

## **A super-family of genes coding for secreted salivary gland proteins from the Hessian fly, *Mayetiola destructor***

Authors: Chen, Ming-Shun, Fellers, John P., Zhu, Yu Cheng, Stuart, Jeffrey J., Hulbert, Scot, et al.

Source: Journal of Insect Science, 6(12) : 1-11

Published By: Entomological Society of America

URL: <https://doi.org/10.1673/2006.06.12.1>

---

BioOne Complete ([complete.BioOne.org](https://complete.BioOne.org)) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](https://www.bioone.org/terms-of-use).

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.



# A super-family of genes coding for secreted salivary gland proteins from the Hessian fly, *Mayetiola destructor*

Ming-Shun Chen<sup>1,2</sup>, John P. Fellers<sup>1,3</sup>, Yu Cheng Zhu<sup>4</sup>, Jeffrey J. Stuart<sup>5</sup>, Scot Hulbert<sup>3</sup>, Mustapha El-Bouhssini<sup>6</sup> and Xiang Liu<sup>2</sup>

<sup>1</sup> USDA-ARS, Plant Science and Entomology Research Unit, 4007 Throckmorton Hall, Manhattan, KS 66506

<sup>2</sup> Department of Entomology, 123 Waters Hall, Kansas State University, Manhattan, KS 66506

<sup>3</sup> Department of Plant Pathology, 4027 Throckmorton Hall, Kansas State University, Manhattan, KS 66506

<sup>4</sup> USDA-ARS-JWDSRC, PO Box 346/141 Exp Stn Rd, Stoneville, MS 38776

<sup>5</sup> Department of Entomology, Purdue University, West Lafayette, Indiana 47907

<sup>6</sup> International Center for Agricultural Research in the Dry Areas, Aleppo, Syria

## Abstract

We have previously characterized a gene coding for the secreted-salivary-gland-protein 11A1 (SSGP-11A1) from the Hessian fly, *Mayetiola destructor* (Say) (Diptera Cecidomyiidae). Here we report the cloning and characterization of three new genes coding for proteins designated SSGP-11B1, SSGP-11C1, and SSGP-11C2, and their relationship with the SSGP-11A1-encoding gene. Based on their structural conservation, similar regulation, and clustered genomic organization, we conclude that the four genes represent a gene superfamily, designated *SSGP-11*, which originated from a common ancestor. Cloning, Southern blot and *in situ* hybridization data suggest that each of the *SSGP-11* families has multiple members that cluster within short chromosome regions. The presence of a secretion signal peptide, the exclusive expression in the larval stage, and the clustered genomic organization indicate that this superfamily might be important for Hessian fly virulence/avirulence.

Keywords: gene superfamily, gene cluster

Abbreviations: SSGP secreted salivary gland protein

Correspondence: ming-shun.chen@gmprc.ksu.edu, scot\_hulbert@wsu.edu, m.bohssini@cgiar.org, xiang@ksu.edu

Received: 6.7.2005 | Accepted: 7.12.2005 | Published: 3.8.2006

Copyright: Creative Commons Attribution 2.5 (<http://creativecommons.org/licenses/by/2.5>)

ISSN: 1536-2442 | Volume 2006, Number 12

Cite this paper as:

Chen M, Fellers JP, Zhu YC, Stuart JJ, Hulbert S, El-Bouhssini M, Liu X. 2006. A super-family of genes coding for secreted salivary gland proteins from the Hessian fly, *Mayetiola destructor*. 13pp. *Journal of Insect Science* 6:12, available online: [insectscience.org/6.12](http://insectscience.org/6.12)

## Introduction

The Hessian fly, *Mayetiola destructor* (Say) (Diptera Cecidomyiidae), is one of the most destructive pests of wheat in Southern Europe, Northern Africa, Central Asia, and North America (Hatchett et al. 1987). Resistance genes in wheat have long been used for controlling this insect (Ratcliff and Hatchett 1997). Host-plant resistance provides an effective, cost-efficient, and environment-friendly way to control many important insect pests (Gracen 1986; Ratcliffe and Hatchett 1997). The challenge for the host-plant resistance strategy is the constant development of new biotypes that can overcome the resistance of deployed genes (Ratcliffe et al. 1994; Ratcliffe et al. 2000). To improve the durability of host-plant resistance, we need to understand how new biotypes arise. At present, little is known about the genetic basis for insect biotype differentiation. Thirty-one resistance genes have been identified and many biotypes have been isolated and are being maintained in laboratory collections of the Hessian fly (Ratcliff and Hatchett 1997; Martin-Sanchez et al. 2003; Williams et al. 2003). The availability of a large collection of host resistance genes and insect biotypes for this insect provides an ideal model system to reveal the genetic basis for biotype differentiation.

Numerous effector proteins from bacterial pathogens have been characterized (Leach et al. 1996; Dent et al. 1997). There is considerable evidence that these effector proteins are essential for microbes to be successful (Bai et al. 2000). Effector proteins that are recognized by specific plant-resistance-gene products and, thus, cause avirulence are designated *Avr* proteins (Baker et al. 1997). Pathogenic fungi also secrete *Avr* proteins into plant tissues, but less is known about this process (Orbach et al. 2000).

The Hessian fly interacts with wheat in a typical gene-for-gene specificity (Hatchett and Gallun 1970; Rider et al. 2002). Like pathogens, Hessian fly larvae apparently inject substances into host plants *via* their salivary glands during feeding (Byers and Gallun 1971; Hatchett et al. 1990). Thus the genetic determinants for Hessian fly biotypes could be those genes that encode secreted salivary gland proteins (SSGP) that are injected into host plants. The injected substances could be determinants for Hessian fly virulence and the variations in these substances could be determinants for biotype differentiation. As the

first step in determining the relationship between SSGP and virulence/avirulence of specific Hessian fly biotypes, we systematically analyzed the genes coding for SSGP from Hessian fly larvae following an expressed sequence tag (EST) approach. Numerous SSGP-encoding genes have been identified as a result of this analysis (Chen et al. 2004; Liu et al. 2004; Chen et al., unpublished). We previously characterized a gene coding for a small (7.1 kDa) protein designated SSGP-11A1 (Liu et al. 2004). Phylogenetic analysis of SSGP-11A1 together with other putative SSGP revealed that there are three other groups of SSGP that belong to the same sublineage group. The three new groups were named SSGP-11B1, SSGP-11C1, and SSGP-11C2, respectively. Here we report the isolation and characterization of the corresponding genes for these three new groups and their evolutionary relationship with the SSGP-11A1-encoding gene.

## Materials and Methods

### Hessian flies

Insects were from a laboratory colony that originated from insects collected in Ellis County, Kansas (Gagne and Hatchett 1989). Since then, the insects were maintained on susceptible wheat seedlings in environmental chambers at 20° C and 12:12 (L:D) photoperiod. The majority of the insects were biotype GP (Great Plains) although biotypes A, B and others were also found in low frequencies (Harris and Rose 1989).

### Library construction, screening, and sequencing

cDNA and BAC library construction, library screening and sequencing were conducted as described previously (Chen et al. 2004; Liu et al. 2004). The nucleotide sequences of genes and cDNAs from this article have been deposited in GenBank under accession nos. AY828552 to AY828563.

