Published July 7, 2016

Candidate Gene Sequence Analyses toward Identifying *Rsv3*-Type Resistance to *Soybean Mosaic Virus*

N. R. Redekar, E. M. Clevinger, M. A. Laskar, R. M. Biyashev, T. Ashfield, R. V. Jensen, S. C. Jeong, S. A. Tolin, and M. A. Saghai Maroof*

Abstract

Rsv3 is one of three genetic loci conferring strain-specific resistance to *Soybean mosaic virus* (SMV). The *Rsv3* locus has been mapped to a 154-kb region on chromosome 14, containing a cluster of five nucleotide-binding leucine-rich repeat (NB-LRR) resistance genes. High sequence similarity between the *Rsv3* candidate genes challenges fine mapping of the locus. Among the five, Glyma14g38533 showed the highest transcript abundance in 1 to 3 h of SMV-G7 inoculation. Comparative sequence analyses were conducted with the five *Rsv3* candidate NB-LRR genes from susceptible (*rsv*-type) soybean [*Glycine max* (L.) Merr.] cultivar Williams 82, resistant (*Rsv3*-type) cultivar Hwangkeum, and resistant lines L29 and RRR. Sequence comparisons revealed that Glyma14g38533 had far more polymorphisms than the other candidate genes. Interestingly, Glyma14g38533 gene from *Rsv3*-type lines exhibited 150 single-nucleotide polymorphism (SNP and six insertion–deletion (InDel) markers relative to *rsv*-type line, Furthermore, the polymorphisms identified in three *Rsv3*-type lines were highly conserved. Several polymorphisms were validated in 18 *Rsv3*-type resistant and six *rsv*-type susceptible lines and were found associated with their disease response. The majority of the polymorphisms were located in LRR domain encoding region, which is involved in pathogen recognition via protein–protein interactions. These findings associating Glyma14g38533 with *Rsv3*-type resistance to SMV suggest it is the most likely candidate gene for *Rsv3*.

Core Ideas

- *Rsv3* locus confers strain-specific resistance to *Soybean mosaic virus*
- The *Rsv3* locus contains a cluster of five NB-LRR resistance genes
- Comprehensive study of five *Rsv3* candidate NB-LRR gene sequences was conducted
- Comparisons were done between *Rsv3*-type resistant vs. *rsv*-type susceptible soybeans
- Glyma14g38533 gene was identified as the most likely candidate gene for *Rsv3*

SOYBEAN MOSAIC VIRUS (genus *Potyvirus*, family *Potyviridae*) is a seedborne and aphid-transmitted virus causing a mosaic disease found in many soybean fields affecting seed quality and yield. Deploying durable virus-resistant soybean varieties can prevent economic

N.R. Redekar, Dep. of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA; E.M. Clevinger, Dep. of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA; M.A. Laskar, Biotechnology Dep., St. Anthony's College, Shillong, India; R.M. Biyashev, Dep. of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA; T. Ashfield, Dep. of Biology, Indiana Univ., Bloomington, IN; R.V. Jensen, Dep. of Biological Sciences, Virginia Tech, Blackburg, VA; S.C. Jeong, Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju, Republic of Korea; S.A. Tolin, Dep. of Plant Pathology, Physiology, and Weed Sciences, Virginia Tech, Blacksburg, VA; M.A. Saghai Maroof, Dep. of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA. Received 21 Sep. 2015. Accepted 22 Feb. 2016. *Corresponding author (smaroof@vt.edu).

Abbreviations: CC, coiled-coil; dN/dS, ratio of nonsynonymous to synonymous changes; hpi, hours postinoculation; mRNA, messenger RNA; NB-LRR, nucleotide-binding leucine-rich repeat; PCR, polymerase chain reaction; RPKM, reads per kilobase per million; SMV, *Soybean mosaic virus*.

Published in Plant Genome
Volume 9. doi: 10.3835/plantgenome2015.09.0088

losses associated with this disease. The SMV strains identified in the United States are classified into seven strain groups, G1 to G7, based on their virulence on a set of differential cultivars (Cho and Goodman, 1979). Three resistance genes, including *Rsv1* on chromosome 13, *Rsv3* on chromosome 14, and *Rsv4* on chromosome 2, have been identified in soybean, which condition strain-specific resistance to SMV (Buss et al., 1997; Buss et al., 1989; Buzzell and Tu, 1984; Hayes et al., 2000; Kiihl and Hartwig, 1979; Liao et al., 2002). All three *Rsv* loci are genetically dominant and confer single-gene resistance to SMV. The *Rsv1* gene confers extreme resistance to SMV-G1 through SMV-G4, but conditions necrosis or mosaic symptoms to SMV-G5 through SMV-G7 (Chen et al., 1991; Tucker et al., 2009). The *Rsv3* gene, on the contrary, confers extreme resistance to SMV-G5 through SMV-G7 but no resistance to SMV-G1 through SMV-G4 (Jeong et al., 2002; Ma et al., 2002). The *Rsv4* gene provides early resistance to all SMV strains; however, the virus may overcome this resistance and exhibit delayed replication and movement, resulting in late susceptibility symptoms (Chen et al., 1993; Hayes et al., 2000; Ma et al., 1995; Saghai Maroof et al., 2010).

The complexity of the SMV–soybean interaction is widely studied. A microarray study with SMV-susceptible cultivar Williams 82 at 7, 14, and 21 h postinoculation (hpi) with SMV-G2 showed downregulation of defense-related genes in early stages of SMV infection (Babu et al., 2008). The proteins belonging to functional categories, such as the defense response, were also identified in SMV-resistant cultivar Kefeng No. 1 at 4 hpi with SMV-JN17 using a proteomics approach (Yang et al., 2011). These studies suggest that the SMV–soybean interactions can trigger different transcriptional and translational responses in susceptible and resistant hosts (Babu et al., 2008; Yang et al., 2011). Recently, type 2C protein phosphatase genes were reported as being associated with *Rsv3*-mediated extreme resistance response in SMV-resistant line L29 at 8 hpi with SMV-G5H (Seo et al., 2014). The SMV–soybean interaction at the *Rsv3* locus activates abscisic acid signaling, which regulates type 2C protein phosphatase and callose biosynthesis to prevent virus movement (Seo et al., 2014). However, the gene at the *Rsv3* locus that initiates these defense responses remains undefined.

The most characterized disease resistance gene family encodes for NB-LRR proteins. The resistance-gene-mediated defense response is initiated when pathogen effectors or the cofactors that bind to these effectors are detected by the host (Bent and Mackey, 2007). This detection is facilitated by the binding specificity of the LRR domain within the NB-LRR protein. Both the *Rsv1* and *Rsv3* loci are associated with a cluster of genes that encode NB-LRR proteins with a coiled-coil N-terminal domain (Hayes et al., 2004; Jeong et al., 2002; Suh et al., 2011; Wang et al., 2011). The 154-kb *Rsv3* chromosomal region (hereafter *Rsv3* locus) contains a cluster of five NB-LRR genes designated as Glyma14g38500, Glyma14g38510, Glyma14g38540, Glyma14g38560, and Glyma14g38590 (Suh et al., 2011). This study will use Glyma.Wm82.a1.v1.1 gene model names for these *Rsv3* candidate NB-LRR genes, which are Glyma14g38500, Glyma14g38516, Glyma14g38533, Glyma14g38561, and Glyma14g38586 (Table 1). Fine mapping of this locus to identify a single NB-LRR gene that encodes *Rsv3*-type resistance is challenging because of high sequence similarity. Despite the sequence similarity, the functional form of NB-LRR genes are expected to differ between resistant and susceptible hosts, since these genes are evolving under selective pressure.

Table 1. *Rsv3* candidate gene annotations.

Glyma.Wm82.a1.v1.0	Glyma.Wm82.a1.v1.1	Glyma.Wm82.a2.v1
Glyma14g38500	Glyma14g38500	Glyma.14g204500
Glyma14g38520	Glyma14g38516	Glyma.14g204600
Glyma14g38540	Glyma14g38533	Glyma.14g204700
Glyma14g38560	Glyma14g38561	Glyma.14g205000
Glyma14g38590	Glyma14g38586	Glyma.14g205300

The current study is focused on reducing the number of *Rsv3* candidate NB-LRR genes within the *Rsv3* locus. Toward this goal, we employed a comparative sequencing approach of *Rsv3* candidate gene sequences from susceptible Williams 82 and three resistant soybean lines, Hwangkeum, L29, and RRR. The *Rsv3* candidate, Glyma14g38533, showed highest transcript abundance during early stages of SMV infection and gene sequence was conserved in the three resistant lines. Several genetic diversity features, in the form of InDel and SNP markers, were identified and validated in 18 *Rsv3*-type resistant and six *rsv*-type susceptible lines. These polymorphisms were located in the LRR domain, which is involved in pathogen recognition. Two allelic forms of the Glyma14g38533 gene, one conserved in resistant lines and another in susceptible lines, were described. Overall, our analyses provide evidence that the Glyma14g38533 gene is the most likely candidate for *Rsv3*.

