae were selected for dissection. At least 300 pairs of salivary glands from a species were pooled together for RNA extraction and library construction. Dissection was achieved in saline buffer 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 internal tissues. Clean salivary glands were then obtained by removing unwanted tissues (Al-Jbory et al. (2018). Salivary glands were put into TRI reagent (Molecular Research, Inc., Cincinnati, OH) as soon as they were dissected from the larvae of each species. The salivary samples in TRI reagent were frozen in liquid nitrogen and were shipped to the USDA Hessian fly research laboratory at Manhattan, KS, United States, for further processing and analyses.

RNA Extraction, cDNA Library Construction, and DNA Sequencing

Each RNA sample was extracted from a pool of 300 pairs of salivary glands from a species. Total RNA was extracted using TRI reagent following the protocol provided by the manufacturer. RNA quality and integrity were analyzed using a TapeStation Bioanalyzer (Agilent Technologies, Santa Clara, CA). The RNA samples were then reverse-transcribed to cDNAs. The cDNA samples were amplified using a 'SMART' library construction kit from Clontech (Palo Alto, CA) as described by Chen et al. (2004). Briefly, amplified cDNA inserts were ligated into the pPCRXL-TOPO plasmid contained in a TOPO TA cloning kit (Invitrogen, Carlsbad, CA). The ligated plasmids were then transformed into individual bacteria. Bacterial clones were picked up individually for plasmid DNA isolation, which were sequenced with the M13 forward and reverse primers via a commercial contract (GENEWIZ, South Plainfield, NJ).

Sequence Analysis

Based on our previous experience with Hessian fly, cDNA clones were sequenced following the traditional Sanger DNA sequencing method because SSGP genes in *Mayetiola* species are unconventionally conserved among family members (Chen et al. 2010, Zhao et al. 2015). Highly diversified coding regions surrounded by highly conserved untranslated regions (UTRs) within a short sequence prevent SSGP transcripts to be assembled correctly from short sequence reads produced from a next generation sequencing approach. In addition, the lack of genome sequences for both barley and oat midges makes Sanger-derived sequences more reliable than contigs assembled from short sequence reads.

After sequencing, vector sequences were trimmed manually from raw reads. Pairwise alignment 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. In our analysis, we have identified unigenes and groups. Each unigene represents a cDNA encoding a unique protein and each unigene may contain multiple redundant clones. Unigenes were further sorted into groups based on sequence similarity.

Local databases were established with the nucleotide sequences of the cDNAs from barley and oat midges, separately. Cluster analyses of these cDNAs were conducted using BlastStation-Local 64 program (https://www.blaststation.com/intl/en/local64.php). To verify the similarity of sequences in each clustered group, multiple alignments of nucleotide sequences were conducted using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Maximum likelihood phylogenetic tree was generated, and representative proteins from each group were aligned using CLUSTAL(W) implemented in MEGA program with 100 bootstrap replicates. The purpose of the phylogenetic tree was merely to further confirm the validity of the groups for each species, and not to measure the genetic distance.

Similarity analyses between sequences from different species were performed using Blastx or Blastp for two sequence comparison (http://www.ncbi.nlm.nih.gov/). 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 of the regions with sequence similarity and gaps within the alignments. Secretion signal peptides were identified using SignalP

version 4 (www.cbs.dtu.dk/services/SignalP/) after individual nucleotide sequences were translated into amino acid sequences (https://www.ncbi.nlm.nih.gov/orffinder/). To identify homologous members of SSGPs groups among the three species, reciprocal blast was conducted using BlastStation-Local 64 program for the local database that was already established for the nucleotide sequences of the three species.

The Genetic Variation of Genes Within and Between Species

To test sequence variations of genes within a species, the mutation rates among SSGP members within a group in either barley midge or oat midge were estimated. Nucleotide sequences of members from each group were selected and aligned using CLUSTAL(W). Total numbers of nonsynonymous and synonymous mutations and *dN/dS* (ω) ratio were calculated using codon-based *Z*-test selection from the MEGA 5.2.2 program. The alternative hypothesis that used in this analysis was positive selection, namely *HA*: $dN > dS$ at *P*-values smaller than 0.05. The variance of the difference was computed using the bootstrap method (100 replicates). Analyses were conducted using the Nei–Gojobori method. This method not only computes the numbers of synonymous and nonsynonymous substitutions per site, but also the numbers of potentially synonymous and potentially nonsynonymous substitutions per site (Nei and Gojobori 1986). Eleven groups from barley midge and nine groups from oat midge were selected for this analysis since each group has multiple unigenes.

Phylogenetic analyses were conducted using amino acid sequences. Members from the four conserved families, SSGP-1, SSGP-4, SSGP-11, SSGP-71, were aligned using CLUSTAL(W). Maximum likelihood trees based on Jones-Taylor-Thornton (JTT) model with 100 bootstrap were generated using the MEGA program. Secreted proteins from salivary glands of wheat midge larvae (*Sitodiplosis mosellana*) were used as reference in this analysis (Al-Jbory et al. 2018).

