Insect Molecular Biology

Insect Molecular Biology (2015) **24**(2), 191–202 doi: 10.1111/imb.12148

Differential expression of candidate salivary effector proteins in field collections of Hessian fly, *Mayetiola destructor*

A. J. Johnson*†, R. H. Shukle*†, M.-S. Chen‡§, S. Srivastava¶, S. Subramanyam, B. J. Schemerhorn*†, P. G. Weintraub††, H. E. M. Abdel Moniem†§§, K. L. Flanders¶¶, G. D. Buntin*** and C. E. Williams*,****

**USDA-ARS Crop Production and Pest Control Research Unit, West Lafayette, IN, USA; †Department of Entomology, Purdue University, West Lafayette, IN, USA; ‡USDA-ARS Hard Winter Wheat Genetics Research Unit, Manhattan, KS, USA; §Department of Entomology, Kansas State University, Manhattan, KS, USA; ¶Department of Statistics, Duke University, Durham, NC, USA; **Department of Agronomy, Purdue University, West Lafayette, IN, USA; ††Agricultural Research Organization, Gilat Research Center, D.N. Negev, Israel; §§Department of Zoology, Faculty of Science, Suez Canal University, Ismailia, Egypt; ¶¶Department of Entomology and Plant Pathology, Auburn University, Auburn, AL, USA; and ***Department of Entomology, University of Georgia College of Agriculture Experiment Stations, Georgia Station, Griffin, GA, USA*

Abstract

Evidence is emerging that some proteins secreted by gall-forming parasites of plants act as effectors responsible for systemic changes in the host plant, such as galling and nutrient tissue formation. A large number of secreted salivary gland proteins (SSGPs) that are the putative effectors responsible for the physiological changes elicited in susceptible seedling wheat by Hessian fly, *Mayetiola destructor* (Say), larvae have been documented. However, how the genes encoding these candidate effectors might respond under field conditions is unknown. The goal of this study was to use microarray analysis to investigate variation in SSGP transcript abundance amongst field collections from different geographical regions (southeastern USA, central USA, and the Middle East). Results revealed significant variation in SSGP transcript abundance amongst the field collections studied. The field collections separated into three distinct groups that corresponded to the wheat classes grown in the different geographical regions as well as to recently described Hessian fly populations. These data support previous reports correlating Hessian fly population structure with micropopulation differences owing to agro-ecosystem parameters such as cultivation of regionally adapted wheat varieties, deployment of resistance genes and variation in climatic conditions.

Keywords: *Mayetiola destructor*, gall midges, secreted salivary effector proteins, wheat, plant–insect interactions.

First published online 22 December 2014.

Correspondence: Richard H. Shukle, USDA-ARS, Department of Entomology, Purdue University, West Lafayette, IN, 47907, USA. Tel.: 765-494-6351; e-mail: shukle@purdue.edu

Introduction

Proteins and other molecules secreted by the salivary glands of phytophagous insects have been proposed to act as ‘effectors’ that, when injected into their hosts, facilitate and enhance penetration by mouthpart stylets, initiate digestion of host-cell contents for ingestion, and suppress host defence responses, resulting in the modification and manipulation of host processes in a manner advantageous to the pest (Hori, 1992; Alfano, 2009; Hogenhout *et al.*, 2009; Hogenhout & Bos, 2011; DeLay *et al.*, 2012). It has recently been hypothesized that, during gall formation, the Hessian fly, *Mayetiola destructor* (Say) (Diptera: Cecidomyiidae), uses an effector-based strategy that is similar to biotrophic plant pathogens (Stuart *et al.*, 2012). This gall midge is a recurrent pest of wheat, *Triticum aestivum* L., in many of the wheat production areas

Published 2014. This article is a U.S. Government work and is in the public domain in the USA.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

worldwide and is the most important insect pest of wheat in the southeastern USA (Ratcliffe & Hatchett, 1997; Cambron *et al*., 2010).

All damage to seedling wheat by Hessian fly is the result of feeding by first-instar larvae. Hatchling larvae enter the whorl, and upon settling near the base of the plant, rapidly induce changes that include the formation of a nutritive tissue that nourishes the developing larvae, a rapid increase in host-cell permeability and stunting of the plant (Harris *et al*., 2006; Saltzmann *et al*., 2008; Williams *et al*., 2011). Further, even if infesting larvae are removed from the seedling plant, normal growth cannot be restored (Byers & Gallun, 1972).

There are three methods to control Hessian fly damage to wheat in the field: avoidance, seed treatment with a systemic insecticide and deployment of genetically resistant wheat. Planting after the historically titled 'fly-free' date is the basis of avoidance; however, in most of the warmer southeastern USA, an effective fly-free date cannot be relied upon to prevent damage to wheat, as temperatures do not remain consistently cold enough to prevent Hessian fly adult emergence. Seed treatment is generally effective for only 2–3 weeks postgermination. Therefore, the most successful method of control is the deployment of genetically resistant wheat. Thirty-five different Hessian fly resistance (*R*) genes in wheat have been identified and characterized (Liu *et al*., 2005; Sardesai *et al*., 2005; Li *et al*., 2013; McDonald *et al*., 2014). This resistance is expressed as antibiosis of first-instar larvae and is controlled by single genes that are dominant or semidominant (Gallun, 1977; Harris *et al*., 2003; Williams *et al*., 2003). In this insect, virulence to *R* genes is controlled by non-allelic recessive genes at single loci and operates on a gene-for-gene basis with resistance (Hatchett & Gallun, 1970; Formusoh *et al*., 1996; Zantoko & Shukle, 1997).

The salivary glands of Hessian fly larvae express hundreds of transcripts that are specific to the Hessian fly and do not show homology to any known genes (Chen *et al*., 2010). Identified through an expressed sequence tag (EST) study, secreted salivary gland proteins (SSGPs) are hypothesized to be effectors that reprogramme the biochemical and physiological pathways of susceptible wheat to benefit the infesting larvae (Chen *et al*., 2004; Liu *et al*., 2007; Zhu *et al*., 2008). SSGPs are identified by three attributes: small size (50–200 amino acids), a secretion signal at the amino terminus and localized expression in the salivary glands (Chen *et al*., 2006). SSGPs are categorized into families that are defined as related proteins that share secretion signal peptides (Chen *et al*., 2006). The genes encoding these small SSGPs are commonly identified in multigenic clusters created by gene duplication and diversification with conserved intergenic regions and highly diversified coding regions (Chen *et al*., 2010). This unusual conservation is a unique feature of SSGPs, suggesting rapid evolution in response to selection pressures (Chen *et al*., 2010).

If SSGPs are the effectors in the wheat–Hessian fly interaction, then investigating their expression in the context of field populations is important to understanding the underlying biology of the Hessian fly. To date, no data are available on the expression of SSGP transcripts in field collections from different geographical regions. There is population data using microsatellite markers that document the structure of Hessian fly populations (Morton *et al*., 2011; Morton & Schemerhorn, 2013). Therefore, the focus of the present study was to compare transcript abundance from four previously studied SSGP families in Hessian fly from different geographical regions (Liu *et al*., 2004; Chen *et al*., 2008, 2010). We hypothesized that the SSGPs, acting as effectors, should vary in transcript abundance amongst field collections of Hessian fly from different geographical regions (southeastern USA, central USA and Israel) owing to biological and ecological parameters associated with the collection sites. Significant variation in SSGP transcript abundance amongst the field collections was observed. SSGP transcript abundance separated field collections into groups that corresponded with the major wheat classes grown in the geographical regions as well as previously described Hessian fly populations.