### RNA isolation and Northern blot analysis

Total RNA was extracted from salivary glands or whole insects using TRI reagent<sup>TM</sup> (Molecular Research Inc., www.mrcgene.com) following the protocol provided by the manufacturer. For Northern blots, equal amounts (5 µg) of total RNA were separated on a 1.2% agarose gel containing formaldehyde and blotted onto GeneScreen membrane (Perkin Elmer, www.perkinelmer.com). The membrane was incubated at 80° C for two

hours. Hybridization and washing conditions are the same as described elsewhere (Liu et al. 2004). For dot blot analysis, 2.5 µg of total RNA was used for each sample.

### Sequence analysis

Open-Reading-Frame (ORF) and sequence similarity analysis were performed using ORF finder and various BLAST programs on the website (<http://www.ncbi.nlm.nih.gov/>) of the National Center for Biotechnology Information (Bethesda, MD). Analysis for secretion signal peptides was carried out using SignalP (Center for Biological Sequence Analysis, Technical University of Denmark, <http://www.cbs.dtu.dk/services/SignalP/>) and PSORT II analysis (Prediction of Protein Sorting Signals and Localization Sites in Amino Acid Sequences, <http://psort.nibb.ac.jp/>). Molecular weight calculations and pI prediction of mature proteins were carried out using the 'Compute pI/Mw tool' ([http://us.expasy.org/tools/pi\\_tool.html](http://us.expasy.org/tools/pi_tool.html)).

### Fluorescent *in situ* hybridization

Polytene chromosomes were isolated from the salivary glands of second instar larvae and prepared as described by Shukle and Stuart (1995). Probes were prepared by labeling DNA (1 µg) with digoxigenin-conjugate dUTP by nick translation. Hybridizations were performed with 40–100 ng of denatured probe DNA on each chromosome preparation in 10 µl of hybridization solution (10% dextran sulfate, 2x SSC, 50% formaldehyde, and 20 µg of herring sperm DNA) at 37° C for 12 hours. Detection was performed using rhodamine-conjugated anti-digoxigenin and Alexa Fluor 488-conjugated streptavidin. Chromosomes were counterstained with DAPI. Digital images were taken with UV optics using an ORCA-ER (Hamamatsu Photonics, [www.hamamatsu.com](http://www.hamamatsu.com)) digital camera mounted on an Olympus BX51 microscope, and MetaMorph (Universal Imaging/Molecular Devices, [www.moleculardevices.com/](http://www.moleculardevices.com/)) imaging software.

## Results

### The SSGP-11B1-encoding gene and related cDNAs

Eighteen full-length cDNA clones were identified that were related to the SSGP-11B1-encoding gene from three thousand clones that were randomly sequenced from a salivary gland cDNA library

(Chen et al. 2004). Eleven of these cDNA clones are identical. The other seven cDNA clones contain minor (less than 2%) sequence variations. The proteins encoded by these seven cDNA clones are very similar to the SSGP-11B1 protein, the changes resulting in single amino acid substitutions (Figure 1A).

An SSGP-11B1 cDNA clone was used to probe a BAC library (Liu et al. 2004) and six BAC clones were identified that cross-hybridized with the cDNA probe. The genomic fragment carrying the SSGP-11B1-encoding gene from one of the BAC clones was subcloned and sequenced. The nucleotide and predicted amino acid sequences of the SSGP-11B1 gene are shown in Fig. 1B. The 5'-noncoding region contains a putative TATA box. Sequence alignment between the gene and the corresponding cDNA revealed that this gene contains a small intron consisting of 120 base pairs in the coding region. A predicted poly(A) addition signal was observed in the 3'-noncoding region. The predicted protein contains 103 amino acids. The first 16 amino acids constitute a putative secretion signal peptide. The gene encodes a mature (predicted) protein of 9.5 kDa with an isoelectric point of 8.9.

### The SSGP-11C1 and SSGP-11C2-encoding genes, and related cDNAs

Five full-length cDNA clones were identified that were predicted to code for a group of proteins designated the SSGP-11C family. cDNAs encoding two of these proteins, designated SSGP-11C1 and SSGP-11C2, were approximately 75% identical in their coding regions. The other three cDNA clones encoded proteins that were similar to one or the other of these two proteins except for several amino acid substitutions and small insertions/deletions (indels) (Figure 2A).

A genomic fragment corresponding to the SSGP-11C1-encoding cDNA was characterized from a BAC clone that hybridized to the cDNA probe. The nucleotide and predicted amino acid sequences of the SSGP-11C1-encoding gene are shown in Figure 2B. The 5'-noncoding region contains a putative TATA box. Sequence alignment between the gene and the corresponding cDNA revealed that this gene contains a small intron consisting of 109 base pairs in the coding region, in a similar position to that of the small intron in the SSGP-11B1-encoding gene. A poly(A) addition signal was found in the 3'-noncoding region. The

## A

SSGP-11B1	1	MQKLTFLVLLVVAIVAVESEGFCGRGNKKRATLEKSSSSRNAANEQATVTGTTTQQDVPS
P18c2	1	MQKLTFLVLLVVAIVAVESEGFCGRGN <b>R</b> KRATLEKSSSSRNAANEQATVTGTTTQQDVPS
L1c8	1	MQKLTFLVLLVVAIVAVESEGFCGRGNKKRATLEKSSSSRNAANEQATVTGTTTQQDVPS
L4e2	1	MQKLTFLVLLVVAIVAVESEGFCGRGNKKRATLEKSSSSRNAANEQATVTGTTTQQDVPS
P8d3	1	MQKLTFLVLLVVAIVAVESEGFCGRGNKKRATLEKSSSSRNAANEQATVTGTTTQQDVPS
P19c4	1	MQKLTFLVLLVVAIVAVESEGFCGRGNKKRATLEKSSSSRNAANEQATVTGTTTQQDVPS
L6h7	1	MQKLTFLVLLVVAIVAVESEGFCGRGN <b>R</b> KRATLEKSSSSRNAANEQ <b>A</b> VTGTTTQQDVPS
P18c10	1	MQKLTFLVLLVVAIVAVESEGFCGRGNKKRATLEKSSSSRNAANEQATVTGTTTQQDVPS