Results

Composition and Classification of Transcripts From Barley Midge and Oat Midge

To identify SSGP-encoding transcripts from both species, cDNA clones were sequenced and only high quality sequences were retained. From the barley midge, 2,570 cDNA clones were sequenced, and 743 cDNA sequences were retained. Among the good quality cDNAs, 458 (61.6%) encode SSGPs and the remaining 285 (38.4%) encode proteins without a typical secretion signal peptide. The 458 SSGP-encoding cDNAs were sorted into 178 unigenes based on sequence similarity, and each unigene encoded a unique protein. Among the 178 unigenes, 102 (57.3%) had sequence similarity to SSGPs from Hessian fly, and remaining 76 (42.7%) with no sequence similarity to any Hessian fly sequences or any known sequences in GenBank (Supp Fig. S1 and Supp Table S1 [online only]).

From the oat midge, 3,226 cDNA clones were sequenced, and 718 cDNA sequences were retained. Among those cDNAs, 450 (62.7%) encode SSGPs and the remaining 268 (37.3%) are non-SSGPs. The 450 SSGP-encoding cDNAs were grouped into 194 unigenes. Among the 194 unigenes, 107 (55%) had sequence similarity to SSGPs from Hessian fly, and remaining 87 (45%) with no sequence similarity to Hessian fly sequences or any known sequences in GenBank (Supp Fig. S2 and Supp Table S2 [online only]).

According to sequence similarity among the cDNAs and derived proteins, SSGP-encoding transcripts from barley midge were classified into 51 groups (Supp Table S1 [online only]). Among the 51 groups, 34 had either a single clone or multiple clones that encode the same protein. The remaining 17 groups had multiple clones that encode at least two different proteins. Like barley midge, SSGP-encoding transcripts from oat midge were classified into 50 groups (Supp Table S2 [online only]). Thirty-one had either a single clone or multiple clones that encoded the same protein. The remaining 19 groups had multiple clones that encoded at least two different proteins.

In both barley and oat midges, proteins within a group had a highly conserved secretion signal peptide (more than 90% identity) and shared at least 30% amino acid identity within the mature protein region. Proteins between different groups had a completely different secretion signal peptide and shared no meaningful ($E > 10^{-3}$) sequence similarity (Supp Figs. S1 and S2 [online only]). Figures 1 and 2 show amino acid sequence alignments of two representative groups from barley midge and oat midge, respectively. These 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 (Chen et al. 2008). Some sequence variation may have also resulted from different alleles of the same gene.

Non-SSGP transcripts from barley midge and oat midge were also identified. There were 285 non-SSGP transcripts from barley midge, 94 (33%) encoded proteins with no sequence similarity to any proteins in GenBank, 58 (20.3%) encoded proteins with sequence similarity to proteins with unknown function, and the remaining 133 (46.7%) encoded proteins with sequence similarity to proteins with various functions. For the transcripts encoding known proteins, 58 (43.6%) were proteins with functions in protein synthesis and the remaining 75 (56.4%) had other housekeeping functions, including energy-metabolic enzymes, structural proteins, and transporters (Supp Table S3 [online only]). From oat midge, there were 268 non-SSGP transcripts, 105 (39.2%) encoded proteins with no sequence similarity to any proteins in GenBank, 55 (20.5%) encoded proteins with sequence similarity to proteins with unknown function, and the remaining 108 (40.3%) encoded proteins with sequence similarity to proteins with various functions. Like barley midge, the transcripts encoding known proteins from oat midge have mainly protein synthesis functions 63 (58%), and the remaining 45 (42%) are housekeeping proteins (Supp Table S4 [online only]).

All EST sequences were deposited to GenBank. The accession number for the barley midge cDNA library is LIBEST_028854. The accession number for the oat midge cDNA library is LIBEST_028855. The accession numbers for barley midge ESTs are JZ977793 to JZ978250. The accession numbers for oat midge ESTs are JZ978251 to JZ978699.

SSGPs Conserved Among the Gall Midge Species

To identify SSGPs conserved among the three species or between two of them, comparative analyses were conducted via local Blast and sequence alignments. Many SSGP groups from both barley and oat midges (Supp Table S5 [online only]) share sequence similarity with the previously identified SSGP families from Hessian fly (Chen et al. 2008, 2010). The conserved families identified among the three species include SSGP-1, SSGP-4, SSGP-11, SSGP-71.

