

Identification and Analyses of Candidate Genes for *Rpp4*-Mediated Resistance to Asian Soybean Rust in Soybean^{1[W][OA]}

Jenelle D.F. Meyer, Danielle C.G. Silva, Chunling Yang, Kerry F. Pedley, Chunquan Zhang, Martijn van de Mortel, John H. Hill, Randy C. Shoemaker, Ricardo V. Abdelnoor, Steven A. Whitham, and Michelle A. Graham*

United States Department of Agriculture-Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, Iowa 50011 (J.D.F.M., R.C.S., M.A.G.); Embrapa Soja, Londrina, Parana, Brazil 86001-970 (D.C.G.S., R.V.A.); Universidade Estadual Paulista, Jaboticabal, Sao Paulo, Brazil 14884-900 (D.C.G.S.); Department of Plant Pathology (C.Y., C.Z., M.v.d.M., J.H.H., S.A.W.) and Department of Agronomy (R.C.S., M.A.G.), Iowa State University, Ames, Iowa 50011; and United States Department of Agriculture-Agricultural Research Service, Foreign Disease-Weed Science Research Unit, Fort Detrick, Maryland 21702 (K.F.P.)

Asian soybean rust is a formidable threat to soybean (*Glycine max*) production in many areas of the world, including the United States. Only five sources of resistance have been identified (*Resistance to Phakopsora pachyrhizi* [*Rpp1*], *Rpp2*, *Rpp3*, *Rpp4*, and *Rpp5*). *Rpp4* was previously identified in the resistant genotype PI459025B and mapped within 2 centimorgans of Satt288 on soybean chromosome 18 (linkage group G). Using simple sequence repeat markers, we developed a bacterial artificial chromosome contig for the *Rpp4* locus in the susceptible cv Williams82 (Wm82). Sequencing within this region identified three *Rpp4* candidate disease resistance genes (*Rpp4C1*–*Rpp4C3* [Wm82]) with greatest similarity to the lettuce (*Lactuca sativa*) RGC2 family of coiled coil-nucleotide binding site-leucine rich repeat disease resistance genes. Constructs containing regions of the Wm82 *Rpp4* candidate genes were used for virus-induced gene silencing experiments to silence resistance in PI459025B, confirming that orthologous genes confer resistance. Using primers developed from conserved sequences in the Wm82 *Rpp4* candidate genes, we identified five *Rpp4* candidate genes (*Rpp4C1*–*Rpp4C5* [PI459025B]) from the resistant genotype. Additional markers developed from the Wm82 *Rpp4* bacterial artificial chromosome contig further defined the region containing *Rpp4* and eliminated *Rpp4C1* (PI459025B) and *Rpp4C3* (PI459025B) as candidate genes. Sequencing of reverse transcription-polymerase chain reaction products revealed that *Rpp4C4* (PI459025B) was highly expressed in the resistant genotype, while expression of the other candidate genes was nearly undetectable. These data support *Rpp4C4* (PI459025B) as the single candidate gene for *Rpp4*-mediated resistance to Asian soybean rust.

Asian soybean rust (ASR) is caused by the fungus *Phakopsora pachyrhizi* and is a formidable threat to world soybean (*Glycine max*) production. ASR was first identified in the Eastern Hemisphere in the early 1900s and has since spread to many countries throughout the world, including the United States (Schneider et al., 2005). In countries where soybean rust is established, soybean yield losses range from 10% to 80% (Ogle

et al., 1979; Bromfield, 1984; Patil et al., 1997). Yield loss severity depends on many factors, including soybean variety, time during the growing season when the rust becomes established, and environmental conditions. Without durable genetic resistance in commercial lines, producers must depend on expensive and time-consuming foliar fungicides for disease management.