◆

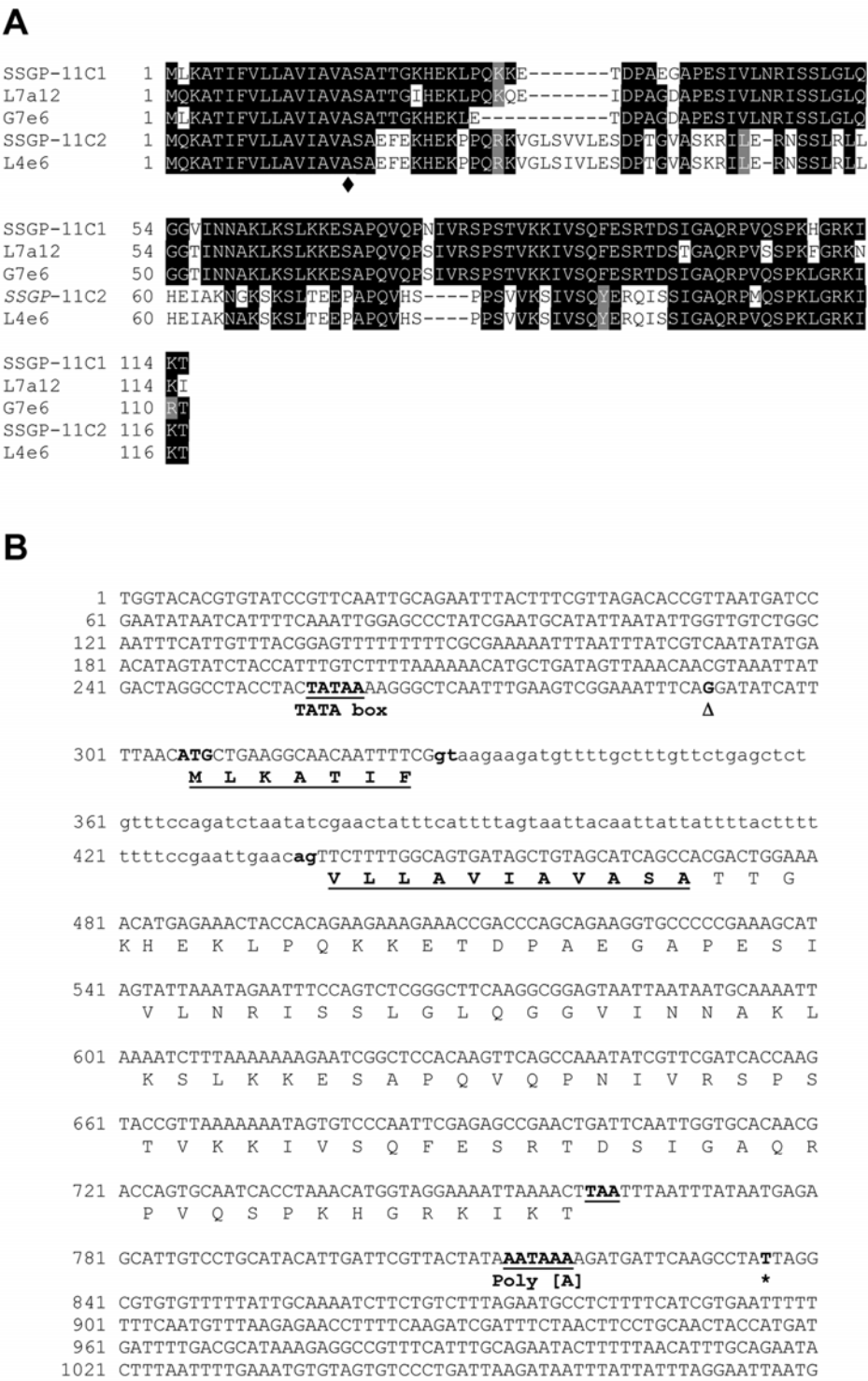
SSGP-11B1	61	EESVTEFAKIH <b>R</b> KISKDPSSFENFERSVSNKNAAGHNAQH
P18c2	61	EESVTEFAKIH <b>R</b> KISKDPSSFENFERSVSNKNAAGHNAQH
L1c8	61	EESVTEFAKIH <b>R</b> KISKDPSSFENFERSVSNKNAAGHNAQH
L4e2	61	EESVTEF <b>V</b> KIH <b>R</b> KISKDPSSFENFERSVSNKNAAGHNAQH
P8d3	61	EESVTEFAKIH <b>R</b> KISKDPSSFENFERSVSNKNAAGH <b>N</b> IQH
P19c4	61	EESVTEFA <b>E</b> I <b>H</b> RKISKDPSSFENFERSVSNKNAAGHNAQH
L6h7	61	EESVTEFAKIH <b>R</b> KISKDPSSFENFERSVSNKNAAGHNAQH
P18c10	61	EESVTEFAK <b>I</b> <b>P</b> RKISKDPSSFENFERSVSNKNAAGHNAQH

## B

```

1  GATCATTGCTAGTGTAACCTGTATCGGCCTGTAATCCGCTCCAATTATCAAGCGATTGAA
61  CTGCTGTCTGTAAAATAAAATATCCTCCGTTAGCAGAATTCGATCTTTTCAGATAATTTT
121 TCACTATAAAAGCCGAATGGAGATCCATTAAATTACATTTCGAATCTTAGAATTGTTTGGC
      TATA box                               Δ
181 ATCATGCAAACTAACTCTATTCGgttaagaatcataataaagtatatatttttgattttc
1      M Q K L T L F
241 taaggctcttctcgaacgccattgtagataaaaattattattatcaaattattaatttg
301 attttaatttcttctctgtgaatagTTCTACTGTAGTAGCTATCGTCGCTGTGGAGTCA
8      V L L V V A I V A V E S
361 GAAGGCTTTGCGGTAGAGGAAATAAAAGAGGGCCACTCTGGAAAAATCGTCCAGTAGC
21  E G F C G R G N K K R A T L E K S S S S
421 AGAAATGCTGCTAATGAACAAGCCACCGTTACAGGCACTACCACTCAACAGGATGTTCCA
41  R N A A N E Q A T V T G T T T Q Q D V P
481 TCAGAAGAATCTGTAACAGAATTTGCAAAATCCACAGAAAAATATCAAAGAAGGATCCA
61  S E E S V T E F A K I H R K I S K K D P
541 TCATCATTTTGGGAGAATTTTGAGAGAAGTGTTAGTAATAAGAACGCTGCAGGACACAAC
81  S S F W E N F E R S V S N K N A A G H N
601 GCTCAACACTTGACTAAAAAGACTATTATCAATGCTTCTTATTCTTGATTGTGAAAATA
101 A Q H                               Poly [A]
661 AAATATTGCTTTTGAATCCGTATAAAAAATGATC
      *
```

**Figure 1.** The SSGP-11B1-encoding gene and related cDNAs. A. Sequence alignment of putative proteins encoded by related cDNA clones. Amino acid differences are high-lighted by white residues. The symbol ◆ indicates the ending of putative secretion signal peptide. B: Nucleotide and predicted amino acid sequences of the SSGP-11B1-encoding gene. The sequence of the intron is shown in small letters. The putative TATA box, cDNA starting and ending positions, intron splicing donor and acceptor, and Poly (A) addition signal sequences are bold and underlined. The symbols Δ and \* represent the starting and ending positions of the corresponding cDNA. Predicted amino acid sequence is shown under the nucleotide sequence. Putative secretion signal peptide of the protein is bold and underlined.



**Figure 2.** The SSGP-11C1- and SSGP-11C2-encoding genes and related cDNAs. A. Sequence alignment of putative proteins encoded by these two genes and related cDNA clones. B: Nucleotide and predicted amino acid sequences of the SSGP-11C1-encoding gene.

