Fine Mapping and Candidate Gene Discovery at the *Rsv3* Locus

Brian Carter Bowman

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
Crop and Soil Environmental Sciences

M. A. Saghai Maroof, Chair
Sue A. Tolin
Brett M. Tyler
Richard E. Veilleux

April 21, 2011

Blacksburg, VA

Keywords: Fine mapping, Leucine rich repeat, Whole genome sequence, *Soybean mosaic virus*, microsatellite, Single nucleotide polymorphism, molecular marker

Fine Mapping and Candidate Gene Discovery at the *Rsv3* Locus

Brian Carter Bowman

Soybean mosaic virus (SMV) is the most common member of the viral genus *Potyvirus* to infect soybeans (*Glycine max* [L.] Merr.) worldwide. SMV has been traditionally controlled by the deployment of single dominant, strain specific resistance genes, referred to as *Rsv* genes. *Rsv1* is the most widely used form of SMV resistance with nine different alleles conferring resistance only to the lower numbered less virulent strains, G1 to G3. *Rsv3* gives resistance to higher numbered more virulent strains G5 to G7. Soybean lines containing *Rsv4*, are resistant to all seven currently recognized North American SMV strains. In this study, the recently released soybean whole genome sequence was used to design molecular markers for fine mapping *Rsv3* to a ~150 kb genomic region containing four coiled-coil nucleotide-binding leucine-rich repeat proteins. In a related study a large population segregating at the *Rsv3* locus was screened for resistance to facilitate future characterization of this region. The markers identified in this study will allow for more accurate marker-assisted selection of *Rsv3.*

Dedication

To my parents Bruce and Ann Bowman; and Uncle George

for his words of wisdom

ACKNOWLEDGEMENTS

During my Master's studies I have had the opportunity to work with many talented people that have helped my scientific pursuits. First, I would like to thank Dr. Saghai Maroof for passing on his first hand experiences in running a molecular breeding program. His lessons will be remembered in the future. I would also like to acknowledge the assistance Dr. Sue Tolin has given me through the use of her greenhouse and her SMV expertise for screening and detection. Additionally, I would like to thank Dr. Brett Tyler for the use of his double barreled gene gun and Dr. Richard Veilleux for exposing me to the plant sciences during my undergraduate years. I want to thank you all for serving on my graduate committee.

Further thanks go out to the undergraduate employees who kept work interesting during the long days in the field: Brenda Arnold, Caitlin Thornley, Bridgett Glover, Anna Molineaux and Kyle Murphy (Too bad you all were not Indians). And to the members of the off campus planting locations at Warsaw and Mt. Holly: Bob Pittman, Lin Barrack, and Bruce Beahm. A special thanks to Tom Pridgen for teaching me the use of the farm equipment at Kentland and his patience.

Finally, I would like to thank the members of the plant genomics lab: Natasha Glover, Jeff Skoneczka, Dominic Tucker, and Christin Kastl for their support and friendship during my Master's studies.

TABLE OF CONTENTS

List of Figures

Chapter 2

Chapter 3

List of Tables

Chapter 1

Literature Review

Resistance to and new challenges for *Soybean Mosaic Virus* Resistance

Soybean mosaic virus (SMV) is a member of the genus *Potyvirus* and causes one of the most commonly found virus diseases of soybean (*Glycine max* [L.] Merr.) worldwide. This virus can reduce yields by as much as 40 % when infection occurs prior to flower development, but is of little importance later in the season because of small losses after pod set (Ren et al., 1997). SMV can spread through seed transmission, mechanical transmission or through aphid vectors. Traditionally SMV has been controlled through the deployment of single dominant resistant genes, known as *Rsv* genes (*Rsv1, Rsv3,* and *Rsv4*). Until recently, SMV has not been considered as important as some of the more devastating diseases such as Soybean Cyst Nematode and Phytophthora root rot (Wrather and Koenning, 2006). The current increase in importance is due to the emergence and establishment of *Aphis glycines* (soybean aphid) populations in the United States soybean producing areas, and the yield losses attributed to co-infection with *Bean pod mottle virus* (BPMV), a virus vectored by native species (Giesler et al., 2002; Ross, 1968).

Bean Pod Mottle Virus and Soybean Mosaic Virus Synergism

Until the year 2000 there had been no soybean colonizing species of aphids which vectored SMV in North America. However, there are 32 species of aphid with the demonstrated ability to transmit SMV throughout a field, including *Aphis glycines* (Domier et al., 2003). The past lack of interest in SMV has stemmed from the low transmissibility rate of virus from infected plants to healthy plants and the inability of native aphids to significantly impact

production. Research has shown that while there are many native species that transmit SMV populations, they do not reach a negative threshold for SMV transmission early enough in the soybean growing season to negatively impact yields (Domier et al., 2003). With the increased transmissibility of SMV in soybean producing areas, additional viral diseases became more important. In work reviewed by Giesler et al. (2002), *Bean pod mottle virus* (BPMV) infection alone was shown to cause 35 to 47 % yield loss, while SMV showed only a 20 percent yield reduction. In comparison, co-infection with both BPMV and SMV showed yield resulted in yield losses of up to 75 %. An effective method of controlling these drastic losses is through the deployment of *Rsv* genes. The yield reductions observed in co-infected York (*Rsv1-y*) were comparable to infection with BPMV alone (Calvert and Ghabrial, 1983). Virus titer calculations showed that during co-infection BPMV titers drastically increase while SMV numbers does not. It is thought that the HC-Pro region of SMV obstructs plant innate virus defense allowing for increased BPMV accumulation (Calvert and Ghabrial, 1983; Giesler et al., 2002). Increased virus titer has also been seen during coinfection with other soybean infecting viruses (Malapi-Nelson et al., 2009).

Rsv1

Molecular Mapping and Characterization of Rsv1 *Resistance Gene*

Rsv1 is the most commonly found gene utilized for SMV resistance and is effective against lower number less virulent strains of SMV, G1- G3. This resistance locus was first described in the soybean line PI 96983 (Kiihl and Hartwig, 1979). Currently there are nine

different alleles of *Rsv1* imparting extreme resistance to different combinations of the seven strains of SMV*; Rsv1* (PI 96983), *Rsv1-t* (Ogden), *Rsv1-m* (Marshall), *Rsv1-y* (York), *Rsv-k* (Kwanggyo), *Rsv1-n* (PI 507389), *Rsv1-s* (LR1), *Rsv1-r* (Raiden) and *Rsv1-sk* (Suweon97) (Chen et al., 1991; Cho and Goodman, 1979; Ma et al., 1995; Saghai Maroof et al., 2008a). While the majority of these alleles convey resistance to a few or all of the SMV strains; one, *Rsv1-n* from PI 507389*,* imparts a lethal necrotic phenotype when challenged by the majority of the SMV strains yet is recessive to all other forms of *Rsv* resistance (Ma et al., 2003; Tucker et al., 2009).

The *Rsv1* locus was first localized to molecular linkage group (MLG) F through the use of restriction fragment length polymorphism (RFLP) clones and the conserved nucleotide binding site (NBS) domain found in many resistance genes (Yu et al., 1996). This region also contains additional resistance genes such as *Rpv1* (resistance to peanut mottle virus), *Rps3* and *Rps8* (resistance to *Phytophthora sojae*), and *Rpg1* (resistance to *Pseudomonas syringae* pv. *glycinea*) found in a large resistance gene cluster (Gordon et al., 2006; Sandhu et al., 2005; Saghai Maroof et al., 2008a*;* Yu et al.; 1996).

Rsv1 was localized to a 0.3 cM region using a population of over one thousand lines created from a cross between PI 96983 (*Rsv1*) and Lee68 (*rsv1*) (Gore et al., 2002). RFPL and AFLP markers previously identified (Hayes and Saghai Maroof, 2000; Hayes et al., 2000a) and microsatellite markers designed specifically for this study, were utilized to identify individuals within the population which contained recombination events at the *Rsv1* and *Rpv1* region and fine map the location of *Rsv1*. In addition, the identified recombinant individuals showed

variation in the phenotype of resistance when challenged by SMV strains (Gore et al., 2002). The phenotypic responses to SMV seen in these "rare recombinants" are the basis for several interesting observations in later studies.

Previously identified NBS probes were used to screen a Williams82 (*rsv1*) lambda phage genomic library for large insert clones containing resistance gene candidates. Subclones were generated for Southern blot analysis and used to probe rare recombinants identified in the previous studies using the PI96983 x Lee68 mapping population (Hayes et al., 2004). Six candidate disease resistance genes were cloned from a genomic library developed from an *Rsv1* isoline (L81-4420) 3gG2, 5gG3, 6gG9, 1eG15, 1eG30 and 1gG4 (Hayes et al., 2004). Sequence analysis of 3gG2, 5gG3 and 6gG9 showed that these three candidate genes are all NBS-LRR resistance genes (Hayes et al., 2004). Southern analysis of the rare recombinant lines identified 3gG2 as co-segregating with *Rsv1*.

Rare recombinant lines were screened for resistance to SMV strains G1-G3 and G5-G7, along with several cultivars containing different alleles of *Rsv1*. PI96983 contains all six candidate resistance genes and shows resistance to G1-G6. Four interesting rare recombinants were identified from this greenhouse screen. Line 800-46 was shown to have only 3gG2, 613-10 and 1044-98 have all candidate genes except 1gG4, and 943-9 lacks only 3gG2 (Hayes et al., 2004). L800-46, 613-10 and 1044-98 showed extreme resistance to SMV strains G1, G2 and G6, and local lesion necrosis to G3 and G5. The reactions observed in these three lines are similar to those of Ogden (*Rsv1-t*) and Marshall (*Rsv1-t*). Line 943-9 showed resistance reaction to strains G1-G3 and susceptibility to strains G5-G7. This pattern of strain resistance is the same

seen in *Rsv1-y* cultivars (Table 1.1). These results indicate that the different alleles found at the *Rsv1* locus may be due to the presence/absence or expression of the six identified resistance gene candidates (Hayes et al., 2004; Saghai Maroof et al., 2008a).

Recently, the ability to identify cultivars or individuals within a population that contains *Rsv1* has been greatly simplified. Previous polymerase chain reaction (PCR) based assays required the amplification of a ~4 kb region and possible sequence analysis to confirm correct amplification. Hybridization assays required two Southern blot assays, one which identified all six of the candidate resistance genes and a second which differentiated 5gG3 from 3gG2 (Hayes et al., 2004). PCR primers (Rsv1-F/R) were developed from 3gG2 which amplified a 341 bp fragment. This has allowed for the rapid identification of *Rsv1* lines. Analysis of several *Rsv1* containing cultivars showed that Rsv1-F/R specifically amplifies 3gG2 from resistant cultivars and fails to amplify any product from susceptible cultivars (Shi et al., 2008). The one allele of *Rsv1* which fails to amplify is *Rsv1-y*. This is consistent with the possibility of *Rsv1-y* cultivars not containing 3gG2, and the observed phenotypic similarities between 943-9 and *Rsv1-y* cultivars (Hayes et al., 2004; Saghai Maroof et al., 2008a).

Molecular Recognition of SMV by Rsv1

Complementary research has also been done on the SMV side of the plant pathogen interaction. Initial investigation into the molecular mechanism of *Rsv1* recognition utilized serial passage of avirulent SMV strain G7 through inoculations on PI 96983 (Hajimorad et al., 2003). SMV-G7 is considered avirulent because it induced a hypersensitive response. This technique enables the virus to mutate and create new strains. It has been shown more recently that the

Table 1.1: Differential responses of rare recombinants and parental lines inoculated with SMV strains (adapted from Saghai Maroof et al., 2008a). LSHR: Lethal Systemic Hypersensitive Response; LLN: Local Lesion Necrosis; SN: Systemic Necrosis; R: Resistant; S: Susceptible.

	SMV Strains						
Cultivar	**3gG2**	**G1**	**G2**	**G3**	**G5**	**G6**	**G7**
PI96983 (*Rsv1*)	Yes	R	R	R	R	R	LSHR
Lee68 (*rsv1*)	No	S	S	S	S	S	S
613-10	Yes	R	R	LLN	LLN	R	SN
1044-98	Yes	R	R	LLN	LLN	R	SN
Ogden	Yes	R	R	LLN	LLN	R	SN
Marshall	Yes	R	R	LLN	LLN	R	SN
943-9	No	R	R	R	S	S	S
York (*Rsv1-y*)	?	R	R	R	S	S	S

mutations observed during SMV serial passage mimic natural variations between strains of SMV (Hajimorad et al., 2011). During the primary inoculation with strain G7, the expected lethal systemic hypersensitive response was observed (Jayaram et al., 1992; Cho and Goodman, 1979). After several rounds of virus passage it was noted that the lethal systemic hypersensitive (LSHR) reaction but that did not occur, systemic mosaic symptoms were observed. Sequence analysis of the original G7 strain and viral RNA isolated from inoculated plants which did not exhibit LSHR, showed that there were numerous nucleotide mutations which lead to seven amino acid substitutions in the P1, P3, HC-Pro, and Coat proteins. It was concluded that one or all of these substitutions enabled the mutated G7 strain to avoid recognition by *Rsv1* in PI96983 (Hajimorad et al., 2003).

Further investigation into the molecular mechanisms of *Rsv1* recognition utilized chimeric infectious clones and site specific mutations. Chimeric infectious clones developed from virulent and avirulent G7 strains of SMV enabled the identification of the P3 region of the SMV genome as the elicitor of LSHR (Hajimorad et al., 2005). Yet, during these studies it was shown that presence or absence of P3 from the virulent clone was not sufficient for avoiding *Rsv*-mediated resistance (Hajimorad et al., 2005). It was later shown through the use of site directed mutagenesis that avoiding *Rsv1* recognition requires mutations on both the P3 and Hc-Pro. The P3 region of the avirulent SMV strain N (or G2) was shown to confer extreme resistance response to G7 when used to challenge PI96983 (G7-N(P3)), yet reciprocal exchanges (N-G7(P3)) did not allow strain G2 to become virulent (Hajimorad et al., 2006). Mutations in three amino acids, R682M, R787I, and A947T, spanning both the P3 and HC-Pro regions of the SMV genome allowed avirulent SMV-N to become virulent (Eggenberger et al., 2008; Hajimorad et al., 2008). An additional mutation in the HC-Pro was later shown to affect symptom development in *Rsv1* containing cultivars (Seo et al., 2011).

Rsv3

Characterization of the Rsv3 Locus

Rsv3 in comparison to *Rsv1* shows resistance to the higher numbered more virulent strains of SMV, G5-G7. There are at least four known sources of *Rsv3* resistance, Hardee, Tousan140, Columbia, and Hourei (Buss et al., 1999; Gunduz et al., 2002; Ma et al., 2002). The first attempts at genetic mapping utilized AFLP, RFLP and SSR molecular markers as well as a

soybean genomic library to determine the linkage group (Jeong et al., 2002b). Briefly, a mungbean clone Mng247, previously identified to be located on linkage group B2 near *Rsv3*, was used to screen a genomic library for clones that should be linked to *Rsv3*. Using this method several sequences were generated which had polymorphic microsatellites. In addition to defining the *Rsv3* locus to a 0.8 cM genetic region, they were able to identify one candidate gene with an LRR domain previously reported in disease resistance genes *Xa21* and *Cf-9* (Jeong et al., 2002b).

Mechanism of Rsv3 *Resistance*

There are several fundamental differences between the resistance conferred by *Rsv1* and *Rsv3*. *Rsv1* has been described as extreme resistance with no observable hypersensitive response. *Rsv3,* on the other hand, is thought to block SMV movement. To further verify these observations, Zhang et al. (2009) used infectious clones of SMV-G7-GUS and SMV-N-GUS to biolistically inoculate cultivar L29 (*Rsv3*) and L78-379 (*Rsv1*). SMV-N is an isolate of G2 and should infect *Rsv3* cultivars but not *Rsv1*. Little GUS expression was observed in L29 (*Rsv3*) leaves inoculated with SMV-G7-GUS but not in the noninoculated leaves, yet was observed in SMV-GUS-N inoculated and noninoculated leaves. In contrast to this, SMV-G7-GUS was observed in both inoculated and noninoculated L78-379 (*Rsv1*), but no GUS was expressed when inoculated with SMV-N-GUS. These results indicate that the mechanisms of *Rsv1* and *Rsv3* resistance are different, and consistent with observed defense response (Zhang et al., 2009).

Chimeric clones were developed from SMV strains G7 and N (an isolate of G2). Clones specifically swapping P3 or HC-Pro regions between SMV-G7 and SMV-N, using site specific recombination, were used to challenge L29 (*Rsv3*), to test the possibility of *Rsv3* recognizing the same regions of the SMV genome as *Rsv1*. No change in resistance response was observed. Only when exchanging the cytoplasmic inclusion (CI) region of SMV-G7 and SMV-N was an alteration in resistance response observed. This region was further mapped to ~100 bp region between nucleotide 3627 and 3790. Three amino acid differences were identified between SMV-G7 and SMV-N within this region. Site-specific mutagenesis was used to further refine the exact mutation which allowed SMV strain N to avoid recognition by *Rsv3*. A change in amino acid 37 (in the CI) from alanine to aspartic acid allowed SMV strain N to avoid recognition by L29 (*Rsv3*) (Zhang et al., 2009). The CI was independently verified as the site recognized by *Rsv3* by a different research group (Seo et al., 2009), utilizing a SMV-G7 isolate which was able to infect *Rsv3* cultivars and a G5 strain of SMV. The authors showed that a single amino acid difference between these two strains was responsible for symptoms observed in L29 when challenged by the infectious SMV-G7 isolate. It was also shown that a mutation in the HC-Pro modulated the symptoms observed in infected cultivar Jinpumkong-2 (which contains *Rsv1* and*Rsv3*) (Seo et al., 2009; Seo et al., 2011).

Rsv4

Molecular Mapping of the Rsv4 Locus

The *Rsv4* locus was first characterized in PI 486355, and imparts resistance to all known North American SMV strains (Ma et al., 1995). Allelism studies using crosses between PI 486355 and *Rsv1* cultivars indicated that this new form of resistance was not located at the *Rsv1* locus, although PI 486355 has an allele of *Rsv1*. This new form of resistance was first genetically mapped using a selected line LR2, developed from a cross between PI 486355 (*Rsv4*) and Essex (*rsv*), crossed to Lee68 (*rsv*). Bulk segregation analysis along with AFLP, RFLP and SSR molecular markers were employed to localize the *Rsv4* locus to a 10 cM region on linkage group D1b (Hayes et al., 2000b). Subsequently, a comparative genomics project utilizing *Lotus japonicus* expressed sequence tags (ESTs) narrowed the genetic distance between flanking markers to 5 cM (Hwang et al., 2006). More recently, the newly released soybean genome sequence (Schmutz et al., 2010) was utilized to specifically develop SSR markers near the *Rsv4* locus. Saghai Maroof et al. (2010) were able to localize this resistance gene to a small 0.7 cM genomic region with a physical distance of 100kb and identified nine candidate genes. This study was also able to develop single nucleotide polymorphism markers in several of the candidate genes identified in the region (Saghai Maroof, unpublished). Additional sources of *Rsv4* resistance have been verified in three other cultivars; Peking, PI 88788, and Columbia (Gunduz, 2004; Ma et al., 2002).