Five major sources of ASR resistance have been identified in soybean: *Resistance to Phakopsora pachyrhizi* (*Rpp1*; Cheng and Chan, 1968; Hidayat and Somaatmadja, 1977; McLean and Byth, 1980; Hartwig and Bromfield, 1983); *Rpp2* (Hidayat and Somaatmadja, 1977); *Rpp3* (Singh and Thapliyal, 1977; Bromfield and Hartwig, 1980); *Rpp4* (Hartwig, 1986); and *Rpp5* (Garcia et al., 2008). *Rpp1* confers an immune response for which there are no visible symptoms in the plant (Miles et al., 2006). Resistance responses mediated by the *Rpp2* to *Rpp5* loci limit fungal growth and sporulation through the formation of visible reddish-brown lesions suggestive of a hypersensitive-like response (HR; Bonde et al., 2006; Garcia et al., 2008). Tan-colored

¹ This work was supported by the U.S. Department of Agriculture-Agricultural Research Service, Embrapa Soja, the Iowa Soybean Association, the North Central Soybean Research Program, the United Soybean Board, and the National Science Foundation (grant no. 0820642).

* Corresponding author; e-mail michelle.graham@ars.usda.gov.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Michelle A. Graham (michelle.graham@ars.usda.gov).

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.108.134551

lesions and fully sporulating uredinia generally indicate a susceptible interaction to ASR (Bromfield and Hartwig, 1980; Bromfield, 1984; Miles et al., 2006). The resistance sources identified are all specific to certain strains of *P. pachyrhizi* (Bonde et al., 2006).

The resistance response governed by *Rpp2* was studied extensively by microarray analyses (van de Mortel et al., 2007). Expression of basal defense pathway genes increased during the first 12 h after inoculation (hai) in both the susceptible and resistant plants. Gene expression in both genotypes returned to mock levels before a second phase of differential gene expression was observed at approximately 72 hai. While this response was detected in both resistant and susceptible interactions, the second phase of differential expression was stronger and detected 1 to 2 d earlier in the resistant plants than in the susceptible plants. Genes included in this biphasic response are associated with transcription, signal transduction, and plant defenses and are consistent with the stronger and more rapid induction of defense genes typically seen in the HR (Tao et al., 2003). The resistance phenotypes conferred by *Rpp3*, *Rpp4*, and *Rpp5* suggest that they mediate responses similar to *Rpp2* and are likely governed by disease resistance genes that mediate gene-for-gene recognition.

Typical disease resistance genes, such as NBS-LRRs (for nucleotide-binding site-Leu rich repeats), receptor-like kinases, and receptor-like proteins (Jones et al., 1994; Dixon et al., 1996, 1998; Song et al., 1997; Dangl and Jones, 2001; Sun et al., 2004), transduce the HR to defend against pathogen attack. An important aspect of defense is the ability of plant *R* genes and pathogen *Avr* genes to coevolve (Michelmore and Meyers, 1998; Jones and Dangl, 2006). Alternating cycles of selection on the plant and pathogen populations are thought to drive this coevolution. When the plant recognizes the pathogen, selection pressure on the pathogen forces it to evolve in order to evade detection. The plant species must follow suit in order to continue detecting the pathogen. It is likely that clustering of *R* gene loci facilitates the generation of novel disease resistance specificities. *R* gene clusters have been identified in many species, including rice (*Oryza sativa*; Song et al., 1997), Arabidopsis (*Arabidopsis thaliana*; Baumgarten et al., 2003; Meyers et al., 2005), maize (*Zea mays*; Hulbert and Bennetzen, 1991), and soybean (Kanazin et al., 1996; Graham et al., 2002; Gao and Bhattacharyya, 2008). Clustering of *R* genes allows evolution to occur via continued rounds of duplication, unequal crossing over, segmental duplication, and/or rearrangements facilitated by transposable elements (Richter and Ronald, 2000; Baumgarten et al., 2003; Meyers et al., 2005). These phenomena contribute to the development of new pathogen specificities or may result in the deletion of genes completely (Leister, 2004). *R* gene evolution is further constrained by plant fitness. If the cost of maintaining the *R* gene without pathogen threat is greater than the cost of disease threat, selection will act to remove the *R* gene (Meyers et al., 2005).