Material and Methods

Soybean Mosaic Virus Inoculation and Tissue Sampling

A culture of SMV-G7 used previously (Saghai Maroof et al., 2008) was maintained on *rsv3*-genotype susceptible soybean cultivars, Hutcheson and Essex, as a source of inoculum. Virus inoculum was prepared as 1:10 (w/v) of SMV-G7-infected leaf tissue in inoculation buffer that was composed of 1% Celite545 in 0.01M sodium phosphate buffer, pH 7.0 (Fisher Scientific Inc.). Fully expanded first trifoliolate leaflets of *Rsv3*-genotype resistant soybean line L29 were inoculated with SMV-G7 by gently rubbing virus inoculum onto each leaflet with a pestle. Samples comprising of trifoliolates from three inoculated plants were collected after 1, 2, and 3 hpi and stored at −80°C until further processing.

Messenger RNA Sequencing and Data Preprocessing

Total RNA was extracted from the above-mentioned leaflet tissue samples using the RNeasy Plant Mini Kit, with on-column DNase digestion (QIAGEN Inc.). The RNA integrity value was estimated for these total RNA samples using an Agilent 2100 Bioanalyzer system (Agilent Technologies Inc.). Total RNA samples with RNA integrity value >8.0 were used to construct messenger RNA (mRNA) libraries. Each library was sequenced with 50-cycle paired-end per lane cycles in two lanes using Illumina HiSeq1000 sequencer (Illumina Inc.) at Virginia Bioinformatics Institute, Virginia Tech. The sequencing data preprocessing and quality control analyses were performed using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit): (i) filtering the sequencing reads with 70% of the bases falling below the threshold quality score (Q = 25); (ii) removing the adaptor sequences and ambiguous bases. These preprocessed sequencing reads were used for two purposes: (i) estimate transcript abundance for *Rsv3* candidate genes and (ii) identify polymorphisms in *Rsv3* candidate gene sequence assemblies. The statistics for the sequencing data analysis is provided in Supplemental Table S1.

Estimating Transcript Abundance of *Rsv3* Candidate Genes

The preprocessed reads were aligned to the reference Williams 82 genome sequence assembly 1.1 and annotation Glyma1.1 using TopHat2 (v2.0.8), a splice-aware aligner tool (Kim et al., 2013). Based on these sequence alignments, read counts (number of reads mapped to a gene) for each gene were estimated using HTSeq-count (Anders et al., 2015). The reads mapping to multiple sites within the reference genome were eliminated while estimating the read counts. These read-count data were normalized to reads per kilobase per million (RPKM), which is a quantitative approximation of the transcript abundance. The RPKM value was estimated as follows: RPKM equals reads mapped to a transcript per unit length of transcript (kb) divided by total reads mapped (millions). Transcript lengths were obtained from Phytozome 9.0v (www.phytozome.net).

Identification of Polymorphisms in *Rsv3* Candidate Gene Sequence Assemblies

The sequencing data from three resistant lines were used: (i) preprocessed sequencing reads from the L29-RNASeq experiment described above, (ii) Hwangkeum genome sequencing reads obtained from Korea Research Institute of Bioscience and Biotechnology, Korea (Chung et al., 2014), and (iii) RRR genomic sequencing reads (data available at NCBI SRA database, ID: SRR2079824) (Saghai Maroof et al., 2008). Hwangkeum, which is a registered cultivar name of a breeding line Suweon97, contains both the *Rsv1* and *Rsv3* resistance genes (Chen et al., 2002; Jeong and Jeong, 2014; Yu et al., 2008). The RRR line, a near isoline of 'Essex', contains *Rsv1* and *Rsv4* as well as *Rsv3* (Saghai Maroof et al., 2008). The sequencing reads from these three resistant lines were separately mapped to the *Rsv3* locus spanning a 154-kb region between nucleotide positions: 47621000 to 47738999 on chromosome 14 of the Williams 82 reference genome sequence. The mapped reads were used to assemble transcript sequences of all five *Rsv3* candidate genes using Geneious Pro version 5.7.7 software (Kearse et al., 2012). The *Rsv3* candidate gene IDs from two annotation versions of the reference genome assembly 1 are as shown in Table 1. These *Rsv3* candidate transcript assemblies of resistant lines were pairwise aligned with the respective transcript sequences of the susceptible Williams 82 genome using ClustalW (Larkin et al., 2007) to identify the nucleotide positions polymorphic between the resistant and susceptible lines.

Validating Genetic Polymorphisms and Genetic Mapping for Del540 Marker

Eighteen *Rsv3*-type (resistant) and six *rsv*-type (susceptible) soybean lines were used for validation of genetic polymorphisms from *Rsv3 candidate* genes as listed in Supplemental Table S2. Genomic DNA was extracted from these soybean cultivars using the CTAB method (Yu et al., 1994). Primers to amplify polymorphic positions from the five *Rsv3* candidate genes were designed using Primer3 software and are listed in Supplemental Table S3 (Koressaar and Remm, 2007; Untergasser et al., 2012). Polymerase chain reaction (PCR) was performed in 50 μL reaction volume comprised of 5 μL 10× reaction buffer, 1.5 μL of 50 mM $MgCl_2$, 8 μL of 2 mM dNTPs, 0.5 μL of 10 mM primers each, and 0.5 μL of Taq polymerase in deionized water (Invitrogen Inc.). Amplification products were purified using QIAquick PCR purification kit (QIAGEN Inc.). Purified PCR products were sequenced by Sanger method and analyzed using Geneious Pro software to identify genotypes at polymorphic sites of *Rsv3* candidate genes. Two segregating populations, L29 (*Rsv3*) × Lee68 (*rsv*) and L29 (*Rsv3*) × Sowon (*rsv*), obtained from Suh et al. (2011), were genotyped using Glyma14g38533 gene-specific InDel marker (Del540) (Suh et al., 2011). The genetic maps were constructed by combining Del540 marker data with the Suh et al. (2011) published marker data using MAPMAKER/EXP 3.0 (Lander et al., 1987; Suh et al., 2011).

Structural Modeling and dN/dS Analysis of the Glyma14g38533 Leucine-Rich Repeat Domain

A structural model for the Glyma14g38533 (L29 allele) LRR domain was generated using default settings of the Modeler comparative modeling software (http://salilab.org/modeller/) accessed remotely using the University of California–San Francisco Chimera package, version 1.10.2 (http://www.cgl.ucsf.edu/chimera/) (Eswar et al., 2007; Pettersen et al., 2004). The known structure of the human TLR3 ectodomain (PDB no. 1ziw) was used as the template (Choe et al., 2005). Sequence alignment and editing was accomplished as described by Ashfield et al. (2014). Four apparent insertions (R565-S580, E887-H901,

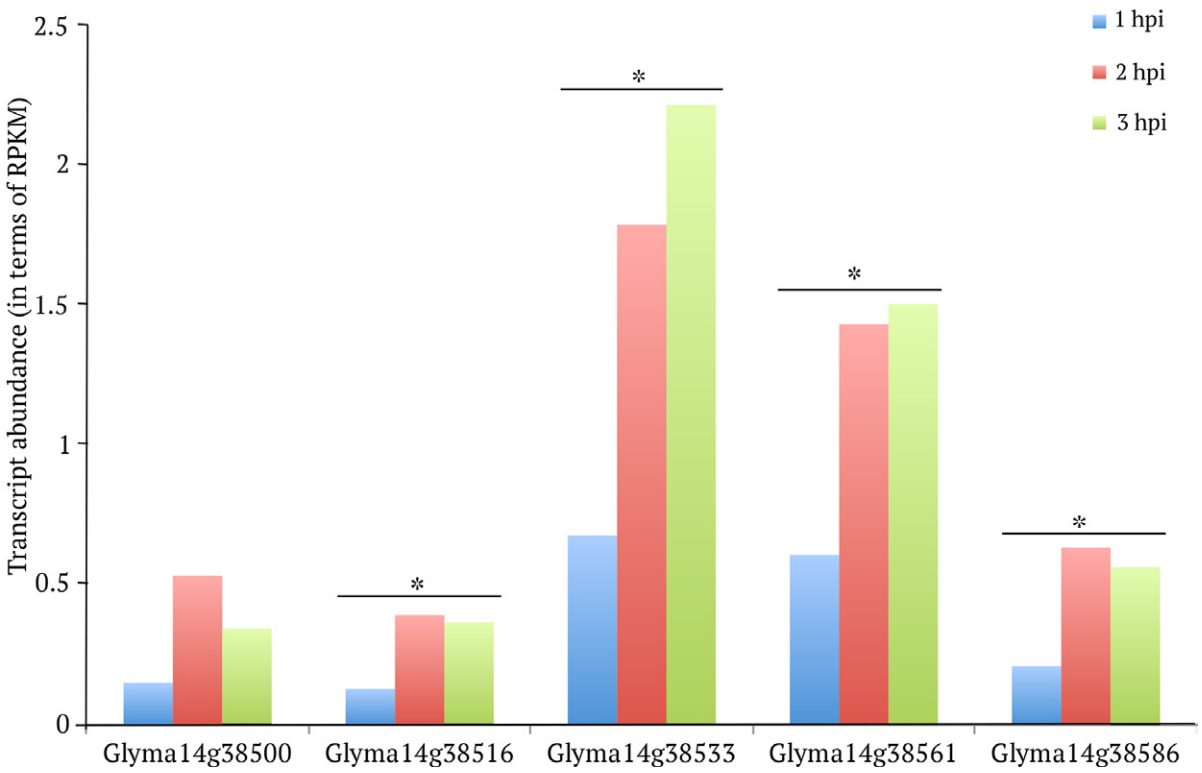

Fig. 1. Relative transcript abundance of *Rsv3* candidate NB-LRR genes. Reads per kilobase per million is estimated from normalized gene counts. Error bars are calculated over two technical replicate samples. One-way ANOVA was performed to identify significant change in gene expression levels between 1 hpi, 2 hpi, and 3 hpi. One asterisk (*) denotes *p*-value <0.05.