Results

Relative abundance amongst field collections of transcripts encoding SSGPs

To document the abundance of SSGP transcripts across different geographical regions, we carried out a microarray experiment. The Affymetrix microarray was composed of 444 probe sets dedicated to SSGP sequences identified from an EST project (Chen *et al*., 2004, 2008). Redundancies in alleles and gene copy number can make analyses of SSGP transcript abundance difficult; therefore, duplicates were removed from the analysis, and four previously described families (*vide supra*) were selected for evaluation.

Within each SSGP family under study, heat maps for transcript fold-change (Fig. 1) were used to visualize the expression of the transcripts for the six field collections relative to the laboratory Biotype Great Plains (GP) that has the lowest frequency of virulent genotypes of any Hessian fly biotype and is thought to represent a nascent state with respect to selection pressure from exposure to *R* genes in wheat (Harris & Rose, 1989). Significant $\log_2$ fold-changes in transcript abundance ≥twofold with significance at $P < 0.05$ in the field collections relative to Biotype GP are documented on the heat maps with an asterisk (*). Families 1 and 11 had the fewest transcripts showing

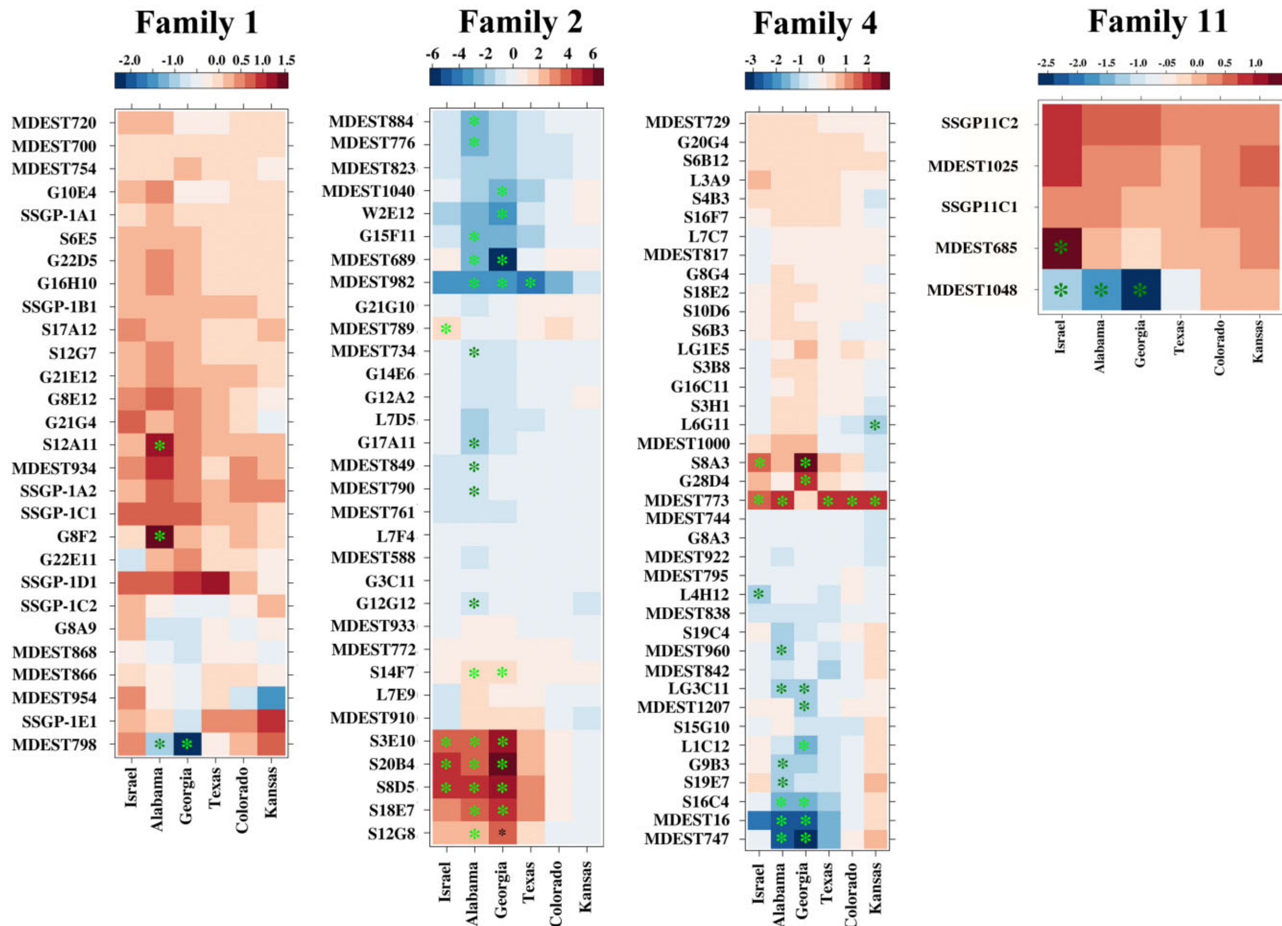

Figure 1. Heat map visualizing probe signal intensities for Hessian fly, *Mayetiola destructor* (Say), secreted salivary gland protein (SSGP) transcripts in Families 1, 2, 4 and 11. Fold-changes are normalized $\log_2$ signal intensities for probes in Hessian fly field collections relative to Biotype Great Plains (GP). $\log_2$ changes ≥twofold with significance at $P \leq 0.05$ are indicated by *. Scale shows colour code for $\log_2$ fold-changes. Positive fold-changes are indicated by red with darker tones indicating larger fold-changes. Negative fold-changes are indicated by blue with darker tones indicating larger fold-changes. Heat maps were drawn using R/BIOCONDUCTOR. GenBank accession numbers for SSGP transcripts are given in Tables S2–S5.

significant variation in abundance relative to Biotype GP, whereas Families 2 and 4 had the greatest (19 significant fold-changes in Family 2 and 14 significant fold-changes in Family 4). Greater decreases in transcript abundance relative to Biotype GP occurred than increases across the four families. Twenty-five genes showed decreased relative transcript abundance, and 13 showed increased abundance. Within the USA, Alabama and Georgia had the greatest number of SSGP transcripts showing significant variation in abundance relative to Biotype GP. In Texas, Colorado and Kansas, most SSGPs were expressed in levels relative to Biotype GP with few significant fold-changes. Israel, the field collection from the Middle East, also had significant variation in the relative abundance of SSGP transcripts in Families 2, 4 and 11. The complete data sets with GenBank accession numbers and *P*-values for significance in variation of transcript abundance for the SSGP genes are given in Tables S1–S4.