Many plant disease resistance genes have been cloned using genetic map-based methods. These approaches can be extremely tedious, given the difficulty in marker development and the size and complexity of *R* gene clusters. Reverse genetic approaches, such as tilling or mutagenesis, have not been used to confirm the identity of candidate resistance genes in crops, since these methods require that time-consuming and expensive experiments be applied to a single genotype of interest. Here, we used forward and reverse genetic approaches, in combination with sequence data from a susceptible soybean genotype, to identify candidate genes controlling *Rpp4*-mediated resistance to ASR. Previously, Silva et al. (2008) used a mapping population derived from PI459025B (*Rpp4* resistant) and BRS184 (ASR susceptible) to map the *Rpp4* locus to soybean chromosome 18 (linkage group [LG] G). *Rpp4* mapped within 1.9 centimorgans (cM) of simple sequence repeat (SSR) marker Satt288. This position is consistent with Garcia et al. (2008), who mapped *Rpp4* within 2.8 cM of Satt288. Using Satt288 and other publicly available SSR markers, we built and anchored a bacterial artificial chromosome (BAC) contig surrounding the *Rpp4* locus in the susceptible cv Williams82 (Wm82). Sequencing of this region identified three *Rpp4* candidate genes. Conserved primers developed from the Wm82 *Rpp4* candidate genes amplified five genes from the resistant genotype (PI459025B). Using a combination of quantitative real-time, reverse transcription (qRT)-PCR, fine mapping with new genetic markers, and virus-induced gene silencing (VIGS), we have identified a single candidate gene for *Rpp4*-mediated resistance to ASR.

RESULTS

Identification of *Rpp4* Candidate Resistance Genes

A BAC contig surrounding the *Rpp4* locus was developed by PCR screening of SSR markers and BAC-end primers against two Wm82 BAC libraries (Supplemental Fig. S1; Supplemental Table S1). BACs GM_WBb0070A12 and GM_WBb0176I01 (M70A12 and M176I01 in Supplemental Figs. S1 and S2; Supplemental Table S1) were selected for complete sequencing because of end sequence similarity to known disease resistance genes. A total of 208,603 bp of contig sequence was generated from BACs GM_WBb0070A12 and GM_WBb0176I01 (GenBank accession nos. FJ225394 and FJ225395). A total of 552 subclones of GM_WBb0070A12 (160,583 bp) were sequenced in both directions, resulting in 5.8× coverage. For GM_WBb0176I01, paired ends were generated for 588 subclones, resulting in 10.4× coverage of the region. Twenty-eight total genes were identified in the sequenced region (Supplemental Fig. S2). Eleven of the gene predictions shared significant sequence similarity to repetitive elements, including two unclassified retrotransposon proteins (Q53QA4 [1E⁻⁶⁰] and

Q53PP0 [1E^{-101}], four Copia-type polyproteins (Q9M2D1 [$0\text{--}3\text{E}^{-8}$]), a retroelement pol polyprotein (Q9S1M3 [1E^{-93}]), an integrase (Q8S8M1 [4E^{-38}]), and a retrotransposon, Tto1 (Q9ZRJ0 [1E^{-22}]). Gene predictions with significant similarity to known genes included a dehydration response family protein (Q6EP94 [1E^{-112}]), two zinc finger proteins (Q2XX4 [2E^{-7} and 2E^{-10}]), a calcineurin-like protein (UPI000016309B [1E^{-53}]), a cellulase protein (Q43105 [3E^{-5}]), and a cyclin-like F-box protein (A2Q5C7 [1E^{-124}]).