K954-I971, and D1167-Q1176) in the Glyma14g38533 sequence that could not be aligned with the TLR3 template structure sequence were subsequently deleted from the model. The Chimera software was used for model editing, image capture, and the production of movies.

The ratio of nonsynonymous to synonymous changes (dN/dS) for an alignment between the Glyma14g38533 alleles from L29 and Williams 82 for the region extending from the start of LRR1 to the end of LRR23 (Supplemental Table S5) was determined using the Yang & Nielsen (2000) method as implemented by yn00 component of the PAML package, version 4.8 (http://web.mit.edu/6.891/www/lab/paml.html) (Yang and Nielsen, 2000; Yang, 2007).

Results

Differential Gene Expression of *Rsv3* Candidate Nucleotide-Binding Leucine-Rich Repeat Genes

Several NB-LRR genes are expressed at a minimal level at any given time to facilitate detection of pathogens or pathogen-derived factors. Many studies have also reported an increase in their expression levels on infection (Kang et al., 2012; Kar et al., 2013; Li et al., 2013). In this study, we compared the relative transcript abundance of *Rsv3 candidate* NB-LRR genes using mRNA sequencing data generated from first trifoliolate leaflets of resistant line L29 collected 1, 2, and 3 hpi with SMV-G7. A total of 1.97 billion paired-end reads were generated from sequencing of all mRNA sample libraries (Supplemental Table S1). Data preprocessing retained more than 921 million paired-end and 399 million single-end reads, which were separately mapped to soybean reference Williams 82 genome. The RPKM values were estimated from mapping data as a measure of transcript abundance for five *Rsv3* candidate NB-LRR genes: Glyma14g38500, Glyma14g38516, Glyma14g38533, Glyma14g38561, and Glyma14g38586 in L29 at 1, 2, and 3 hpi with SMV-G7 (Fig. 1). The *Rsv3* candidate NB-LRR genes showed differential expression in SMV-G7-inoculated L29 leaves between these sampling times. The transcript abundance of all *Rsv3* candidate genes, except Glyma14g38500, was significantly different at three sampling stages. At 3 hpi, the transcript levels declined in Glyma14g38500, Glyma14g38516, and Glyma14g38586 as opposed to increased levels of Glyma14g38533 and Glyma14g38561. The Glyma14g38533 gene showed highest transcript abundance at each time point with a total of 3.3-fold (highest) increase during first 3 h after inoculation. The higher transcript abundance of Glyma14g38533 gene may be associated with or required for the *Rsv3*-type resistance; however, the current experimental design was unable to provide conclusive evidence. Although gene expression information alone is not sufficient to narrow down the list of *Rsv3* candidates, differences in the transcript abundance suggest dominance of Glyma14g38533 over the other *Rsv3* candidates.

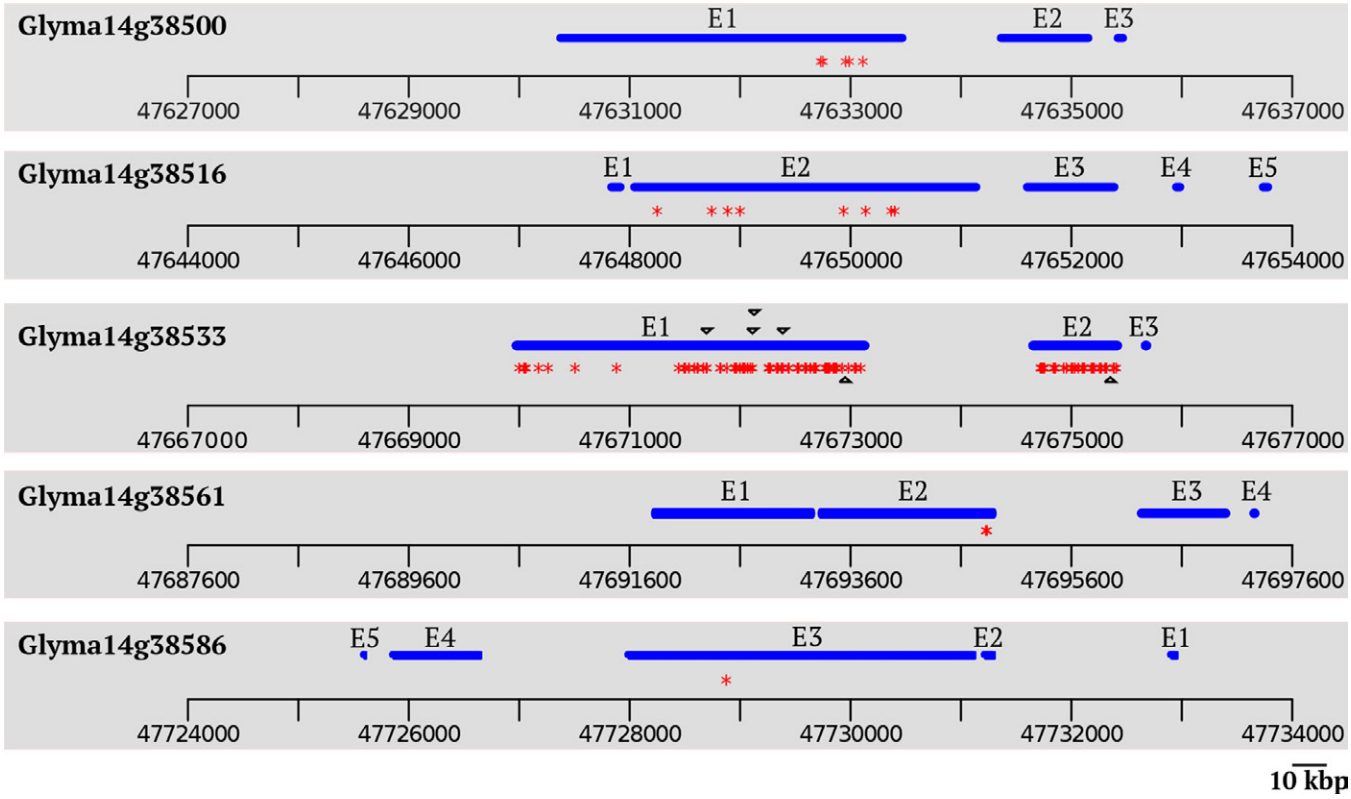

Fig. 2. Location of polymorphisms identified in five *Rsv3* candidate genes. Each gray box shows *Rsv3* candidate gene within an inclusive 10 kb genomic region. Blue bars indicate coding regions as per Glyma1.1 gene models. Individual exons are labeled as E1, E2, etc. Red asterisks indicate single-nucleotide polymorphism sites, while the black triangles indicate insertion–deletion marker sites within the coding regions. Downward triangles represent deletions, while upward triangles represent insertions within coding region of resistant lines.

Comparison of *Rsv3* Candidate Nucleotide-Binding Leucine-Rich Repeat Gene Sequences

The NB-LRR genes typically exist in clusters with high sequence similarity among the members within a cluster (Marone et al., 2013). The susceptible and resistant lines often exhibit different structural forms of the cluster. Any allelic features shared between multiple resistant lines may contribute to the resistance phenotype. The allelic features of *Rsv3* candidate NB-LRR genes were first identified by comparing susceptible (Williams 82) and resistant (Hwangkeum) lines. Several InDels and SNPs were identified in the five *Rsv3* candidate NB-LRR genes as summarized in Fig. 2. The InDel-type polymorphisms were observed in Glyma14g38533 gene only. All SNPs identified in four *Rsv3* candidates—Glyma14g38500, Glyma14g38516, Glyma14g38561, and Glyma14g38586—were present in a single exon only, namely exon 1, 2, 2, and 3, respectively (Fig. 2). Table 2 represents 17 nonsynonymous SNPs identified by comparing Hwangkeum and Williams 82 sequences of these four *Rsv3* candidates. These candidate genes did not show any InDel-type polymorphisms.