Family SSGP-1 (family 1) is the most abundant group of SSGPs in Hessian fly and it includes SSGP-1A, SSGP-1B, SSGP-1C, SSGP-1D,

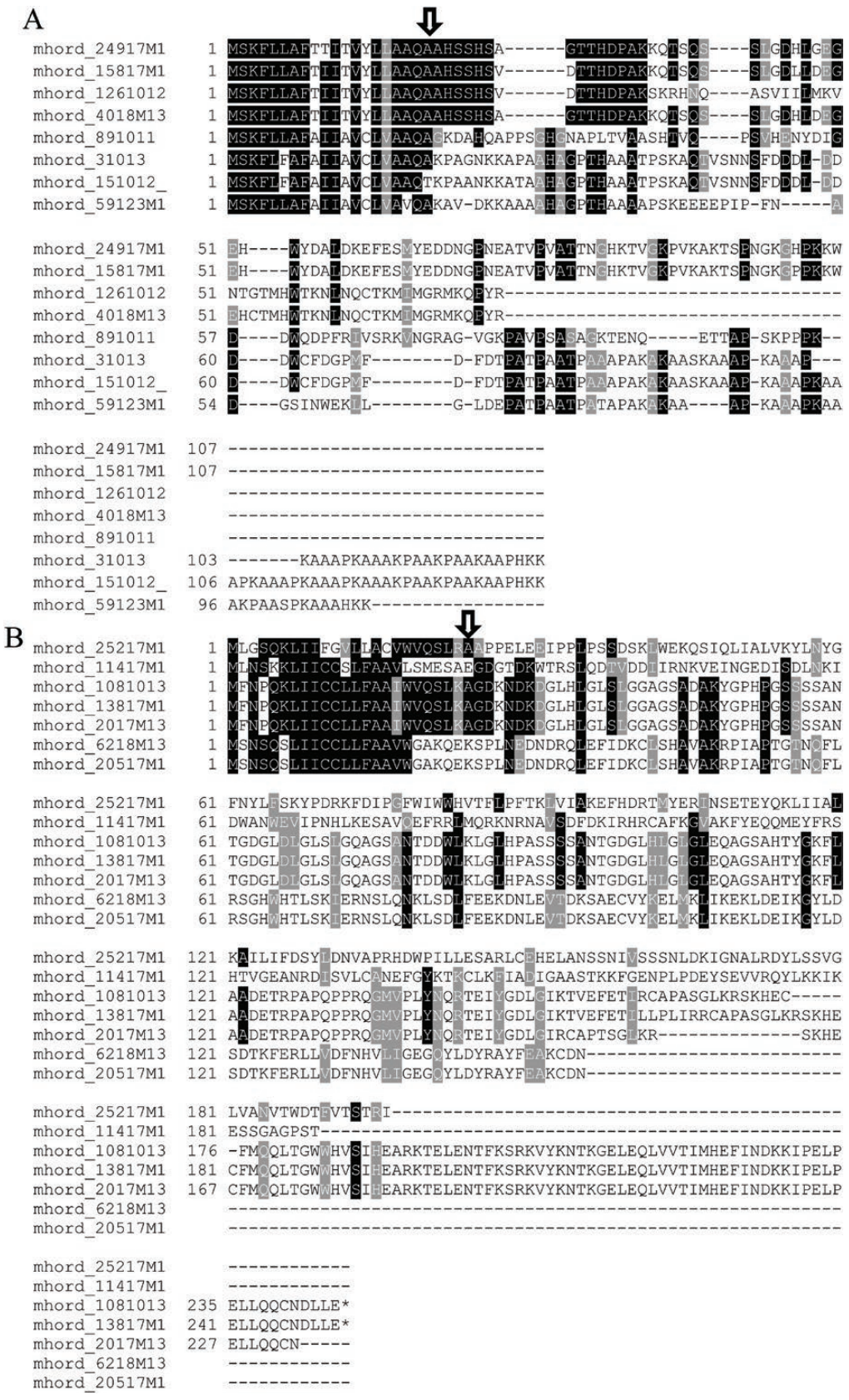

Fig. 1. Amino acid sequence alignments of two representative groups from barley midge. The boundary between predicted signal peptide and mature proteins is indicated by an arrow. Black shading indicates similarity of amino acids residues. (A) Sequence alignment of predicted amino acid sequences from unigenes in group 1. (B) Sequence alignment of predicted amino acid sequences from unigenes in group 28.