In addition, three *Rpp4* candidate genes (*Rpp4C1* [Wm82], *Rpp4C2* [Wm82], and *Rpp4C3* [Wm82]) with significant sequence similarity (1E^{-110}) to the lettuce *Resistant Gene Candidate2* (RGC2; Meyers et al., 1998; Shen et al., 2002) family of disease resistance genes were identified. The *Rpp4* candidate genes belong to the coiled-coil (CC), NBS, and LRR family of disease resistance genes (Fig. 1; Supplemental Fig. S3). The NBS is required for ATP/GTP binding and acts as a signaling molecule (Tameling et al., 2002), as does the CC domain (Rairdan et al., 2008). In contrast, the LRR provides a potential binding surface for protein-protein interactions (Jones and Jones, 1997).

Like the lettuce RGC2 genes, the *Rpp4* candidate genes are quite large. *Rpp4C1* (Wm82), *Rpp4C2* (Wm82), and *Rpp4C3* (Wm82) are 17,528, 19,706, and

22,665 bp in length, respectively. Similarly, the predicted proteins range in size from 3,055 to 3,693 amino acids. Nucleotide identity in the coding sequence ranges from 87% to 95%. BLASTN (Altschul et al., 1997) searches against the Dana Farber Soybean Gene Index (Lee et al., 2005) failed to identify any ESTs with significant homology ($E < 10\text{E}^{-4}$) to the *Rpp4* candidate genes. Since the genes were not represented by ESTs, which are the foundation for the soybean microarrays, they are also not represented on any publicly available microarray.

To identify other candidate *R* genes in the *Rpp4* locus outside of the sequenced BACs, the soybean whole genome sequence scaffold corresponding to the *Rpp4* BAC contig was identified (scaffold_21, version Glyma0; United States Department of Energy Joint Genome Institute, 2008). A BLASTX (Altschul et al., 1997) comparison of the genome scaffold identified no additional *R* genes in the sequence corresponding to the *Rpp4* BAC contig. BLASTN ($E < 10\text{E}^{-4}$) analyses were used to identify genes homologous to *Rpp4* candidate genes elsewhere in the soybean genome. These analyses identified a single *Rpp4-like gene* (RLG) located on scaffold_48 (version Glyma0; United States Department of Energy Joint Genome Institute, 2008), which corresponds to soybean chromosome 1 (LG D1A). The coding region of RLG shares 74% nucleotide identity with the candidate *Rpp4* genes.

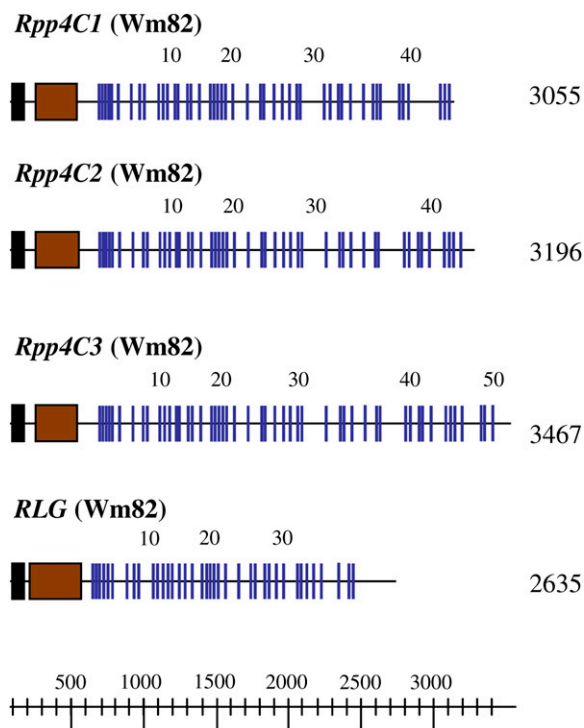


Figure 1. Motif identification in *Rpp4C1* (Wm82), *Rpp4C2* (Wm82), *Rpp4C3* (Wm82), and RLG (Wm82). The predicted amino acid sequences of the *Rpp4* cluster as well as *Rpp4L* were used for motif prediction. Identified motifs included CC domains (black boxes), NBS domains (brown boxes), and LRR domains (blue vertical lines). Full amino acid length is indicated at right and by the scale at bottom.