The comparison of the fifth *Rsv3* candidate NB-LRR, Glyma14g38533, gene sequence between Hwangkeum and Williams 82 identified nearly 146 SNPs and several InDels (Supplemental Table S4), suggesting a remarkable gene variation between the two lines. Further, we performed de novo assembly of the Glyma14g38533 sequence using L29 transcriptome sequences and RRR genomic resequencing data. The Glyma14g38533 gene

Table 2. Nonsynonymous mutations in *Rsv3* candidate nucleotide-binding leucine-rich repeat genes.

		Williams 82		Hwangkeum	
Gene ID	Physical positions†	Base	AA	Base	AA
Glyma14g38500	47632717	C	Pro	T	Leu
	47632741	T	Phe	C	Ser
	47632942	T	Phe	G	Cys
	47632978	T	Leu	A	Gln
	47633100	G	Gly	A	Arg
Glyma14g38516	47648236	G	Glu	C	Gln
	47648733	G	Glu	C	Asp
	47648875	A	Arg	G	Gly
	47648989	A	Thr	G	Ala
	47649926	T	Leu	C	Ser
	47650126	G	Val	A	Met
	47650357	A	Lys	C	Gln
	47650390	A	Lys	G	Glu
Glyma14g38561	47694814	A	Lys	G	Glu
	47694820	GC	Ala	CG	Arg
	47694826	AT	Ile	TA	Tyr
Glyma14g38586	47728863	C	Lys	A	Asn

† Physical positions (start) correspond to the genome assembly Wm82.a1.

Table 3. Summary of single-nucleotide polymorphisms (SNPs) identified in Glyma14g38533 coding sequence.

Glyma14g38533	Length	SNPs	Synonymous	Nonsynonymous
	bp			
Exon 1	3153	98	9	89
Exon 2	763	52	5	47
Exon 3	8	0	0	0
Total	3924	150	14	136

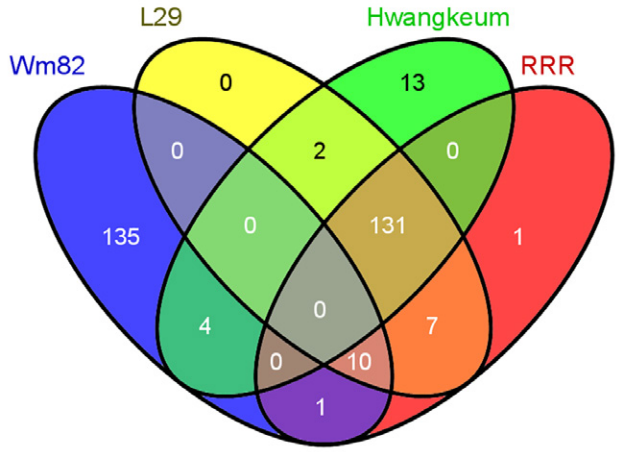

Fig. 3. Overlap of Glyma14g38533 SNPs between Williams 82, L29, Hwangkeum, and RRR soybean lines. Total of 146, 140, and 139 nucleotide positions within Glyma14g38533 gene were polymorphic in Hwangkeum, L29, and RRR when compared against Williams 82 (Wm82), respectively.

assembly sequence of L29 and RRR, compared with that of Williams 82, led to identification of 140 and 139 SNPs, respectively. Altogether 150 single nucleotide base positions were polymorphic between the resistant lines (Hwangkeum, L29, and RRR) and the susceptible Williams 82 line, and 136 of these polymorphic sites were nonsynonymous (Table 3; Supplemental Table S4). Figure 3 summarizes the SNP overlap between Williams 82, L29, Hwangkeum, and RRR. One hundred and thirty-one SNPs were a perfect match between the three resistant lines, suggesting substantial gene sequence homology and conserved residues between these resistant lines. In addition to SNPs, four in-frame deletions (sizes 3, 9, 12, and 39 bp) and two in-frame insertions (sizes 6 and 21 bp) were identified in the Glyma14g38533 coding sequence of resistant lines, predominantly in exons 1 and 2 (Table 4; Fig. 2; Supplemental Table S4). No polymorphisms were identified in exon 3, which is merely 8 bp long. While additional resistance alleles of this Glyma14g38533 gene could exist, the current study suggests the presence of at least two allelic forms of the Glyma14g38533 gene: one observed in the resistant lines L29, Hwangkeum, and RRR and the other observed in susceptible lines such as Williams 82.

Table 4. Conserved insertion–deletion markers (InDels) from Glyma14g38533 transcript of Hwangkeum, L29, and RRR.

Physical positions†	InDel genotypes	Size
		bp
Deletions		
47671691‡	AGCTATAATTTCCTTAGGCGTGAATTGAACAAGGCATGT	39
47672110‡	CTG	3
47672119‡	ATGATGGAAGGC	12
47672378	GATCTGCTG	9
Insertions		
47672947	ATTCACATCAATTTCCTTAAT	21
47675349	TAAAAG	6

† Physical positions (start) correspond to genome assembly Wm82.a1.

‡ InDels verified in 18 resistant and six susceptible lines.

Validation of Polymorphisms in *Rsv3* Candidate Nucleotide-Binding Leucine-Rich Repeat Genes

The genetic alleles conferring resistance are usually conserved in resistant lines. Therefore, the nucleotide positions that are polymorphic between susceptible and resistant lines but conserved in multiple resistant lines, which are more likely to be associated with resistance. The polymorphic sites in *Rsv3* candidate NB-LRR genes, identified by sequence comparisons between susceptible and resistant lines, were then genotyped in additional resistant and susceptible lines. Disease responses to SMV-G1 and SMV-G7 and genotype information for the soybean lines used for this analysis are reported in Supplemental Table S2. The SNP verification results for Glyma14g38500, Glyma14g38516, Glyma14g38561, and Glyma14g38586 are reported in Table 5. The five polymorphic sites of Glyma14g38500 gene were conserved in all resistant lines and were distinct from susceptible lines except for Archer, in which the SNP genotype was the same as that of resistant lines. We also compared our Archer genotypes with those estimated using Archer exome sequence data obtained from McHale et al. (2012). All SNPs but one at position 47632978 in Glyma14g38500 gene were a perfect match between the two datasets. With Archer as an outlier, the data show that the genotype of polymorphic sites in Glyma14g38500 gene is highly correlated with the disease phenotype (resistant or susceptible). This genotype–phenotype correlation, suggesting association with resistance, does not overshadow the fact that the Glyma14g38500 gene expression level remains unaffected with SMV-G7 inoculation as shown in Fig. 1. No perfect genotype–phenotype correlation was observed with Glyma14g38561 and Glyma14g38586 genes. For the Glyma14g38516 gene, six out of eight nucleotide positions were genotyped in a sufficient number of lines. Among these, the position at 47648733 showed perfect genotype–phenotype correlation, while the position at 47649926 showed imperfect correlation with Archer line an outlier (Table 5).

Table 5. Single-nucleotide polymorphism validation from different *Soybean mosaic virus* resistant and susceptible soybean lines.

		Genetic positions and genotypes																
		Glyma14g38500					Glyma14g38516								Glyma14g38561			Glyma14g38586
Cultivars	Type†	47632717	47632741	47632942	47632978	47633100	47648236	47648733	47648875	47648989	47649926	47650126	47650357	47650390	47694814	47694820	47694826	47728863
Williams 82	S	C	T	T	T	G	G	G	A	A	T	G	A	A	A	GC	AT	C
Lee 68	S	C	T	T	T	G	G	G	A	A	T	–	–	–	A	GC	AT	A
Essex	S	C	T	T	T	G	–	G	A	A	T	–	–	–	A	GC	AT	A
Archer ‡	S	T	C	G	A	A	G	–	–	–	C	–	–	–	G	GC	AT	A
Hutcheson	S	C	C	T	T	G	G	G	G	G	T	G	–	–	G	CT	TA	A
York	S	C	C	T	T	G	G	G	G	G	T	G	–	–	A	GC	AT	A
L29	R	T	C	G	A	A	G	C	G	G	C	T	–	–	G	GC	AT	A
Suweon 97	R	T	C	G	A	A	C	C	G	G	C	A	C	G	G	CG	TA	A
RRR	R	T	C	G	A	A	G	–	–	–	C	T	–	–	G	CC	AT	A
Tousan 140	R	T	C	G	A	A	G	C	G	G	C	A	–	–	G	GC	AT	A
Harosoy	R	T	C	G	A	A	–	C	G	G	C	T	–	–	G	CA	AT	A
Columbia	R	T	C	G	A	A	G	C	G	G	C	T	–	–	G	CA	AT	A
Hourei	R	T	C	G	A	A	G	C	G	G	C	A	–	–	G	CG	TA	A
Hardee	R	T	C	G	A	A	G	C	G	G	C	T	–	–	G	CA	AT	A
PI 91346	S	T	C	G	A	A	–	–	–	–	C	–	–	–	A	CC	AT	A
PI 61947	R	T	C	G	A	A	–	C	G	G	C	A	–	–	G	GG	TT	A
VIR 5532	R	T	C	G	A	A	–	–	–	–	C	–	–	–	G	CA	AT	A
Paoting	R	T	C	G	A	A	–	–	–	–	C	–	–	–	G	CG	TA	A
PI 323555	R	T	C	G	A	A	–	–	–	–	C	–	–	–	G	CG	TA	A
PI 323556	R	T	C	G	A	A	–	–	–	–	C	–	–	–	G	CG	TA	A
PLSO-63	R	T	C	G	A	A	–	–	–	–	C	–	–	–	G	CC	AT	A
PLSO-70	R	T	C	G	A	A	–	C	G	G	C	A	–	–	G	CC	AT	A
OCB 81	R	T	C	G	A	A	–	C	G	G	C	A	–	–	G	CA	AT	–
Graine Jaune Unie	R	T	C	G	A	A	–	–	–	–	C	–	–	–	A	CC	AT	A

† S, susceptible (*rsv3*); R, resistance (*Rsv3*) response to SMV-G7.