Within the three geographical regions that the field collections were made from, the fold-change patterns within each of the SSGP families showed similar trends. In particular, the fold-change patterns for collections from Kansas and Colorado were extremely similar for transcripts across the four SSGP families (Fig. 1). The southeastern USA collections also showed similar trends although fold-changes in Georgia were not often as statistically significant as in Alabama (Fig. 1). Fold-change patterns for transcript abundance in Texas, although not as close as between Kansas and Colorado, were similar to Kansas and Colorado across many transcripts within the four SSGP families. Fold-change patterns for transcripts across the four families in Israel showed three significant variations in transcript abundance (*MDEST789*, *L4H12* and *MDEST685*) that were distinct from geographical locations within the USA. Fold-change patterns of transcripts within a family

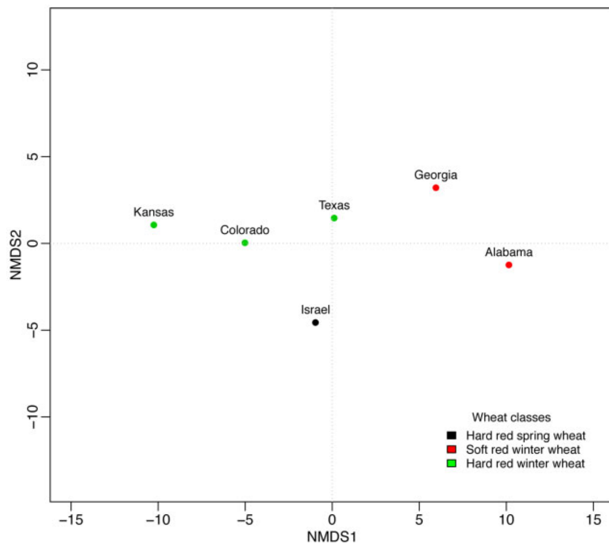

Figure 2. Ordination plot using nonmetric multidimensional scaling (NMDS). The plot depicts the relationships between Hessian fly, *Mayetiola destructor* (Say), field collections from the three different geographical locations (southeastern USA – Georgia and Alabama; central USA – Kansas, Colorado, and Texas; Middle East – Israel) as grouped by the three different wheat classes (ie soft-red-winter, hard-red-winter, hard-red-spring) predominantly grown at these locations. The 104 secreted salivary gland protein gene expression data from the microarray results for the Hessian fly collections correlated to the three different wheat classes in the analysis.

were also fairly consistent across geographical regions; however, small differences were present that could correspond to transcripts that might be suitable for further exploration relative to differences in agro-ecosystem parameters.

Hessian fly field collections from all three geographical locations were found to be significantly different ($P < 0.05$) in gene expression rates measured as $\log_2$ fold-changes when grouped according to the three different wheat classes cultivated at the geographical locations. The nonmetric multidimensional scaling (NMDS) ordination plot (Fig. 2) shows the grouping of these field collections. The first axis (NMDS1) separated field collections from the southeastern soft-red-winter wheat geographical locations (Georgia and Alabama) from field collections from the central hard-red-winter wheat locations (Kansas, Colorado and Texas). The second axis (NMDS2) separated the Middle Eastern (Israel) field collection from hard-red-spring wheat from the collections from the USA of soft-red and hard-red winter wheat.

Relative transcript abundance from the microarray analysis was further supported by quantitative real-time PCR (qRT-PCR) for three SSGP sequences within each family. Significant $\log_2$ fold-changes ≥twofold in comparison to Biotype GP are listed beside the microarray values (Table 1). Significance was scored at $P < 0.05$ and is indicated by grey highlighted boxes. The abundance trends identified on the microarray (equivalent, decreased and increased) are similar to those found with qRT-PCR.

Phylogenetic analyses

Phylogenetic trees were constructed to show the evolutionary relationships within each SSGP family. As members of each family share identical or highly similar secretion signals as well as 5′ and 3′ noncoding regions, diversity is often found within the coding sequence. SSGPs that share high sequence identity are commonly found in arrays of tandem repeats; thus, phylogenies may reveal SSGPs with increased copy number.

For Family 1, the phylogenetic tree shows two clades (Fig. S1). Although the general trend for Family 1 showed an increase in transcript abundance, only two SSGP sequences showed a significant increase in transcript abundance (*S12A11* and *G8F2*). Additionally, SSGP sequence *MDEST798* showed a significant decrease in transcript abundance in the collections from Alabama and Georgia. The three SSGP sequences *S12A11, G8F2* and *MDEST798* were located within the first clade.

In Family 2, there was a single large clade containing most of the genes in this family and three smaller clades (Fig. 3). Although, in general, the trend for Family 2 showed lower abundance in comparison to Biotype GP, the fourth clade contained a unique branch. Transcript abundance for SSGP sequences *S20B4*, *S3E10* and *S8D5* was significantly increased in the collections from Israel, Alabama and Georgia whereas transcript abundance for SSGP sequences *S18E7* and *S12G8* was significantly increased only in the collections from Alabama and Georgia. Although not statistically significant, the collection from Texas also showed a trend toward an increase in transcript abundance for *S20B4*, *S3E10* and *S8D5*, whereas Colorado and Kansas were equivalent to the Biotype GP reference. BLAST revealed that all of the five sequences were located on scaffold X1Random.8 at the same location in the Hessian fly genome. Two additional SSGP sequences showed an increase in transcript abundance outside of clade four (*S14F7* in Alabama and Georgia and *MDEST789* in Israel).

Family 4 (Fig. S2) also showed two clades; however, the SSGP sequences showing significant changes in transcript abundance were dispersed throughout the tree. Two SSGP sequences in which transcript abundance varied significantly relative to that in Biotype GP (*MDEST685* and *MDEST1048*) grouped together in Family 11 (Fig. S3). However, no other correlations between phylogenetic groups and transcript abundance within families were documented in the current study.

Table 1. Validation of secreted salivary gland protein (SSGP) transcript abundance from the microarray analysis by quantitative real-time PCR (qRT-PCR). Three SSGP sequences (expressed sequence tags, ESTs) were selected from each family; one in which transcript abundance was equivalent to that in Biotype Great Plains (GP), one in which transcript abundance decreased in one or more field collections, and one in which transcript abundance increased in one or more field collections. Significant fold changes ($P < 0.5$) are bolded

Family 1

EST	Location	Log_2 fold change microarray	Log_2 fold change qRT-PCR
MDEST700	Israel	0.077733468	0.379
	Alabama	0.022047536	0.308
	Georgia	0.042207697	0.355
	Texas	–0.017914909	0.053
	Colorado	0.060313076	0.361
	Kansas	–0.032557243	–0.070
MDEST798	Israel	0.382834123	0.428
	Alabama	**–1.120323992**	–0.326
	Georgia	**–2.117407727**	**–5.904**
	Texas	–0.224268014	0.493
	Colorado	0.147308884	0.023
	Kansas	0.721278658	0.525
G8F2	Israel	0.036373589	0.971
	Alabama	**1.328591004**	**2.448**
	Georgia	0.147813982	1.266
	Texas	–0.020539627	0.467
	Colorado	0.198159697	0.654
	Kansas	–0.024036948	1.026

Family 2

EST	Location	Log_2 fold change microarray	Log_2 fold change qRT-PCR
L7D5	Israel	–0.426608955	0.0123
	Alabama	–1.833044979	–0.174
	Georgia	–0.532575579	–0.159
	Texas	–0.555516459	–0.067
	Colorado	0.052065782	–0.113
	Kansas	0.055497373	0.094
MDEST689	Israel	0.358062704	0.430
	Alabama	**–2.606239269**	–1.587
	Georgia	**–5.383489305**	**–2.966**
	Texas	–0.260119533	0.333
	Colorado	0.327699189	0.439
	Kansas	0.734141467	0.316
S20B4	Israel	**4.812618901**	**5.514**
	Alabama	**3.764416323**	**5.298**
	Georgia	**5.957124208**	**6.561**
	Texas	2.485906705	2.726
	Colorado	0.865846488	–0.091
	Kansas	0.01447477	0.438