## C

```

1  TTTGAAAAAAGCTACTACCCAAGCGCGTGTCTGGCAACGAGCAACATTTTTTGGGTT
61  TTAGTTGGAATATTTGGTACACTTGTGTTTCGTTTCAGATACGAAATTTACTTTCGTTAGAC
121 AGCAATAATGATTTGAATATAATCATTTTCAAATTGGAGCCCTATCGAATGCATATTTAT
181 ATTGGTTGTCTGGCAATTCATTGTTTACGGAGTTTTCTACCGAAAAGAATATCGTTTATC
241 ATCAATATTCGTAAATATATTTATCATTTTGGTGTTCAAAACACCTGATAGTAAAGCTGC
301 GTAAATTATTAGGTACTATATAAAGGCTCAATATGAAGTCGGAAATTTCAAATATTATTTT
      TATA box                                Δ
361 ATCATGCAAAAGGCAACAATTTTCGgttaagataatgTTTTgctgtgTTTTgagctctgtt
      M Q K A T I F
421 ttcagatctaatatcgattattatttccgaattgaacagTTCTTTTGGCAGTGATAGCTG
      V L L A V I A
481 TAGCATCAGCCGAATTTGAAAAACATGAGAAACCACCACAGAGGAAAGTAGGATTATCAG
      V A S A E F E K H E K P P Q R K V G L S
541 TTGTATTGGAAAGCGATCCAACAGGAGTTGCATCCAAAAGAATATTAGAAAGAAATTTCCA
      V V L E S D P T G V A S K R I L E R N
601 GTCTCCGACTATTGCACGAAATAGCTAAAAATGGAAAATCTAAATCTTTAACAGAAGAAC
      S L R L L H E I A K N G K S K S L T E E
661 CGGCTCCACAAGTTCACCTCTCCACCAAGTGTGCGTTAAAGCATAGTGTCCCAATACGAGA
      P A P Q V H S P P S V V K S I V S Q Y E
721 GACAAATAAGTTCAATTGGTGCCCAACGACCAATGCAATCACCTAACTTGGTAGGAAAA
      R Q I S S I G A Q R P M Q S P K L G R K
781 TTAAACTTAAATTTAATTTAAATGAGAGCATTGTGTTGCATTATTGATTTCGATACTAT
      I K T
841 AAATTAAACATGATTCAAGCGTATTAGGCGTATGTTTTTATTGAAAATGTTTCTGTCTTT
      Poly [A] *
901 AGAATGCTCTTTTCATCGTAAATTATTTTACTGTCCAAAAAACCTTTTCAAGATCGATT
961 TCTTACTTTCTGCAACTATCATGTGCAATGTATGTGCATTTTATTATTTAGTTAAAGTT
1021 TCTTGACAAATTTATCTATTTATTTATAGTAAATTGGAGAACATTTGTCCCAACAAATAC

```

**Figure 2 (con't).** C: Nucleotide and predicted amino acid sequences of the SSGP-11C2-encoding gene. The symbols were the same as described in Figure 1.

predicted protein contains 115 amino acids. The first 18 amino acids constitute a putative secretion signal peptide. The gene encodes a mature (predicted) protein of 10.5 kDa with an isoelectric point of 10.34.

Similarly, the nucleotide and predicted amino acid sequences of the SSGP-11C2-encoding gene are shown in Figure 2C. There is a putative TATA box. The gene contains a small intron consisting of 74 base pairs in the same position as that of the small intron in the SSGP-11C1-encoding gene. The predicted protein contains 117 amino acids. The

first 18 amino acids constitute a putative secretion signal peptide. The gene encodes a mature (predicted) protein of 11.0 kDa with an isoelectric point of 10.13.

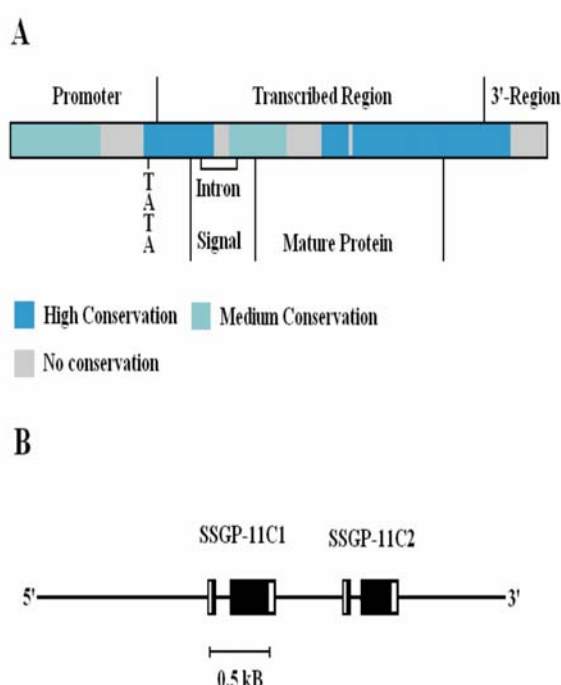
### Structural conservation between the SSGP-11C1- and SSGP-11C2-encoding genes

The proteins encoded by the SSGP-11C1 and SSGP-11C2-encoding genes share structural similarity despite significant divergence. Except for one amino acid difference, the two proteins share an otherwise identical secretion signal peptide (Figure 2A). Similarly, there is only one residue

difference between the two proteins in the nineteen amino acids at the C-terminal. In addition, localized conservation can also be found in the diversified N-terminal portion of the mature protein.

At the nucleotide level, sequence comparison revealed that most parts of the SSGP-11C1 and SSGP-11C2-encoding genes are conserved (Figure 3A). The four major regions that are not conserved are located immediately before the TATA box, within the intron, in the N-terminal region of the mature-protein-coding region, and in the 3'-untranscribed region, respectively.

Structural conservation between SSGP-11C1 and SSGP-11C2 indicated that the two genes arose by gene duplication. Consistent with this possibility, analysis of the genomic organization revealed that the SSGP-11C1 and SSGP-11C2-encoding genes are tandem repeats located within a 1.5 Kb DNA fragment (Figure 3B).



**Figure 3.** Genomic organization and sequence similarity between the SSGP-11C1 and SSGP-11C2-encoding genes. A: Sequence similarity. The sequence alignment was generated by the BLASTN program. The bars with different colors represent different levels of conservation between these two genes. The alignment is divided into different regions according to their functions, including promoter, transcribed region (regions coding for signal peptide and mature protein), and 3'-region. A single intron is within the region coding for the signal peptide. "TATA" indicates where the putative TATA box is located. B: Genomic organization. The boxed regions represent exons with the coding region in black.