‡ Archer sequence from McHale et al. (2012) exhibited nucleotide T at position 47632978 in Glyma14g38500 gene.

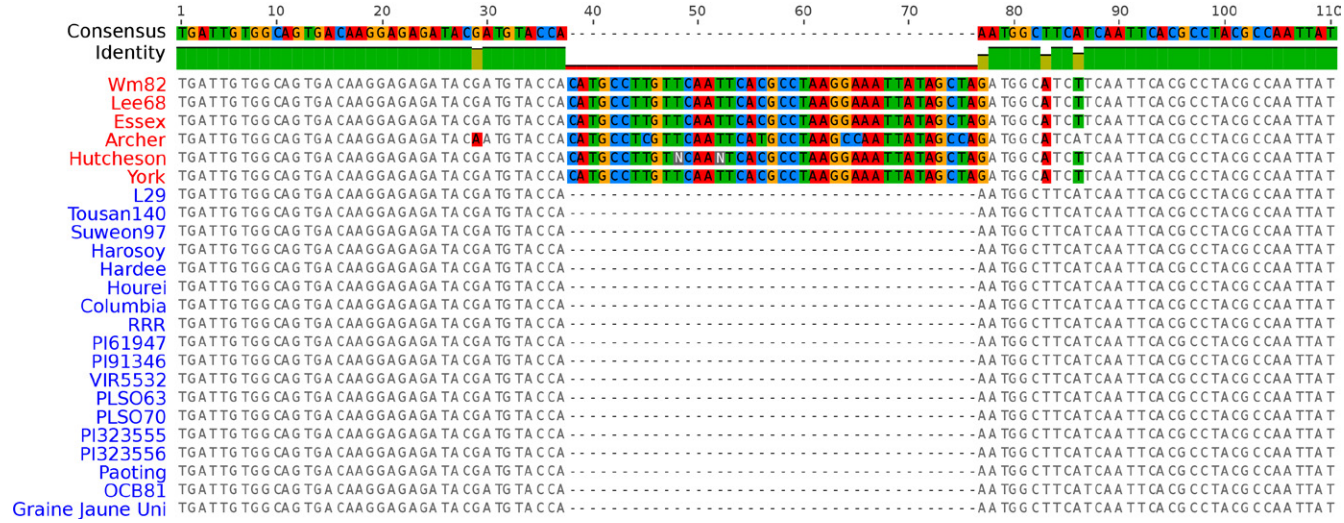

Fig. 4. Validation of 39-bp deletion region within Glyma14g38533 gene from multiple resistant soybean lines. Only resistant lines (blue) showed this deletion when compared with susceptible lines (red).

Unlike the single correlation in the above *Rsv3* candidates, the Glyma14g38533 gene showed the highest number of genotype–phenotype correlations. Eighteen soybean lines expected to contain *Rsv3* resistance gene, viz. L29, Tousan140, Suweon97, RRR, Hourei, 'Harosoy', 'Columbia', 'Hardee', VIR 5532, Paoting, PLSO-70, PLSO-63, PI 323555, PI 323556, OCB-81, PI 91346, PI 61947, and Graine Jaune Unie, showed the presence of 39-, 3-, and 12-bp InDels when compared with six *rsv*-type susceptible lines, viz. Williams 82, 'Lee68', Essex, Archer, Hutcheson, and 'York'. Many of these resistant soybean lines were reported by Li et al. (2010) to originate from more than five different countries and to possess the *Rsv3* gene (Li et al., 2010). Figure 4 shows the 39-bp InDel region from six susceptible and 18 resistant soybean lines we tested. The SNPs surrounding this deletion were exactly identical in all 18 resistant lines. When we mapped this InDel in L29 (*Rsv3*) × Lee68 (*rsv3*) and L29 (*Rsv3*) × Sowon (*rsv3*) segregating populations from Suh et al. (2011), we found that the 39-bp InDel (or Del540 marker) cosegregates with the *Rsv3* gene position in the genetic map (Supplemental Fig. S1). In summary, the allele preservation in multiple resistant lines and perfect genotype–phenotype correlation suggests that the Glyma14g38533 gene may be the most promising candidate encoding *Rsv3*.

Variations in Glyma14g38533 Protein Domains

Nucleotide-binding leucine-rich repeat proteins contain three domains: a variable N-terminal domain (either toll/interleukin-1 receptor or coiled-coil [CC]), a central NB-ARC domain with a NB site, and a C-terminal LRR domain (Heil and Baldwin, 2002; McHale et al., 2006). The *Rsv3* candidate genes encode NB-LRR proteins with a CC domain (Suh et al., 2011). Previous studies have suggested that the CC and NB-ARC domains are involved in protein conformation change and defense signal transduction, on pathogen perception by LRR domain (McHale et al., 2006). We studied the distribution of SNPs and InDels identified in Glyma14g38533 gene across different domains, to estimate their impact on domain function. The CC and NB-ARC domains of Glyma14g38533 contained <5% of the total polymorphisms identified, while the remaining 95%, including all InDels, were present within the LRR domain, which is often involved in pathogen effector recognition via protein–protein interactions. This suggests that the Glyma14g38533 gene alleles from resistant and susceptible lines may differ in LRR domain specificity to recognize pathogen effectors.

To further explore this possibility, we searched for LRRs within the Glyma14g38533 protein sequence from L29 and Williams 82 using the LRR motif consensus LxxLxLxx[N/L]xL and LxxLxLxx[C/L]xxL, where L is Leu (L)/Ile (I)/Val (V)/Phe (F), N is Asn (N)/Thr (T)/Ser (S)/Cys (C), C is Cys (C)/Ser (S), and x is any amino acid (Kajava, 1998; Kajava et al., 2008; Kobe, 2001; Matsushima et al., 2010; Ohyanagi and Matsushima, 1997). We allowed for one mismatch within conserved residues of consensus pattern so as to identify LRR variants, if any. Twenty-three and 22 LRRs were identified in the Glyma14g38533 protein of L29 and Williams 82, respectively (Table 6; Fig. 5; Supplemental Table S5). These LRRs are labeled from the N-terminal of the domain as LRR1 to LRR23 and LRR1 to LRR22 for L29 and Williams 82, respectively. The Williams 82 protein sequence corresponding to LRR6 of L29 protein did not match the motif search criteria because of polymorphisms, and therefore, no LRR was identified in this sequence. Therefore, LRR7 of L29 corresponds to LRR6 of Williams 82 and so on. Eight LRRs in L29, including LRR3, LRR6, LRR8, LRR14, LRR16, LRR17, LRR18, and LRR23, showed a mismatch within the conserved residues of the LRR motif consensus (Fig. 5). Several residues were polymorphic between L29 and Williams 82 for all LRRs except LRR16 (Table 6). To predict their three-dimensional distribution, these polymorphisms were mapped

Table 6. Comparison of leucine-rich repeats (LRR) in Glyma14g38533 protein sequence between L29 and Williams 82.