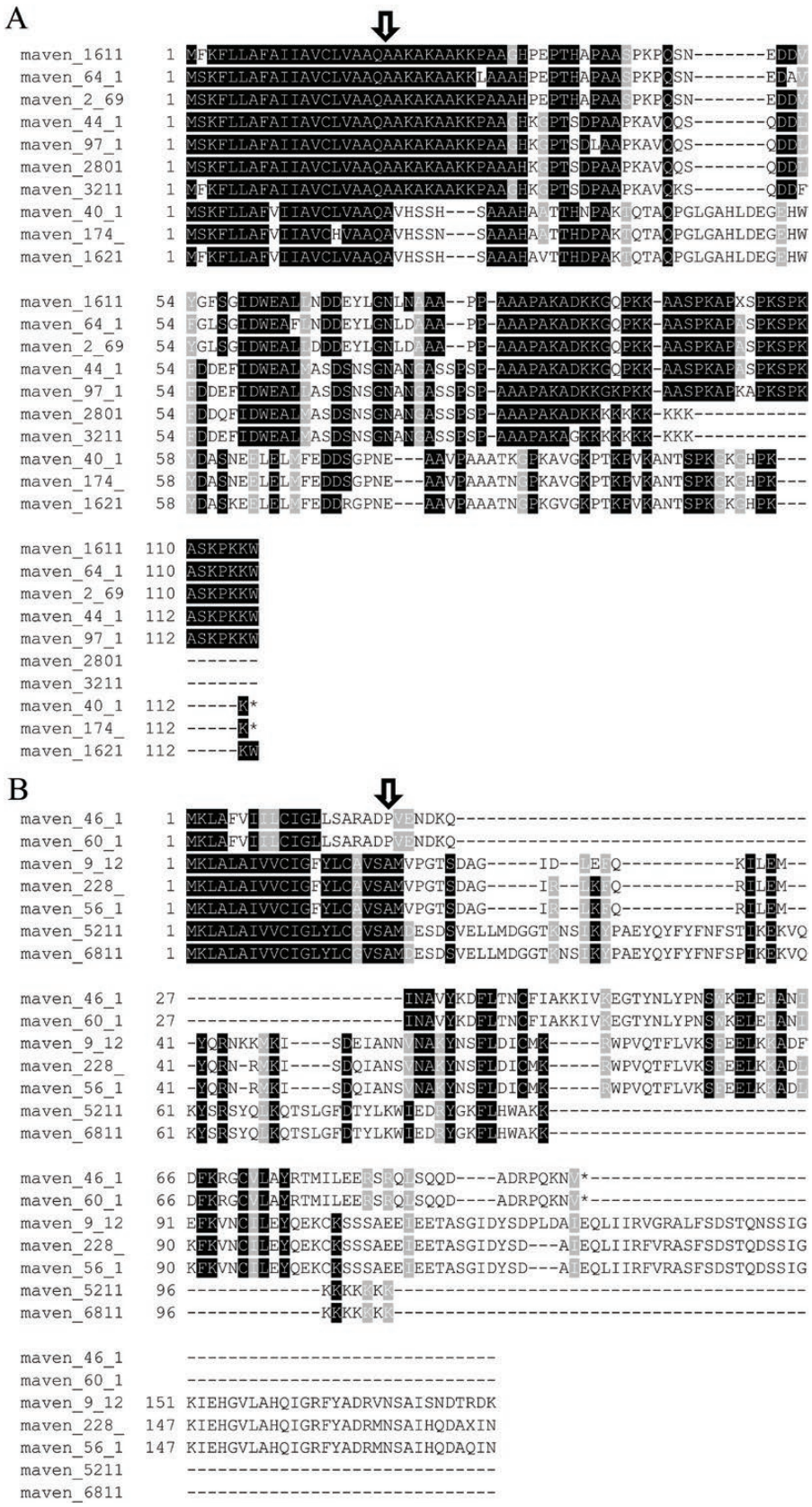

Fig. 2. Amino acids sequence alignments of two representative groups from oat midge. The boundary between predicted signal peptide and mature proteins is indicated by an arrow. Black shading indicates similarity of amino acids residues. (A) Sequence alignment of predicted amino acid sequences from unigenes in group 3. (B) Sequence alignment of predicted amino acid sequences from unigenes in group 17.

and SSGP-1E (Chen et al. 2010). In this study, we identified many SSGPs from barley and oat midges that share sequence similarity to members belonging to the Hessian fly SSGP-1 family. There are 26 unigenes from barley midge that share high similarity ($E < 10^{-3}$) to Hessian fly SSGP-1C1, and five unigenes encoding proteins with similarity ($E < 10^{-3}$) to SSGP-1D1. From oat midge, there are 18 unigenes encoding proteins that share high similarity ($E < 10^{-5}$) to Hessian fly SSGP-1C1, and five unigenes encoding proteins that share high similarity ($E < 10^{-4}$) to SSGP-1D1 (Supp Table S5 [online only]). Sequence alignments revealed identical or nearly identical signal peptides in SSGP-1C1 (Supp Fig. S3-1A [online only]) and SSGP-1D1 (Supp Fig. S3-1B [online only]) sequences from the three gall midges, and over 80% and 60% identities, respectively, among the mature proteins.