Family 4

EST	Location	Log_2 fold change microarray	Log_2 fold change qRT-PCR
MDEST817	Israel	–0.309475207	0.014
	Alabama	–0.076787128	–0.48
	Georgia	–0.006210489	–0.551
	Texas	–0.085317029	–0.512
	Colorado	–0.127456972	–0.461
	Kansas	–0.193863967	–0.618
MDEST747	Israel	–0.287153977	0.161
	Alabama	**–2.724266649**	–0.621
	Georgia	**–2.955720298**	**–2.213**
	Texas	–1.393828877	–0.573
	Colorado	–0.118984453	–0.306
	Kansas	0.733232293	0.258
S8A3	Israel	**1.486031054**	**2.290**
	Alabama	0.634000592	1.000
	Georgia	**2.513540395**	**2.139**
	Texas	0.943275829	0.759
	Colorado	0.210749525	0.169
	Kansas	–0.773248978	–0.943

Family 11

EST	Location	Log_2 fold change microarray	Log_2 fold change qRT-PCR
SSGP–11C1	Israel	0.319793191	0.059
	Alabama	0.211247182	0.135
	Georgia	0.172950997	0.061
	Texas	0.175970821	–0.414
	Colorado	0.416042074	0.052
	Kansas	0.444192339	0.111
MDEST1048	Israel	**–1.337873533**	**–1.526**
	Alabama	**–1.743631184**	**–0.967**
	Georgia	**–2.406430858**	**–1.331**
	Texas	–0.764068986	–0.551
	Colorado	0.009765357	0.115
	Kansas	–0.000425817	0.172
MDEST685	Israel	1.227504498	6.039
	Alabama	0.042517634	2.051
	Georgia	–0.111158045	0.931
	Texas	0.118463963	0.836
	Colorado	0.139080735	1.124
	Kansas	0.234566853	1.347

Discussion

The microarray analysis revealed significant differential expression of SSGP transcripts, the candidate effectors in the Hessian fly–wheat interaction, amongst field collections from different geographical regions relative to SSGP transcript abundance of the Biotype GP reference. The field collections under study and the Biotype GP reference were reared on the same variety of susceptible wheat (cv. Newton, carrying no genes for resistance). Thus, the variations in expression documented are not because of different wheat genotypes. Therefore, the documented variation in expression of SSGP genes is associated with genetic adaptations that accumulated over time from environmental and agro-ecosystem selection pressures (Morton *et al.*, 2011; Morton & Schemerhorn, 2013). These selection pressures could influence population structure and evolution in the field and influence the expression of effectors.

In the current study, SSGP transcript abundance in the six field collections fell into three geographical groups based on similarity of SSGP expression and wheat class

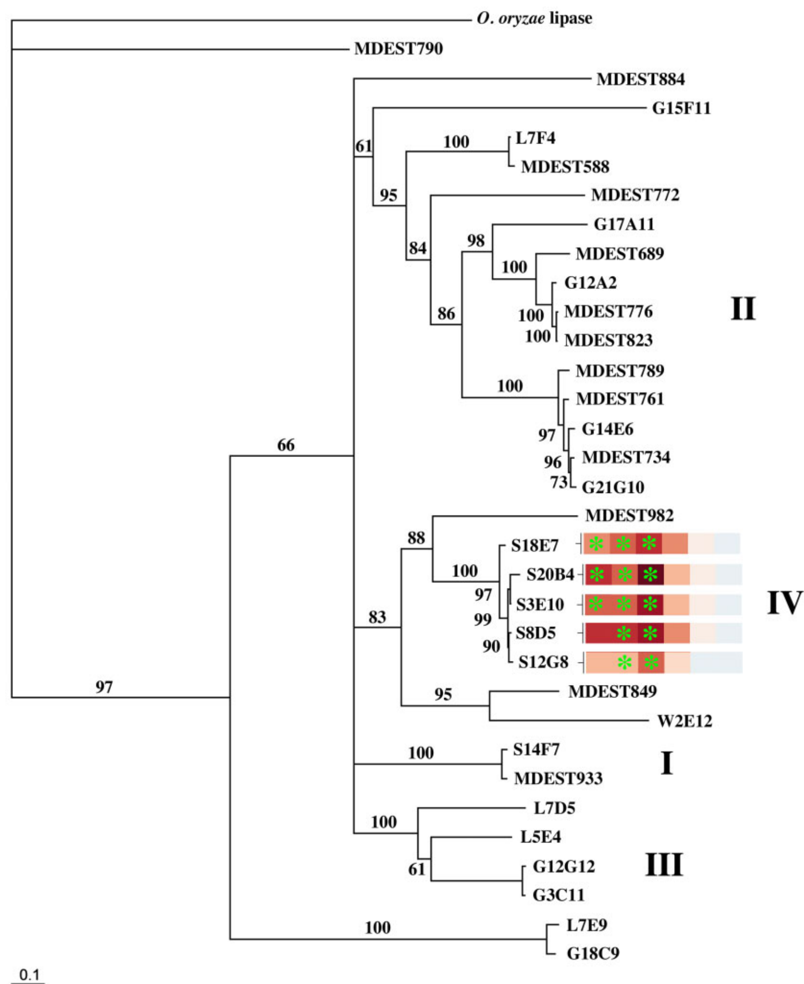

Figure 3. Bayesian phylogenetic tree of secreted salivary gland protein (SSGP) transcripts in Family 2. The phylogenetic reconstruction is rooted using the secreted salivary lipase-like gene from the Asian rice gall midge, *Orseolia oryzae* (Wood-Mason) as an outgroup, posterior probability values are located at the nodes and clades are indicated by Roman numerals. Although the most significant variations in transcript abundance are located within Family 2, there is only a single branch that shows five related genes (*S20B4*, *S3E10*, *S18E7*, *S12G8* and *S8D5*) with similar transcript abundance patterns as shown in the heat map insert for SSGP transcript probe intensities. These genes have a high degree of nucleotide similarity and as such may represent alleles or paralogues.

grown in the geographical region: (1) southeastern USA; (2) central USA; and (3) the Middle East. These groupings were also in agreement with a previously published population survey that revealed the worldwide structure of Hessian fly populations using microsatellite markers (Morton & Schemerhorn, 2013). Alabama and Georgia are located in the southeastern USA, where soft-red-winter wheat varieties are grown, and multiple *R* genes (*H*3, *H*5, *H*6, *H7H*8, *H*9 and *H*13) have been deployed in adapted wheat varieties (Cambron *et al.*, 2010). Kansas, Colorado and Texas are in the central USA, where hard-red-winter wheat cultivars are primarily grown, and *R* genes have not been deployed to the same extent as in the southeastern USA (Garcés-Carrera *et al.*, 2014). In Israel *R* genes are not commercially deployed and hard-red-spring wheat is predominantly grown (Johnson *et al.*, 2012). Our microarray analysis of SSGP sequences further support the findings that Hessian fly populations across multiple locations within the USA have low levels of local adaptation that are the result of the sharing of agro-ecosystem pressures over large geographical areas (Black *et al.*, 1990; Morton *et al.*, 2011). These local adaptations result in micropopulations that vary within the larger overall population.

The equivalency in abundance of transcripts encoding SSGPs between Kansas and Biotype GP is not surprising. The laboratory Biotype GP reference used in the present study was derived from a field collection made in Ellis County, Kansas, and maintained under greenhouse conditions since 1986 (Harris & Rose, 1989). The microarray analysis indicates that the diversity in SSGP transcript abundance in the laboratory Biotype GP reference and in the current field collection from Ellis County, Kansas, are essentially identical. Although deployment of *R* genes *H*3 and *H*6 has occurred in recent years, their usage is neither consistent nor widespread in Kansas. This indicates that field conditions over the last 25 years have resulted in little significant variation between the current Ellis County collection and the reference Biotype GP from agro-ecosystem pressures. The similarity in abundance of SSGP transcripts between Kit Carson County, Colorado, and Ellis County, Kansas, located 200 linear miles apart,

could also be associated with similarity in environmental and agro-ecosystem selection pressures between the field collection sites in eastern Colorado and central Kansas.