L29 repeats	Sequence†	Consensus‡	L29 repeats	Sequence†	Consensus
LRR1	L29: LEILLFHS**T**EV; W82: LEILLFHS***P***EV	LxxLxLxxNxL	LRR13	L29: LE**EL**N**IGF**CDKL; W82: LE**KL**M**VER**CDKL	LxxLxLxxCxxL
LRR2	L29: IKILAILTSS**L**; W82: IKILAILTSS***Y***	LxxLxLxxNxL	LRR14	L29: LETLRLT**E**LPNL; W82: LETLRLT**Q**LPNL	LxxLxLxxLxxL
LRR3	L29: LHTLCLRG***H***IL; W82: LHTLCLRG***YE***L	LxxLxLxxNxL	LRR15	L29: VRR**V**MIIDSDL; W82: VRR***G***MIIDSDL	LxxLxLxxNxL
LRR4	L29: LEVLDLR**N**SSFI; W82: LEVLDLR**G**SSFI	LxxLxLxxCxxL	LRR16	L29: LCSVT*T*TFNQL; W82: LCSVT*T*TFNQL	LxxLxLxxNxL
LRR5	L29: LKLLDLF**NCV**I; W82: LKLLDLF**HCS**I	LxxLxLxxNxL	LRR17	L29: LR**HLQLYGL**GV; W82: LR**ELGLSGV**GV	LxxLxLxxLxL
LRR6§	L29: LQL***N***ELYLCIYL; W82: ***M***QL*M*ELYL**SIPS**	LxxLxLxxCxxL	LRR18	L29: LAPLNLDL**IY***A*; W82: LAPLNLDL***TH****A*	LxxLxLxxLxL
LRR7	L29: L**T**FLIL**E**DCPEI; W82: L**I**FLIL**H**DCPEI	LxxLxLxxCxxL	LRR19	L29: LDVI**Y**VNRCPKL; W82: LDVI**N**VNRCPKL	LxxLxLxxCxxL
LRR8	L29: LVIL**R**LY**E**LDNL; W82: LVIL**S**LY**G**LDNL	LxxLxLxxLxxL	LRR20	L29: L**RTLEITH**CEEL; W82: L**GRLQIID**CEEL	LxxLxLxxCxxL
LRR9	L29: LEEL**S**IE**S**CRQL; W82: LEEL**T**IE**R**CRQL	LxxLxLxxCxxL	LRR21	L29: L**H**YI**C**V**E**KCNKL; W82: L**Y**YI**S**V**K**KCNKL	LxxLxLxxCxxL
LRR10	L29: LK**F**LTI**DH**CPML; W82: LK**S**LTI**RD**CPML	LxxLxLxxCxxL	LRR22	L29: L**IA**LEI**K**DCS**Q**L; W82: L**SK**LEI**E**DCS**E**L	LxxLxLxxCxxL
LRR11	L29: LEQV**T**IS**D**C**F**EL; W82: LEQV**R**IS**E**C**Y**EL	LxxLxLxxCxxL	LRR23	L29: LL**RIR**LS**R**LPNF; W82: LL**YIT**LS**S**LPNF	LxxLxLxxLxxL
LRR12	L29: LRTLTI**LR**C**H**SL; W82: LRTLTI**RG**C**R**SL	LxxLxLxxCxxL			

† Amino acid residues different between L29 and Williams 82 (W82) LRR repeats are in bold, while the amino acid residues that fail to match the LRR motif consensus sequence are represented with italic letters.

‡ Residues in the LRR motif consensus LxxLxLxx[N/L]xL and LxxLxLxx[C/L]xxL, where L is Leu (L)/Ile (I)/Val (V)/Phe (F); N is Asn (N)/Thr (T)/Ser (S)/Cys (C); C is Cys (C)/Ser (S), and x is any amino acid.

§ Williams 82 sequence corresponding to L29 repeat LRR6 sequence does not match the consensus LRR motif, therefore this sequence in Williams 82 is not considered as LRR repeat.

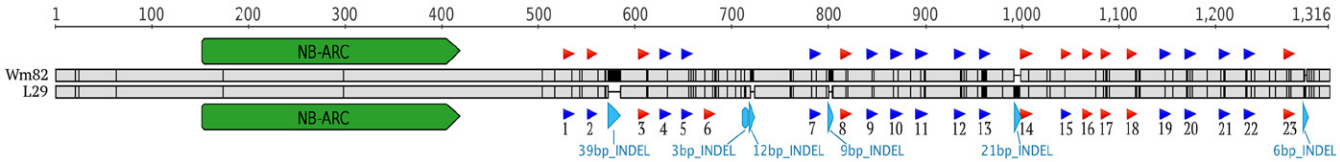

Fig. 5. Differences in Glyma14g38533 protein domains from susceptible Williams 82 and resistant L29 line. Gray boxes in the center represent Glyma14g38533 protein sequences from Williams 82 (Wm82) and L29. Black bars within gray boxes represent positions where the amino acid residues differ between Wm82 and L29. Dark blue triangles indicate LRRs whose sequences exactly match to leucine-rich repeat (LRR) motif consensus, LxxLxLxx[N/L]xL and LxxLxLxx[C/L]xxL; red triangles indicate LRRs whose sequences showed a mismatch with the LRR motif consensus. The LRR sequences of L29 are labeled 1–23 (refer to Table 6 for more information). Light blue triangles indicate the position of insertion–deletion markers identified in Glyma14g38533 transcript (refer to Table 4 for more information).

on a structural model of the Glyma14g38533 LRR domain (Fig. 6; Supplemental Movie S1, S2). Interestingly, many of the polymorphic sites are solvent exposed and arranged in clusters on the β strands, or flanking loops, that constitute the structure's concave surface. The resulting highly polymorphic surfaces are consistent with the Glyma14g38533 alleles in L29 and Williams 82 encoding distinct pathogen recognition specificities. Also suggestive of functional divergence, the dN/dS ratio found in the L29–Williams 82 LRR nucleotide alignment is 2.65, indicative of positive selection having acted on this region. We also observed that most of the LRR sequences were conserved between L29, Hwangkeum, and RRR. Since the L29 line is the source of the *Rsv3* gene in RRR, all the LRRs were identical between L29 and RRR. However, LRR8 and LRR12 exhibited differences in the nonconserved residues within the consensus sequence between L29 and Hwangkeum (Supplemental Table S6). In summary, based on the distribution of polymorphisms described above, the pathogen recognition specificity of the LRR domain is likely different in the Glyma14g38533 NB-LRR proteins encoded by the susceptible and resistant lines.

Discussion

The current study is focused on narrowing the list of candidate genes in the *Rsv3* locus. We performed mRNA sequencing with trifoliolate leaflets of resistant soybean L29 taken 1, 2, and 3 hpi with SMV-G7. The *Rsv3* candidate NB-LRR genes were differentially expressed between early stages of SMV-G7 infection. Among five *Rsv3* candidates, the Glyma14g38533 gene showed highest transcript abundance of over threefold. Although our current experimental setup fails to provide a comparison of *Rsv3* candidate gene expression profiles between susceptible and resistant lines, a significant increase in Glyma14g38533 transcript abundance on infection cannot be ignored, as it may be associated with enhancing the resistance reaction. These expression profiles revealed by our RNA sequencing experiment were somewhat consistent to those assayed by quantitative real-time PCR in 'Dabaima', which contains the *Rsc4* gene locus whose genetic mapping interval coincides with that of the current *Rsv3* locus (Wang et al., 2011). The *Rsc4* locus (<100 kb) conferring resistance to SMV strain SC4 from China contains three NB-LRR genes, viz., Glyma14g38500, Glyma14g38516, and Glyma14g38533 (Wang et al., 2011). The Glyma14g38516 and Glyma14g38533 genes showed significantly different expression in resistant and

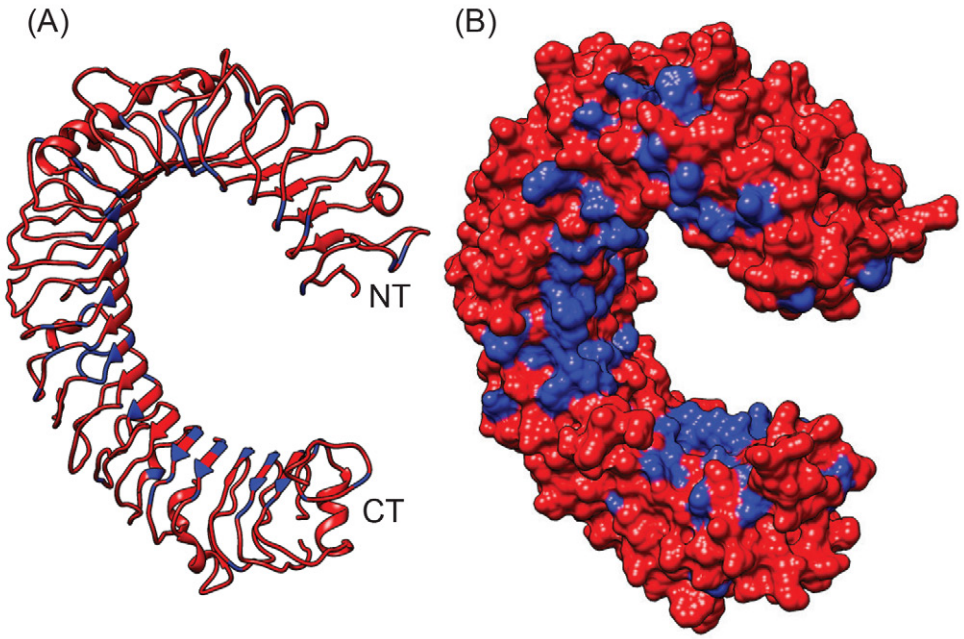

Fig. 6. Distribution of Glyma14g38533 polymorphic sites shown on a structural model of the LRR domain. Shown is a structural model of the Glyma14g38533 LRR domain represented as a ribbon structure either without (Panel A) or with (Panel B) the predicted surface shown. Residues that are conserved in Glyma14g38533 alleles from the susceptible Williams 82 and resistant L29 lines are shown in red. Polymorphic sites are shown in blue. The model's N terminus (NT) and C terminus (CT) are indicated in Panel A.

susceptible parents, whereas the Glyma14g38500 gene was not even detected in either resistant Dabaima or susceptible 'Nannong1138-2' lines at 1 and 2 hpi with SMV-SC4 strain (Wang et al., 2011).