We have also identified SSGPs which are conserved either between barley midge and Hessian fly or between oat midge and Hessian fly. There were 47 unique genes from barley midge encoding proteins that share sequence similarity to Hessian fly SSGP-1C2 (Supp Table S5 and Supp Fig. S3-2A [online only]), and 67 unigenes from oat midge encoding proteins that share similarity with Hessian fly SSGP-1A2 (Supp Table S5 and Supp Fig. S3-2B [online only]).

Family SSGP-4 (family 4) from Hessian fly has 11 identified groups: SSGP-4A, SSGP-4B, SSGP-4C, SSGP-4D, SSGP-4E, SSGP-4F, SSGP-4G, SSGP-4H, SSGP-4I, SSGP-4J, SSGP-4K (Chen et al. 2008). In this study, SSGPs were identified from both barley midge and oat midge that are homologous to family SSGP-4 members from Hessian fly. Eight unigenes from barley midge and three unigenes from oat midge encoded proteins with sequence similarity to SSGP-4 Hessian fly members (Supp Table S5 [online only]). All these unigenes are homologous to group SSGP-4A (Supp Fig. S3-3 [online only]).

Family SSGP-11 from Hessian fly includes three groups: SSGP-11A, SSGP-11B, and SSGP-11C (Chen et al. 2006, 2008). In this study, we identified 11 unigenes from barley midge and eight unigenes from oat midge encoding proteins homologous to SSGP-11 (Supp Table S5 [online only]). All these unigenes are homologous to group SSGP-11B, except one unigene from oat midge, which exhibited similarity to group SSGP-11C. Sequence alignments of SSGP-11B members from the three midges showed that those proteins share high similarity in both the signal peptide and mature protein regions (Supp Fig. S3-4 [online only]).

Family SSGP-71 has 426 members identified from Hessian fly. Among the Hessian fly SSGPs, 14% of these proteins are partially orthologous to other arthropods, whereas 86% have no homology to any organisms (Zhao et al. 2015). Members belonging to this family have been found to encode larger proteins (over 400 amino acids). In this study, we identified three unigenes from barley midge and two unigenes from oat midge (Supp Table S5 [online only]) that are homologous to Hessian fly SSGP-71 proteins (Supp Fig. S3-5A and 5B [online only]).

Eight groups of SSGPs were found conserved between only two species. These are groups 19, 20, 23, 24, 29, 32, 40, and 51 from barley midge and groups 8, 15, 24, 39, 23, 21, 30, and 35 from oat midge (Supp Table S6 [online only]). In these eight groups, 19 (10.7%) unigenes from the barley midge and 25 (12.9%) from oat midge were found to be conserved between the barley and oat midges, and no homologues were found in Hessian fly. Based on blast results against the NCBI database, the functions of these eight groups are not known, except groups 24 and 51 from barley midge and groups 30 and 35 from oat midge, which have been identified as larval cuticle protein and peptidyl-prolyl cis-trans isomerase, respectively (Supp Table S6 [online only]).

SSGPs Unique to Each Species

There were 37 groups with 56 unigenes that were unique to barley midge. Among these groups, eight groups (groups 4, 5, 6, 13, 15, 18, 33, 34) had at least two or more unigenes. The remaining groups had only one unigene in each group (Supp Table S1 [online only]). There were 37 groups with 62 unigenes that were unique to oat midge. Among the 37 groups, 10 groups (groups 2, 3, 4, 6, 9, 12, 14, 16, 17, and 19) had at least two unigenes. The remaining groups had only one unigene (Supp Table S2 [online only]). All these groups from both barley and oat midges were identified with unknown functions based on blast results against the NCBI database, with the exception of group 35 from barley midge, which encodes a kinase domain protein (Supp Table S1 [online only]).

Sequence Variations Within Group Members of Each Species

Mutation rate of the selected groups from barley midge and oat midge were estimated using a MEGA program. Among the 11 groups from barley midge that were analyzed, the *dN*/*dS* ratio indicated positive selection in six of the groups. They are groups 1, 6, 19, 31, 33, 34 (Table 1). Although the $\omega > 1$ indicated positive selection in these groups, the overall average *P*-value was only significant for three groups. The *P*-values for the remaining groups were significant individually based on sequence pair alignments of *Z*-test combinations.

The reason behind this result is that members in these groups are highly diversified with only 30–90% sequence identity within the mature protein. High variations among tested sequences could not produce correct alignments for analyzing *dN*/*dS* ratio. When we tested these sequences for neutral selections, *P*-values were not significant neither in the overall average nor in sequence pair alignments. Evidence of positive selection could not be efficiently detected using MEGA tools due to too much diversification. However, the fact that the coding regions are hard to align (like random sequences) is itself evidence for fast evolution by positive selection (Chen et al. 2010).

Among the nine groups that were tested from oat midge, the *dN*/*dS* ratio indicated positive selection in five groups, these are groups 1, 3, 4, 15, 17 (Table 2).