In the central USA additional agro-ecosystem parameters that can affect Hessian fly populations are: a low number of generations per year and the lack of successive deployment of multiple *R* genes over time. Generally, there are two generations per year of Hessian fly (autumn and spring) that can be controlled by planting after the 'fly-free' date. Therefore, the deployment of *R* genes for control of Hessian fly has not been as extensive as in the southeastern USA. With a limited number of generations per year and dispersed local deployment of resistant cultivars, the number of virulent Hessian fly in the field is slow to accumulate and perpetuates the repeated use of a resistant cultivar (Gould, 1986). Recently, low levels of virulence have been identified in Kansas (Chen *et al*., 2009). A new survey from Texas has shown that virulence in the field is increasing as the repeated, annual deployment of multiple *R* genes increases (Garcés-Carrera *et al*., 2014). However, neither shows the widespread high proportion of virulence seen in the southeastern USA (Cambron *et al*., 2010).

In the southeastern USA, climate, availability of alternative hosts and the successive deployment of *R* genes can affect the biology of Hessian fly. The growing of wheat for forage and the presence of alternative host plants increases the availability of host plants during the warm, wet growing season before the 'fly-free date' leading to multiple generations (six to eight; Buntin & Raymer, 1989a; Buntin & Chapin, 1990; Buntin *et al*., 1992; Flanders *et al*., 2014). Together, these factors negate the avoidance practice, as host plants are always readily available for each generation and aid in populations rapidly overcoming resistant wheat cultivars.

The greatest variation in expression amongst all four of the SSGP families under study occurred in collections from Alabama and Georgia. Successive deployment of wheat cultivars carrying *R* genes has resulted in a decline in *R* gene efficacy, an increase in field populations of Hessian fly that can overcome formerly resistant wheat and the highest proportion of local adaptation to *R* genes in the USA (Cambron *et al*., 2010; Ratcliffe, 2013). Between 1986 and 2000, deployment of the *R* genes *H3*, *H5*, *H6* and the gene combination *H7H8* led to the evolution of Hessian fly from being moderately virulent to *H3* to 100% virulent to all four of the deployed genes (Buntin & Raymer, 1989b; Alabama Cooperative Extension System, 2013). This successive deployment of *R* genes could also be a factor influencing SSGP expression in field populations from the southeastern USA.

Populations near the centre of origin for a species can but do not always show the most significant genetic diversity (Harlan, 1974). The Israeli field collection showed neither the greatest variation in relative abundance of SSGP transcripts nor the greatest virulence to known *R* genes (Johnson *et al*., 2012). Hessian fly is thought to have coevolved with the genus *Triticum* in the Fertile Crescent, and high frequencies of Hessian fly virulence to the identified *R* genes have been documented in field collections of the fly from Syria (Ratcliffe & Hatchett, 1997; El Bouhssini *et al*., 2009). Climatic differences in temperature and moisture that drive the generational cycle have impacted Hessian fly population structure in Israel (Johnson *et al*., 2012). In Israel >95% of the wheat currently grown is hard-red-spring and has replaced cultivation of local land races and durum wheat [*Triticum turgidum* L. ssp. *durum* (Desf.)]. As *R* genes for Hessian fly resistance have not been deployed in Israel, indigenous wild wheat as well as alternative grass hosts could be the sources of *R* gene exposure for Israeli Hessian fly populations. Additionally, lack of migration resulting in low gene flow and isolation separates the Israeli populations from neighbouring populations such as those in Syria. Thus, the class of wheat cultivated coupled with very different environmental conditions, isolation and low gene flow, and sporadic exposure to *R* genes could be factors resulting in the differentiation of SSGP expression between the Israeli collection and those from the southeastern and central USA.

Fitness costs associated with virulence and adaptive responses should play an important role in plant–parasite coevolution (Montarry *et al*., 2010). Reproductive fitness costs have been associated with Hessian fly virulence to resistance genes *H9* and *H13* in wheat (Zhang *et al*., 2011). Most of the decreases in relative abundance of transcripts encoding SSGPs were found in the southeastern USA populations and this could be associated with fitness costs associated with these SSGPs. However, a clearer understanding of the significance of the differential expression of SSGPs reported here requires knowledge of the role of the SSGPs during interactions with both susceptible and resistant wheat, respectively. Currently, this knowledge is lacking, and this is a hindrance to fully understanding the diversity in expression of SSGPs amongst Hessian fly populations documented here.

Differential expression of SSGPs could also be attributed to variation in copy number of tandem repeats. Within Family 2, one branch in a clade of related SSGP sequences showed similarity in relative abundance significantly greater than in Biotype GP for flies from Alabama and Georgia (*S3E10*, *S20B4*, *S8D5*, *S18E7* and *S12G8*) as well as Israel (*S3E10*, *S20B4* and *S8D5*). In Colorado and Kansas, the relative abundance of these transcripts was equivalent to that in Biotype GP. However, although the collections from Texas were not significantly different from Biotype GP, they did show a trend toward an

increase. Although the BLAST results for the Hessian fly genome sequence positioned all five sequences (*S18E7*, *S20B4*, *S3E10*, *S8D5* and *S12G8*) on the same scaffold (X1Random.8), problems with the assembly of the Hessian fly genome sequence mean that SSGP sequences are often positioned at a single locus owing to sequence similarity. Further, no sequenced Bacterial Artificial Chromosome (BAC) clones were available to resolve whether or not these five sequences occupy the same location. However, an analysis of the alignments for these five transcripts suggests that variation amongst the transcripts is greater than would be expected for alleles and could represent tandem repeats that have diverged over time (ie paralogues). Future sequencing of BAC clones in this region of the genome should resolve this question.

Conclusion

A microarray-based study documented significant variation in transcript abundance within a set of four SSGP families amongst Hessian fly field collections from three distinct geographical regions by the wheat class predominantly grown in the regions. These data support findings from previous studies indicating that ecological and agro-ecosystem dynamics within the three geographical regions exert different selection pressures associated with the different geographical regions and influence Hessian fly population structure.

Experimental procedures

Field collection of insect material

Hessian fly field collections of autumn infestations were made from five localities in the USA (Pike County, Georgia; Limestone County, Alabama; Brazos County, Texas; Ellis County, Kansas; Kit Carson County, Colorado) and one locality in the Middle East (northern Negev, Magen, Israel). Within the USA, the collection localities represented the southeastern and central geographical regions. The Magen, Israel, collection is from the Middle East where Hessian fly and the genus *Triticum* are proposed to have coevolved (Ratcliffe & Hatchett, 1997). The laboratory Biotype GP that is defined as having a low frequency of virulence to the known *R* genes (Harris & Rose, 1989) was used as a reference biotype for comparison of transcript abundance.

Field collections were made by randomly harvesting approximately 500 infested plants from three to five different areas within an infested field (Johnson *et al.*, 2012). Collections of flies from the different areas within a field were pooled and treated as one sample. Field collections underwent one cycle of increase in the greenhouse under conditions documented to retain genetic diversity (Foster *et al.*, 1988; Black *et al.*, 1990). Adults were allowed to emerge, mate and oviposit under mesh tents on flats of Cultivar 'Newton', that carries no Hessian fly resistance genes. When infesting larvae reached the third-instar within puparia, the flats were sifted to remove soil, and the infested plant material was placed into cold storage at 4 °C. Under these conditions larvae retain their viability for up to a year. Infested plant material was removed from cold storage to allow adult emergence in order to infest Newton wheat in pots for SSGP expression studies.