Amplification of *Rpp4* Candidate Genes from the Resistant Genotype (PI459025B)

To determine copy number of the *Rpp4* candidate genes from the resistant genotype (PI459025B), primers were developed from conserved regions of the *Rpp4* candidate genes from Wm82 (*Rpp4_NB_F/R*; Supplemental Table S3). These primers were used to amplify PCR products from PI459025B genomic DNA. Sequencing of 192 cloned PCR products identified six unique genes, five of which had greater than 90% identity to the *Rpp4* candidate genes (Supplemental Fig. S4). *Rpp4C1*, *Rpp4C2*, and *Rpp4C3* (PI459025B) share the greatest homology with their Wm82 counterparts (*Rpp4C1*, *Rpp4C2*, and *Rpp4C3* [Wm82]), with greater than 95%, 99%, and 99% sequence identity, respectively. However, *Rpp4C4* (PI459025B) is closely related to both *Rpp4C2* (Wm82) and *Rpp4C3* (Wm82; greater than 98% nucleotide identity). The fifth identified gene, *Rpp4C5* (PI459025B), shares the most sequence identity (greater than 92%) with *Rpp4C1* (Wm82). The sixth gene was identified as RLG (PI459025B) because it shares greater than 98% identity with its Wm82 counterpart, RLG (Wm82).

VIGS Confirms That *Rpp4* Is Encoded by One of the PI459025B *Rpp4C* Genes

Given the similarity (greater than 92% identity) between *Rpp4* candidate genes from PI459025B and

Wm82, we developed VIGS constructs using the *Rpp4* candidate genes from Wm82. If *Rpp4* was encoded by a gene orthologous to one of the Wm82 *Rpp4C* genes, we expected that it would be silenced in the resistant genotype (PI459025B), which would become susceptible to *P. pachyrhizi* isolate LA04-1. The similarity between genes made it impossible to develop VIGS constructs specific to each gene; therefore, two *Rpp4* candidate gene VIGS constructs were developed that could silence all members of this gene cluster. The first was developed from the NBD region using primers BPMV_NBD_F/R (Supplemental Table S2), and the second was developed from the LRR region using primers BPMV_LRR_F/R (Supplemental Table S2). To perform the VIGS experiments, 14-d-old PI459025B plants were subjected to one of five pretreatments: no treatment, mock inoculation with buffer and carborundum, inoculation with a BPMV vector lacking an insert, or inoculation with one of two BPMV VIGS vectors targeting the NBD or the LRR (Fig. 2, A–E). At 21 d after BPMV inoculation, all plants were inoculated with a spore suspension from *P. pachyrhizi* isolate LA04-1. Silencing with both the NBD and LRR constructs caused the PI459025B *Rpp4* plants to exhibit a susceptible phenotype at 14 d after inoculation with *P. pachyrhizi* isolate LA04-1 (Fig. 2, D and E). In addition to the tan lesions, fully sporulating uredinia were visible on the upper and lower leaf surfaces. Silencing of the *Rpp4C* cluster among three independent biological replicates was verified by Taqman RT-PCR. *Rpp4* candidate gene cluster mRNAs were reduced by an average of 1.51- ± 0.76-fold and 2.43- ± 0.36-fold for the NBD and LRR constructs, respectively, when compared with the empty vector control. The loss of resistance was in contrast to the reaction of PI459025B plants that received the control pretreatment of no treatment, mock inoculation, and BPMV empty vector inoculation (Fig. 2, A–C). As expected, these control plants developed only red-brown lesions when challenged with *P. pachyrhizi* isolate LA04-1, indicative of a resistant HR. To confirm that *Rpp4*-mediated resistance was broken, fungal growth was measured using Taqman RT-PCR in the BPMV empty vector plants and plants that were pretreated with BPMV VIGS vectors targeting the *Rpp4C* cluster NBD and LRR domains. The amount of *P. pachyrhizi* α -tubulin transcript in the *Rpp4C* NBD- and LRR-silenced plants was 2- to 3-fold greater than in the BPMV empty vector control, demonstrating that ASR growth increased as would be expected if resistance was broken (Fig. 2F). These results indicate that *Rpp4* is encoded by a gene orthologous to the *Rpp4C* genes from Wm82.