Sequence Variations of Conserved Members Among the Three Species

Genetic distance of homologous members among the three species was tested by phylogenic trees for members from each of the four conserved families (Supp Fig. S4 [online only]). Based on the genetic distance, the homologous genes from each family appear to have a common ancestor among the three midges.

Discussion

Like other plant-sucking insects, gall midges secrete effector proteins into their host, inducing various forms of plant outgrowth (galls) (Dieleman 1969, Hori 1992). In the genus *Mayetiola*, the three stem midges, Hessian fly, barley midge, and oat midge, do not induce typical outgrowth galls, but they can inhibit plant growth and induce nutritive tissues at the feeding site. The three gall midges share the same feeding mechanism and cause identical symptoms to host plants. However, each species infests a different host plant. The genetic mechanisms behind this type of peculiar feeding mechanism and molecular determinants for different host specificity are not known, but presumably are linked with effectors injected into host plants during feeding (Hatchett et al. 1990). Like other insect species, secreted proteins from salivary glands are the major source of effectors for plant manipulation and host selection (Chen et al.

Table 1. Mutation ratio ω among representative members from barley midge groups

Group no.	Σ*dN*	Σ*dS*	ω = *dN*/*dS*	*P*-value 0.05
1	102.18	99.902	1.022	0.049**
4	11.978	4.739	0.395	1.000
6	37.896	29.678	1.276	0.013**
18	10.229	3.934	0.384	1.000
19	31.64	25.108	1.260*	0.086
20	44.363	50.791	0.873	1.000
28	166.47	184.22	0.903	1.000
31	68.515	54.79	1.250*	0.106
32	2.847	6.713	0.424	1.000
33	21.554	8.512	2.532	0.001**
34	11.228	9.448	1.188*	1.000

**dN*/*dS* > 1, and *P*-value is individually significant based on in sequence pairs values of the *Z*-test.

***P*-value is significant based on overall average value of the *Z*-test.

Table 2. Mutation ratio ω among representative members from the oat midge groups

Group no.	Σ*dN*	Σ*dS*	ω = *dN*/*dS*	*P*-value 0.05
1	67.171	48.248	1.394*	0.094
3	18.838	9.586	1.964	0.023**
4	13.744	11.071	1.241*	1.000
5	51.729	94.829	0.545	1.000
8	4.283	5.880	0.728	1.000
15	68.463	52.094	1.314*	0.118
17	12.373	5.334	2.319	0.007**
22	130.48	146.03	0.893	1.000
24	219.57	223.31	0.983	0.49

**dN*/*dS* > 1, and *P*-value is individually significant based on in sequence pairs values of the *Z*-test.

***P*-value is significant based on the overall average value of the *Z*-test.

2008, 2010). Accordingly, identification and comparative analyses of SSGPs from the three gall midge species may lead to discovery of key effectors responsible for manipulating host plants, and molecular determinants that allow a specific gall midge species to live on a specific host plant.

In this study, we first identified SSGPs from barley midge and oat midge via a transcriptomic approach. Over 60% of total transcripts from both species were SSGP-encoding transcripts, which is similar to that observed in Hessian fly (Chen et al. 2008). In addition to the high percentage of SSGP-encoding transcripts, there was also high percentage of transcripts encoding proteins that were either involved directly in protein synthesis or in housekeeping functions related to protein synthesis (Supp Tables S2 and S4 [online only]). This suggested that the larval salivary glands of the three stem gall midges are specialized tissues for synthesis of proteins for host injection. This high proportion of SSGPs have also revealed from two other gall midges, the wheat midge (*S. mosellana*) (Al-Jbory et al. 2018), and the Asian rice gall midge (*Orseolia oryzae*) (Chen et al. 2010).

Our comparative analysis revealed additional similarities among SSGPs from the three species. Similar features of SSGPs from the three gall midge species included 1) high proportion of conserved SSGPs; 2) small sizes for the majority of putative SSGPs; 3) lack of similarity to other proteins in GenBank database; and 4) the existence of the so-called unconventional conservation pattern among gene family members (Chen et al. 2010). The differences among SSGPs from the three midges are mainly the diversity of the mature proteins.

Among SSGPs conserved in the three gall midges are members from family SSGP-1. SSGP-1 family members encode small proteins with 80–100 amino acids, and represent the most abundant groups in terms of transcripts in the salivary glands of first instar larvae for all the three gall midges. Based on this study, 78 out of 178 unigenes from barley midge, and 90 of 194 unigenes from oat midge were from the SSGP-1 family (Supp Table S5 [online only]). Based on a previous study, around 30% of total salivary gland transcripts from Hessian fly were from this family (Chen et al. 2008). Members from the SSGP-1 family have also the consensus motif that is similar to those observed in some pathogen effectors (Jiang et al. 2008, Petre and Kamoun 2014). SSGP-1 members are also exclusively expressed in the crucial larval stage for host infestation (Chen et al. 2008, 2010). All these characteristics suggest that members in the SSGP-1 family may play important roles in host manipulation and virulence of the three midges.