RNA extraction

Twenty seedlings of Newton wheat were grown in a 10-cm pot containing a sterile mixture of soil and potting mix. When plants reached the 1.5 leaf stage, they were infested with five gravid females by confining them under a plastic cup covering the pot (Foster *et al.*, 1988). Four-day-old first-instar larvae were released from the plants by dissecting the crown with forceps in deionized water. Preliminary analyses documented that abundance of transcripts encoding SSGPs generally peaks in first-instar larvae 4 days after egg hatching (R. Shukle, unpubl. results). Infestations were carried out with each of the six field collections in triplicate to produce three biological replicates (collections of larvae) for transcript abundance studies. Total RNA was extracted from the collected larvae using an RNAqueous-4PCR kit (Life Technologies, Grand Island, NY, USA) according to the manufacturer's protocol. Extracted RNA was frozen at −80 °C until further analysis. The RNA samples were used to carry out the microarray hybridization as well as the qRT-PCR analysis.

Microarray hybridizations

A custom microarray (Affymetrix, Santa Clara, CA, USA) containing probes for 444 previously identified Hessian fly SSGP sequences was used in the current study. Microarray processing and hybridization were carried out in the Integrated Gene Expression Facility at Kansas State University following the procedures described in Liu *et al.* (2007). An Ovation RNA Amplification System V2 kit (NuGEN Technologies, San Carlos, CA, USA) was used to convert 50 ng of RNA to antisense cDNA that was used for hybridization. The Minelute PCR purification kit (Qiagen, Valencia, CA, USA) was used to isolate single-stranded cDNAs, which were quantified using a Nanodrop-ND-1000 spectrophotometer (Thermoscientific, Waltham, MA, USA). The purified cDNA (3.75 μg) was fragmented and labelled using a FL-Ovation cDNA Biotin module V2 kit (NuGEN Technologies). Labelled fragments were checked for integrity by running the fragmented cDNA through a RNA nano-chip in an Agilent Bioanalyzer (Santa Clara, CA, USA). The hybridization mixture was prepared following the protocol included in the FL-Ovation cDNA Biotin module V2 kit and was then injected into the microarrays. After 18 h of incubation in a GeneChip oven (Affymetrix), standard protocol was followed to wash the microarrays, and they were stained with streptavidin phycoerythrin in a GeneChip fluidic station 450 (Affymetrix). A GeneChip scanner 3000-7G (Affymetrix) was used to scan the microarrays and GeneChip operating software version 1.4 generated the initial image (.dat) and scaled image (.cel) files.

Microarray analyses

The microarray data from the .cel files were analysed using R (R Development Core Team, 2013) and Bioconductor (Gentleman *et al.*, 2004). The .cel files were imported into R using Affy software, and microarray data were corrected for technical variation using the Robust Multi-array Average (RMA) procedure (Irizarry *et al.*, 2003; Gautier *et al.*, 2004). A total of 444 probes

sets was assayed on 20 microarrays that were hybridized with DNA from the six Hessian fly field collections plus Biotype GP as the reference.

Differentially expressed genes were identified in the six field collections using hypothesis testing based on a probe-wise modified two-sample *t*-test; therefore, 444 hypothesis tests were simultaneously performed for each of the six field collections with Biotype GP as the reference (Efron, 2010). As a two-sample *t*-test is an unreliable estimation of noise variance resulting from the limited number of biological replicates in the microarray data, a modified two-sample *t*-test that has better statistical properties for testing differential expression of probes in microarrays was used (Smyth, 2004). Using the modified two-sample *t*-test, *P*-values were necessary for the differential correction for multiple comparisons to control the number of false positives (the probes that are falsely declared as showing differential signals). False discovery rate (FDR) has greater statistical power than family-wise error rate procedures and has optimal properties for simultaneous hypotheses tests in analysis of microarrays in which only a small fraction of transcripts are differentially expressed (Efron, 2010). The *P*-values from the modified two-sample *t*-test were adjusted using Benjamini & Hochberg's (1995) procedure for controlling FDR and obtaining adjusted *P*-values. These *P*-values can be directly compared with the standard cut-off of 0.05. For each field collection, the transcripts with abundance level changes having adjusted *P*-values less than 0.05 were considered to be differentially expressed.

Validation of microarray results by qRT-PCR

To validate the fold-change data observed in the microarray analysis for SSGP transcript abundance in the field collections relative to Biotype GP, three genes from each of the four SSGP families were selected for qRT-PCR analyses. These genes were selected on the basis of equivalent expression across all populations, decreased expression and increased expression. One μg of DNase-treated RNA was used as a template for synthesis of first-strand cDNA with random hexamers using a Tetro cDNA synthesis kit (Bioline, Taunton, MA, USA) according to the manufacturer's instructions. As the relative standard curve method (ABI User Bulletin #2, http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_040980.pdf) was used, cDNA concentrations were quantified using a Nanodrop-ND-1000 spectrophotometer and diluted to 10 ng/μl.

The software PRIMER EXPRESS v. 3.0 (Applied Biosystems, Foster City, CA, USA) was used to design gene-specific qRT-PCR primers that would amplify a 50–75 bp fragment between 58–62 °C (Table S5). qRT-PCR was performed on a LightCycler 480 (Roche Diagnostics, Indianapolis, IN, USA) with SensiFAST SYBR no-ROX chemistry (Bioline). The total qRT-PCR volume of 20 μl contained 10 μl 2× SensiFAST SYBR No-ROX mix, 10 μM of a forward and a reverse gene-specific primer, and 40 ng cDNA template per reaction. No-template samples were included in each PCR plate as negative controls. PCR parameters were as follows: 95 °C for 2 min; 40 cycles of 95 °C for 5 s, 55 °C for 10 s and 72 °C for 20 s. To determine the specificity of the reaction, a melt curve analysis was carried out following qRT-PCR, confirming amplification of a single product. The reactions were set up in triplicate for each of the three biological replicates in a 384-well plate. 18S ribosomal RNA (National Center for Biotechnology Information accession no. KC177284.1) was used as an internal reference for transcript normalization. Transcript abundance data were calculated according to the relative standard curve method. Relative expression values (REVs) were presented as $\log_2$ fold-change relative to transcript abundance in Biotype GP.

Significant differences in mean REVs associated with transcript abundance in the field collections relative to the Biotype GP reference were identified using a Dunnett multiple comparisons test (Dunnett, 1955, 1964). Differences were considered significant at $P < 0.05$.

Phylogenetic analyses

CLUSTAL X v. 2.1 was used to create an alignment file for the nucleotide sequences (Larkin *et al*., 2007). The best-fit model of nucleotide substitution was calculated using JMODELTEST2 (Guindon & Gascuel, 2003; Darriba *et al*., 2012). Bayesian maximum likelihood trees were constructed under the general time-reversible model with invariable sites and gamma distribution (GTR + I + G) using MRBAYES 3.2.1, and the analyses were computed using an excess of 1 000 000 generations until the split frequency deviation was less than or equal to 0.01 (Ronquist *et al*., 2012). TREEVIEW 1.6.6 was used to display the phylogenetic trees (Page, 1996). All trees were rooted with a lipase-like SSGP outgroup from the Asian rice gall midge *Orseolia oryzae* (Wood-Mason) (GenBank accession no: FJ196713) that is a homologue of a lipase-like SSGP for Hessian fly and encodes a protein with a secretion signal (Shukle *et al*., 2009).