The NB-LRR genes show high sequence similarity and coevolve with pathogens to accumulate distinct haplotypes or genetic features (Bent and Mackey, 2007; Marone et al., 2013). The genetic features conferring fitness (in this case, resistance) become fixed during the course of evolution (Marone et al., 2013). We conducted comparative sequence analyses between resistant and susceptible lines with the five *Rsv3* candidate NB-LRR genes, which are 83 to 92% similar to each other. Four of the five *Rsv3* candidates, Glyma14g38500, Glyma14g38561, Glyma14g38516, and Glyma14g38586, showed ~17 SNP-type differences between a susceptible and a resistant line (Table 2). Most of the SNP positions were validated in several other susceptible and resistant lines (Table 5); however, the genotype–phenotype correlation existed only for one or two SNP alleles. The *Rsv3* candidate Glyma14g38533 gene, on the other hand, showed high degree of polymorphisms (150 SNPs and six in-frame InDels) between three resistant lines (L29, Hwangkeum, and RRR) and the susceptible Williams 82 with >87% of these genetic features conserved between the resistant lines. This suggested perfect genotype–phenotype correlation for Glyma14g38533 gene. Validation of these genetic features in 18 *Rsv3*-type resistant and six *rsv*-type susceptible lines suggested that the Glyma14g38533 gene allele in resistant lines differs considerably from its counterpart in susceptible lines.

The N-terminal region of the *Rsv3* candidate gene-encoded CC-NB-LRR protein, comprised of the CC and NB-ARC domains, is likely involved in protein activation and signal transduction postpathogen recognition, whereas the LRR domain is likely involved in recognizing the pathogen via a protein–protein interaction (McHale et al., 2006). The nucleotides encoding LRR domains tend to have a higher rate of variability than those encoding nucleotide-binding regions so as to recognize evolving pathogens (Kuang et al., 2004; Marone et al., 2013; Mondragon-Palomino et al., 2002). About 95% of the polymorphic genetic features identified in Glyma14g38533 were found within the LRR domain, with an increased ratio of nonsynonymous to synonymous nucleotide substitution (Supplemental Table S4). These observations are consistent with diversifying selection maintaining variation in the solvent-exposed residues within the LRR domain (McDowell et al., 1998; Michelmore and Meyers, 1998). The LRR domain variations also included all of the InDels polymorphic between resistant and susceptible lines (Fig. 5). It is likely that the variability in the LRR domain of Glyma14g38533 gene may lead to differential pathogen recognition and further affect resistance specificity in susceptible vs. resistant lines.

The LRR domain in NB-LRR genes consists of 20 to 30 LRR repeating units, which are frequently interrupted by non-LRR island regions (Matsushima et al., 2010). Each LRR repeating unit is defined by a highly conserved motif, the 11 and 12 residues LRR signature sequence and a variable sequence (Kajava, 1998; Matsushima et al., 2010). We identified 23 and 22 LRR repeating units within the Glyma14g38533 protein in L29 and Williams 82,

respectively. More than 57% of the SNPs identified in the Glyma14g38533 gene were accumulated across 22 LRR repeating units in L29. This means that only one repeating unit matched between susceptible and resistant lines. This is in agreement with the inference that the majority of the polymorphisms identified in Glyma14g38533 tend to modify the specificity of the LRR domain.

Evidence presented here indicates that the *Rsv3* candidate NB-LRR gene Glyma14g38533 is expressed in early events of virus infection. The Glyma14g38533 gene was highly abundant during 1 to 3 h of SMV-G7 infection compared with other *Rsv3* candidates. While the Glyma14g38533 gene sequence was greatly diverse between the resistant and susceptible lines, the genetic features were noticeably preserved between the two groups. The allelic differences in the Glyma14g38533 gene sequence between resistant and susceptible lines were accumulated within the LRR domain, a region in NB-LRR protein known to confer pathogen recognition specificity. Overall, these observations suggest that the Glyma14g38533 gene, unlike the other *Rsv3* candidate NB-LRR genes, is the most likely candidate for functional assay, which will be necessary to validate its role in conferring *Rsv3*-type resistance.

Supplemental Information Available

Six supplemental tables, one supplemental figure, and two supplemental movies are available with this manuscript.

Acknowledgments

This work was funded in part by the Virginia Soybean Board. Additional funding was provided by the Virginia Tech Agricultural Experiment Station Hatch Program. We thank Dr. John McDowell for his constructive comments on earlier drafts. We would like to thank support team of Advanced Research Computing (ARC) server and Translational Plant Sciences' MAGYK server at Virginia Tech. The study in the SCJ lab was supported by a grant from the Next-Generation BioGreen 21 Program (PJ01101503) by the Rural Development Administration. Publication of this article was partially supported by Virginia Tech's Open Access Subvention Fund.

References

Anders, S., P.T. Pyl, and W. Huber. 2015. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169. doi:10.1093/bioinformatics/btu638

Ashfield, T., T. Redditt, A. Russell, R. Kessens, N. Rodibaugh, L. Galloway, Q. Kang, R. Podicheti, and R.W. Innes. 2014. Evolutionary relationship of disease resistance genes in soybean and Arabidopsis specific for the *Pseudomonas syringae* effectors AvrB and AvrRpm1. Plant Physiol. 166:235–251. doi:10.1104/pp.114.244715

Babu, M., A.G. Gagarinova, J.E. Brandle, and A.M. Wang. 2008. Association of the transcriptional response of soybean plants with soybean mosaic virus systemic infection. J. Gen. Virol. 89:1069–1080. doi:10.1099/vir.0.83531-0

Bent, A.F., and D. Mackey. 2007. Elicitors, effectors, and R genes: The new paradigm and a lifetime supply of questions. Annu. Rev. Phytopathol. 45:399–436. doi:10.1146/annurev.phyto.45.062806.094427

Buss, G.R., G. Ma, P. Chen, and S.A. Tolin. 1997. Registration of V94-5152 soybean germplasm resistant to soybean mosaic potyvirus. Crop Sci. 37:1987–1988. doi:10.2135/cropsci1997.0011183X003700060068x

Buss, G.R., C.W. Roane, S.A. Tolin, and P. Chen. 1989. Inheritance of reaction to soybean mosaic-virus in 2 soybean cultivars. Crop Sci. 29:1439–1441. doi:10.2135/cropsci1989.0011183X002900060022x

Buzzell, R.I., and J.C. Tu. 1984. Inheritance of soybean resistance to soybean mosaic-virus. J. Hered. 75:82.

Chen, P., G.R. Buss, C.W. Roane, and S.A. Tolin. 1991. Allelism among genes for resistance to soybean mosaic-virus in strain-differential soybean cultivars. Crop Sci. 31:305–309. doi:10.2135/cropsci1991.0011183X003100020015x

Chen, P., G.R. Buss, and S.A. Tolin. 1993. Resistance to soybean mosaic-virus conferred by 2 independent dominant genes in Pl-486355. J. Hered. 84:25–28.

Chen, P.Y., G.R. Buss, S.A. Tolin, I. Gunduz, and M. Cicek. 2002. A valuable gene in Suweon 97 soybean for resistance to soybean mosaic virus. Crop Sci. 42:333–337. doi:10.2135/cropsci2002.0333

Cho, E.K., and R.M. Goodman. 1979. Strains of soybean mosaic virus: Classification based on virulence in resistant soybean cultivars. Phytopathology 69:467–470. doi:10.1094/Phyto-69-467

Choe, J., M.S. Kelker, and I.A. Wilson. 2005. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. Science 309:581–585. doi:10.1126/science.1115253

Chung, W.H., N. Jeong, J. Kim, W.K. Lee, Y.G. Lee, S.H. Lee, W. Yoon, J.H. Kim, I.Y. Choi, H.K. Choi, J.K. Moon, N. Kim, and S.C. Jeong. 2014. Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. DNA Res. 21:153–167. doi:10.1093/dnares/dst047

Eswar, N., B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, and A. Sali. 2007. Comparative protein structure modeling using MODELLER. Curr. Protoc. Protein Sci. Chapter 2: Unit 2.9. doi:10.1002/0471140864.ps0209s50

Hayes, A.J., S.C. Jeong, M.A. Gore, Y.G. Yu, G.R. Buss, S.A. Tolin, and M.A. Maroof. 2004. Recombination within a nucleotide-binding-site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to soybean mosaic virus in soybeans. Genetics 166:493–503. doi:10.1534/genetics.166.1.493

Hayes, A.J., G.R. Ma, G.R. Buss, and M.A.S. Maroof. 2000. Molecular marker mapping of *RSV*4, a gene conferring resistance to all known strains of soybean mosaic virus. Crop Sci. 40:1434–1437. doi:10.2135/cropsci2000.4051434x

Heil, M., and I.T. Baldwin. 2002. Fitness costs of induced resistance: Emerging experimental support for a slippery concept. Trends Plant Sci. 7:61–67. doi:10.1016/S1360-1385(01)02186-0