Molecular Characterization of Rsv4 Resistance

Were as the molecular mechanisms of recognition have been studied extensively in *Rsv1* and *Rsv3* cultivars, *Rsv4* resistance has not been well characterized. This difficulty is apparent in that the seven characterized strains of SMV are unable to systemically invade *Rsv4* lines, only

locally at the site of inoculation (Gunduz et al., 2004). All strains of SMV found in the United States are unable to overcome *Rsv4* resistance when tested, but several isolates from Asia have been shown to infect *Rsv4* cultivars (Choi et al., 2005). Some insights into the mechanism of *Rsv4* resistance can be gained through the use of whole leaf immunoblots. It was observed in previous studies that *Rsv4* does not exhibit the extreme resistance phenotype observed with *Rsv1* cultivars. Instead, *Rsv4* seems to limit the movement of the virus as does *Rsv3*. This is apparent in the late susceptible phenotype observed in previous studies with individuals heterozygous at the *Rsv4* locus (Gunduz, 2004; Saghai Maroof et al., 2008; Saghai Maroof et al., 2010). More recently, mutations in the SMV P3 region were shown to allow isolates of SMV to overcome *Rsv4* resistance (Chowda-Reddy et al., 2011).

Recent Advancements in Soybean Genomic Resources

In recent years the genomic resources available to soybean breeders and researchers have vastly improved with the sequencing of multiple domesticated (*Glycine max*) and wild (*Glycine soja*) soybean lines (Kim et al., 2010; Lam et al., 2010; Schmutz et al., 2010). Prior to the release of the soybean genome sequence, researchers were limited to publically available databases containing published genetic and physical maps and publically available molecular markers (Grant et al., 2010; Shultz et al., 2006). The first of these maps consisted almost exclusively of RFLP markers, and could not properly define the 20 soybean chromosomes (Keim et al., 1989; Keim et al., 1990). Later, with the availability of additional sequence data, microsatellite and single nucleotide polymorphism (SNP) markers were shown to be more highly

polymorphic than RLFPs, of higher abundance in the soybean genome and easier to analyze (Akkaya et al., 1992; Akkaya et al., 1995; Marino et al., 1995; Van et al., 2005; Zhu et al., 2003). Microsatellites have become invaluable to molecular breeders as was seen in pyramiding the three *Rsv* genes. Previously, based on conventional breeding approaches, combining all the Rsv genes in a single cultivar would have been difficult. *Rsv4* would have masked both *Rsv1* and *Rsv3,* and extensive back crossing and disease assays with multiple SMV isolates would have been required (Saghai Maroof et al., 2008b). With the availability of microsatellites and SNPs, comprehensive genetic linkage maps with multiple markers suitable for genetic mapping studies have become available (Cregan et al., 1999; Song et al., 2004) .

While the previously mentioned resources were extremely helpful, the lack of adequate sequence data hindered the development of additional molecular markers. Map-based cloning efforts relied on the availability of sufficient polymorphic markers within a locus of interest to define the region, or the ability to develop additional markers which would target a specific region or subset of genes. One of the more widely used techniques for developing additional markers is analysis of bacterial artificial chromosome (BAC) end sequences and expressed sequence tags (ESTs) (Choi et al., 2007; Shoemaker et al., 2008b; Zhang et al., 2004). Several disease resistance studies have developed molecular markers based on shared homology between many resistance genes (Godwin et al., Unpublished; Kanazin et al., 1996; Yu et al., 1996).

Recently, several studies have utilized the available genetic linkage maps to specifically target regions of interest for sequencing using molecular marker anchored-BACs (Wawrzynski et al., 2008). This technique has been used to identify candidate resistance genes to both SMV and Asian Soybean Rust. Meyer et al. (2009) was able to physically map and sequence the susceptible allele of the *Rpp4* (resistance to rust) locus from the susceptible cultivar Williams82.

The Williams82 genome sequence enabled primers to be designed that would amplify the four candidate resistance genes from the rust resistant cultivar PI 459025B. After sequencing the candidate genes from the resistant line, virus induced gene silencing (Purkayastha and Dasgupta, 2009) vectors were developed to specifically silence the resistant gene cluster and confirm the identification of the four candidate genes (Meyer et al., 2009). Another interesting outcome from this experiment was the similarity seen between sequenced candidate resistance genes of resistant and susceptible lines, indicating that there was a high level of sequence conservation between Williams82 and PI 459025B at the *Rpp4* locus. It addition, it was shown that sequence information from the susceptible source was useful for identifying candidate genes in the resistant line. In another study, the birth and death of resistance genes found at the *Rsv1* and *Rpg1* (resistance to *Pseudomonas syringae* pv. *glycinea*) were analyzed by comparing recently duplicated regions in the soybean genome and syntenic regions in closely related species. There was a high retention of resistance genes between legume species and after the most recent soybean genome duplication event (Innes et al., 2008). This study demonstrated the importance of sequencing multiple cultivars to properly investigate clusters of disease resistance genes.

When the unassembled soybean genome was published in 2008, several studies utilized the sequence data to identify genes of interest for various traits. This initial genome release consisted of scaffolds which were not assigned to chromosomes or to known molecular linkage groups. However, the available sequence information was useful to researchers. Blast analysis of previously published molecular markers enabled the identification of scaffolds which could potentially contain a gene or QTL of interest. Bioinformatic software such as SSRIT (Temnykh et al., 2001) and Primer3 (Rozen and Skaletsky, 1999) could then be used to identify and design microsatellite markers to further refine the region for the locus of interest. These techniques

were recently used before the final assembly of the soybean genome was released to discover several genes of agronomic importance (Saghai Maroof et al., 2009, 2010; Skoneczka et al., 2009). Using various bioinformatic tools Saghai Maroof et al. (2009) showed a single SNP, in a multi drug resistance – associated protein (MRP) on linkage groups L and N, to be responsible for low levels of phytic acid in the soybean line CX1834; and Skoneczka et al. (2009) classified a mutation in *rsm1* (raffinose synthase) which could be used to select for lines low in the sugar raffinose. Saghai Maroof et al. (2010) developed several markers from the initial scaffold release of the soybean genome to fine-map the *Rsv4* resistance locus, then used primers designed from the susceptible cv. Williams82 to sequence candidate resistance genes. Access to the soybean genome allowed researchers to narrow the physical distance harboring *Rsv4* from approximately 1 Mb to 100 kb, a region with only nine genes of interest, according to the Williams82 reference genome (Saghai Maroof et al., 2010).

Research Objectives

In the next two chapters the soybean genome sequence will be used to develop molecular markers for fine mapping *Rsv3* and further analysis of the *Rsv3* loci.

- **Chapter 2:** "The Rsv3 Locus Conferring Resistance to Soybean Mosaic Virus is Associated with a Cluster of Coiled-Coil Nucleotide-Binding Leucine-Rich Repeat Genes". Markers were developed to further refine the *Rsv3* locus, and candidate nucleotide-binding leucine rich repeat genes were identified.

- **Chapter 3:** A large population segregating for *Rsv3* will be assayed for SMV resistance to facilitate future resistance studies. The *Rsv3* locus will be further analyzed to identify syntenic regions within the soybean genome and in the closely related model organism Medicago.

References:

Akkaya M.S., Bhagwat A.A., Cregan P.B. (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Akkaya M.S., Shoemaker R.C., Specht J.E., Bhagwat A.A., Cregan P.B. (1995) Integration of simple sequence repeat dna markers into a soybean linkage map. Crop Sci. 35:1439-1445.

Buss G.R., Ma G., Kristipati S., Chen P., Tolin S.A. (1999) A new allele at the *Rsv3* locus for resistance to soybean mosaic virus, in: H. E. Kauffman (Ed.), Proc. World Soybean Res. Conf. VI., Superior Printing, Champaign, IL. p. 490.

Buzzell R.I., Tu J.C. (1989) Inheritance of a soybean stem-tip necrosis reaction to soybean mosaic virus. J. Hered. 80:400-401.

Calvert L.A., Ghabrial S.A. (1983) Enhancement by soybean mosaic virus of bean pod mottle virus titer in doubly infected soybean. Phytopathology 73:992-997.

Chen P., Buss G.R., Roane C.W., Tolin S.A. (1991) Allelism among genes for resistance to soybean mosaic virus in strain-differential soybean cultivars. Crop Sci. 31:305-309.

Cho E.K., Goodman R.M. (1979) Strains of soybean mosaic virus: classification based on virulence in resistant soybean cultivars. Phytopathology 69:467-470.

Choi B.K., Koo J.M., Ahn H.J., Yum H.J., Choi C.W., Ryu K.H., Chen P., Tolin S.A. (2005) Emergence of *Rsv*-resistance breaking soybean mosaic virus isolates from Korean soybean cultivars. Virus Res. 112:42-51.

Choi I.Y., Hyten D.L., Matukumalli L.K., Song Q., Chaky J.M., Quigley C.V., Chase K., Lark K.G., Reiter R.S., Yoon M.S., Hwang E.Y., Yi S.I., Young N.D., Shoemaker R.C., van Tassell C.P., Specht J.E., Cregan P.B. (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685-696.

Chowda-Reddy R.V., Sun H.Y., Chen H.Y., Poysa V., Ling H., Gijzen M., Wang A.M. (2011) Mutations in the P3 protein of soybean mosaic virus G2 isolates determine virulence on Rsv4-genotype soybean. Mol. Plant Microbe In. 24:37-43.

Cregan P.B., Jarvik T., Bush A.L., Shoemaker R.C., Lark K.G., Kahler A.L., Kaya N., VanToai T.T., Lohnes D.G., Chung L., Specht J.E. (1999) An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464-1490.

Domier L.L., Latorre I.J., Steinlage T.A., McCoppin N., Hartman G.L. (2003) Variability and transmission by Aphis glycines of North American and Asian soybean mosaic virus isolates. Arch. Virol. 148:1925-1941.

Edwards K., Johnstone C., Thompson C. (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. 19:1349.

Eggenberger A.L., Hajimorad M.R., Hill J.H. (2008) Gain of virulence on Rsv1-genotype soybean by an avirulent soybean mosaic virus requires concurrent mutations in both P3 and HC-Pro. Mol. Plant Microbe In. 21:931-936.

Giesler L.J., Ghabrial S.A., Hunt T.E., Hill J.H. (2002) Bean pod mottle virus: A threat to U.S. soybean production. Plant Dis. 86:1280-1289.

Godwin M., Tucker D.M., Jeong S.C., Hayes A.J., Schmidt D., Chaky J.M., Han F., Saghai Maroof M.A. Molecular mapping of disease-resistance related sequence tags and resistance gene analogues in soybean. (Unpublished).

Gordon S.G., St Martin S.K., Dorrance A.E. (2006) Rps8 maps to a resistance gene rich region on soybean molecular linkage group F. Crop Sci. 46:168-173.

Gore M.A., Hayes A.J., Jeong S.C., Yue Y.G., Buss G.R., Maroof S. (2002) Mapping tightly linked genes controlling potyvirus infection at the Rsv1 and Rpv1 region in soybean. Genome 45:592-599.

Grant D., Nelson R.T., Cannon S.B., Shoemaker R.C. (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res. 38:843-846.

Gunduz I., Buss G.R., Ma G., Chen P., Tolin S.A. (2001) Genetic analysis of resistance to soybean mosaic virus in OX670 and Harosoy soybean. Crop Sci. 41:1785-1791.

Gunduz I., Buss G.R., Chen P., Tolin S.A. (2002) Characterization of SMV resistance genes in Tousan 140 and Hourei Soybean. Crop Sci. 42:90-95.

Gunduz I., Buss G.R., Chen P., Tolin S.A. (2004) Genetic and phenotypic analysis of soybean mosaic virus resistance in PI 88788 soybean. Phytopathology 94:687-692.

Hajimorad M.R., Eggenberger A.L., Hill J.H. (2003) Evolution of soybean mosaic virus-G7 molecularly cloned genome in Rsv1-genotype soybean results in emergence of a mutant capable of evading Rsv1-mediated recognition. Virology 314:497-509.

Hajimorad M.R., Eggenberger A.L., Hill J.H. (2005) Loss and gain of elicitor function of Soybean mosaic virus G7 provoking Rsv1-mediated lethal systemic hypersensitive response maps to P3. J. Virol. 79:1215-1222.

Hajimorad M.R., Eggenberger A.L., Hill J.H. (2006) Strain-specific P3 of soybean mosaic virus elicits Rsv1-mediated extreme resistance, but absence of P3 elicitor function alone is insufficient for virulence on Rsv1-genotype soybean. Virology 345:156-166.

Hajimorad M.R., Eggenberger A.L., Hill J.H. (2008) Adaptation of soybean mosaic virus avirulent chimeras containing P3 sequences from virulent strains to Rsv1-genotype soybeans is mediated by mutations in HC-Pro. Mol. Plant Microbe In. 21:937-946.

Hajimorad M.R., Wen R.H., Eggenberger A.L., Hill J.H., Maroof M.A. (2011) Experimental adaptation of an RNA virus mimics natural evolution. J. Virol. 85:2557-2564.

Hayes A.J., Saghai Maroof M.A. (2000) Targeted resistance gene mapping in soybean using modified AFLPs Theor. Appl. Genet. 100:1279-1283.

Hayes A.J., Y.G. Y., Saghai Maroof M.A. (2000a) Expression of two soybean resistance gene candidates shows divergence of paralogous single copy genes. Theor. Appl. Genet 101:789-795.

Hayes A.J., Ma G., Buss G.R., Saghai Maroof M.A. (2000b) Molecular marker mapping of Rsv4, a gene conferring resistance to all known strains of soybean mosaic virus. Crop Sci. 40:1434-1437.

Hayes A.J., Jeong S.C., Gore M.A., Yu Y.G., Buss G.R., Tolin S.A., Saghai Maroof M.A. (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.

Hwang T. Y., Moon J. K., Yu S., Yang K., Mohankumar S., Yu Y.H., Lee Y.H.o., Kim HSig., Kim H. Mook., Saghai Maroof M. A., Jeong S.C. (2006) Application of comparative genomics in developing molecular markers tightly linked to the virus resistance gene Rsv4 in soybean. Genome 49:380-388.

Innes R.W., Ameline-Torregrosa C., Ashfield T., Cannon E., Cannon S.B., Chacko B., Chen N.W., Couloux A., Dalwani A., Denny R., Deshpande S., Egan A.N., Glover N., Hans C.S., Howell S., Ilut D., Jackson S., Lai H., Mammadov J., Del Campo S.M., Metcalf M., Nguyen A., O'Bleness M., Pfeil B.E., Podicheti R., Ratnaparkhe M.B., Samain S., Sanders I., Segurens B., Sevignac M., Sherman-Broyles S., Thareau V., Tucker D.M., Walling J., Wawrzynski A., Yi J., Doyle J.J., Geffroy V., Roe B.A., Maroof M.A., Young N.D. (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol. 148:1740-1759.

Jayaram C., Hill J.H., Miller W.A. (1992) Complete nucleotide sequences of two soybean mosaic virus strains differentiated by response of soybean containing the Rsv resistance gene. J. Gen. Virol. 73:2067-2077.

Jeong S.C., Kristipati S., Hayes A.J., Maughan P.J., Noffsinger S.L., Gunduz I., Buss G.R., Maroof M.A. (2002) Genetic and sequence analysis of markers tightly linked to the Soybean mosaic virus resistance gene, Rsv3. Crop Sci 42:265-270.

Kanazin V., Marek L.F., Shoemaker R.C. (1996) Resistance gene analogs are conserved and clustered in soybean. Proc. Natl. Acad. Sci. U. S. A. 93:11746-11750.

Keim P., Shoemaker R.C., Palmer R.G. (1989) Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-792.

Keim P., Diers B.W., Olson T.C., Shoemaker R.C. (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Kiihl R.A.S., Hartwig E.E. (1979) Inheritance of reaction to soybean mosaic virus in soybeans. Crop Sci. 19:372-375.

Kim M.Y., Lee S., Van K., Kim T.H., Jeong S.C., Choi I.Y., Kim D.S., Lee Y.S., Park D., Ma J., Kim W.Y., Kim B.C., Park S., Lee K.A., Kim D.H., Kim K.H., Shin J.H., Jang Y.E., Kim K.D., Liu W.X., Chaisan T., Kang Y.J., Lee Y.H., Moon J.K., Schmutz J., Jackson S.A., Bhak J., Lee S.H. (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. Proc. Natl. Acad. Sci. U. S. A. 107:22032-22037.

Lam H.M., Xu X., Liu X., Chen W., Yang G., Wong F.L., Li M.W., He W., Qin N., Wang B., Li J., Jian M., Wang J., Shao G., Sun S.S., Zhang G. (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 42:1053-1059.

Ma G., Chen P., Buss G.R., Tolin S.A. (1995) Genetic characteristics of two genes for resistance to soybean mosaic virus in PI 486355 soybean. Theor. Appl. Genet. 91:907-914.

Ma G., Chen P., Buss G.R., Tolin S.A. (2002) Complementary action of two independent dominant genes in Columbia soybean for resistance to soybean mosaic virus. J Hered 93:179-184.

Ma G., Chen P., Buss G.R., Tolin S.A. (2003) Genetic study of a lethal necrosis to soybean mosaic virus in PI 507389 soybean. J. Hered. 94:205-211.

Malapi-Nelson M., Hajimorad M.R., Wen R.H., Ownley B.H. (2009) Co-infection of soybean with soybean mosaic virus and alfalfa mosaic virus results in disease synergism and alteration in accumulation level of both viruses. Plant Dis. 93:1259-1264.

Marino M.A., Turni L.A., Delrio S.A., Williams P.E., Cregan P.B. (1995) The analysis of simple sequence repeat DNA in soybean by capillary gel-electrophoresis. Appl. Theor. Electroph. 5:1-5.

Meyer J.D., Silva D.C., Yang C., Pedley K.F., Zhang C., van de Mortel M., Hill J.H., Shoemaker R.C., Abdelnoor R.V., Whitham S.A., Graham M.A. (2009) Identification and analyses of candidate genes for rpp4-mediated resistance to Asian soybean rust in soybean. Plant Physiol. 150:295-307.

Purkayastha A., Dasgupta I. (2009) Virus-induced gene silencing: A versatile tool for discovery of gene functions in plants. Plant Physiol. Bioch. 47:967-976.