Ordination and analysis of Hessian fly field collections by wheat classes

A NMDS approach was used to group the Hessian fly field collections based on variation in wheat classes (ie soft-red-winter, hard-red-winter and hard-red-spring) as a function of the 104 gene expression results within each collection from the microarray analysis. Gene expression data were standardized and a Euclidian distance matrix was calculated as a proximity matrix. To test the statistical significance of the field collection groupings, a permutational multivariate analysis of variance using the same proximity matrix (*vide supra*) was conducted using the function 'adonis' from the R package 'vegan' 2.0.1 (Oksanen *et al*., 2013). The statistical significance was calculated after 99 999 permutations.

Acknowledgements

This is a joint contribution of the USDA-ARS and Purdue University and was supported through USDA-ARS CRIS number 3602-22000-016-00D. Mention of a commercial or proprietary product does not constitute an endorsement by the USDA.

References

Alabama Cooperative Extension System (2013) *Effectiveness of Hessian fly resistance genes* (online). Available at: http://www.aces.edu/dept/grain/Geneeffectiveness.php. Accessed on 1 March 2014.

Alfano, J.R. (2009) Roadmap for future research on plant pathogen effectors. *Mol Plant Pathol* **10**: 805–813.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* **57**: 289–300.

Black, W.C.I.V., Hatchett, J.H. and Krchma, L.J. (1990) Allozyme variation among populations of the Hessian fly (*Mayetiola destructor*) in the United States. *J Hered* **81**: 331–337.

Buntin, G.D. and Chapin, J.W. (1990) Hessian fly (Diptera: Cecidomyiidae) biology in the southeastern United States: geographic variation and temperature-dependent phenology. *J Econ Entomol* **85**: 1015–1024.

Buntin, G.D. and Raymer, P.L. (1989a) Hessian fly (Diptera: Cecidomyiidae) damage and forage production of winter wheat. *J Econ Entomol* **82**: 301–306.

Buntin, G.D. and Raymer, P.L. (1989b) Susceptibility of winter wheat and triticale to Hessian fly. *Geo Agr Exp St Bull* **389**.

Buntin, G.D., Ott, S.L. and Johnson, J.W. (1992) Integration of plant resistance, insecticides, and planting date for management of the Hessian fly (Diptera: Cecidomyiidae) in winter wheat. *J Econ Entomol* **85**: 530–538.

Byers, R.A. and Gallun, R.L. (1972) Ability of Hessian fly to stunt winter wheat. *J Econ Entomol* **65**: 955–959.

Cambron, S.E., Buntin, G.D., Weisz, R., Holland, J.D., Flanders, K.L., Schemerhorn, B.H. *et al*. (2010) Virulence in Hessian fly (Diptera: Cecidomyiidae) field collections from the Southeastern United States to 21 resistance genes in wheat. *J Econ Entomol* **103**: 2229–2235.

Chen, M.S., Fellers, J.P., Stuart, J.J., Reese, J.C. and Liu, X. (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., Fellers, J.P., Zhu, Y.C., Stuart, J.J., Hulber, S., El Bouhssini, M. *et al*. (2006) A super-family of genes coding for secreted salivary gland proteins from the Hessian fly, *Mayetiola destructor*. *J Insect Sci* **6**: 12.

Chen, M.S., Zhao, H.X., Zhu, Y.C., Scheffler, B., Liu, X. and Liu, X. (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., Echegaray, E., Whitworth, R.J., Wang, H., Sloderbeck, P.E., Knutson, A. *et al*. (2009) Virulence analysis of Hessian fly populations from Texas, Oklahoma, and Kansas. *J Econ Entomol* **102**: 774–780.

Chen, M.S., Liu, X., Yang, Z., Zhao, H., Shukle, R.H., Stuart, J.J. *et al*. (2010) Unusual conservation among genes encoding small secreted salivary gland proteins from a gall midge. *Evol Biol* **10**: 296. doi:10.1186/1471-2148-10-296.

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**: 772.

DeLay, B., Mamidal, P., Wijeratne, A., Wijeratne, S., Mittapalli, O., Wang, J. *et al*. (2012) Transcriptome analysis of the salivary glands of Potato Leafhopper, *Empoasca fabae*. *J Insect Physiol* **58**: 1626–1634.

Dunnett, C.W. (1955) A multiple comparison procedure for comparing several treatments with a control. *J Am Stat Assoc* **50**: 1096–1121.

Dunnett, C.W. (1964) New tables for multiple comparisons with a control. *Biometrics* **20**: 482–491.

Efron, B. (2010) *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press, Cambridge.

El Bouhssini, M., Chen, M.S., Lhaloui, S., Zharmukhamedova, G. and Rihawi, R. (2009) Virulence of Hessian fly (Diptera: Cecidomyiidae) in the Fertile Crescent. *J Appl Entomol* **133**: 381–385.

Flanders, K.L., Reisig, D.D., Buntin, G.D., Winslow, M., Herbert, D.A., Jr. and Johnson, D.W. (2014) Biology and management of Hessian fly in the Southeast. *Al Coop Ext Syst* **ANR-1069** (online). Available at: http://www.aces.edu/pubs/docs/A/ANR-1069/ANR-1069.pdf. Accessed on 1 March 2014.

Formusoh, E.S., Hatchett, J.H., Black, W.C. and Stuart, J.J. (1996) Sex-linked inheritance of virulence against wheat resistance gene *H9* in the Hessian fly (Diptera: Cecidomyiidae). *Ann Entomol Soc Am* **89**: 428–434.

Foster, J.E., Araya, J.E., Safranski, G.G., Cambron, S.E. and Taylor, P.L. (1988) Rearing Hessian fly for use in the development of resistant wheat. *Purdue Univ Agr Exp St Bull* **536**.

Gallun, R.L. (1977) Genetic basis of Hessian fly epidemics. *Ann N Y Acad Sci* **287**: 223–229.

Garcés-Carrera, S., Knutson, A., Wang, H., Giles, K.L., Huang, F., Whitworth, R.J. *et al*. (2014) Virulence and biotype analyses of Hessian fly (Diptera: Cecidomyiidae) populations from Texas, Louisiana, and Oklahoma. *J Econ Entomol* **107**: 417–423.

Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A. (2004) Affy – analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**: 307–315.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S. *et al*. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80.

Gould, F. (1986) Simulation models for predicting durability of insect-resistant germ-plasm: Hessian fly (Diptera: Cecidomyiidae) resistant winter wheat. *Environ Entomol* **15**: 11–23.

Guindon, S. and Gascuel, O. (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst Biol* **52**: 696–704.

Harlan, J.R. (1974) Agricultural origins: centers and noncenters. *Science* **174**: 468–474.

Harris, M.O. and Rose, S. (1989) Temporal changes in egg laying behavior of the Hessian fly. *Entomol Exp Appl* **53**: 17–29.

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

Harris, M.O., Freeman, T.P., Rohfritsch, O., Anderson, K.G., Payne, S.A. and Moore, J.A. (2006) Virulent Hessian fly (Diptera: Cecidomyiidae) larvae induce a nutritive tissue during compatible interactions with wheat. *Ann Entomol Soc Am* **99**: 305–316.

Hatchett, J.H. and Gallun, R.L. (1970) Genetics of the ability of the Hessian fly, *Mayetiola destructor*, to survive on wheats having different genes for resistance. *Ann Entomol Soc Am* **63**: 1400–1407.

Hogenhout, S.A. and Bos, J.I.B. (2011) Effector proteins that modulate plant-insect interactions. *Curr Opin Plant Biol* **14**: 1–7.