## Structural conservation among gene families

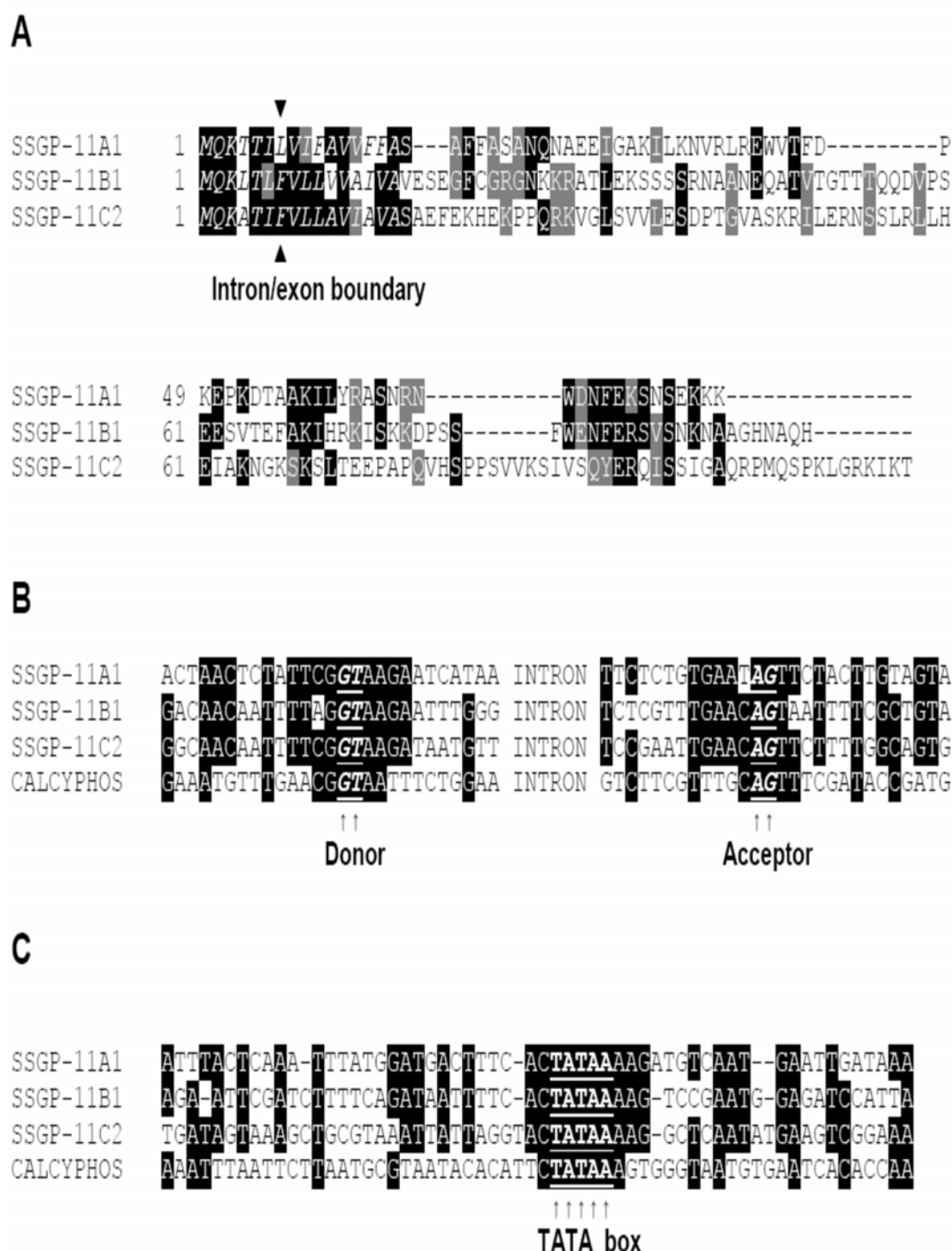
According to the genomic organization and structural conservation, SSGP-11C1, SSGP-11C2, and related cDNAs can be considered a gene family. SSGP-11B1 and the previously characterized SSGP-11A1 (Liu et al. 2004) represent two other families. Sequence analysis indicated that there was structural conservation among these gene families. First, the number, the location, and the phase of the intron are conserved. There is a single intron in all genes and the location of the intron is the same in terms of protein encoding (Fig. 4A). Second, the secretion signal peptide, which is different from those found in other SSGPs (Chen et al. 2004, Liu et al. 2004), is conserved among all three families (Figure 4A). In addition to the conserved secretion signal peptide, there is also localized conservation in the mature protein among the three sequences (Figure 4A).

At the nucleotide level, there are also conservations around the intron/exon boundaries (Figure 4B) and in the putative promoter region (Fig. 4C). In the sequence alignment of the intron/exon boundaries with the three genes together with an unrelated Hessian fly gene encoding a calcyphosine-like protein, there are 29, 30, and 33 residues conserved in the SSGP-11A1, SSGP-11B1, and SSGP-11C2-encoding genes, respectively (Figure 4B). In comparison, only 19 residues are conserved in the calcyphosine-encoding gene. Similar observations can also be found in the alignment of the putative promoter region. There are 29–30 residues conserved in all the SSGP genes, while only 18 residues are conserved in the calcyphosine gene (Figure 4C). The structural conservation among these gene families suggests that they may have arisen from the same origin through gene duplications. The intron/exon boundaries and the promoter region of the duplicated genes diversified at a lower rate than the rest of the gene, which could be due to regulatory roles in RNA splicing and in gene expression.

## Gene expression analysis

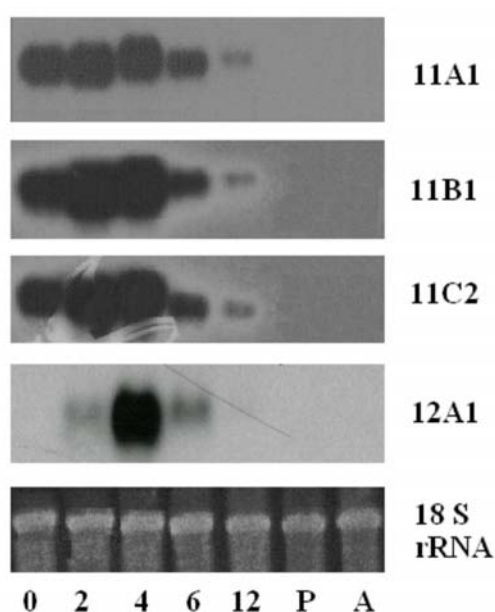
The conservation in the promoter region suggested that these gene families might have similar regulatory mechanisms controlling their expression. Northern blot analysis indeed demonstrated very similar or the same expression profiles among the SSGP-11A1, SSGP-11B1, and SSGP-11C2-encoding genes (Figure 5). This was in comparison with a very different expression profile for a different gene, the SSGP-12A1-encoding gene, which was only transiently expressed in 2- to 6-day





**Figure 4.** Structural similarity among the SSGP-11A1, SSGP-11B1, and SSGP-11C2-encoding genes. A: Amino acid sequence alignment. The first 16 amino acids (italic) constitute a putative secretion signal peptide. Black residues indicate identical amino acids between at least two proteins. Gray residues represent conserved substitutions. White residues represent amino acid differences. Dashed lines indicate absence of amino acids in specific proteins. The conserved location of the intron (between F and V) is indicated by the symbols ▲ and ▼. B: Conservation in nucleotide sequence at intron/exon splicing boundaries. The alignment was made using standard program (see Experimental Protocol section). CALCYPHO represents calcyphosine, an unrelated gene from the Hessian fly. Black areas represent residues identical at least 3 sequences. Intron splicing donor and acceptor are bold and underlined and indicated by arrows. C: Conservation in the promoter region. Putative “TATA” box is bold and underlined and indicated by arrows.

old larvae (Liu et al. 2004). All the SSGP-11-encoding genes were abundantly expressed in freshly hatched larvae and reached a maximum in 2- to 4-day old larvae. The RNA level was significantly lower in 6-day old larvae and became barely detectable in 12-day old larvae. No RNA could be detected in pupae and adults. The first 4 days are the first instar, which is the critical stage to determine if a larva lives or will die in compatible or incompatible interactions with wheat plants (Hatchett and Gallun 1970; Byers and Gallun 1971). The predominant expression of these genes in the first instar and all of the four genes peaked in 4-day old larvae suggested that they may play a critical role in Hessian fly virulence/avirulence.