Ren Q., Pfeiffer T.W., Ghabrial S.A. (1997) Soybean mosaic virus incidence level and infection time: interaction effects on soybean. Crop Sci. 37:1712-1718.

Ross J.P. (1968) Effect of single and double infections of soybean mosaic and bean pod mottle viruses on soybean yield and seed characters. Plant Dis. Repor**.** 52:344-348

Saghai Maroof M.A., Biyashev R.M., Yang G.P., Zhang Q., Allard R.W. (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. (USA) 91:5466-5470.

Saghai Maroof M.A., Tucker D.M., Tolin S.A. (2008a) Genomics of viral soybean interactions, in: G. Stacey (Ed.), Soybean Genomics, Springer Press. Ch. 16 p. 293-319

Saghai Maroof M.A., Jeong S.C., Gunduz I., Tucker D.M., Buss G.R., Tolin S.A. (2008b) Pyramiding of Soybean mosaic virus resistance genes by marker-assisted selection. Crop Sci 48:517-526.

Saghai Maroof M.A., Glover N.M., Biyashev R.M., Buss G.R., Grabau E.A. (2009) Genetic basis of the low-phytate trait in the soybean line CX1834. Crop Sci. 49:69-76.

Saghai Maroof M.A., Tucker D.M., Skoneczka J.A., Bowman B.C., Tripathy S., Tolin S.A. (2010) Fine mapping and candidate gene discovery of the soybean mosaic virus resistance gene, Rsv4. Plant Gen. 3:14-22.

Sandhu D., Schallock K.G., Rivera-Velez N., Lundeen P., Cianzio S., Bhattacharyya M.K. (2005) Soybean phytophthora resistance gene Rps8 maps closely to the Rps3 region. J. Hered. 96:536-41.

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., Jackson S.A. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Seo J.K., Lee S.H., Kim K.H. (2009) Strain-specific cylindrical inclusion protein of soybean mosaic virus elicits extreme resistance and a lethal systemic hypersensitive response in two resistant soybean cultivars. Mol. Plant Microbe In. 22:1151-1159.

Seo J.K., Sohn S.H., Kim K.H. (2011) A single amino acid change in HC-Pro of soybean mosaic virus alters symptom expression in a soybean cultivar carrying Rsv1 and Rsv3. Arch. Virol. 156:135-141.

Shi A.N., Chen P.Y., Zheng C.N., Hou A.F., Zhang B. (2008) A PCR-based marker for the Rsv1 locus conferring resistance to soybean mosaic virus. Crop Sci. 48:262-268.

Shoemaker R.C., Grant D., Olson T., Warren W.C., Wing R., Yu Y., Kim H., Cregan P., Joseph B., Futrell-Griggs M., Nelson W., Davito J., Walker J., Wallis J., Kremitski C., Scheer D., Clifton S.W., Graves T., Nguyen H., Wu X., Luo M., Dvorak J., Nelson R., Cannon S., Tomkins J., Schmutz J., Stacey G., Jackson S. (2008) Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. Genome 51:294-302.

Shultz J.L., Kurunam D., Shopinski K., Iqbal M.J., Kazi S., Zobrist K., Bashir R., Yaegashi S., Lavu N., Afzal A.J., Yesudas C.R., Kassem M.A., Wu C., Zhang H.B., Town C.D., Meksem K., Lightfoot D.A. (2006) The Soybean Genome Database (SoyGD): a browser

for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of Glycine max. Nucleic Acids Res 34:758-765.

Skoneczka J.A., Saghai Maroof M.A., Shang C., Buss G.R. (2009) Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. Crop Sci. 49:247-255.

Song Q.J., Marek L.F., Shoemaker R.C., Lark K.G., Concibido V.C., Delannay X., Specht J.E., Cregan P.B. (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122-128.

Suh S.J., Bowman B.C., Jeong N., Yang K., Kaslt C., Tolin S.A., Saghai Maroof M.A., Jeong S.-C. (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 Gen. 4:55-64.

Tucker D.M., Saghai Maroof M.A., Jeong S.C., Buss G.R., Tolin S.A. (2009) Validation and interaction of the Soybean mosaic virus lethal necrosis allele, Rsv1-n, in PI 507389. Crop Sci 49:1277-1283.

Van K., Hwang E.Y., Kim M.Y., Park H.J., Lee S.H., Cregan P.B. (2005) Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea. J. Hered. 96:529-535.

Van Ooijen J.W., Voorrips R.E. (2001) JoinMap 3.0 software for the calculation of genetic linkage maps. Plant Research International,Wageningen, The Netherlands.

Wawrzynski A., Ashfield T., Chen N.W., Mammadov J., Nguyen A., Podicheti R., Cannon S.B., Thareau V., Ameline-Torregrosa C., Cannon E., Chacko B., Couloux A., Dalwani A., Denny R., Deshpande S., Egan A.N., Glover N., Howell S., Ilut D., Lai H., Del Campo S.M., Metcalf M., O'Bleness M., Pfeil B.E., Ratnaparkhe M.B., Samain S., Sanders I., Segurens B., Sevignac M., Sherman-Broyles S., Tucker D.M., Yi J., Doyle J.J., Geffroy V., Roe B.A., Saghai Maroof M.A., Young N.D., Innes R.W. (2008) Replication of nonautonomous retroelements in soybean appears to be both recent and common. Plant Physiol. 148:1760-1771.

Wrather J.A., Koenning S.R. (2006) Estimates of disease effects on soybean yields in the United States 2003 to 2005. J. Nematol. 38:173-180.

Yu Y.G., Buss G.R., Saghai Maroof M.A. (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc. Natl. Acad. Sci. (USA) 93:11751-11756.

Zhang C., Hajimorad M.R., Eggenberger A.L., Tsang S., Whitham S.A., Hill J.H. (2009) Cytoplasmic inclusion cistron of Soybean mosaic virus serves as a virulence determinant on Rsv3-genotype soybean and a symptom determinant. Virology 391: 240-248

Zhang W.K., Wang Y.J., Luo G.Z., Zhang J.S., He C.Y., Wu X.L., Gai J.Y., Chen S.Y. (2004) QTL mapping of ten agronomic traits on the soybean (Glycine max L. Merr.) genetic map and their association with EST markers. Theor. Appl. Genet. 108:1131-1139.

Zheng C., Chen P., Gergerich R. (2005) Effect of temperature on the expression of necrosis in soybean infected with soybean mosaic virus. Crop Sci. 45:916-922.

Zhu Y.L., Song Q.J., Hyten D.L., Van Tassell C.P., Matukamalli L.K., Grimm D.R., Hyatt S.M., Fickus E.W., Young N.D., Cregan P. (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123-1134.

Chapter 2

The *Rsv3* Locus Conferring Resistance to Soybean Mosaic Virus is Associated with a Cluster of Coiled- Coil Nucleotide-Binding Leucine-Rich Repeat Genes

Su Jeoung Suh[1,2], Brian C. Bowman[1,3], Namhee Jeong[2], Kiwoung Yang[2], Christin Kastl[3], Sue A. Tolin[4], M.A. Saghai Maroof[3], and Soon-Chun Jeong[2]*

Author Contribution:

Dr. Su Jeoung Suh, Namhee Jeong, Kiwoung Yang, and Dr. Soon-Chun Jeong all contributed to marker development, candidate gene discovery, disease assays and mapping for the L29 by Sowon population. They also prepared the first draft of the manuscript based on three mapping populations. Christin Kastl aided in microsatellite analysis for the L29 by Lee68 and Tousan140 by Lee68 populations. Dr. Sue Tolin contributed the virus isolates and soybean mosaic virus expertise, as well as editing of the manuscript together with Dr. Saghai Maroof. Dr. Saghai Maroof originally developed the L29 by Lee68 and Tousan140 by Lee68 populations and offered advice on marker design and genetic mapping. Dr. Soon-Chun Jeong and Dr. Saghai Maroof were the principle investigators for this project and provided expertise in all areas of the research. They together published the initial *Rsv3* mapping paper based on the L29 by Lee68 and Tousan140 by Lee68 populations. Dr. Su Jeoung Suh, Dr. Sue Tolin, Dr. Saghai Maroof, Dr Soon-Chun Jeong and I, aided in the writing and editing of this manuscript. Finally, I contributed towards designing nine of the molecular markers, identifying and analyzing candidate genes, genetic mapping of *Rsv3*, data analysis and all topics mentioned above.

[1] *Equally Contributed*
[2] *Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Cheongwon, Chungbuk 363-883, Republic of Korea; B.C.*
[3] *Dep. of Crop & Soil Environmental Sciences, Virginia Tech, Blacksburg, VA, 20461*
[4] *Dep. of Plant Pathology, Physiology & Weed Science, Virginia Tech, Blacksburg, VA 24061*
[*] *Author for correspondence: scjeong@kribb.re.kr*

Chapter 2

Abbreviations: Avr, avirulence; CC, coiled-coil; GI, gene identification; LRR, leucine-rich repeat; LL, 'L29' × 'Lee68'; LS, 'L29' × 'Sowon'; MLG, molecular linkage group; NB, nucleotidebinding; PCR, polymerase chain reaction; R genes, resistance genes; RLK, receptor-like kinase; SMV, Soybean mosaic virus; TIR, Toll/ interleukin-1 receptor; TL, 'Tousan 140' × 'Lee68'.

The *Rsv3* Locus Conferring Resistance to *Soybean Mosaic Virus* is Associated with a Cluster of Coiled-Coil Nucleotide-Binding Leucine-Rich Repeat Genes

Su Jeoung Suh, Brian C. Bowman, Namhee Jeong, Kiwoung Yang, Christin Kastl, Sue A. Tolin, M.A. Saghai Maroof, and Soon-Chun Jeong*

Abstract

The *Soybean mosaic virus* (SMV) resistance locus, *Rsv3*, previously mapped between markers A519F/R and M3Satt in the soybean molecular linkage group B2 (chromosome 14), has been characterized by examination of the soybean genome sequence. The 154 kbp interval encompassing *Rsv3* contains a family of closely related coiled-coil nucleotide-binding leucine-rich repeat (CC-NB-LRR) genes. Tightly linked to this region are additional CC-NB-LRR genes and several leucine-rich repeat receptor-like kinase (LRR-RLK) genes, thereby indicating that members of both multigene families constitute a heterogeneous cluster at the *Rsv3* chromosomal region. To further confirm the sequence and genetic map concordance, we developed 16 markers from the genomic sequence including predicted CC-NB-LRR genes and their flanking sequences. Mapping of the resultant markers in three populations showed parallel alignment between the genetic and sequence maps in the *Rsv3*-containing region. Phylogenetic analysis of five CC-NB-LRR genes including a pseudogene showed they were highly similar to each other and formed a subclade within a CC-NB-LRR gene clade with representatives from several plant families including legume species. These results demonstrate that the *Rsv3* locus is associated with this cluster of CC-NB-LRR genes, thereby suggesting that the *Rsv3* gene most likely encodes a member of this gene family. In addition, information from this study should facilitate marker-assisted selection and pyramiding of resistance genes.

Published in The Plant Genome 4:55–64. Published 16 Mar. 2011.
doi: 10.3835/plantgenome2010.11.0024

5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

IN PLANTS, most characterized disease resistance genes (*R* genes) encode nucleotide-binding leucine-rich repeat (NB-LRR) proteins characterized by nucleotide-binding (NB) and leucine-rich repeat (LRR) structures as well as variable amino- and carboxy-terminal domains (Collier and Moffett, 2009). These genes tend to cluster in short chromosomal regions (Michelmore and Meyers, 1998). On the basis of their N-termini, two subfamilies of NB-LRR resistance proteins are known: the first is characterized by the Toll/interleukin-1 receptor (TIR)-domain homologous to the *Drosophila* Toll and mammalian interleukin-1 receptors and the second is characterized by a coiled-coil (CC) structure. Truncated versions of NB-LRR genes exist that encode proteins lacking either a domain near the N-terminus of NB or the LRR region or that consist only of a TIR domain. The two subfamilies cluster separately in phylogenetic analysis using their NB domains (McHale et al., 2006).

Resistance genes mediate dominant resistance to pathogens possessing corresponding avirulence (Avr) genes (Collier and Moffett, 2009; Jones and Dangl, 2006). Single dominant *R* genes typically respond differentially to pathogen strains or races. The typical dominant resistance response is associated with several defense-related events, including rapid induction of reactive oxygen species, phytoalexin accumulation,

S.J. Suh, N. Jeong, K. Yang, and S.C. Jeong, Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Cheongwon, Chungbuk 363-883, Republic of Korea; B.C. Bowman, C. Kastl, and M.A. Saghai Maroof, Dep. of Crop & Soil Environmental Sciences, Virginia Tech, Blacksburg, VA, 20461; S.A. Tolin, Dep. of Plant Pathology, Physiology & Weed Science, Virginia Tech, Blacksburg, VA 24061; Su Jeong Suh and Brian C. Bowman contributed equally to this work. Received 11 Nov. 2010. *Corresponding author (scjeong@kribb.re.kr).

Abbreviations: Avr, avirulence; CC, coiled-coil; GI, gene identification; LRR, leucine-rich repeat; LL, 'L29' × 'Lee68'; LS, 'L29' × 'Sowon'; MLG, molecular linkage group; NB, nucleotide-binding; PCR, polymerase chain reaction; *R* genes, resistance genes; RLK, receptor-like kinase; SMV, *Soybean mosaic virus*; TIR, Toll/interleukin-1 receptor; TL, 'Tousan 140' × 'Lee68'.

and activation of salicylic acid biosynthesis and pathogenesis-related genes, which often results in localized necrotic response (Hammond-Kosack and Jones, 1996).

Rsv3, one of three known genes that confer resistance to the *Soybean mosaic virus* (SMV) in soybean [*Glycine max* (L.) Merr.], is unlike the well-characterized *Rsv1* alleles in terms of the patterns of resistance to seven SMV strain groups (G1–G7 classified on the basis of their virulence; Cho and Goodman, 1979). Various alleles of the *Rsv1* locus generally confer extreme resistance to the lower numbered (G1 through G4) strain groups and condition necrotic or mosaic reactions to higher numbered (G5 through G7) groups (Chen et al., 1991). *Rsv3* alleles from diverse soybean cultivars including 'Columbia', 'Hardee', 'Tousan 140', and 'Harosoy' confer extreme resistance to the higher numbered strain groups (G5 through G7) and condition stem-tip necrosis and/or mosaic symptoms to the lower numbered groups (Tu and Buzzell, 1987; Buzzell and Tu, 1989; Bowers et al., 1992; Gunduz et al., 2002; Ma et al., 2002). Although stem-tip necrosis was proposed as a representative symptom conditioned by *Rsv3* in a line derived from Columbia × Harosoy at the time of the first description of this gene (Tu and Buzzell, 1987; Buzzell and Tu, 1989), this symptom has not been observed in 'L29', a 'Williams' isoline derived from Hardee (Bernard et al., 1991; Gunduz et al., 2000). Pyramiding the *Rsv3* gene from L29 with *Rsv1* conferred resistance to all strains of SMV, demonstrating the value of this gene for developing durable SMV resistant soybean lines (Saghai Maroof et al., 2008). *Rsv1* is associated with a NB-LRR gene cluster on the soybean molecular linkage group (MLG) F (chromosome 13) (Hayes et al., 2004) where multiple disease resistance genes have been identified and there are multiple NB-LRR clusters (Innes et al., 2008; Wawrzynski et al., 2008). Jeong et al. (2002) mapped the *Rsv3* locus between markers A519F/R and M3Satt on MLG B2 (chromosome 14). Although the disease responses of the *Rsv3* gene against SMV, including extreme resistance and stem-tip necrosis, were typical of those conditioned by NB-LRR genes, the molecular nature of the *Rsv3* gene is largely unknown. Interestingly, one end sequence of the restriction fragment length polymorphism marker M1a, which is closely linked to the *Rsv3* gene, was reported (Jeong et al., 2002) to contain an LRR consensus sequence highly similar to that of the extracellular LRR domain of resistance genes, *Cf-9* and *Xa21* (Jones et al., 1994; Song et al., 1995).

In 2008, preliminary soybean whole genome shotgun sequence assembly was released (version "Glyma0") by the USDOE-Joint Genome Institute Community Sequencing Program (www.phytozome.net/soybean.php [verified 7 Jan. 2011]) and then an improved version Glyma1.0 was released and reported in 2010 (Schmutz et al., 2010). Integration of the soybean sequence and physical maps with the dense genetic marker map would allow the association of mapped phenotypic effectors with the causal DNA sequence (Jackson et al., 2006). In this study, using a sequence-based marker development strategy in three populations, we determined that the *Rsv3* gene cosegregates with a cluster of the coiled-coil nucleotide-binding leucine-rich repeat (CC-NB-LRR) resistance genes, which is located in the middle of a heterogeneous cluster containing multiple CC-NB-LRR and leucine-rich repeat receptor-like kinase (LRR-RLK) genes.

Materials and Methods

Plant Genetic Materials and Disease Reactions

A BC_3F_2 population of 188 individuals from a cross between L29 (*Rsv3*) and 'Sowon' (*rsv3*) (hereafter referred to as the LS population) was used to investigate the genetic linkage relationship between *Rsv3* and molecular markers. This LS population was previously used to develop an *Rsv3*-linked sequence-based marker, and disease reactions of its F_2 individuals were determined by inoculation with the SMV strain G6 (Yu et al., 2005).

To substantiate the genetic relationship between *Rsv3* and the microsatellite markers, two additional populations were used: an F_2 population of 183 individuals from a cross between L29 (*Rsv3*) and 'Lee68' (*rsv3*) (hereafter referred to as the LL population) and an F_2 population of 61 individuals from a cross between Tousan 140 (*Rsv3*) and Lee68 (*rsv3*) (hereafter referred to as the TL population). These LL and TL populations were previously used to locate the *Rsv3* gene in the context of soybean molecular linkage groups (Jeong et al., 2002). The LL and TL populations were previously screened for resistance to SMV strain G7, as described by Jeong et al. (2002) and Gunduz et al. (2002). Briefly, SMV G7 cultures were maintained on the soybean cultivar York, which is resistant to SMV G1 but susceptible to SMV G7, to ensure a uniform source of inoculum. Twenty plants for each $F_{2:3}$ line were inoculated at the unifoliate stage using a carborundum rub method and scored at 14 and 28 d post inoculation. Plants were designated resistant if no symptoms were present and susceptible if mosaic symptoms appeared.

Alignment of Sequences of *Rsv3*-Linked Markers against the Soybean Whole Genome Sequence

To locate the *Rsv3* region in the soybean whole genome sequence, BLASTN searches of sequences of molecular markers that had been mapped near the *Rsv3* locus were performed initially against the whole genome shotgun sequence release Glyma0 (http://www.phytozome.net). Sequence analysis was subsequently repeated against Glyma1.0. The markers used for BLASTN searches included A519F/R, M3Satt, M1a, Satt063, Satt560, and Gm-r-Z20a (Jeong et al., 2002; Yu et al., 2005). The predicted gene models from the region delimited by these markers were retrieved from the soybean gene annotation database (accessible at Phytozome v5.0, http://www.phytozome.net, accessed April 2010) for further analysis. Open-reading frame and conserved protein domains were obtained from the Glyma1.0 annotations with the help of the "GBrowse" function of Phytozome (Stein et al., 2002).