Jeong, N., and S.C. Jeong. 2014. Multiple genes confer resistance to soybean mosaic virus in the soybean cultivar Hwangkeum. Plant Genet. Resour. 12:S41–S44. doi:10.1017/S1479262114000227

Jeong, S.C., S. Kristipati, A.J. Hayes, P.J. Maughan, S.L. Noffsinger, I. Gunduz, G.R. Buss, and M.A.S. Maroof. 2002. Genetic and sequence analysis of markers tightly linked to the soybean mosaic virus resistance gene, Rsv3. Crop Sci. 42:265–270. doi:10.2135/cropsci2002.0265

Kajava, A. 1998. Structural diversity of leucine-rich repeat proteins. J. Mol. Biol. 277:519–527. doi:10.1006/jmbi.1998.1643

Kajava, A.V., M. Anisimova, and N. Peeters. 2008. Origin and evolution of GALA-LRR, a new member of the CC-LRR subfamily: From plants to bacteria? PLoS ONE 3:E1694. doi:10.1371/journal.pone.0001694

Kang, Y.J., K.H. Kim, S. Shim, M.Y. Yoon, S. Sun, M.Y. Kim, K. Van, and S.H. Lee. 2012. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol. 12:139. doi:10.1186/1471-2229-12-139

Kar, B., S. Nanda, P.K. Nayak, S. Nayak, and R.K. Joshi. 2013. Molecular characterization and functional analysis of *CzR1*, a coiled-coil-nucleotide-binding-site-leucine-rich repeat *R*-gene from *Curcuma zedoaria* Loeb. that confers resistance to *Pythium aphanidermatum*. Physiol. Mol. Plant Pathol. 83:59–68. doi:10.1016/j.pmpp.2013.05.003

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies, and A. Drummond. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kiihl, R.A.S., and E.E. Hartwig. 1979. Inheritance of reaction to soybean mosaic-virus in soybeans. Crop Sci. 19:372–375. doi:10.2135/cropsci1979.0011183X001900030024x

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S.L. Salzberg. 2013. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36. doi:10.1186/gb-2013-14-4-r36

Kobe, B. 2001. The leucine-rich repeat as a protein recognition motif. Curr. Opin. Struct. Biol. 11:725–732. doi:10.1016/S0959-440X(01)00266-4

Koressaar, T., and M. Remm. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291. doi:10.1093/bioinformatics/btm091

Kuang, H., S.-S. Woo, B.C. Meyers, E. Nevo, and R.W. Michelmore. 2004. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell Online 16:2870–2894. doi:10.1105/tpc.104.025502

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181. doi:10.1016/0888-7543(87)90010-3

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. doi:10.1093/bioinformatics/btm404

Li, D., P. Chen, J. Alloatti, A. Shi, and Y.F. Chen. 2010. Identification of new alleles for resistance to soybean mosaic virus in soybean. Crop Sci. 50:649–655. doi:10.2135/cropsci2009.06.0302

Li, J., Q.Y. Zhang, Z.H. Gao, F. Wang, K. Duan, Z.W. Ye, and Q.H. Gao. 2013. Genome-wide identification and comparative expression analysis of NBS-LRR-encoding genes upon *Colletotrichum gloeosporioides* infection in two ecotypes of *Fragaria vesca*. Gene 527:215–227. doi:10.1016/j.gene.2013.06.008

Liao, L., P. Chen, G.R. Buss, Q. Yang, and S.A. Tolin. 2002. Inheritance and allelism of resistance to soybean mosaic virus in Zao18 soybean from China. J. Hered. 93:447–452. doi:10.1093/jhered/93.6.447

Ma, G., P. Chen, G.R. Buss, and S.A. Tolin. 1995. Genetic-characteristics of 2 genes for resistance to soybean mosaic-virus in P1486355 soybean. Theor. Appl. Genet. 91:907–914.

Ma, G., P. Chen, G.R. Buss, and S.A. Tolin. 2002. Complementary action of two independent dominant genes in Columbia soybean for resistance to soybean mosaic virus. J. Hered. 93:179–184. doi:10.1093/jhered/93.3.179

Marone, D., M.A. Russo, G. Laido, A.M. De Leonardis, and A.M. Mastrangelo. 2013. Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: Active guardians in host defense responses. Int. J. Mol. Sci. 14:7302–7326. doi:10.3390/ijms14047302

Matsushima, N., H. Miyashita, T. Mikami, and Y. Kuroki. 2010. A nested leucine rich repeat (LRR) domain: The precursor of LRRs is a ten or eleven residue motif. BMC Microbiol. 10:235. doi:10.1186/1471-2180-10-235

McDowell, J.M., M. Dhandaydham, T.A. Long, M.G.M. Aarts, S. Goff, E.B. Holub, and J.L. Dangl. 1998. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of Arabidopsis. Plant Cell Online 10:1861–1874. doi:10.1105/tpc.10.11.1861

McHale, L.K., W.J. Haun, W.W. Xu, P.B. Bhaskar, J.E. Anderson, D.L. Hyten, D.J. Gerhardt, J.A. Jeddeloh, and R.M. Stupar. 2012. Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol. 159:1295–1308. doi:10.1104/pp.112.194605

McHale, L., X. Tan, P. Koehl, and R.W. Michelmore. 2006. Plant NBS-LRR proteins: Adaptable guards. Genome Biol. 7:212. doi:10.1186/gb-2006-7-4-212

Michelmore, R.W., and B.C. Meyers. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8:1113–1130.

Mondragon-Palomino, M., B.C. Meyers, R.W. Michelmore, and B.S. Gaut. 2002. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genome Res. 12:1305–1315. doi:10.1101/gr.159402

Ohyanagi, T., and N. Matsushima. 1997. Classification of tandem leucine-rich repeats within a great variety of proteins. FASEB J. 11:A949.

Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin. 2004. UCSF Chimera: A visualization system for exploratory research and analysis. J. Comput. Chem. 25:1605–1612. doi:10.1002/jcc.20084

Saghai Maroof, M.A., S.C. Jeong, I. Gunduz, D.M. Tucker, G.R. Buss, and S.A. Tolin. 2008. Pyramiding of soybean mosaic virus resistance genes by marker-assisted selection. Crop Sci. 48:517–526. doi:10.2135/cropsci2007.08.0479

Saghai Maroof, M.A., D.M. Tucker, J.A. Skoneczka, B.C. Bowman, S. Tripathy, and S.A. Tolin. 2010. Fine mapping and candidate gene discovery of the soybean mosaic virus resistance gene, *Rsv4*. Plant Genome 3:14–22. doi:10.3835/plantgenome2009.07.0020

Seo, J.K., S.J. Kwon, W.K. Cho, H.S. Choi, and K.H. Kim. 2014. Type 2C protein phosphatase is a key regulator of antiviral extreme resistance limiting virus spread. Sci. Rep. 4:5905.

Suh, S.J., B.C. Bowman, N. Jeong, K. Yang, C. Kastl, S.A. Tolin, M.A.S. Maroof, and S.C. Jeong. 2011. The *Rsv3* locus conferring resistance to *soybean mosaic virus* is associated with a cluster of coiled-coil nucleotide-binding leucine-rich repeat genes. Plant Genome 4:55–64. doi:10.3835/plantgenome2010.11.0024

Tucker, D.M., M.A.S. Maroof, S.C. Jeong, G.R. Buss, and S.A. Tolin. 2009. Validation and interaction of the *soybean mosaic virus* lethal necrosis Allele, *Rsv1-n*, in PI 507389. Crop Sci. 49:1277–1283. doi:10.2135/cropsci2008.08.0477

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen. 2012. Primer3–new capabilities and interfaces. Nucleic Acids Res. 40:E115. doi:10.1093/nar/gks596

Wang, D.G., Y. Ma, N. Liu, Z.L. Yang, G.J. Zheng, and H.J. Zhi. 2011. Fine mapping and identification of the soybean R_{SC4} resistance candidate gene to soybean mosaic virus. Plant Breed. 130:653–659. doi:10.1111/j.1439-0523.2011.01888.x

Yang, H., Y.P. Huang, H.J. Zhi, and D.Y. Yu. 2011. Proteomics-based analysis of novel genes involved in response toward soybean mosaic virus infection. Mol. Biol. Rep. 38:511–521. doi:10.1007/s11033-010-0135-x

Yang, Z.H. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591. doi:10.1093/molbev/msm088

Yang, Z., and R. Nielsen. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17:32–43. doi:10.1093/oxfordjournals.molbev.a026236

Yu, Y., M. Saghai Maroof, G. Buss, P. Maughan, and S. Tolin. 1994. RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology 84:60–64. doi:10.1094/Phyto-84-60

Yu, Y.H., H. Yu, J. Jeong, H. Park, D. Song, C. Kim, S. Kim, Y. Lee, J. Park, and K. Park. 2008. A general survey of korean legume cultivars. (In Korean) National Institute of Crop Science, Suwon, Korea.