Other SSGPs conserved among the three species were from families SSGP-4, SSGP-11, and SSGP-71. Members from SSGP-4 and SSGP-11 families encode small proteins with similar characteristics as in SSGP-1 family members except with less abundant transcripts and lack the consensus motif. These conserved members may also play roles in the conserved feeding mechanism among the three gall midges. On the other hand, members from the SSGP-71 family encode relatively large proteins with more than 400 amino acid residues (Zhao et al. 2015). It is believed that SSGP-71 proteins are a novel class of F-box-LRR (leucine-rich repeats) mimics, which allow larvae of Hessian fly to control the plant proteasome to induce the nutritive cells and ultimately stunt plant growth (Zhao et al. 2015). The existence of SSGP-71 homologs in barley and oat midges suggests similar effector roles of these conserved members in barley and oat midges, which are critical in interactions between gall midges and host plants. Some SSGPs were found conserved only between barley and oat midges with no corresponding homologues in Hessian fly (Supp Table S6 [online only]). These SSGPs are likely absent in Hessian fly since the Hessian fly genome has been sequenced and the Hessian fly salivary gland transcriptome has been analyzed extensively. We postulate that the SSGPs conserved between barley and oat midges contribute to the ability of these two insects to attack barley and oat plants, which share a closer evolutionary relationship than to wheat (Kellogg 1998).

Most of the identified SSGP transcripts from barley and oat midges encode proteins with 50–180 amino acid residues. All predicted small SSGPs are unique with no meaningful ($E > 0.001$) sequence similarity to any known proteins in GenBank. Therefore, the functions of these SSGPs remained to be determined. Small proteins are used as effectors by other insects, such as MpC002 and Me10 from aphids, which are less than 150 kD (Thorpe et al. 2016). Small secreted proteins have been also found to play crucial roles in interactions between pathogens and plants (Lyu et al. 2016). One apparent advantage for small effectors is that they may be easier to penetrate into host tissues once they are injected into host plants.

One of the common features of SSGPs from the three gall midges is that many gene families exhibit unconventional conservation among family members. Unconventional conservation was first discovered in SSGP genes from Hessian fly, which exhibited highly conserved 5′- and 3′-UTRs, regions encoding signal peptides, and introns, but highly diversified regions that encode mature proteins (Chen et al. 2010, Zhao et al. 2015). In this study, the unconventional conservation was also found among SSGP gene family members from barley midge (Supp Fig. S5-1 [online only]) and oat midge

(Supp Fig. S5-2 [online only]). Sequence alignments of the 5′- and 3′-UTRs for family 1 members revealed that these regions are not only highly conserved within a midge, but are also highly conserved among the three midges (Supp Fig. S5-3 [online only]). The reason for the unconventional conservation pattern remains unknown. One possibility is that the UTRs are conserved for regulation of gene expression, or that the UTRs are highly conserved because these genes arose through gene duplications relatively recently, and have not enough time to be diversified under random mutation, whereas the coding regions are highly diversified under strong positive selection pressure (Chen et al. 2010). The unconventional conservation has been also found in some of the SSGP cDNAs from the Asian rice gall midge *O. oryzae* (Chen et al. 2010), but not from the wheat midge *S. mosellana* (Al-Jbory et al. 2018).

Despite conservation of UTRs among homologous SSGP gene families among the three gall midges, the mature SSGP proteins are highly diversified. Our analyses showed that none of the identified SSGPs share 90% or greater sequence identity between proteins obtained from different insect species. The diversification may reflect positive selection for mutations in effectors to avoid host recognition for defense responses. Positive selection appears to be one of the forces driving diversifications in the coding regions of some of the SSGP groups among the three species based on *dN/dS* ratio from this study (Tables 1 and 2), and previous studies (Chen et al. 2004, 2010). The fast-evolving nature of SSGP-encoding genes in all three species is another indicator that these genes were under high selection pressure for functional adaptation during coevolution with the host plant (Thompson 1998).

In addition to the SSGPs conserved among the three species or between two of them, 32% of unigenes from barley midge and oat midge share no homology to SSGPs from Hessian fly or any other species (Supp Tables S1 and S2 [online only]). Considering that each gall midge species parasitizes different host plants, these unique putative effectors might be responsible for host specificities. However, the uniqueness of SSGPs in different species could be due to the fact that some of them have not been identified due to low coverage of transcripts of barley and oat midges. The reason we adapted a traditional library construction and sequencing strategy is because SSGP genes in *Mayetiola* species are unconventionally conserved among family members (Chen et al. 2010, Zhao et al. 2015), which prevent correct assembly of short sequence reads obtained from a high-throughput sequencing approach (for example, Illumina sequencing). More comprehensive studies should lead to the identification of more conserved and unique SSGPs from barley midge and oat midge.