Marker Development

To confirm the genetic and physical concordance of the region of the soybean genome sequence corresponding to the predicted *Rsv3*-residing region, several novel markers were

generated from a single NB-LRR gene (Glyma14 g38500.1) and microsatellite repeat sites in the sequence region delimited by A519F/R and M3Satt and from a single LRR-RLK gene (Glyma14 g38650.1) and microsatellite repeat sites flanking the region. For the NB-LRR gene and LRR-RLK gene sequences, primer sets were designed to amplify genomic DNA of the parental soybean lines Sowon and L29 to detect sequence polymorphisms between them (Supplemental Table S1). Polymerase chain reaction (PCR) products were prepared for sequencing by excising a band of expected size from an agarose gel followed by purification by an Accuprep Gel Purification Kit (Bioneer, Daejeon, Korea). When necessary, a given PCR product was subcloned into a plasmid and multiple clones were sequenced. Primers were designed using the Primer3 (http://frodo.wi.mit.edu/primer3/input.htm [verified 7 Jan. 2011]) program. Amplification by means of microsatellite primer sets (Supplemental Table S2) was performed as described by Jeong and Saghai Maroof (2004). Polymerase chain reaction products were resolved by using 3% agarose or 6.5% polyacrylamide gel electrophoresis. Sequence analysis was performed using the BioEdit program (Hall, 1999). Linkage analysis of the markers was performed using MapMaker 3.0b (Lander et al., 1987).

Phylogenetic Analysis

The 1600 NB sequences used in phylogenetic analyses to subdivide NB-LRR proteins from the diverse plant texa into the functionally distinct TIR-domain-containing and CC-domain-containing subfamilies (McHale et al., 2006) were downloaded to serve as a local protein database. BLASTP searches of NB-LRR proteins located at the *Rsv3*-residing chromosomal region were performed against the local protein database using the "Local Blast" option implemented in BioEdit (Hall, 1999). The NB sequence of NB-LRR proteins near or cosegregating with the *Rsv3* locus, the hit sequences, and their close relatives were used to construct a gene tree.

Full-length amino acid sequences of 194 LRR-RLKs and their phylogeny, representing most of the LRR-RLK genes in the *Arabidopsis thaliana* (L.) Heynh. genome (Gou et al., 2010), were downloaded to serve as a local protein database. TBLASTN searches of LRR-RLK proteins located near the *Rsv3* locus were performed as above against this database (Hall, 1999).

For multiple sequence alignment and phylogenetic analysis, protein sequences were analyzed by using ClustalW and the Neighbor-joining and bootstrap methods implemented in MEGA 4 (Kumar et al., 2008). The weighing matrix used for ClustalW alignment was BLOSUM with the penalty of gap opening 10 and gap extension 0.2. The bootstrap consensus trees were inferred from 1000 replicates.

Results

Phytozome Annotation Map

Sequences of molecular markers that have been mapped in three *Rsv3*-segregating populations (Jeong et al., 2002; Yu et al., 2005) were positioned on soybean chromosome 14 sequence (pseudomolecule) using BLASTN searches against the soybean genome sequence database. The sequential order of the markers determined by genetic maps was concordant to the physical positions of the markers on the soybean chromosome 14 sequence (Fig. 1). Inspection of the soybean gene annotation database revealed that the *Rsv3* chromosomal region contains multiple members of two gene families: NB-LRR and LRR-RLK (Fig. 2; see Table 1 for the gene and marker annotation). Therefore, the sequence region (154 kbp) between A519F/R and M3Satt that brackets the *Rsv3* locus on the soybean MLG B2 (chromosome 14) (Jeong et al., 2002; Yu et al., 2005) and the surrounding regions were further analyzed by sequence comparison and sequence-based marker development.

Sequence Evaluation of the Chromosomal Region between A519F/R and M3Satt through New Marker Development

To further substantiate that the soybean genome sequence region delimited by A519F/R and M3Satt is correctly assembled and corresponds to the predicted *Rsv3*-residing region, six new markers were developed based on the regional sequence information. Four markers, S156a, S156b, S156c, and S156e, were microsatellite based and two (NB500pro1 and NB500pro2) were designed from the gene model Glyma14 g38500.1 (see below). The locations of the markers were confirmed by mapping in three populations: LS, LL, and TL. The microsatellite markers S156a and S156b cosegregated with *Rsv3* in the LS population (Fig. 1). The markers S156a, S156b, and S156c mapped 0.3 cM away from *Rsv3* and S156e cosegregated with *Rsv3* in the LL population (Fig. 1). In the TL population, markers S156b and S156c cosegregated with *Rsv3* (Fig. 1). Marker BARCSOYSSR_14_1417 (hereafter referred to as BS1417), which was in silico identified from the soybean genome sequence (Song et al., 2010), was mapped 0.3 cM away from *Rsv3* in the LS population and cosegregated with *Rsv3* in the LL population (Fig. 1).

Of the four NB-LRR genes located between A519F/R and M3Satt, Glyma14 g38500.1 was genetically mapped by using two markers generated from its promoter region. A set of primers was designed to PCR amplify the promoter and 5′-end-coding region (Supplemental Table S1). The PCR products amplified from the soybean parental lines using this primer set gave an expected size of 2.3 kbp and their end sequences were aligned, as expected, to the Glyma14 g38500.1 sequence region with greater than 99% similarity. Then, sequences of promoter parts of the PCR products from the parental lines L29 and Sowon were determined and then aligned. The comparison showed several single nucleotide polymorphic sites. Two of the polymorphic sites were used to generate the sequence-based markers NB500pro1 and NB500pro2 (Supplemental Table S2). The markers cosegregated with *Rsv3* in the LS population (Fig. 1).

Sequence Evaluation of the Chromosomal Regions Flanking the A519F/R to M3Satt Region through New Marker Development

Alignment of the soybean genome sequence north of the *Rsv3*-containing chromosomal region delimited by A519F/R

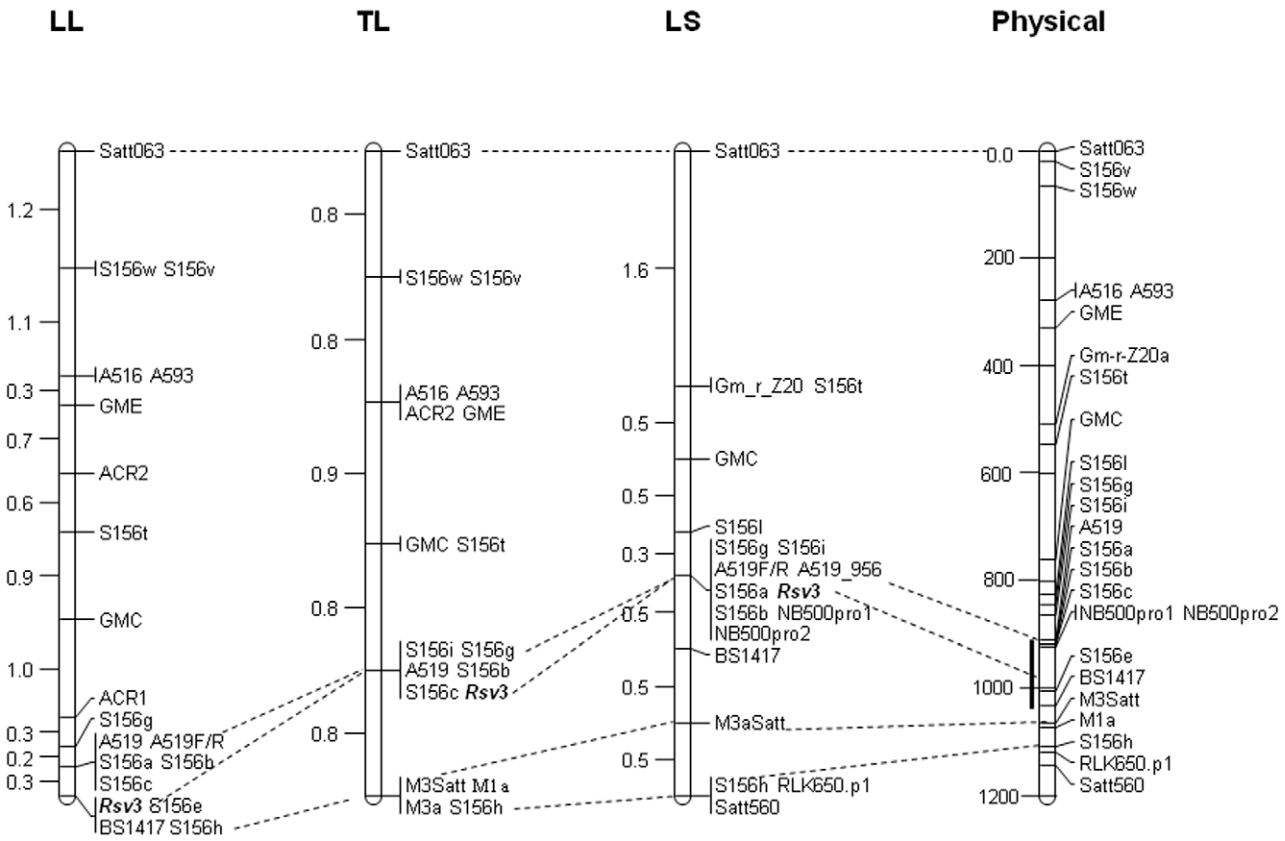

Figure 1. Genetic and sequence (physical) maps in the vicinity of the *Soybean mosaic virus* resistance gene, *Rsv3*, on the soybean chromosome 14 (molecular linkage group B2). Markers were mapped in the BC_3F_2 population 'L29' (*Rsv3*) × 'Sowon' (*rsv3*) (LS), the F_2 population L29 (*Rsv3*)' × 'Lee68' (*rsv3*) (LL), and the F_2 population 'Tousan 140' (*Rsv3*) × Lee68 (*rsv3*) (TL). The resultant genetic maps were aligned with each other and then were aligned with a map constructed from the positions of the markers on the sequence of soybean chromosome 14. Values on the left side of the genetic maps are map distances in cM. Values on the left side of the sequence (physical) map are physical distance from Satt063 in kbp. The bar on the left side of the sequence map indicates the *Rsv3* locus predicted on the basis of comparison of the current genetic maps. Satt063, A519-derived markers, *Rsv3*, M3Satt, and S156h are connected by dotted lines to show the *Rsv3*-containing chromosomal region.

and M3Satt was examined through mapping microsatellite markers generated from its corresponding sequence region. Three microsatellite markers, S156g, S156i, and S156l, were generated from the A519F/R-flanking sequence within 100 kbp (Supplemental Table S2). S156g and S156i cosegregated in the LS population with A519F/R, and S156l mapped 0.3 cM away from A519F/R (Fig. 1). S156g was 0.2 cM away from A519F/R in the LL population. Markers S156g and S156i cosegregated with A519 in the TL population. Additional microsatellite markers located over 100 kbp away from A519F/R were developed: GMC, S156t, GME, S156v, and S156w (Supplemental Table S2). All of the additional markers mapped in the three populations at the genetic locations predicted by the soybean genome sequence.

Alignment of the soybean genome sequence south of the *Rsv3*-containing chromosomal region delimited by A519F/R and M3Satt was examined by mapping microsatellite markers generated from its corresponding sequence region. One microsatellite marker, S156h, generated from the M3Satt-flanking sequence was within 100 kbp. Marker S156h mapped 0.5 cM south of M3Satt in the LS population, cosegregated with *Rsv3* in the LL population, and cosegregated with M3Satt in the TL population (Fig. 1). One of the LRR-RLK genes, Glyma14 g38650.1, which is located near the M1a-hit Glyma14 g38630.1 gene, was genetically mapped using a marker generated from one of its intron–exon junction regions. A set of primers was designed to PCR amplify the intron–exon region (Supplemental Table S1). Sequences of the PCR products from L29 and Sowon were aligned, as expected, to the Glyma14 g38650.1 sequence region with greater than 99% similarity. One sequence-based marker RLK650.p1, which was generated from a T/C single nucleotide polymorphism site between L29 and Sowon, cosegregated with S156h and Satt560 in the LS population (Fig. 1).

Construction of a Phylogenetic Tree of NB-LRR Genes

Examination of the list of the gene models for the sequence region between A519F/R and M3Satt on Gm14 indicated that the region contains four full-length NB-LRR genes and one NB-LRR pseudogene, which are members of the disease resistance gene superfamily (Table 1). An additional three NB-LRR genes were observed outside of the A519F/R to M3Satt sequence region (Fig. 2). The NB sequences of the NB-LRR proteins near or cosegregating with the *Rsv3* locus, the sequences BLAST-hit against the NB sequence database

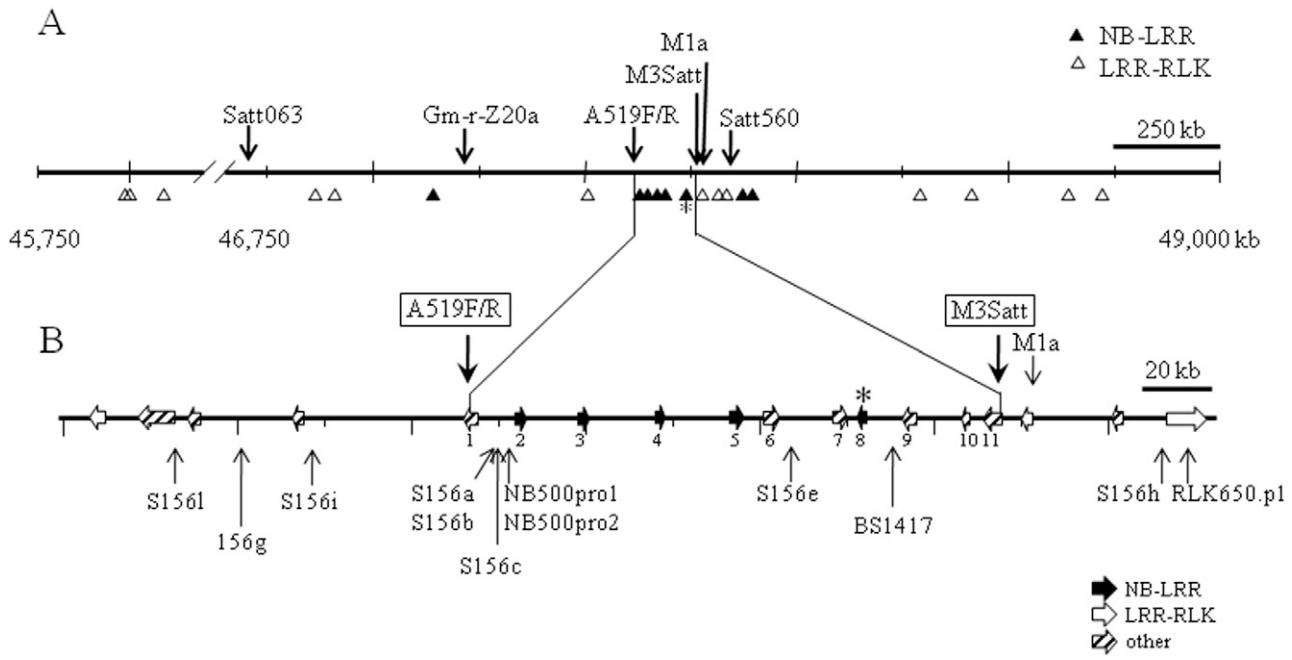

Figure 2. Sequence map of the soybean chromosome 14 in the vicinity of the soybean *Rsv3* gene presumed to be located between molecular markers A519F/R and M3Satt. (A) Sequence map and gene annotations of chromosome 14 between 45,750 and 49,000 kbp positions. Locations of markers are indicated above the chromosome line and locations of nucleotide-binding leucine-rich repeat (NB-LRR) (filled triangle) and leucine-rich repeat receptor-like kinase (LRR-RLK) (open triangle) genes are indicated below the line. Genes other than the NB-LRR and LRR-RLK genes are not presented. (B) Sequence map and gene annotations of the chromosome 14 between A519F/R and M3Satt. The predicted NB-LRR genes are indicated by filled rectangular arrows with orientations (from 5′ to 3′), the predicted LRR-RLK genes by open rectangular arrows, and the other genes by hatched rectangular arrows. Newly developed markers between A519F/R and M3Satt are indicated below the chromosome line. Numbers below the chromosome line are the annotated genes numbered consecutively from A519F/R to M3Satt. The five NB-LRR genes: 2, Glyma14 g38500.1; 3, Glyma14 g38510.1; 4, Glyma14 g38540.1; 5: Glyma14 g38560.1; 8, Glyma14 g38590.1 (pseudogene; highlighted by *). The other six genes: 1, Glyma14 g38490.1; 6, Glyma14 g38570.1; 7, Glyma14 g38580.1; 9, Glyma14 g38600.1; 10, Glyma14 g38610.1; 11, Glyma14 g38620.1.

created by McHale et al. (2006), and their close relatives were used to construct a gene tree. The previously described full-length soybean CC-NB-LRR proteins were also included to determine their relationships with the *Rsv3*-associated CC-NB-LRR proteins: 3gG2, 5gG3, and 6gG9 associated with the *Rsv1* locus on chromosome 13 (Hayes et al., 2004), Rpg1-b encoded by *Rpg1-b* on chromosome 13 (Ashfield et al., 2004), and Rps1-k-1 and Rps1-k-2 associated with the *Rps1-k* locus on chromosome 3 (Gao and Bhattacharyya, 2008). Overall amino-acid-sequence identity between the *Rsv3*-associated NB-LRR proteins and the previously described soybean CC-NB-LRR proteins was less than 25% and amino-acid-sequence identity between their N-terminal domains ranged from 20 to 50%. The NB sequences of the five NB-LRR genes (including the one pseudogene) found in the sequence between A519F/R and M3Satt are highly similar to each other and formed a subclade in the Neighbor-joining tree (Fig. 3). The subclade was designated as the *Rsv3*-associated NB in Fig. 3. The NB sequences of two (Glyma14 g38700.1 and Glyma14 g38740.1) of the three NB-LRR genes outside of the A519F/R to M3Satt sequence region are sisters to the *Rsv3*-associated NB subclade. Collectively, the seven NB sequences inside and outside of the A519F/R to M3Satt sequence region formed a well-supported monophyletic group with a bootstrap support of 98%. Branch lengths indicated that the *Rsv3*-associated NBs are probably the consequence of recent duplications. We defined the clade as the GmCC-NB I (the red box in Fig. 3). The gene model Glyma14 g37860.1 is an outlier out of the three NB-LRR genes outside of the A519F/R to M3Satt sequence region (below the red box in Fig. 3). In BLAST searches, N-terminal and C-terminal parts of the NB domain in Glyma14 g37860.1 best hit, respectively, different groups of NB sequences that belong to two distantly related nonlegume clades, in the phylogenetic tree of McHale et al. (2006). The best-hit sequences (GenBank gene identification [GI] number 16933577 and GenBank GI number 53680944) for each of the two groups were included in the present phylogenetic tree. The six previously described full-length-cloned soybean CC-NB-LRR proteins formed a monophyletic clade, which was defined as the GmCC-NB II (the blue box in Fig. 3). Thus, the results from phylogenetic analysis of NB-LRR genes indicate that the *Rsv3*-associated CC-NB-LRR genes appear to be members of a novel CC-NB-LRR class that has not been functionally characterized in soybean.