**Figure 5.** Northern blot analysis. The numbers 0, 2, 4, 6, and 12 underneath the figure represent samples extracted from 0-day (freshly hatched), 2-day, 4-day, 6-day, and 12-day old larvae, respectively, while the letters P and A represent samples extracted from pupae and adults. A blot was hybridized to a probe derived from the SSGP-11A1 gene (11A1). After the radioactivity stripped or decayed (no image by 48 hours exposure), the same blot was rehybridized sequentially to probes specific to SSGP-11B1 (11B1), SSGP-11C2, and SSGP-12A1, an unrelated SSGP gene (Liu et al. 2004). An image of the 18 S rRNA for the blot was included to show that the loading amount of RNA in each lane was about the same.

### Southern blot analysis

Southern blot analysis with genomic DNA isolated from Hessian fly larvae revealed restriction fragments with specific probes (Figure 6). A probe specific to the SSGP-11A1-encoding gene detected three fragments with both *Eco*RI and *Bam*HI restriction enzymes (Figure 6A). Since there are no

*Eco*RI or *Bam*HI sites in the gene and the size of the gene is less than 1 kb (Liu et al. 2004), the three fragments indicated that there are at least three SSGP-11A1 related genes in the Hessian fly genome. Alternatively, there are allelic variations in the SSGP-11A1 locus that produce different restriction fragments on the Southern blots.

Multiple fragments were also detected in Southern blots with a probe specific to the SSGP-11B1-encoding gene (Fig. 6B). There is an *Eco*RI and a *Bam*HI site in the SSGP-11B1-encoding gene, but the *Eco*RI site is not within the region where the probe was generated. The single *Bam*HI site in the SSGP-11B1 probe should only detect two hybridization bands if there is only one gene. The multiple fragments with sizes greater than 1 kb (the SSGP-11B1-encoding gene is less than 1 kb) on Southern blots again indicated that there are multiple genes related to SSGP-11B1 in the Hessian fly genome, or that there are allelic variations in the SSGP-11B1 locus.

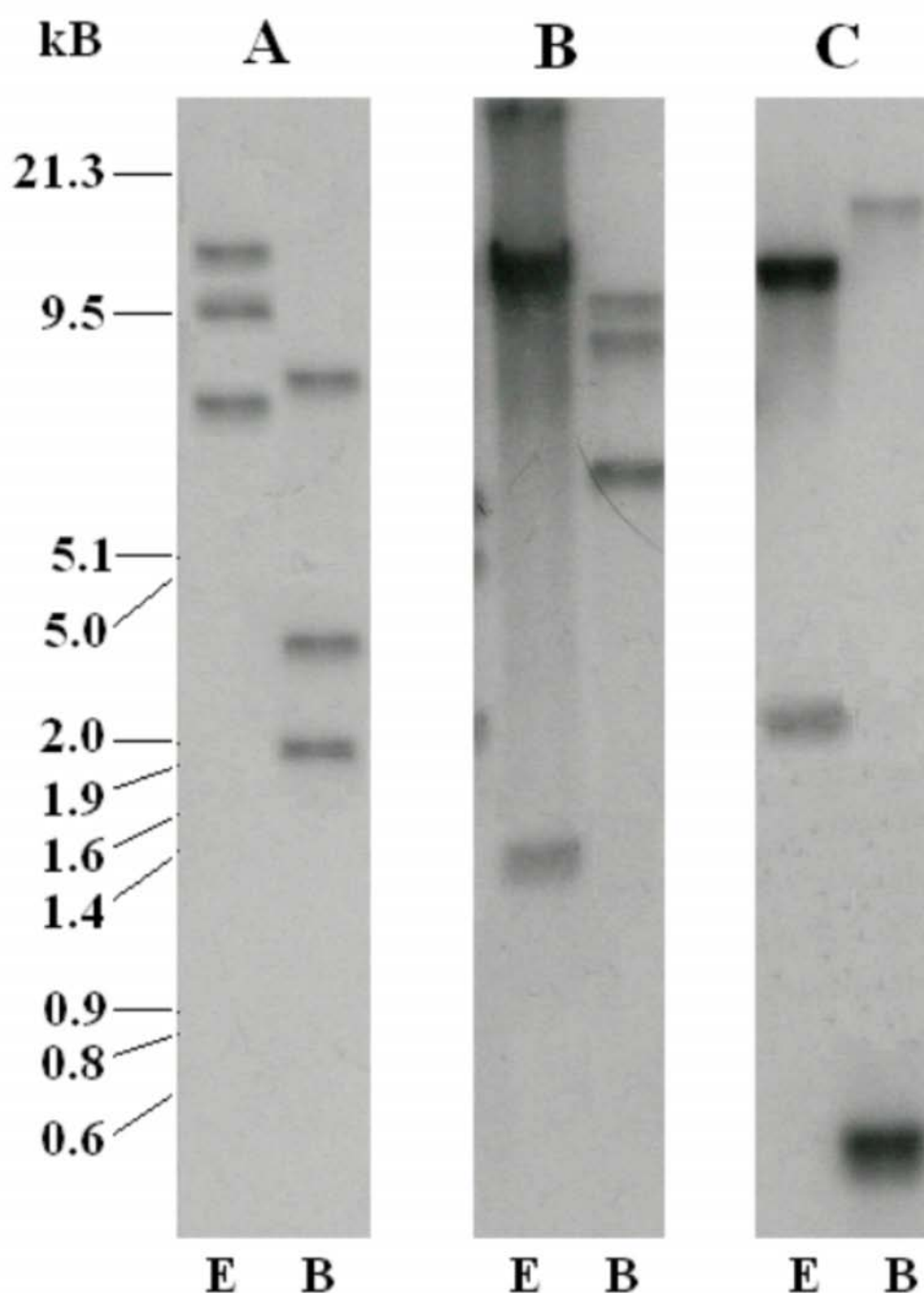
Two fragments were detected on Southern blots with a probe generated from the SSGP-11C2 encoding cDNA (Fig 6C). Since SSGP-11C2 has sequence similarity with SSGP-11C1, the probe may also hybridize to both subfamilies of genes. However, there are no *Eco*RI or *Bam*HI sites in the genomic fragment that carries both the SSGP-11C1 and SSGP-11C2-encoding genes. The two bands detected by the probe indicated that there are other related gene(s) in addition to the SSGP-11C1 and SSGP-11C2 encoding genes unless there are allelic variations in these two loci.

### Chromosomal localization

Fluorescence *in situ* hybridization (FISH) was performed to determine the relative physical locations of the three gene families using a two color system. Figure 7A shows the result of FISH using BAC clones 3k23 (containing SSGP-11A1) and 8p1 (containing SSGP-11C1 and SSGP-11C2), while Figure 7B is the image of FISH using BAC clones 8p1 and 8p24 (containing SSGP-11B1). According to the FISH results, SSGP-11A1 and SSGP-11C1 are in the distal region of the short arm of chromosome A1. SSGP-11B1 is in the proximal region of the same chromosome arm (Figure 7C).