Hogenhout, S.A., Van der Hoorn, R.A.L., Terauchi, R. and Kamoun, S. (2009) Emerging concepts in effector biology of plant-associated organisms. *Mol Plant Microbe Interact* **22**: 115–122.

Hori, K. (1992) Insect secretions and their effect on plant growth. In *Biology of Insect-Induced Galls* (Shorthouse, J.D. and Rohfritsch, D., eds), pp. 118–140. Oxford University Press, New York.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. *et al*. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249–264.

Johnson, A.J., Weintraub, P.G., Katoch, R., Schemerhorn, B.J. and Shukle, R.H. (2012) Biological and molecular characterization of Hessian fly (Diptera: Cecidomyiidae) from Israel. *B Entomol Res* **102**: 632–643.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. *et al*. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Li, C., Chen, M.S., Chao, S., Yu, J. and Bai, G. (2013) Identification of a novel Hessian fly resistance gene, H34, in wheat using recombinant inbred lines and single nucleotide polymorphism markers. *Theor Appl Genet* **126**: 2065–2071.

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

Liu, X., Bai, J., Huang, L., Zhu, L., Liu, X., Weng, N. *et al*. (2007) Gene expression of different wheat genotypes during attack by virulent and avirulent Hessian fly (*Mayetiola destructor*) larvae. *J Chem Ecol* **33**: 2171–2194.

Liu, X.M., Brown-Guedira, G.L., Hatchett, J.H., Owuoche, J.O. and Chen, M.S. (2005) Genetic characterization and molecular mapping of a Hessian fly-resistance gene transferred from *T. turgidum* ssp. *dicoccum* to common wheat. *Theor Appl Genet* **111**: 1308–1315.

McDonald, M.J., Ohm, H.W., Rinehart, K.D., Giovanini, M.P. and Williams, C.E. (2014) *H33*: a wheat gene providing Hessian fly resistance for the Southeastern United States. *Crop Sci* **54**: 1–9.

Montarry, J., Hamelin, F.M., Glais, I., Corbière, R. and Didier, A. (2010) Fitness costs associated with unnecessary virulence factors and life history traits: evolutionary insights from the potato late blight pathogen *Phytophthora infestans*. *BMC Evol Biol* **10**: 283.

Morton, P.K. and Schemerhorn, B.J. (2013) Population structure and colonization route of one of the oldest North American invasive insects: stories from the worn road of the Hessian fly, *Mayetiola destructor* (Say). *PLoS ONE* **8**: e59833.

Morton, P.K., Foley, C.J. and Schemerhorn, B.J. (2011) Population structure and spatial influence of agricultural variables on Hessian fly populations in the Southeastern United States. *Environ Entomol* **40**: 1303–1316.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. *et al*. (2013) *vegan: Community Ecology Package*. R package version 2.0-10. Available at: http://CRAN.R-project.org/package=vegan. Accessed on 1 March 2014.

Page, R.D.M. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**: 357–358.

R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ratcliffe, R.H. (2013) Breeding for Hessian fly resistance in wheat. In *Radcliffe's IPM World Textbook* (Radcliffe, E.D., Hutchison, W.D. and Cancelado, R.E., eds). University of Minnesota, St. Paul, MN (online). Available at: http://ipmworld.umn.edu/chapters/ratcliffe.htm. Accessed on 1 March 2014.

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

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S. *et al*. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**: 539–542.

Saltzmann, K.D., Giovanini, M.P., Zheng, C. and Williams, C.E. (2008) Virulent Hessian fly larvae manipulate the free amino acid content of host wheat plants. *J Chem Ecol* **34**: 1401–1410.

Sardesai, N., Nemacheck, J.A., Subramanyam, S. and Williams, C.E. (2005) Identification and mapping of *H32*, a new wheat gene conferring resistance to Hessian fly. *Theor Appl Genet* **111**: 1167–1173.

Shukle, R.H., Mittapalli, O., Morton, P.K. and Chen, M.S. (2009) Characterization and expression analysis of a gene encoding a secreted lipase-like protein expressed in the salivary glands of the larval Hessian fly, *Mayetiola destructor* (Say). *J Insect Physiol* **55**: 105–112.

Smyth, G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: 3.

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

Williams, C.E., Collier, C.C., Sardesai, N., Ohm, H.W. and Cambron, S.E. (2003) Phenotype assessment and mapped markers for *H31*, a new wheat gene conferring resistance to Hessian fly (Diptera: Cecidomyiidae). *Theor Appl Genet* **107**: 1516–1523.

Williams, C.E., Nemacheck, J.A., Shukle, J.T., Subramanyam, S., Saltzmann, K.D. and Shukle, R.H. (2011) Induced epidermal permeability modulates resistance and susceptibility of wheat seedlings to herbivory by Hessian fly larvae. *J Exp Bot* **62**: 4521–4531.

Zantoko, L. and Shukle, R.H. (1997) Genetics of virulence in the Hessian fly to resistance gene *H13* in wheat. *J Hered* **88**: 120–123.

Zhang, H., Anderson, K.M., Stuart, J.J., Cambron, S. and Harris, M.O. (2011) A reproductive fitness cost associated with Hessian fly (Diptera: Cecidomyiidae) virulence to gene-for-gene resistance. *J Econ Entomol* **104**: 1055–1064.

Zhu, L., Liu, X., Liu, X., Jeannote, R., Reese, J.C., Harris, M. *et al*. (2008) Hessian fly (*Mayetiola destructor*) attack causes dramatic shift in carbon and nitrogen metabolism in wheat. *Mol Plant Microbe Interact* **21**: 70–78.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Bayesian phylogenetic tree of secreted salivary gland protein (SSGP) transcripts in Family 1. The phylogenetic reconstruction is rooted using the secreted salivary lipase-like gene from the Asian rice gall midge, *Orseolia oryzae* (Wood-Mason) as an outgroup, posterior probability values are located at the nodes and clades are indicated by Roman numerals. SSGPs in this family separated into two clades. However, there is no correlation between transcript abundance and phylogeny.

Figure S2. Bayesian phylogenetic tree of secreted salivary gland protein (SSGP) transcripts in Family 4. The phylogenetic reconstruction is rooted as in Fig. S1; posterior probability values and clades are indicated. Significant variability in transcript abundance is dispersed throughout the tree, and no pattern between transcript abundance and phylogeny is shown.

Figure S3. Bayesian phylogenetic tree of secreted salivary gland protein (SSGP) transcripts in Family 11. The phylogenetic reconstruction is rooted as in Fig. S1; posterior probability values and clades are indicated. No correlation of transcript abundance variation and phylogeny could be seen within this small family of SSGPs.

Table S1. Microarray data expressed in $\log_2$ fold change in comparison to Biotype Great Plains (GP) for Family 1. Adjusted *P*-values are listed. EST, expressed sequence tag.

Table S2. Microarray data expressed in $\log_2$ fold change in comparison to Biotype Great Plains (GP) for Family 2. Adjusted *P*-values are listed. EST, expressed sequence tag.

Table S3. Microarray data expressed in $\log_2$ fold change in comparison to Biotype Great Plains (GP) for Family 4. Adjusted *P*-values are listed. EST, expressed sequence tag.

Table S4. Microarray data expressed in $\log_2$ fold change in comparison to Biotype Great Plains (GP) for Family 11. Adjusted *P*-values are listed. EST, expressed sequence tag.

Table S5. Primers for quantitative real-time PCR (qRT-PCR) including the melting temperature (Tm). Forward primers are labelled F, and reverse primers are labelled R.