The Cluster of NB-LRR Genes Cosegregating with *Rsv3* is Located in the Middle of an LRR-RLK Gene Cluster

The restriction fragment length polymorphism marker M1a is tightly linked to *Rsv3* (TL in Fig. 1), and the end sequence of M1a contains the extracellular LRR domain (Jeong et al., 2002). Because the consensus sequence of the M1a LRR is

Table 1. Gene and marker annotations of soybean chromosome 14 between 45,750,000 bp and 49,000,000 bp[†].

Gene name	Marker name	Position	Gene annotation
Glyma14g36630.1		45988517..45993144	Leucine-rich repeat receptor-like kinase
Glyma14g36660.1		46001094..46004267	Leucine-rich repeat receptor-like kinase
Glyma14g36810.1		46091207..46093272	Leucine-rich repeat receptor-like kinase
	Satt063	46705813..46705956	
	S156v	46726494..46727093	
	S156w	46772356..46772567	
Glyma14g37590.1		46863817..46868269	Leucine-rich repeat receptor-like kinase
Glyma14g37630.1		46908176..46914671	Leucine-rich repeat receptor-like kinase
	A516	46983593..46984402	
	A593	46983608..46984399	
	GME	47034693..47035079	
Glyma14g37860.1		47141015..47143938	Coiled-coil nucleotide-binding leucine-rich repeat protein
	Gm-r-Z20a	47214338..47215135	
	S156t	47253102..47253291	
	GMC	47466562..47466808	
Glyma14g38390.1		47508501..47512747	Leucine-rich repeat receptor-like kinase
Glyma14g38420.1		47522200..47532508	Nucleotidyltransferase, putative
	S156l	47531967..47532194	
Glyma14g38430.1		47535945..47538954	Expansin 45, endoglucanase-like
	S156g	47550957..47551193	
Glyma14g38460.1		47566439..47569354	Transcription factor RF2b, putative
	S156i	47570884..47571083	
Glyma14g38490.1		47615588..47619943	Transcriptional factor B3, putative
	A519F/R	47616322..47617670	
	S156a	47623366..47623586	
	S156b	47623609..47623789	
	S156c	47624441..47624659	
	NB500pro1	47628799..47628967	
	NB500pro2	47628799..47628967	
Glyma14g38500.1		47630519..47633501	Coiled-coil nucleotide-binding leucine-rich repeat protein
Glyma14g38510.1		47648153..47651727	Coiled-coil nucleotide-binding leucine-rich repeat protein
Glyma14g38540.1		47670152..47672833	Coiled-coil nucleotide-binding leucine-rich repeat protein
Glyma14g38560.1		47691826..47695095	Coiled-coil nucleotide-binding leucine-rich repeat protein
Glyma14g38570.1		47701161..47706197	DNA double-strand break repair RAD50 ATPase
	S156e	47709881..47710153	
Glyma14g38580.1		47721318..47725350	Cinnamate 4-hydroxylase, putative
Glyma14g38590.1		47728270..47730998	Coiled-coil nucleotide-binding leucine-rich repeat protein (partial pseudogene)
	BarcSoySSR_14_1417	47738893..47739065	
Glyma14g38600.1		47741179..47745790	Translation initiation factor IF5
Glyma14g38610.1		47758657..47760317	AP2 domain transcription factor
Glyma14g38620.1		47764961..47770195	Ubiquitin-conjugating enzyme E2, putative
	M3Satt	47769743..47770460	
Glyma14g38630.1		47775271..47778947	Leucine-rich repeat receptor-like kinase
	M1a	47778362..47779054	
	S156h	47814988..47815220	
Glyma14g38640.1		47800987..47804143	Root phototropism protein, putative
Glyma14g38650.1		47816591..47827883	Leucine-rich repeat receptor-like kinase
	RLK650.p1	47823861..47824095	
Glyma14g38670.1		47836300..47846561	Leucine-rich repeat receptor-like kinase
	Satt560	47849427..47849504	
Glyma14g38700.1		47872961..47876633	Coiled-coil nucleotide-binding leucine-rich repeat protein
Glyma14g38740.1		47897628..47899939	Coiled-coil nucleotide-binding leucine-rich repeat protein
Glyma14g39180.1		48293479..48297011	Leucine-rich repeat receptor-like kinase
Glyma14g39290.1		48411264..48415224	Leucine-rich repeat receptor-like kinase
Glyma14g39550.1		48641632..48644930	Leucine-rich repeat receptor-like kinase
Glyma14g39690.1		48721778..48724358	Leucine-rich repeat receptor-like kinase

[†]Gene annotation information was retrieved from the soybean gene annotation database, Glyma1.0 (accessible at Phytozome v5.0, http://www.phytozome.net [verified 13 Jan. 2011]). Only nucleotide-binding leucine-rich repeat (NB-LRR) and leucine-rich repeat receptor-like kinase (LRR-RLK) genes are presented for the chromosomal region outside of the region delimited by S1561 and RLK650.p1.

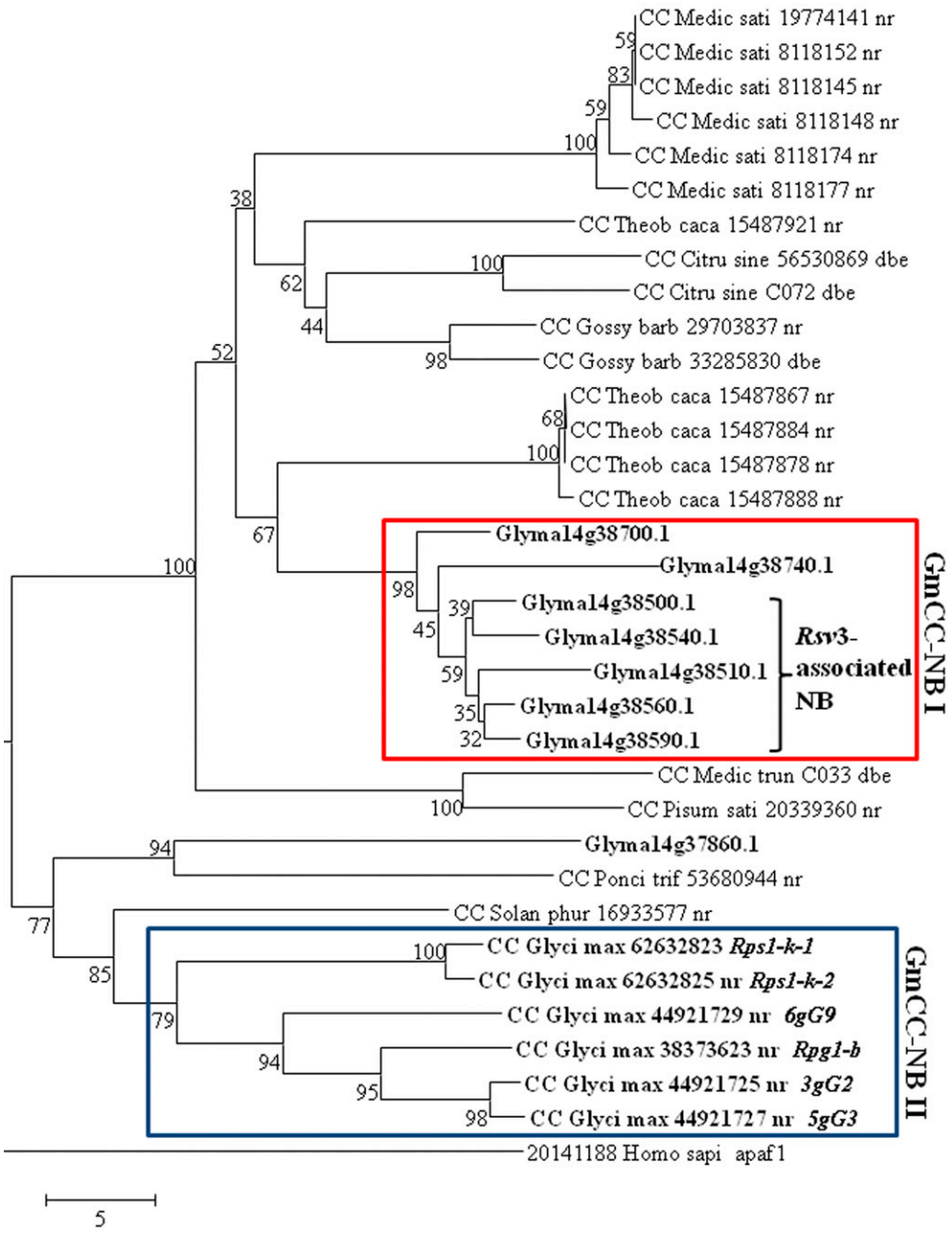

Figure 3. Neighbor-joining phylogenetic tree of nucleotide-binding leucine-rich repeat (NB-LRR) genes. The tree was constructed using only the nucleotide-binding (NB) sequences. The NB sequence of NB-LRR proteins near or cosegregating with the *Rsv3* locus, the hit sequences and their close relatives retrieved from BLASTP searches against the 1600 NB sequences collected by McHale et al. (2006) with the NB-LRR proteins in the vicinity of the *Rsv3* locus, and the full-length-cloned soybean coiled-coil nucleotide-binding leucine-rich repeat (CC-NB-LRR) proteins (3 gG2, 5 gG3, 6 gG9, Rpg1-b, Rps1-k-1, and Rps1-k-2) were used to construct a gene tree. The designations of sequences except the soybean sequences are as used by McHale et al. (2006). The soybean sequences are in bold. The tree was rooted using human APAF1 protein. The scale bar represents five amino acid differences. The boxed regions represent two soybean CC-NB-LRR subclades GmCC-NB I and GmCC-NB II. Medic sati, *Medicago sativa* L.; Theob caca, *Theobroma cacao* L.; Citru sine, *Citrus sinensis* (L.) Osbeck; Gossy barb, *Gossypium barbadense* L.; Pisum sati, *Pisum sativum* L.; Ponci trif, *Citrus trifoliata* L.; Solan phur, *Solanum phureja* × *Solanum stenotomum*; Glyci max, *Glycine max* (L.) Merr.; Homo sapi, *Homo sapiens*.

identical to that of the extracellular LRR resistance genes *Cf-9* and *Xa21* (Jones et al., 1994; Song et al., 1995) and because the clustering of disease resistance genes has been reported in many plants (Michelmore and Meyers, 1998), it was hypothesized that the *Rsv3* region might contain a member of the extracellular LRR class of disease resistance genes (Jeong et al., 2002). The BLASTN search of the M1a end sequence (GenBank accession no. AF348333) against the soybean genome sequence identified significantly several chromosomal regions, including Gm11 (32.6 Mbp position), Gm18 (4.4 Mbp), Gm02 (45.6 Mbp), Gm14 (47.8 Mbp), and Gm20 (23.1 Mbp), with an E value smaller than $1.3e^{-113}$. The M1a sequence showed the highest similarity (99.9% identity) to a part of Gm11 and 80.6% similarity to a part of Gm14. The

Table 2. Comparison of the 13 soybean leucine-rich repeat receptor-like kinase (LRR-RLK) genes in the vicinity of *Rsv3* to their *Arabidopsis thaliana* (L.) Heynh. homologues.

Query sequence	Subject sequence†	E value	Subfamily‡
Glyma14g36630	AT5G58300	0	LRR III
Glyma14g36660	AT1G72180	$3e^{-20}$	LRR XI
Glyma14g36810	AT1G56140	$3e^{-40}$	LRR VIII-2
Glyma14g37590	AT1G74360	$3e^{-20}$	LRR X
Glyma14g37630	AT4G18640	e^{-114}	LRR VI
Glyma14g38390	AT4G29450	$4e^{-10}$	LRR I
Glyma14g38630	AT5G58300	0	LRR III
Glyma14g38650	AT1G06840	0	LRR VIII-1
Glyma14g38670	AT1G06840	0	LRR VIII-1
Glyma14g39180	AT1G56145	$6e^{-71}$	LRR VIII-2
Glyma14g39290	AT1G66150	0	LRR IX
Glyma14g39550	AT1G48480	0	LRR III
Glyma14g39690	AT1G56120	$7e^{-52}$	LRR VIII-2

†*A. thaliana* gene that showed the best match in a TBLASTN search against the local protein database, which contains 194 Arabidopsis LRR-RLK sequences reported by Gou et al. (2010), using a predicted soybean LRR-RLK amino acid sequence.

‡Name of subfamily in the phylogenetic tree constructed using the 194 *A. thaliana* LRR-RLK sequences, to which the *A. thaliana* gene homologous to the soybean gene belongs.

M1a-hit Gm14 sequence was located near the M3Satt locus and outside of the sequence region delimited by M3Satt and A519F/R (Fig. 2A and 2B). The M1a-containing full-length gene (Glyma14 g38630.1; Glyma1.0 release at http://www.phytozome.net [verified 13 Jan. 2011]) is a member of the LRR-RLK gene family. Interestingly, members of this gene family are repeated 13 times in the vicinity of the *Rsv3* locus (Fig. 2; Table 1), thereby supporting the previous thought that extracellular LRR domain-containing sequences might be clustered in this region of the chromosome (Jeong et al., 2002). However, none of these genes appears to be located between M3Satt and A519F/R. The predicted full-length amino acid sequences of the 13 LRR-RLK genes were compared using TBLASTN searches against a local protein database, which contains 194 *A. thaliana* LRR-RLK sequences reported by Gou et al. (2010). All the sequences hit the *A. thaliana* LRR-RLK sequences with an E value smaller than $3e^{-20}$. However, the *G. max* sequences dispersed into eight subfamilies of the phylogenetic tree constructed using the 194 *A. thaliana* LRR-RLK sequences (Table 2).

Discussion

More than 40 *R* genes have been functionally characterized over the past two decades, the majority of which belong to the NB-LRR family (Lukasik and Takken, 2009). The NB-LRR genes tend to cluster in many plant genomes (Michelmore and Meyers, 1998). In this study, we showed that a cluster of the four NB-LRR genes is cosegregating with the *Rsv3* locus in our three mapping populations segregating for *Rsv3*. Despite the lack of physical mapping, parallel alignment between the genetic maps (constructed using public and novel markers) and the genome sequence map (constructed by placing the marker sequences on the soybean genome sequence) are strong evidence that the NB-LRR genes or their variants (as Williams 82 is SMV susceptible) are candidate(s) for *Rsv3*. Furthermore, none of the other types of genes in the sequence region delimited by A519F/R and M3Satt have been reported to be involved in classical disease resistance response mechanisms (Table 1). Although some members of the LRR-RLK gene family have been reported to be disease resistance genes (Parniske and Jones, 1999; Song et al., 1997), our genetic mapping results indicated that these sequences are outside of the sequence region delimited by A519F/R and M3Satt and are unlikely candidate genes for *Rsv3*.

The *Rsv3*-residing chromosomal region is of great interest with respect to the evolution of multigene clusters because members of the CC-NB-LRR and LRR-RLK multigene families constitute a heterogeneous cluster. Nucleotide-binding leucine-rich repeat (NB-LRR) or LRR-RLK genes often occur in clusters that consist of several copies of homologous gene sequences arising from a single gene subfamily (simple clusters) or colocalized gene sequences derived from two or more unrelated subfamilies (complex clusters) and may also contain unrelated single genes interspersed between the homologs (Shiu and Bleecker, 2001; Friedman and Baker, 2007). It has been suggested that intergenic unequal crossover and intragenic mispairing contribute to altered gene copy number within the cluster (e.g., Parniske and Jones, 1999; Kuang et al., 2004, 2005; for a review, see Friedman and Baker, 2007). Although it has been reported that a single *Prf* gene, a member of the NB-LRR superfamily, is embedded within a cluster of five *Pto* kinase homologs (Salmeron et al., 1996), colocalization of multiple members of both of the two disease resistance gene superfamilies has not been reported to the best of our knowledge. Surprisingly, the members of the two gene superfamilies appear to be interspersed with each other in the *Rsv3*-residing chromosomal region on Gm14. Our results suggest that the NB-LRR genes in the *Rsv3*-containing chromosomal region likely arose from a single gene subfamily (Fig. 3) and that the LRR-RLK genes likely arose from eight unrelated subfamilies reported by Gou et al. (2010) (Table 2). Feature(s) of this chromosomal region that resulted in the coevolution of the two gene clusters at the same chromosomal region or interactions between the two gene clusters during the evolution remain unclear. Interestingly, our examination of the duplication blocks at the SoyBase Browser (http://soybase.org [verified 13 Jan. 2011]) indicated that only LRR-RLK genes cluster on the region on chromosome 2 that is homeologous to the *Rsv3* region (data not shown). Although our genetic analysis clearly excluded an *Rsv3* candidacy of LRR-RLK genes, the possibility of interaction between NB-LRR and LRR-RLK proteins in conferring resistance to SMV at the *Rsv3* locus cannot be dismissed in light of the case of the *Prf* and *Pto* interaction (Salmeron et al., 1996). Further analyses including crosses and sequence comparison between soybean cultivars and heterologous expression will help resolve these issues.

The LRR domain structure in the NB-LRR genes would lead to the expectation that, in determining recognition specificity, either Avr proteins or recognition cofactors bind to this

domain. At the same time, an accumulating body of evidence suggests that the N-terminal domains of NB-LRR proteins also play a role in Avr recognition (reviewed by Collier and Moffett, 2009). *Rsv1*, one of the three known SMV resistance genes, is associated with the CC-NB-LRR gene cluster on the soybean chromosome 13 (MLG F) (Jeong et al., 2001; Hayes et al., 2004), the NB domains of which belong to the CC-NB-LRR superfamily (Ashfield et al., 2004). Because the findings of this study suggest that *Rsv3* may also encode a member of the CC-NB-LRR gene family, which is distantly related to the *Rsv1* locus-associated CC-NB-LRR genes, it is hypothesized that the different disease responses of *Rsv1* and *Rsv3* against a spectrum of SMV strains may be due to the different structures of the CC-NB-LRR genes. It will be interesting to further elucidate what structural difference between the *Rsv1*-associated and *Rsv3*-associated CC-NB-LRR genes make the *Rsv1* and *Rsv3* genes confer different disease responses to the same SMV strain.