### Discussion

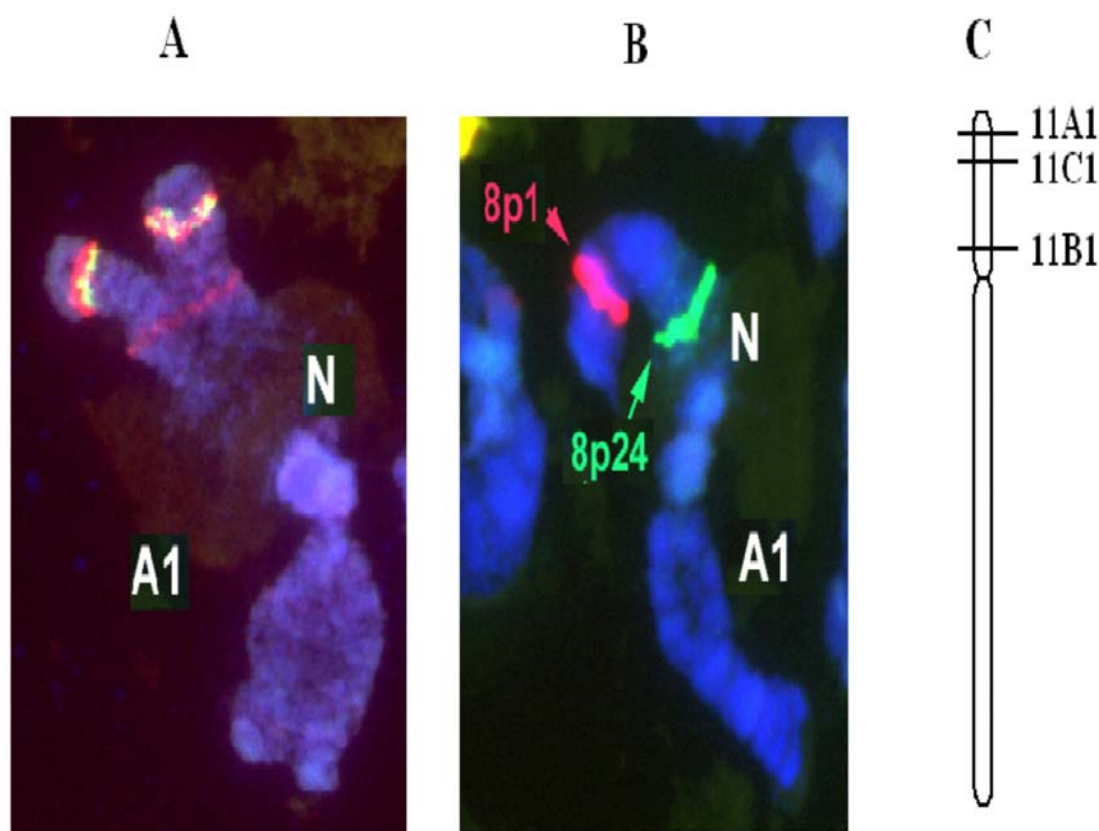
The SSGP-11C1 and SSGP-11C2 encoding genes share relatively high sequence similarity and are located in a 1.5 kb DNA fragment as tandem repeats. This suggests that they originated from a



**Figure 6.** Southern blot analysis of genomic DNA. Genomic DNA was isolated from first instar larvae and digested with restriction enzymes EcoRI (E) or BamHI (B). The blot was hybridized to the SSGP-11A1 cDNA (panel A). After the blot was stripped or decayed, the blot was rehybridized sequentially to probes specific to SSGP-11B1 (panel B) and SSGP-11C2 (panel C). DNA size markers were indicated on the left of the figure.

gene duplication event followed by sequence diversification. The conservation in most of the mature protein (Figure 2A) indicates functional similarity between these two proteins. The above evidence, suggests that SSGP-11C1 and SSGP-11C2 are two members of a gene family, the SSGP-11C family. On the other hand, the high degree of

sequence diversification between the SSGP-11C family and the SSGP-11A1 and SSGP-11B1-encoding genes indicated they are distinct gene families that are likely to perform different functions, or possibly have different specificities.



**Figure 7.** Fluorescent *in situ* hybridization (FISH) analysis. A: Relative positions of the SSGP-11A1 and SSGP-11C1 genes. BAC clones 3k23 (containing SSGP-11A1) and 8p1 (containing SSGP-11C1 and -11C2) were labeled with red and green fluorescent dyes, respectively, and hybridized to polytene chromosomes prepared from larval salivary glands. A yellow image was observed in the overlapping region, indicating that these two BACs were located next to each other. In addition to a strong image at the distal region of chromosome A1 with the 3k23 (red) probe, a weak image was also detected in the proximal region of the same chromosome. This weak image appeared to be in the same region where the SSGP-11B gene family (see panel B), and might be caused by related hybridization. B: Relative positions of the SSGP-11B1 and SSGP-11C1 genes. BAC clones 8p24 (containing SSGP-11B1) and 8p1 (containing SSGP-11C1 and -11C2) were labeled with green and red fluorescent dyes, respectively, and hybridized to polytene chromosomes prepared from larval salivary glands. A single location for each BAC probe was observed in the Hessian fly genome. C: A physical map of the three SSGP genes. A map was constructed to show the relative chromosomal localization of the SSGP-11A1, SSGP-11B1, and SSGP-11C1 (and -11C2) genes according to the FISH analysis.

Although the DNA sequences have little noticeable sequence similarity, and very little homology exists in the predicted amino acid sequences (data not shown), structural conservation among members from the three families can be found. For example, a single intron was found in all gene families (Figures 1 and 2, Liu et al. 2004) and the location of the intron was exactly the same, namely in the eighth codon of the putative signal peptide (Figure 3B and Figure 4A). Not only the position was the same, the phase of the intron was also identical in all these genes. It is unlikely that the same position and identical phase of the intron was achieved through convergence of genes of different origin. In fact, different numbers of introns were found in different locations and in different phases in other

SSGP-encoding genes (Liu et al. 2004; Chen et al., unpublished) regardless of their expression patterns or protein structures. The conservation in the number and the location of the intron also suggests that members from the three gene families arose from a common origin through gene duplications.

Consistent with this hypothesis, localized conservation exists among these genes at the nucleotide and amino acid levels. At the amino acid level, the secretion signal peptide and a short region at the C-terminal are relatively conserved (Figure 4A). Such conservation could not be found when comparing these proteins with other unrelated SSGPs (data not shown). At the

nucleotide level, conservations can also be found in the intron/exon boundaries (Figure 4B) and in the promoter region (Figure 4C) in addition to the signal peptide coding region. The localized conservations were consistent with the possibility that the signal peptide coding region, the intron/exon boundaries, and the promoter region diversified at a slower rate after gene duplication because of a functional requirement for secretion, RNA splicing, and expression regulation. The other parts of the genes diversified at a higher rate to gain new functions for the proteins. The conservation in the promoter region could explain that the expression profiles were essentially the same for all the three families, and that they were very different from those of other unrelated SSGP genes (Chen et al. 2004; Liu et al. 2004).