In summary, we conducted transcriptomics analyses on genes expressed in the salivary glands of first instars of the barley midge and oat midge for the first time. Our analyses have identified many putative effectors that are produced in the saliva of these two species. When putative effectors from barley midge and oat midge were compared to those from Hessian fly, many of them were conserved among the three midges, indicating conserved roles that are likely associated with feeding mechanisms. We have also identified many putative effectors that are unique to each midge. We postulate that some of these unique effectors might play roles in host specification. Identification of conserved and unique putative effector genes provides a foundation for further characterization for the roles of these genes in gall midges–host plant interactions. We believe that getting insight about the functions of these effectors could be useful to reveal feeding mechanisms of these destructive pests.

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 66506. Z.A.-J. conducted experiments and analyzed data. Z.A.-J., M.-S.C., and M.E.-B. wrote the paper. M.E.-B. provided dissected salivary glands. M.-S.C and M.E.-B. conceived the research and designed experiments.

References Cited

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

Al-Jbory, Z., K. M. Anderson, M. O. Harris, O. Mittapalli, R. J. Whitworth, and M. Chen. 2018. Transcriptomic analyses of secreted proteins from the salivary glands of wheat midge larvae. J. Insect Sci. 18: 17.

Barnes, H. F. 1956. Gall midges of economic importance. Vol. VII: gall midges of cereal crops. Crosby Lockwood, London, UK.

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., J. P. Fellers, Y. C. Zhu, J. J. Stuart, S. Hulbert, M. El-Bouhssini, and X. Liu. 2006. A super-family of genes coding for secreted salivary gland proteins from the Hessian fly, *Mayetiola destructor*. J. Insect Sci. 6: 1–13.

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.

Dieleman, F. 1969. Effects of gall midge infestation on plant growth and growth regulating substances. Entomol. Exp. Appl. 12:745–749.

Gagné, R. J. 1989. The plant-feeding gall midges of North America. Cornell Univ. Press, Ithaca, NY.

Gagné, R. J. 2010. Update for a catalog of the Cecidomyiidae (Diptera) of the world. Entomological Society of Washington, Washington, DC.

Gagné, R. J., J. Hatchett, S. Lhaloui, and M. El Bouhssini. 1991. Hessian fly and barley stem gall midge, two different species of Mayetiola (Diptera: Cecidomyiidae) in Morocco. Ann. Entomol. Soc. Am. 84:436–443.

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

Hatchett, J., G. Kreitner, and R. Elzinga. 1990. Larval mouthparts and feeding mechanism of the Hessian fly (Diptera: Cecidomyiidae). Ann. Entomol. Soc. Am. 83: 1137–1147.

Hori, K. 1992. Insect secretions and their effect on plant growth, with special reference to hemipterans, pp. 157–170. *In* Biology of insect-induced galls. Oxford Univ. Press, NY.

Jiang, R. H., S. Tripathy, F. Govers, and B. M. Tyler. 2008. RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. Proc. Natl. Acad. Sci. USA. 105: 4874–4879.

Johnson, A., M. S. Chen, P. Morton, S. Cambron, and R. H. Shukle. 2009. Analyzing the diversity of secreted salivary gland transcripts in Hessian fly populations from Israel and the United States. NCB-ESA North Central Branch Entomological Society of America, 15–18 March, Hilton at the Ballpark in St. Louis, MO.

Kellogg, E. A. 1998. Relationships of cereal crops and other grasses. Proc. Natl. Acad. Sci. USA. 95: 2005–2010.

Lhaloui, S. 1995. Biology, host preference, host suitability and plant resistance studies of the barley midge and Hessian fly (Diptera: Cecidomyiidae) in Morocco. Ph.D. dissertation, Kansas State University, Manhattan, KS.

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

Lyu, X., C. Shen, Y. Fu, J. Xie, D. Jiang, G. Li, and J. Cheng. 2016. A small secreted virulence-related protein is essential for the necrotrophic interactions of *Sclerotinia sclerotiorum* with its host plants. PLoS Pathog. 12: e1005435.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3: 418–426.

Petre, B., and S. Kamoun. 2014. How do filamentous pathogens deliver effector proteins into plant cells? PLoS Biol. 12: e1001801.

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

Stuart, J. J., M. S. Chen, R. Shukle, and M. O. Harris. 2012. Gall midges (Hessian flies) as plant pathogens. Annu. Rev. Phytopathol. 50: 339–357.

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

Thorpe, P., 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.

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. Avirulence gene mapping in the Hessian fly (*Mayetiola destructor*) reveals a protein phosphatase 2C effector gene family. J. Insect Physiol. 84: 22–31.