Recently the elicitors or pathogenic determinants governing the disease reactions of *Rsv1*- or *Rsv3*-genotype soybeans to different SMV strains have been identified. Chimeric clones, constructed by exchanging genomic sequences from virulent and avirulent SMV strains, were used to map the region within the SMV genome that induces a defense response in *Rsv1*- and *Rsv3*-containing cultivars. It was shown that the helper component-protease (HC-Pro) and P3 proteins are independently recognized by *Rsv1*-genotype soybean and elicit the extreme resistance phenotype (Hajimorad et al., 2006; Eggenberger et al., 2008). Similarly, the cytoplasmic inclusion (CI) protein was shown to be the elicitor or pathogenic determinant recognized by *Rsv3*-genotype soybean (Seo et al., 2009; Zhang et al., 2009) and that a single amino acid substitution was responsible for strains that can avoid *Rsv3* recognition to become avirulent (Seo et al., 2009). Thus, two lines of evidence, phylogenetic difference between *Rsv1* and *Rsv3* shown in this work and distinct viral genomic regions as elicitors and as determinants of pathogenicity, demonstrate the complexity of the interactions in the soybean–SMV pathosystem that are yet to be elucidated.

Alleles of *R* genes confer different resistance reactions to pathogens (e.g., Jones and Dangl, 2006). For example, when inoculated with SMV strain G1, *Rsv1-n*, an allele of *Rsv1*, confers a severe or lethal necrotic reaction that is a typical resistance reaction but is not desirable in a commercial cultivar (Tucker et al., 2009). To remove or to replace these genes in a soybean cultivar, markers located in the middle of the resistance gene clusters, which span over several centiMorgans in many cases, would be essential for marker-assisted selection of desirable line(s) from a breeding population. Among the three known *R* genes conferring resistance to SMV, several studies have developed molecular markers tightly linked to *Rsv1* (e.g., Gore et al., 2002; Shi et al., 2010) and *Rsv4* (Hwang et al., 2006; Saghai Maroof et al., 2010). In particular, Saghai Maroof et al. (2010) used the soybean genome sequence to develop numerous *Rsv4*-linked molecular markers as well as to show that the *Rsv4* gene likely belongs to a new class of resistance genes. The molecular markers developed in this study, and prediction of the molecular nature of the *Rsv3* gene, should provide additional tools for pyramiding SMV-resistance genes to obtain durable SMV-resistant soybean cultivars and for elucidating the structure and function of the *Rsv3* gene.

Acknowledgments

This work was supported by a grant from the BioGreen 21 Project (code no. 20080401034011), Rural Development Administration, the Republic of Korea, and, in part, by the KRIBB Research Initiative Program. The work was also supported in part by the Virginia Agricultural Experiment Station. We thank two anonymous reviewers for their constructive comments on an earlier version of the manuscript.

References

Ashfield, T., L.E. Ong, K. Nobuta, C.M. Schneider, and R.W. Innes. 2004. Convergent evolution of disease resistance gene specificity in two flowering plant families. Plant Cell 16:309–318.

Bernard, R.L., R.I. Nelson, and C.R. Cremeens. 1991. USDA soybean genetic collection: Isoline collection. Soybean Genet. Newsl. 18:27–57.

Bowers, G.R., E.H. Paschal, R.L. Bernard, and R.M. Goodman. 1992. Inheritance of resistance to soybean mosaic virus in 'Buffalo' and HLS soybean. Crop Sci. 32:67–72.

Buzzell, R.I., and J.C. Tu. 1989. Inheritance of soybean stem-tip necrosis reaction to soybean mosaic virus. J. Hered. 80:400–401.

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.

Cho, E.K., and R.M. Goodman. 1979. Strains of soybean mosaic virus: Classification based on virulence in resistant soybean cultivars. Phytopathology 69:469–470.

Collier, S.M., and P. Moffett. 2009. NB-LRRs work a "bait and switch" on pathogens. Trends Plant Sci. 14:521–529.

Eggenberger, A.L., M.R. Hajimorad, and J.H. Hill. 2008. Gain of virulence on *Rsv*1-genotype soybean by an avirulent *Soybean mosaic virus* requires concurrent mutations in both P3 and HC-Pro. Mol. Plant Microbe Interact. 21:206–210.

Friedman, A.R., and B.J. Baker. 2007. The evolution of resistance genes in multi-protein plant resistance systems. Curr. Opin. Genet. Dev. 17:493–499.

Gao, H., and M.K. Bhattacharyya. 2008. The soybean-*Phytophthora* resistance locus *Rps1*-k encompasses coiled coil-nucleotide binding-leucine rich repeat-like genes and repetitive sequences. BMC Plant Biol. 8:29.

Gore, M.A., A.J. Hayes, S.C. Jeong, Y.G. Yue, G.R. Buss, and M.A. Saghai Maroof. 2002. Mapping tightly linked genes controlling potyvirus infection at the *Rsv*1 and *Rpv*1 region in soybean. Genome 45:592–599.

Gou, X., K. He, H. Yang, T. Yuan, H. Lin, S.D. Clouse, and J. Li. 2010. Genome-wide cloning and sequence analysis of leucine-rich repeat receptor-like protein kinase genes in *Arabidopsis thaliana*. BMC Genomics 11:19.

Gunduz, I., G.R. Buss, P. Chen, and S.A. Tolin. 2002. Characterization of SMV resistance genes in Tousan 140 and Hourei soybean. Crop Sci. 42:90–95.

Gunduz, I., G.R. Buss, G. Ma, P. Chen, and S.A. Tolin. 2000. Genetic analysis of resistance to *Soybean mosaic virus* in OX670 and Harosoy soybean. Crop Sci. 41:1785–1791.

Hajimorad, M.R., A.L. Eggenberger, and J.H. Hill. 2006. Strain-specific P3 of Soybean mosaic virus elicits *Rsv*1-mediated extreme resistance, but absence of P3 elicitor function alone is insufficient for virulence on *Rsv*1-genotype soybean. Virology 345:156–166.

Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41:95–98.

Hammond-Kosack, K.E., and J.D. Jones. 1996. Resistance gene-dependent plant defense responses. Plant Cell 8:1773–1791.

Hayes, A.J., S.C. Jeong, M.A. Gore, Y.G. Yu, G.R. Buss, S.A. Tolin, and M.A. Saghai 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.

Hwang, T.-Y., J.-K. Moon, S. Yu, K. Yang, S. Mohankumar, Y.H. Yu, Y.H. Lee, H.S. Kim, H.M. Kim, M.A. Saghai Maroof, and S.-C. Jeong. 2006. Application of comparative genomics in developing molecular markers tightly linked to the virus resistance gene *Rsv*4 in soybean. Genome 49:380–388.

Innes, R.W., C. Ameline-Torregrosa, T. Ashfield, E. Cannon, S.B. Cannon, B. Chacko, N.W.G. Chen, A. Couloux, A. Dalwani, R. Denny, S. Deshpande, A.N. Egan, N. Glover, C.S. Hans, S. Howell, D. Ilut, S. Jackson, H. Lai, J. Mammadov, S. Martin del Campo, M. Metcalf, A. Nguyen, M. O'Bleness, B. E. Pfeil, R. Podicheti, M.B. Ratnaparkhe, S. Samain, I. Sanders, B. Ségurens, M. Sévignac, S. Sherman-Broyles, V. Thareau, D.M. Tucker, J. Walling, A. Wawrzynski, J. Yi, J.J. Doyle, V. Geffroy, B.A. Roe, M.A. Saghai Maroof, and N.D. Young. 2008. Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol. 148:1740–1759.

Jackson, S.A., D. Rokhsar, Gary Stacey, R.C. Shoemaker, J. Schmutz, and J. Grimwood. 2006. Toward a reference sequence of the soybean genome: A multiagency effort. Crop Sci. 46:S55–S66.

Jeong, S.C., A.J. Hayes, R.M. Biyashev, and M.A. Saghai Maroof. 2001. Diversity and evolution of a non-TIR-NBS sequence family that clusters to a chromosomal "hotspot" for disease resistance genes in soybean. Theor. Appl. Genet. 103:406–414.

Jeong, S.C., S. Kristipati, A.J. Hayes, P.J. Maughan, S.L. Noffsinger, I. Gunduz, G.R. Buss, and M.A. Saghai Maroof. 2002. Genetic and sequence analysis of markers tightly linked to the *soybean mosaic virus* resistance gene, *Rsv*3. Crop Sci. 42:265–270.

Jeong, S.C., and M.A. Saghai Maroof. 2004. Detection and genotyping of SNPs tightly linked to two disease resistance loci, *Rsv*1 and *Rsv*3, of soybean. Plant Breed. 123:305–310.

Jones, D.A., C.M. Thomas, K.E. Hammond-Kosack, P.J. Balint-Kurti, and J.D.G. Jones. 1994. Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266:789–793.

Jones, J.D., and J.L. Dangl. 2006. The plant immune system. Nature 444:323–329.

Kuang, H., F. Wei, M.R. Marano, U. Wirtz, X. Wang, J. Liu, W.P. Shum, J. Zaborsky, L.J. Tallon, W. Rensink, S. Lobst, P. Zhang, C.-E. Tornqvist, A. Tek, J. Bamberg, J. Helgeson, W. Fry, F. You, M.-C. Luo, J. Jiang, C.R. Buell, and B. Baker. 2005. The *R1* resistance gene cluster contains three groups of independently evolving, type I R1 homologues and shows substantial structural variation among haplotypes of *Solanum demissum*. Plant J. 44:37–51.

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 16:2870–2894.

Kumar, S., M. Nei, J. Dudley, and K. Tamura. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. 9:299–306.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, J.M. Daly, S.E. Lincoln, and L. Newberg. 1987. Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Lukasik, E., and F.L.W. Takken. 2009. STANDing strong, resistance proteins instigators of plant defense. Curr. Opin. Plant Biol. 12:427–436.

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.

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

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.

Parniske, M., and J.D. Jones. 1999. Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc. Natl. Acad. Sci. USA 96:5850–5855.

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.

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, *Rsv*4. Plant Gen. 3:14–22.

Salmeron, J.M., G.E. Oldroyd, C.M. Rommens, S.R. Scofield, H.S. Kim, D.T. Lavelle, D. Dahlbeck, and B.J. Staskawicz. 1996. Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86:123–133.

Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.C. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, and S.A. Jackson. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178–183.

Seo, J.K., S.-H. Lee, and K.-H. Kim. 2009. Strain-specific cylindrical inclusion protein of *Soybean mosaic virus* elicits extreme resistance and a lethal systemic hypersensitive response in two resistant soybean cultivars. Mol. Plant Microbe Interact. 22:1151–1159.

Shi, A., P. Chen, R. Vierling, C. Zheng, D. Li, D. Dong, E. Shakiba, and I. Cervantez. 2010. Multiplex single nucleotide polymorphism (SNP) assay for detection of soybean mosaic virus resistance genes in soybean. Theor. Appl. Genet. doi:10.1007/s00122-010-1459-6.

Shiu, S.-H., and A.B. Bleecker. 2001. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. Proc. Natl. Acad. Sci. USA 98:10763–10768.

Song, Q., G. Jia, Y. Zhu, D. Grant, R.T. Nelson, E.Y. Hwang, D.L. Hyten, and P.B. Cregan. 2010. Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci. 50:1950–1960.

Song, W.Y., L.Y. Pi, G.L. Wang, J. Gardner, T. Holsten, and R.C. Ronald. 1997. Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9:1279–1287.

Song, W.-Y., G.-L. Wang, L.-L. Chen, H.-S. Kim, L.-Y. Pi, T. Holsten, J. Gardner, B. Wang, W.-X. Zhai, L.-H. Zhu, C. Fauquet, and P. Roland. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804–1806.

Stein, L.D., C. Mungall, S.Q. Shu, M. Caudy, M. Mangone, A. Day, E. Nickerson, J.E. Stajich, T.W. Harris, A. Arva, and S. Lewis. 2002. The generic genome browser: A building block for a model organism system database. Genome Res. 12:1599–1610.

Tu, J.C., and R.I. Buzzell. 1987. Stem-tip necrosis: A hypersensitive, temperature dependent, dominant gene reaction of soybean to infection by soybean mosaic virus. Can. J. Plant Sci. 67:661–665.

Tucker, D.M., M.A. Saghai 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.

Wawrzynski, A., T. Ashfield, N.W.G. Chen, J. Mammadov, A. Nguyen, R. Podicheti, S.B. Cannon, V. Thareau, C. Ameline-Torregrosa, E. Cannon, B. Chacko, A. Couloux, A. Dalwani, R. Denny, S. Deshpande, A.N. Egan, N. Glover, S. Howell, D. Ilut, H. Lai, S. Martin del Campo, M. Metcalf, M. O'Bleness, B.E. Pfeil, M.B. Ratnaparkhe, S. Samain, I. Anders, B. Ségurens, M. Sévignac, S. Sherman-Broyles, D.M. Tucker, J. Yi, J.J. Doyle, V. Geffroy, B.A. Roe, M.A. Saghai Maroof, N.D. Young, and R.W. Innes. 2008. Recent activity of both autonomous and non-autonomous retrotransposons continues to shape the soybean genome. Plant Physiol. 148:1760–1771.

Yu, S., J.-K. Moon, T.-Y. Hwang, Y.-H. Kim, H.M. Kim, and S.C. Jeong. 2005. Genetic and sequence analysis of the RAPD fingerprint containing a *Rsv*3-linked marker rapidly identified by using RILs in soybean. Kor. J. Genet. 27:127–132.

Zhang, C., M.R. Hajimorad, A.L. Eggenberger, S. Tsang, S.A. Whitham, and J.H. Hill. 2009. Cytoplasmic inclusion cistron of *Soybean mosaic virus* serves as a virulence determinant on *Rsv*3-genotype soybean and a symptom determinant. Virology 391:240–248.

Supplemental Material

Supplemental Table 1. Primers designed to amplify parts of Glyma14g38500.1 and Glyma14g38650.1.

Target	Primer specificity	Primer sequence (5'->3')
Promoter and 5' region	Forward	TCCAATCTGCAACGTCAACT
of Glyma14g38500.1	Reverse	TGTAATTGCAGAGTCCTAAT
Promoter of	Forward	TTCAGTTTTGTCTTTTATCC
Glyma14g38500.1	Reverse	TGTAATTGCAGAGTCCTAAT
An intron-exon of	Forward	CTGGATGAGTTTACATCAAT
Glyma14g38650.1	Reverse	GCGTACTCTCAAAATAAGAA

Supplemental Table 2. Attributes of microsatellite and single-nucleotide polymorphism markers generated from the *Rsv*3-containing genomic DNA sequence.

Marker name	Chromosome range of template sequence	Description of template sequence and accession number[†]	Type of polymorphism	Predicted product size (bp)[†]	Primer specificity[‡] and sequence (5'→3')
S156a	47623366..47623586	A microsatellite site between A519F/R and M3Satt	AT repeat	240 (W)	F AAATTGAGAATTAGACCAAA
					R GGGGTATATCTTAATTGGTT
S156b	47623609..47623789	A microsatellite site between A519F/R and M3Satt	AT repeat	181 (W)	F GCGGTTCGTTTATTATTAC
					R TATTGCTTTCATCACTTACC
S156c	47624441..47624659	A microsatellite site between A519F/R and M3Satt	AT repeat	219 (W)	F TATTATGATGAATTCCAAGG
					R AGTGCTCTAAAACTGAGACA
S156e	47709881..47710153	A microsatellite site between A519F/R and M3Satt	AT repeat	296 (W)	F GCGTATTGGAATTAAGCTAAAAA
					R GCGATATGTCAGTGGAAAATCAG
S156g	47550957..47550957	A microsatellite site north of	AT repeat	237(W)	F TCTAGTTGCCAATAATTTTC

	51193	A519F/R			R	TTGAGGATAAGTGTGTTTTC
S156i	47570884..47571083	A microsatellite site north of A519F/R	AT repeat	200(W)	F	ATGAAAGCGATGTAAACTAA
					R	TGATGTGCATAAATGATAAA
S156l	47531967..47532194	A microsatellite site north of A519F/R	GA repeat	228(W)	F	CAAATTCGTGTCCAATACCT
					R	GTCCAGAGACTCATTCGGTA
S156h	47814988..47815220	A microsatellite site south of M3Satt	AT repeat	233(W)	F	AAAATAAAATGTCCTTGAGAAGAAAT
					R	GGCCCAAGTGAAGTTTGACA
S156t	47253102..47253291	A microsatellite site north of A519F/R	AT repeat	190(W)	F	GAAAGGTTTGGTGTGGTTCAA
					R	AAGCCCTTAGCTTGTAGTTTGG
S156v	46726494..46727093	A microsatellite site north of A519F/R	TAA repeat	ND	F	CTCAACCTGTAAGTTGGGTGA
					R	CCGGCCACTTATTTAACCTG
S156w	46772356..46772567	A microsatellite site north of A519F/R	GA repeat	212(W)	F	TTGTGTACGTTGTGAAACATGC
					R	GACCGATTCATCTCCCAAAA
GMC	47466562..47466808	A microsatellite site north of A519F/R	AT repeat	247(W)	F	GATTTGCTGTACAGAATCATGAAAA
					R	AAACCATCTGTTGCCACCAT

GME	47034693..47035079	A microsatellite site north of A519F/R	AT repeat	387(W)	F TCAAGTTACGGGTAGTTTGTTGG R CCAAATTTCTTACTAGGCGATCC
NB500pro1	47628799..47628967	Promoter and 5' coding region of Glyma14g38500.1; HR298283 (L) and HR298285 (S)	SNP G(S)/T(L)	169	F S-sp AATGCTACTACATTGAATACTGG F L-sp AATGCTACTACATTGAATACTGT Com R GGATCAAAATTACGGATCAA
NB500pro2	47628799..47628967	Promoter and 5' coding region of Glyma14g38500.1; HR298283 (L) and HR298285 (S)	SNP G(S)/T(L)	181	F S-sp CTATATTGATGGAAACACATGC F L-sp CTATATTGATGGAAACACATGT Com R TCACATGATCAAAAAGTGACA
RLK650.p1	47823861..47824095	Intron-exon region of Glyma14g38650.1; HR298284 (L) and HR298286 (S)	SNP T(L)/C(S)	235	F L-sp GCTGTATAATATAAACTGTAAT F S-sp GCTGTATAATATAAACTGTCGC Com R GCGTACTCTCAAAATAAGAA

[†]L, L29; S, Sowon; W, Williams 82; ND, not determined.

[‡]F, forward; R, reverse; Com, common; sp, specific.