Based on above evidence we conclude that SSGP-11C1 and SSGP-11C2 are two members of a gene family, the SSGP-11C family. Each gene family appeared to have multiple members according to genomic Southern blot analysis. In addition, there were multiple cDNAs that encoded proteins with sequence variations in each gene family. Even though some of the cDNA variations could be caused by polymorphisms, many of them were likely to represent paralogs because of the predicted amino acid divergence and the presence of relatively large indels (Fig. 2B).

The two genes (SSGP-11C1 and SSGP-11C2) in the SSGP-11C family are tandem repeats located within a 1.5 Kb DNA fragment (Fig. 3B). Evidence also indicated that multiple members in other families were also clustered within a short chromosome region. For example, Southern blot analysis of a 145 Kb BAC clone produced the same hybridization pattern (data not shown) revealed by genomic Southern blot analysis, indicating that multiple genes from the SSGP-11B family were contained within this BAC clone. In addition to clustered family members, the SSGP-11A and SSGP-11C encoding gene families are also located next to each other on the distal region of the short arm of chromosome 1A. The localization of the SSGP-11B family in the proximal region of the same chromosome arm might be caused by a chromosome rearrangement.

It is not clear at present what kind of function this group of proteins may perform. Sequence comparison using various programs detected no significant homology with genes or proteins in known databases including the GenBank. To

elucidate the exact biological functions of these genes, we are currently generating recombinant proteins and producing antibodies. Our future direction is to determine where these proteins are located once they are injected into plants and which proteins from host plants they are interacting with.

## Notes

### Disclaimer

Mention of commercial or proprietary product does not constitute endorsement by the USDA.

## Acknowledgments

The authors thank Drs Kun-Yan Zhu and Yoon-Seong Park for reviewing an earlier version of the manuscript. This paper is contribution No 05-221-J from the Kansas Agricultural Experiment Station. Hessian fly voucher specimens (No. 150) are located in the KSU Museum of Entomological and Prairie Arthropod Research, Kansas State University, Manhattan, Kansas. This work was supported by a grant from the National Research Initiative Competitive Grant program, the U.S. Department of Agriculture (USDA 04-35607-14861), and a grant from USDA regional IPM, and funds from Kansas State University Plant Biotechnology Center.

## References

- Bai JF, Choi SH, Ponciano G, Leung H, Leach JE. 2000. *Xanthomonas oryzae* pv. *oryzae* avirulence genes contribute differently and specifically to pathogen aggressiveness. *Molecular Plant-Microbe Interactions* 13: 1322-1329.
- Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar SP. 1997. Signaling in plant-microbe interactions. *Science* 276: 726-733.
- Byers RA, Gallun RL. 1971. Ability of the Hessian fly to stunt winter wheat. I. Effect of larval feeding on elongation of leaves. *Journal of Economic Entomology* 65: 955-958.
- Chen MS, Fellers JP, Stuart JJ, Reese JC, Liu XM. 2004. A group of related cDNAs encoding secreted proteins from Hessian fly *Mayetiola destructor* (Say) salivary glands. *Insect Molecular Biology* 13: 101-108.
- Dent JA, Davis MW, Avery L. 1997. *Avr-15* encodes a chloride channel subunit that mediates inhibitory glutamatergic neurotransmission and ivermectin sensitivity in *Caenorhabditis elegans*. *EMBO Journal* 16: 5867-5879.
- Gagne RJ, Hatchett JH. 1989. Instars of the Hessian fly (Diptera: Cecidomyiidae). *Annals of the Entomological Society of America* 82: 73-79.

- Gracen VE. 1986. Host plant resistance for insect control in some important crop plants. *CRC critical reviews in plant sciences* 4: 277-291.
- Harris MO, Rose S. 1989. Temporal changes in egg-laying behavior of the Hessian fly. *Entomologia experimentalis applicata* 53: 17-29.
- Hatchett JH, Gallun RL. 1970. Genetics of the ability of Hessian fly, *Mayetiola destructor*, to survive on wheats having different genes for resistance. *Annals of Entomological Society of America* 63: 1400-1407.
- Hatchett JH, Kreitner GL, Elzinga RJ. 1990. Larval mouthparts and feeding mechanism of the Hessian fly (Diptera: Cecidomyiidae). *Annals of Entomological Society of America* 83: 1137-1147.
- Hatchett JH, Starks KJ, Webster, JA. 1987. Insect and mite pests of wheat. In: *Wheat and Wheat improvement*. pp. 625-675. Agronomy Monograph No. 13.
- Leach JE, White FF. 1996. Bacterial avirulence genes. *Annual Review of Phytopathology* 34: 153-179.
- Liu XM, Fellers JP, Wilde GE, Stuart JJ, Chen MS. 2004. Characterization of two genes expressed in the salivary glands of the Hessian fly, *Mayetiola destructor* (Say). *Insect Biochemistry and Molecular Biology* 34: 3229-237.
- Maniatis T, Fritsch E, Sambrook J. 1982. *Molecular cloning – A laboratory manual*. pp 90-91. New York: Cold Spring Harbor Press.
- Martin-Sanchez JA, Gomez-Colmenarejo M, Del Moral J, Sin E, Montes MJ, Gonzalez-Belinchon C, Lopez-Brana I, Delibes A. 2003. A new Hessian fly resistance gene (h40) transferred from the wild grass *Aegilops triuncialis* to hexaploid wheat. *Theoretical and Applied Genetics* 106: 1248-1255.
- Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. 2000. A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *Plant Cell* 12: 2019-2032.
- Ratcliffe RH, Cambron SE, Flanders KL, Bosque-Perez NA, Clement SL, Ohm HW. 2000. Biotype composition of Hessian fly (Diptera: Cecidomyiidae) populations from the southeastern, Midwestern, and northwestern United States and virulence to resistance genes in wheat. *Journal of Economic Entomology* 93: 1319-1328.
- Ratcliffe RH, Hatchett JH. 1997. Biology and genetics of the Hessian fly and resistance in wheat. In: K. Bondari, editor. *New Developments in Entomology*. pp. 47-56. Trivandrum, India: Research Signpost, Scientific information guild.
- Ratcliffe RH, Safranski GG, Patterson FL, Ohm HW, Taylor PL. 1994. Biotype status of Hessian fly (Diptera: Cecidomyiidae) populations from the eastern United States and their response to 14 Hessian fly resistance genes. *Journal of Economic Entomology* 87: 1113-1121.
- Rider SD, Sun W, Ratcliffe RH, Stuart JJ. 2002. Chromosome landing near avirulence gene vH13 in the Hessian fly. *Genome* 45: 812-822.
- Shukle RH, Stuart JJ. 1995. Physical mapping of DNA sequences in the Hessian fly, *Mayetiola destructor*. *J Hered* 86: 1-5.
- Williams CE, Collier CC, Sardesai N, Ohm HW, Cambron SE. 2003. Phenotypic assessment and mapped markers for *h41*, a new wheat gene conferring resistance to Hessian fly (Diptera: Cecidomyiidae). *Theoretical and Applied Genetics* 107: 1516-1523.