Chapter Three

Additional Information and Future Research into *Rsv3* Resistance

Abstract

Rsv3 has recently been mapped to a 154 kb physical region and four full length candidate resistance genes have been identified. The populations used in the most recent mapping effort had multiple molecular markers co-segregate with *Rsv3*. Development of populations which would allow for further fine mapping, sequencing and additional marker development, could help to narrow the physical distance in which *Rsv3* is found and possibly designating a single gene as the source of resistance. It was also shown that the candidate genes residing at the *Rsv3* locus are unique to previously cloned soybean resistance genes. It is believed that soybean has undergone two rounds of whole genome duplication. Comparison of syntenic regions between soybean chromosomes and the related legume *Medicago*, should give insight into when the *Rsv3* locus formed in the soybean genome.

Introduction

Extreme Resistance and Limited Virus Movement

There are four sources for *Rsv3* resistance, Hardee, Tousan140, Columbia, and Hourei (Gunduz et al., 2001; Gunduz et al., 2002; Ma et al., 2002). The cultivar L29 (*Rsv3*) is a Williams near-isogenic line of Hardee (Buss et al., 1999). Fayad (2003) was able to distinguish

two unique alleles of *Rsv3,* using morphological and serological assays. It was shown that cultivar Harosoy (*Rsv3*) showed local infection restricted the movement of SMV strains G6 and G7, in contrast to L29 (*Rsv3*) which showed extreme resistance. Upon inoculation with the two highest numbered strains of SMV (SMV-G6 and SMV-G7), Harosoy developed chlorotic lesions at inoculation site, but no virus was detected in the younger trifoliate leaves. These results indicate that there was limited cell-to-cell virus movement at the site of infection, but no systemic movement. These lesions were detected both visually and by whole leaf immunoprints. Conversely, L29 showed no symptoms and virus was not detectable in the inoculated leaf. These differences in resistance led to the tentative designation of two alleles, $Rsv3^{hs}$ (Harosoy source) and $Rsv3^{hd}$ (Hardee source in L29) (Fayad, 2003). It is yet to be determined whether these different reactions are due to genetic background or to differences at the *Rsv3* locus. Additional experiments will need to be performed to answer this question.

Additional Rsv3 Populations

Three different populations were used in Chapter 2 to fine map Rsv3; Sowon (*Rsv3*) x Lee68 (*rsv*) population size 188, L29 (*Rsv3*) x Lee68 (*rsv*) population size 183, and Tousan140 (*Rsv3*) x Lee68 (*rsv*) with a population of 61. A weakness in this study is the size of each population. The L29 (*Rsv3*) x Sowon (*rsv*) population is the largest population, yet it contained less than 200 individuals. An indication that the populations were inefficient is that there were multiple molecular markers that co-segregated with *Rsv3;* eight in L29 x Sowon, five in Tousan140 x Lee68 and three in the L29 x Lee68 population. The fact that all of these markers co-segregated with *Rsv3* limits the use of these populations to develop additional markers.

Several additional populations were developed. One of these populations consisting of 283 $F_{2:3}$ families of a cross between L29 and Lee68, hereafter called the New LL population, was screened for resistance to facilitate further characterization of the *Rsv3* locus.

Syntenic Regions in Soybean and Medicago

It has recently been reported that soybean has undergone two rounds of genome duplications, one at 59 million years ago and a more recent duplication 13 million years ago, since divergence from an ancient ancestor (Schmutz et al., 2010). Analyses of predicted genes indicate that around 60% occur in paralogous pairs (Schmutz et al., 2010). Furthermore, analysis of a resistance gene cluster on molecular linkage group F (chromosome 13), with respect to its homologous duplicated region on linkage group E (chromosome 15), and showed that nearly 75% of the genes in this region were retained. Resistance genes were shown to be retained yet highly divergence in sequence and copy number (Innes et al., 2008). It has also been shown that *Medicago truncatula* and soybean share abundant microsyntenic regions, yet lack abundant large regions of macrosynteny. This can be explained by gene loss and chromosomal rearrangements which occur after genome duplication events (Shoemaker et al., 2008).

Materials and Methods

Population Development and Experimental Design

Initial crossing for the development of these populations was done during the summer of 2008. Potential true crosses were harvested from the field as individual pods and planted in a

greenhouse during the 2008/2009 winter under 14 hours of constant lighting. DNA was extracted as previously described (Edwards et al., 1991) and molecular markers were used to identify true crosses. Twenty nine true F_1 individuals were identified, each of which resulted in the generation of over 500 F_2 seeds per plant. Three of these populations were planted in the Virginia Tech research farm at Whitethorn, one in 2009 and two in 2010. Newly emergent trifoliate leaf tissue was collected in the 2009 and 2010 fields from each of the respective population(s), and stored at -80^{0}C. During the summer of 2010 the first population, here after called the New LL and grown in 2009, was screened for resistance to SMV.

The parental lines L29 (*Rsv3*), Lee68 (*rsv*), SMV-G7 infecting differential checks Hutcheson (*Rsv1-y*), York (*Rsv1-y*) and a population containing 283 $F_{2:3}$ families from the New LL, of which 243 were assayed for their response to SMV strain G7. Twenty five to thirty seeds from each parental and check lines; and each of the $F_{2:3}$ families were evenly distributed among five plastic pots (7.6 cm diameter, 7.6 cm in depth) filled with MetroMix360®, in the greenhouse for testing. Individual lines were distributed across the greenhouse. The New LL population was divided into ten sets of 25 to 30 families each and the experiment was performed over a 4 month period. Each set of 25 to 30 families was inoculated in a single day, with one pot not inoculated to act as a control. Sixteen grams/pot of Osmocote® 18%N, 6%P 12%K, (Scotts-Sierra Horticultural Products Co., Marysville, OH) was added to the surface of each pot as a supplemental fertilizer prior to inoculation as needed. If a specific $F_{2:3}$ family had less than 15 plants, an additional replication was performed to ensure a minimum of 15 to 20 plants were inoculated (Jeong et al., 2002). Additional bulk leaf tissue from approximately 25 F_3 plants was collected from each line screened for virus resistance to ensure adequate genetic material to

complete future studies, and to supplement F_2 tissue collected in 2009. Tissue samples were stored at -80 ^{o}C. DNA was extracted using a modified CTAB protocol (Edwards et al., 1991). Low yielding or poor quality DNA samples were either recollected from inoculated plants or replaced with F_2 DNA collected during the 2009 field season.

Strain Verification and Disease Screening

The SMV strain G7 was used for this study is the same used in previous studies (Gunduz et al., 2002; Jeong et al., 2002; Saghai Maroof et al., 2008). SMV strain G7 does not occur naturally in The United States, yet it is key for differentiating the 3 *Rsv* resistance genes. Identity of strains G1 and G7 was confirmed prior to screening by inoculating the differential cultivars Essex (*rsv*), PI 96983 (*Rsv1*), Ogden (*Rsv1-t*), Kwanggyo (*Rsv1-k*), Marshall (*Rsv1-m*), Hutcheson (*Rsv1-y*), PI 507389 (*Rsv1-n*), L29 (*Rsv3*), and V94-5152 (*Rsv4*) (Table 3.1). This was necessary to ensure that control Hutcheson, Lee68 and L29 seed sources were correct. SMV-G7 cultures were periodically transferred to Hutcheson (*Rsv1-y*) or York (*Rsv1-y*) to ensure a constant inoculum source. All inoculations were performed 10 days after planting, at the fully expanded unifoliate stage. Inoculum was prepared by grinding infected trifoliates in a 0.01 M neutral sodium phosphate buffer (1:10; w:v) in a mortar and pestle (Saghai Maroof et al., 2010). The unifoliates of ten day old plants, previously dusted with 600 mesh carborundum, were rub inoculated with the prepared inoculum and subsequentially rinsed with water (Chen et al., 1991). Inoculated New LL families were scored at 7, 10, and 14 days post inoculation. Individual plants were scored as resistant if they showed no mosaic symptoms when challenged by SMV strain G7 or susceptible if mosaic symptoms appeared at any of the scoring dates (Fig.

3.1). New LL F_2's were scored based on the reaction of their progeny. Susceptible or resistant families were determined depending on the development of virus symptoms or not. Heterozygous F_2's were designated if their inoculated progeny contained both resistant and susceptible individuals segregating in a 3:1 phenotypic ratio.

Immunoblot Assays

Immuno analysis was performed as described by Chang et al. (2010). Selected leaves were rolled and torn to expose potentially infected cellular fluid; then gently pressed onto nitrocellulose membrane (0.45 μm pore) to transfer the sap. Membranes were cleaned with a brush to remove excess tissue then treated in 5% Triton X-100 ((octylphenolpoly(ethyleneglycolether)x) for 5 to 10 minutes, then rinsed in potassium phosphate saline (1x KPS) (0.02MK2HPO4, 0.15MNaCl, pH 7.4) Tween-20 (5 g/L) for 3 minutes. Nonspecific antibody binding was reduced with 5% milk powder and 0.5% bovine serum albumin in 1x KPS-Tween for 30 minutes. Rabbit anti-SMV-antiserum, was added to bind SMV to particles, membranes were then exposed to secondary goat antibody to rabbit-AP (Alkaline Phosphatase) conjugate, both at 1:10,000 dilution, in 1x KPS for 60 minutes, then rinsed with Tris buffer (TBS) (0.05MTris base, 0.15M NaCl, pH 7.6) for 10 minutes three times. After rinses in TBS, membranes were visualized through exposure to Zymed® substrate solution containing 1 part NBT (Nitro Blue Tetrazolium), 1 part BCIP (5-Bromo-4-Chloro-3- Indolyl Phosphate) and 8 parts water until precipitation was seen. Positive blots were determined based on intensity of the precipitate.

Molecular Marker Analysis

Microsatellite data, for markers s156c (from Chapter 2) and BARCSOYSSR_14_1418, was generated with the use of Applied Biosystems 3130x Genetic Analyzer (ABI) (Foster City, CA) or by radioactive labeled products. Each ABI PCR reaction contained ~60 ng of genomic DNA, 0.25 mM dNTP, 1.8 mM $MgCl_2$, 0.96 µM M13 tagged forward primer, 0.72 µM reverse primer, 0.72 M13 dye labeled primer, 0.9 units of Invitrogen™ *Taq* DNA Polymerase (recombinant) and water to a total volume of 11 µL. The general PCR cycles were altered to account for primer annealing temperatures, and are as follows: 94°C for 3 min; four cycles of 94°C for 45 s, 51°C to 56°C for 45 s, 68°C for 45 s; 12 cycles of 94°C for 45 s, 49°C to 54°C for 45 s, 68°C for 45 s; and 30 cycles of 94°C for 45 s, 47°C to 52°C for 45 s; and 68°C for 1 min. GeneMarker software (Softgenetics®) was used to visualize fluorescently tagged microsatellite data. Radioactive labeled microsatellites (GMC and S156h) were resolved through polyacrylamide gel electrophoresis and visualized by autoradiography, as previously described (Saghai Maroof et al., 1994). The PCR reaction mixture consisted of: 60 ng genomic DNA; 0.2 µM forward and reverse primers; 200 µM dATP, dGTP, dTTP, and 200 5 µM dCTP ; 1.5mM $MgCl_2$; 10X buffer; 1 µCi of α-^{32}P dCTP; 0.5 units of Taq polymerase and water was added to a final reaction volume of 11 µL. PCR conditions varied depending on primer annealing temperatures as described above. Gels were exposed to x-ray film for 1 to 5 h before developing. Linkage analysis was calculated using Joinmap® (Van Ooijen and Voorrips, 2001) at default settings.

Analysis of Genomic Regions

Further analysis of the *Rsv3* region was undertaken to identify syntenic regions within soybean and between Medicago and soybean. BLASTp was used to initially identify the best homologs in soybean and Medicago; Soybase (Soybase.org) provided up-to-date gene annotations and 'recently duplicated regions' tract (Grant et al., 2010); and Phytozome (Phytozome.net) identified peptide homologies between soybean and Medicago (Schmutz et al., 2010). Soybase and Phytozome were used to independently confirm the designation of syntenic regions identified by BLASTp prior to the 'recently duplicated regions' tract and peptide homologies.

Results and Discussion

The New LL families screened, segregated in the expected 1:2:1 ratio (52 resistant, 127 heterozygous and 64 susceptible) for resistance to SMV (X^2 =1.68, p = 0.43). While all susceptible individuals displayed mosaic symptoms; necrotic flecks, necrotic spots, chlorotic veins, and chlorotic spots were also observed (Fig. 3.1). Interestingly, only severe mosaic symptoms were seen during the months of July and August. It was concluded that the temperature might have altered the development of additional symptoms, as previously reported (Zheng et al., 2005). Even though there were multiple symptoms observed, they did not alter the genetic ratio.

Molecular marker S156c, the closest microsatellite 0.3 cM to the 5' of the *Rsv3* locus, and BARCSOYSSR_14_1418, which flanks the 3', were used to determine if the observed

resistance was segregating for *Rsv3*. BARCSOYSSR_14_1417, from the mapping study in Chapter 2, could not be used for this assay as it was not polymorphic in the New LL population. A possible explanation for this, is that L29 source used for developing the New LL population is more than ten years advanced than the L29 source used for the original *Rsv3* mapping (Buss et al., 1999). This discrepancy in L29 polymorphism was also observed in several other molecular markers during the study in Chapter 2 (data not shown). Linkage analysis confirmed that the population was segregating at the *Rsv3* locus and that molecular markers flanked *Rsv3*. Even though this population has 95 more individuals than the LS population, only 5 recombination events occurred between the flanking markers, ~200 kb apart. In a previous *Rsv1* mapping study using a population of over 1,000 flanking markers only identified 14 recombinants within a ~300 kb physical region. These 14 individuals enabled the identification of three resistance genes which contributed to virus resistance (Hayes et al., 2004). If this method of gene identification is to be pursued for *Rsv3*, additional populations segregating for *Rsv3* will need to be assayed. Recombinant lines at the *Rsv3* locus can then be screened for phenotypic variation in resistance to the seven SMV strains.

Initially, it was determined that the four NB-LRR resistance genes at the *Rsv3* locus (Glyma14g38500, Glyma14g38510, Glyma14g38540 and Glyma14g38560) were unique to the soybean genome. The highest blastp hits were localized on soybean chromosomes six and twelve, with only 53 to 57 % homology. None of the genes flanking the resistance gene cluster, identified in Chapter 2, were located within these regions. Non-NB-LRR genes within a 200 kb region surrounding *Rsv3* (Glyma14), from bases 47,569,000 to 47,769,000, were used for BLASTn analysis against both the soybean and Medicago genomes. The top soybean hits,

besides the genes themselves, all occurred within a small 125 kb genomic region of chromosome 2 (Glyma02), bases 45,443,800 to 45,568,400. And, in Medicago the top hits were found within a 115 kb region on chromosome 5 (Medtr5), bases 31,540,000 to 31,655,500. (Fig 3.2) Investigation using Phytozome homologous gene tract, which uses peptide similarities, reinforced these findings. The only gene which was not in agreement was Glyma14g38620, an Ubiquitin-conjugating enzyme, which had two homologs in Medicago with 99 % similarity. The designation of syntenic regions was further confirmed by Soybase 'recently duplicated regions tract' and personal communication with Steven Cannon (Schlueter et al., 2008). The initial analysis of this region was done prior to the public release of the peptide homology and 'recently duplicated regions' tract, on Phytozome and Soybase, respectively.

Analysis of the identified syntenic regions indicates that the *Rsv3* resistance gene cluster is likely a recent addition to the soybean genome. The syntenic region on chromosome two is lacking any NB-LRR resistance genes as is Medicago (Fig 3.2). Medicago has a common ancestor with soybean and the combination of a missing resistance gene cluster in both these genomes insinuates that the *Rsv3* cluster has developed since their divergence. Support for this inference is also seen in the inversion of genes Glyma14g38570 and Glyma14g38580 (Fig. 3.2). This inversion could have occurred during a resistance gene transposition event, creating the *Rsv3* locus, which occurred after the most recent soybean genome duplication; 13 million years ago (Schmutz et al., 2010). Alternatively, the resistance gene locus may have developed after the first duplication, yet been lost on soybean chromosome two after the second duplication due to redundant gene loss. Further evaluation of the non-CC-NB-LRR genes shows that there is much greater similarity between the soybean paralogs than the *Medicago* homologs, although

there is still a high percentage of similarity between *Medicago* and soybean (Table 3.2). This evidence makes it unlikely that the *Rsv3* locus was present before either of the soybean whole genome duplication events, and is due to recent genome evolution. Annotations are based on Joint Genome Institute soybean genome release Gma1.01.

Table 3.1: Reactions of soybean differential lines to mosaic virus strains G1 and G7 at 19 days post inoculation

	Soybean Lines								
SMV Strain	**Essex** ***rsv***	**PI96983** ***Rsv1***	**Hutcheson** ***Rsv1-y***	**Marshall** ***Rsv1-m***	**Ogden** ***Rsv1-t***	**PI507389** ***Rsv1-n***	**Kwanggyo** ***Rsv1-k***	**L29** ***Rsv3***	**V94-5152** ***Rsv4***
G1	SvMos	NS	NS	NS	NS	TN	NS	Mos	NS
G7	Mos	LSHR	Mos	TN	VNc	Mos	NT	NS	NS

Abbreviations**:** LSHR, Lethal Systemic Hypersensitive Response; Mos, Mosaic; NS, No Symptoms; SvMos, Severe Mosaic; TN, Tip necrosis; and VNc, Veinal Necrosis**.**

Table 3.2: ClustalX identity matrix of Non-NB-LRR syntenic peptides within a 200 kb region surrounding *Rsv3* (Glyma14) showing higher similarities between genes on soybean chromosomes 14 and 2 (Glyma02) when compared to Medicago chromosome 5 (Medtr5). Gene numbers are the same as in Figure 3.2.

Gene 1	**HY5 (LONG HYPOCOL5 protein)**		
	Glyma14g38460	**Glyma02g40270**	**Medtr5g083200**
Glyma14g38460	100	94	80
Glyma02g40270	-	100	80
Medtr5g083200	-	-	100

Gene 2	**B3 DNA binding domain**		
	Glyma14g38490	**Glyma02g40280**	**Medtr5g083230**
Glyma14g38490	100	87	61
Glyma02g40280	-	100	64
Medtr5g083230	-	-	100

Gene 3	**KIP1-like protein**		
	Glyma14g38570	**Glyma02g40300**	**Medtr5g083340**
Glyma14g38570	100	93	75
Glyma02g40300	-	100	76
Medtr5g083340	-	-	100

Gene 4	**Cytochrome P450**		
	Glyma14g38580	**Glyma02g40290**	**Medtr5g083280**
Glyma14g38580	100	98	88
Glyma02g40290	-	100	88
Medtr5g083280	-	-	100

Table 3.2 Cont.

Gene 5 — **Translation initiation factor**

	Glyma14g38600	Glyma02g40310	Medtr5g083360
Glyma14g38600	100	97	85
Glyma02g40310	-	100	86
Medtr5g083360	-	-	100

Gene 6 — **AP2 domain**

	Glyma14g38610	Glyma02g40320	Medtr5g083380
Glyma14g38610	100	89	61
Glyma02g40320	-	100	60
Medtr5g083380	-	-	100

Gene 7 — **Ubiquitin-conjugating enzyme**

	Glyma14g38620	Glyma02g40330	Medtr5g083390
Glyma14g38620	100	100	99
Glyma02g40330	-	100	99
Medtr5g083390	-	-	100

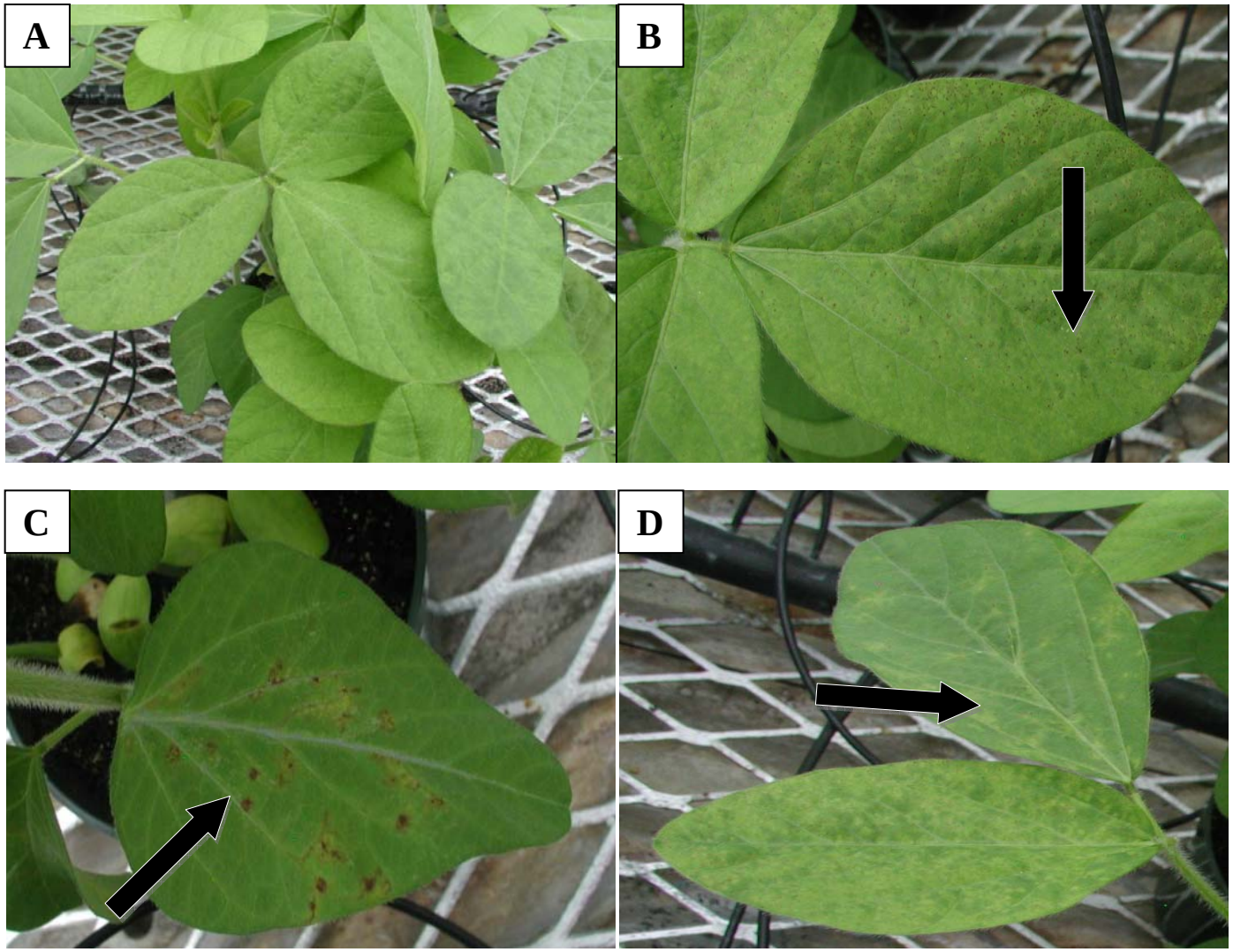

Figure 3.1: Symptoms observed in susceptible plants of the New LL population: A) Mosaic (Family 7 of 283, susceptible; B) Necrotic flecking (Family 13, heterozygous); C) Necrotic spots (Family 20, susceptible); and D) Chloritic spots and Chlorotic veins (Family 14, heterozygous). Missing leaves (D) were sampled for immunoassays.

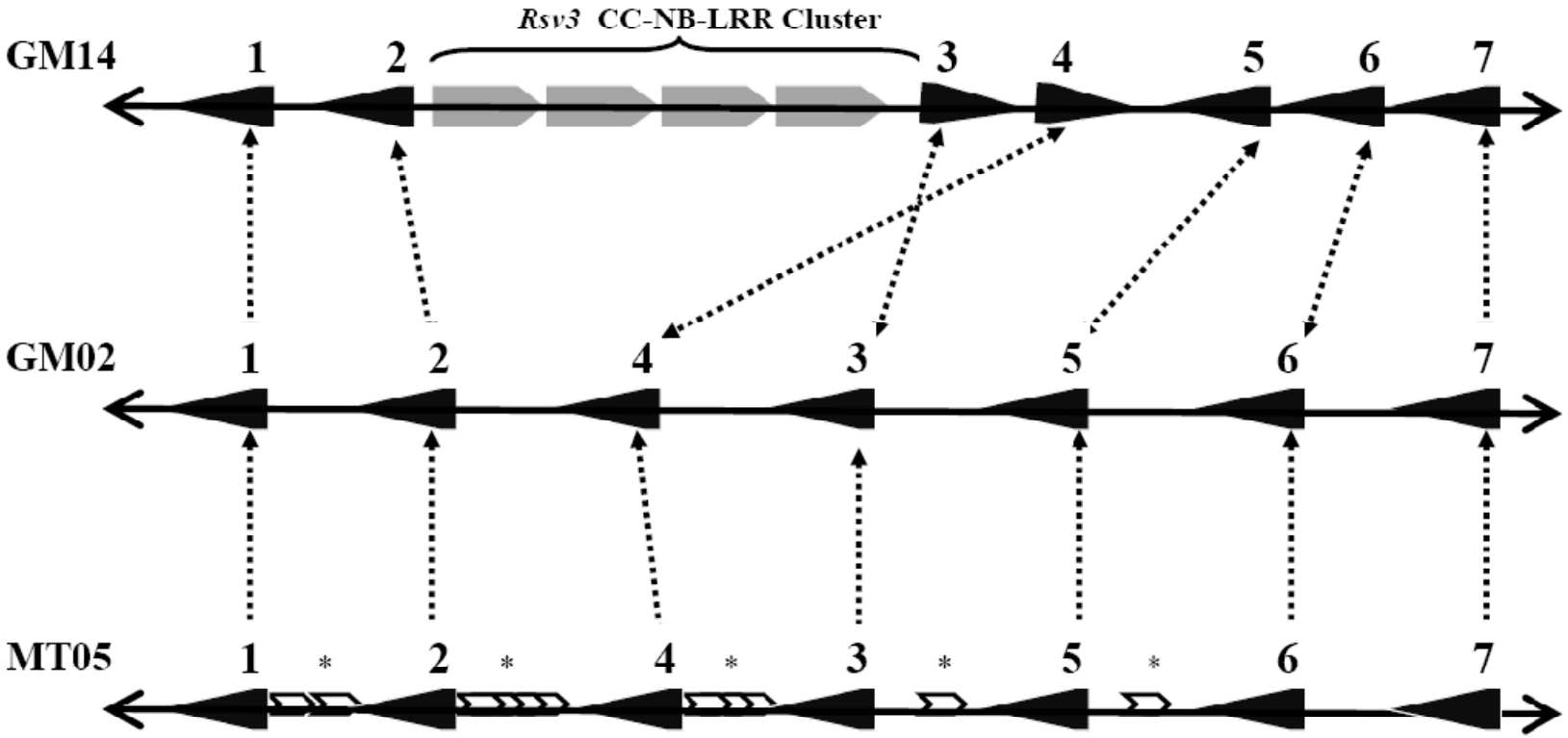

Figure 3.2: GM14, the 200 kb region harboring *Rsv3* locus on soybean chromosome 14 from nucleotide 47,569,000 to 47,769,000; GM02 paralogous 125 kb genomic region of soybean chromosome 2 from nucleotide 45,443,800 to 45,568,400; and MT05, homologous 115 kb region of *Medicago trunculata* chromosome 5 from nucleotide 31,540,000 to 31,655,500. Description based on Phytozome ontologies: 1, HY5 (long hypocotly5 protein); 2, B3 DNA binding domain; 3, KIP1-like protein; 4, Cytochrome p450; 5, Eukaryotic translation initiation factor; 6, AP2 domain containing protein; and 7, Ubiquitin-conjugating enzyme. '*' Designates short hypothetical Medicago peptides or peptides with weak homology to known genes, based on Phytozome ontologies. "*Rsv3* CC-NB-LRR cluster" shows the position of the identified candidate resistance genes in relation to surrounding putative genes. An alternative GBrowse view of these regions can be found in Figure 3.3.

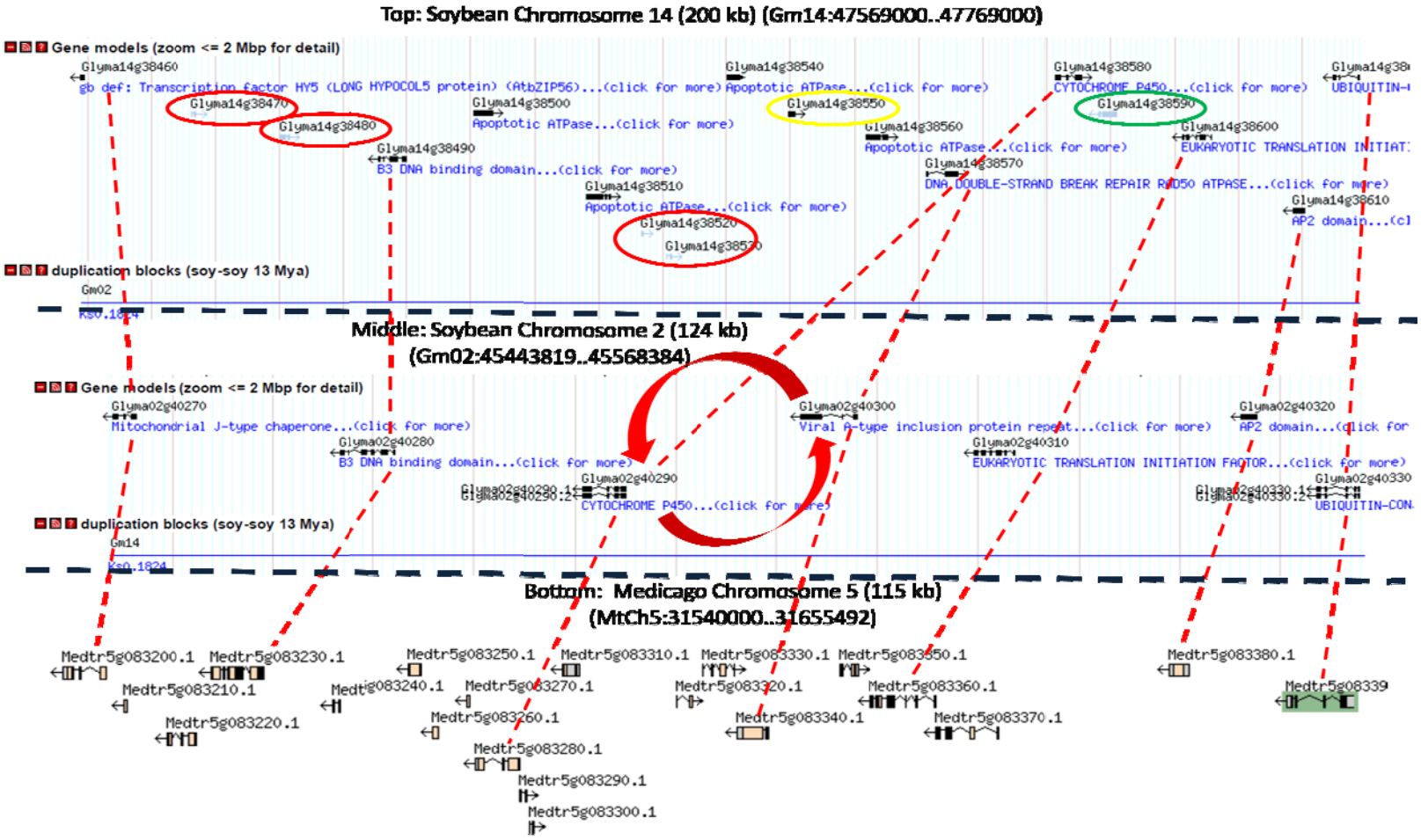

Figure 3.3: Alternative GBrowse view of syntenic regions within soybean and between soybean and *Medicago*. *Rsv3* locus on chromosome 14 is on the top, syntenic region in soybean chromosome 2 is in the middle, and syntenic region on *Medicago* chromosome 5 is on the bottom. Red circles indicate low confidence transcripts, the yellow circle is a small hypothetical unannotated peptide, and the green circled gene is a low confidence CC-NB-LRR. Candidate disease resistance genes are annotated as Apoptotic ATPase. In Medicago there are numerous small unannotated peptides dispersed within the region.

References:

Buss G.R., Ma G., Kristipati S., Chen P., Tolin S.A. (1999) A new allele at the *Rsv3* locus for resistance to soybean mosaic virus, in: H. E. Kauffman (Ed.), Proc. World Soybean Res. Conf. VI., Superior Printing, Champaign, IL. 490.

Chang P.S., McLaughlin W.A., Tolin S.A. (2011) Tissue blot immunoassay and direct RT-PCR of cucumoviruses and potyviruses from the same nitropure nitrocellulose membrane. J. Virol. Methods 171:345-351.

Chen P., Buss G.R., Roane C.W., Tolin S.A. (1991) Allelism among genes for resistance to Soybean mosaic virus in strain-differential soybean cultivars. Crop Sci. 31:305-309.

Edwards K., Johnstone C., Thompson C. (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. 19:1349.

Fayad, A. (2003) Interactions of soybean *Rsv* genes and Soybean mosaic virus. Virginia Tech Electronic Thesis and Dissertation Database.

Grant D., Nelson R.T., Cannon S.B., Shoemaker R.C. (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 38:843-846.

Gunduz I., Buss G.R., Ma G., Chen P., Tolin S.A. (2001) Genetic analysis of resistance to soybean mosaic virus in OX670 and Harosoy soybean. Crop Sci. 41:1785-1791.

Gunduz I., Buss G.R., Chen P., Tolin S.A. (2002) Characterization of SMV resistance genes in Tousan140 and Hourei soybean. Crop Sci 42:90-95.

Hayes A.J., Jeong S.C., Gore M.A., Yu Y.G., Buss G.R., Tolin S.A., Saghai Maroof M.A. (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.

Innes R.W., Ameline-Torregrosa C., Ashfield T., Cannon E., Cannon S.B., Chacko B., Chen N.W., Couloux A., Dalwani A., Denny R., Deshpande S., Egan A.N., Glover N., Hans

C.S., Howell S., Ilut D., Jackson S., Lai H., Mammadov J., Del Campo S.M., Metcalf M., Nguyen A., O'Bleness M., Pfeil B.E., Podicheti R., Ratnaparkhe M.B., Samain S., Sanders I., Segurens B., Sevignac M., Sherman-Broyles S., Thareau V., Tucker D.M., Walling J., Wawrzynski A., Yi J., Doyle J.J., Geffroy V., Roe B.A., Saghai Maroof M.A., Young N.D. (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol. 148:1740-1759.

Jeong S.C., Kristipati S., Hayes A.J., Maughan P.J., Noffsinger S.L., Gunduz I., Buss G.R., Saghai Maroof M.A. (2002) genetic and sequence analysis of markers tightly linked to the Soybean mosaic virus resistance gene, *Rsv3*. Crop Sci 42:265-270.

Ma G., Chen P., Buss G.R., Tolin S.A. (2002) Complementary action of two independent dominant genes in Columbia soybean for resistance to soybean mosaic virus. J. Hered. 93:179-184.

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

Saghai Maroof M.A., Biyashev R.M., Yang G.P., Zhang Q., Allard R.W. (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. (USA) 91:5466-5470.

Schlueter J.A., Scheffler B.E., Jackson S., Shoemaker R.C. (2008) Fractionation of synteny in a genomic region containing tandemly duplicated genes across Glycine max, Medicago truncatula, and Arabidopsis thaliana. J. Hered. 99:390-395.

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C.,

Jackson S.A. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Shoemaker R.C., Grant D., Olson T., Warren W.C., Wing R., Yu Y., Kim H., Cregan P., Joseph B., Futrell-Griggs M., Nelson W., Davito J., Walker J., Wallis J., Kremitski C., Scheer D., Clifton S.W., Graves T., Nguyen H., Wu X., Luo M., Dvorak J., Nelson R., Cannon S., Tomkins J., Schmutz J., Stacey G., Jackson S. (2008) Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. Genome 51:294-302.

Suh S.J., Bowman B.C., Jeong N., Yang K., Kaslt C., Tolin S.A., Saghai Maroof M.A., Jeong S.-C. (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 Gen. 4:55-64.

Van Ooijen J.W., Voorrips R.E. (2001) JoinMap 3.0 software for the calculation of genetic linkage maps. Plant Research International,Wageningen, The Netherlands.

Zheng C., Chen P., Gergerich R. (2005) Effect of temperature on the expression of necrosis in soybean infected with soybean mosaic virus. Crop Sci. 45:916-922.

Future Research:

1. Verify candidate *Rsv3* resistance genes are expressed through gene expression assays. Determining which of the candidate disease resistance genes are expressed could eliminate several of the genes as candidates.

2. Identify specific *Rsv3* candidate gene responsible for resistance through gene silencing and additional mapping. There are unique regions within the 4 full length candidate genes which could be exploited for targeted silencing.

3. Screen additional LL populations to identify recombinants within the *Rsv3* locus to screen for alterations in symptom development when challenged by different SMV strains.

4. Compare sequence level differences between soybean cultivars which have *Rsv3* resistance

5. Clone the five candidate resistance genes at the *Rsv3* locus for transformation into a susceptible cultivar to identify what each gene contributes to resistance.