SSR Markers Define a Region Containing the Candidate *Rpp4* Resistance Gene

Two SSR markers (sc21_3360 and sc21_3420; Fig. 3; Supplemental Table S4), located within BAC GM_

WBb0176I01 and separating the three *Rpp4* candidate genes in the Wm82 cluster, were identified as polymorphic between the resistant (PI459025B) and susceptible (BRS184) parents. These two markers were mapped using the *Rpp4* mapping population reported by Silva et al. (2008), and they were found to flank *Rpp4* and define *Rpp4C2* (PI459025B) as the candidate *Rpp4* resistance gene on soybean chromosome 18 (LG G; Fig. 3). However, since our sequence information from PI459025B is limited to PCR products, we were unable to develop markers specifically defining *Rpp4C4* (PI459025B) or *Rpp4C5* (PI459025B). The high level of recombination (11 recombination events detected between *Rpp4* and sc21_3360 and five events detected between *Rpp4* and sc21_3420; Fig. 3) suggested that *Rpp4C4* (PI459025B) and *Rpp4C5* (PI459025B) are also located between these markers. These results suggest that *Rpp4C2* (PI459025B), *Rpp4C4* (PI459025B), and *Rpp4C5* (PI459025B) could all be candidate genes for *Rpp4*-mediated resistance.

Rpp4C4 (PI459025B) Is Highly Expressed in PI459025B Regardless of ASR Inoculation

qRT-PCR was used to determine relative expression levels of *Rpp4C1* to *Rpp4C3* (Wm82) and *Rpp4C1* to *Rpp4C5* (PI459025B) in the susceptible line (Wm82) and the resistant line (PI459025B) following infection with *P. pachyrhizi* isolate LA04-1 and mock inoculation. The *Rpp4F/R* primers (Supplemental Table S3) were designed to amplify all *Rpp4* candidate genes with approximately the same efficiency. In the ASR-infected samples, differences of 6.49-, 3.92-, 4.82-, and 5.02-fold were detected between the resistant and susceptible samples at 12, 24, 72, and 216 hai, respectively (Table I). Similarly, in the mock-inoculated samples, differences of 5.38-, 2.84-, 2.86-, and 3.01-fold were detected at 12, 24, 72, and 216 hai, respectively. Successful inoculation of plants was confirmed by Taqman RT-PCR. The amount of *P. pachyrhizi* α -tubulin was measured and normalized relative to the amount of soybean ubiquitin. Plants inoculated with *P. pachyrhizi* were compared with mock-inoculated plants. The RT-PCR revealed 18-, 50-, 145-, and 545-fold change in *P. pachyrhizi* α -tubulin transcript in PI459025B plants at 12, 24, 72, and 216 hai and 16-, 23-, 86-, and 1,997-fold change in Wm82 plants at the same time points, indicating successful inoculation with *P. pachyrhizi*. These results indicate that *Rpp4* candidate gene expression is greater in the resistant samples, regardless of ASR infection.

To determine which genes were expressed in these samples, the RT-PCR products of primers *Rpp4_NB_F/R* and *Rpp4F/R* were cloned and sequenced (Table I; Supplemental Table S3; Supplemental Fig. S4). The *Rpp4_NB_F/R* primers were chosen because they amplified approximately 1,200-bp fragments including single nucleotide polymorphisms that distinguish all five PI459025B *Rpp4* candidate genes. RT-PCR products were cloned from RNA isolated 12 and 72 hai

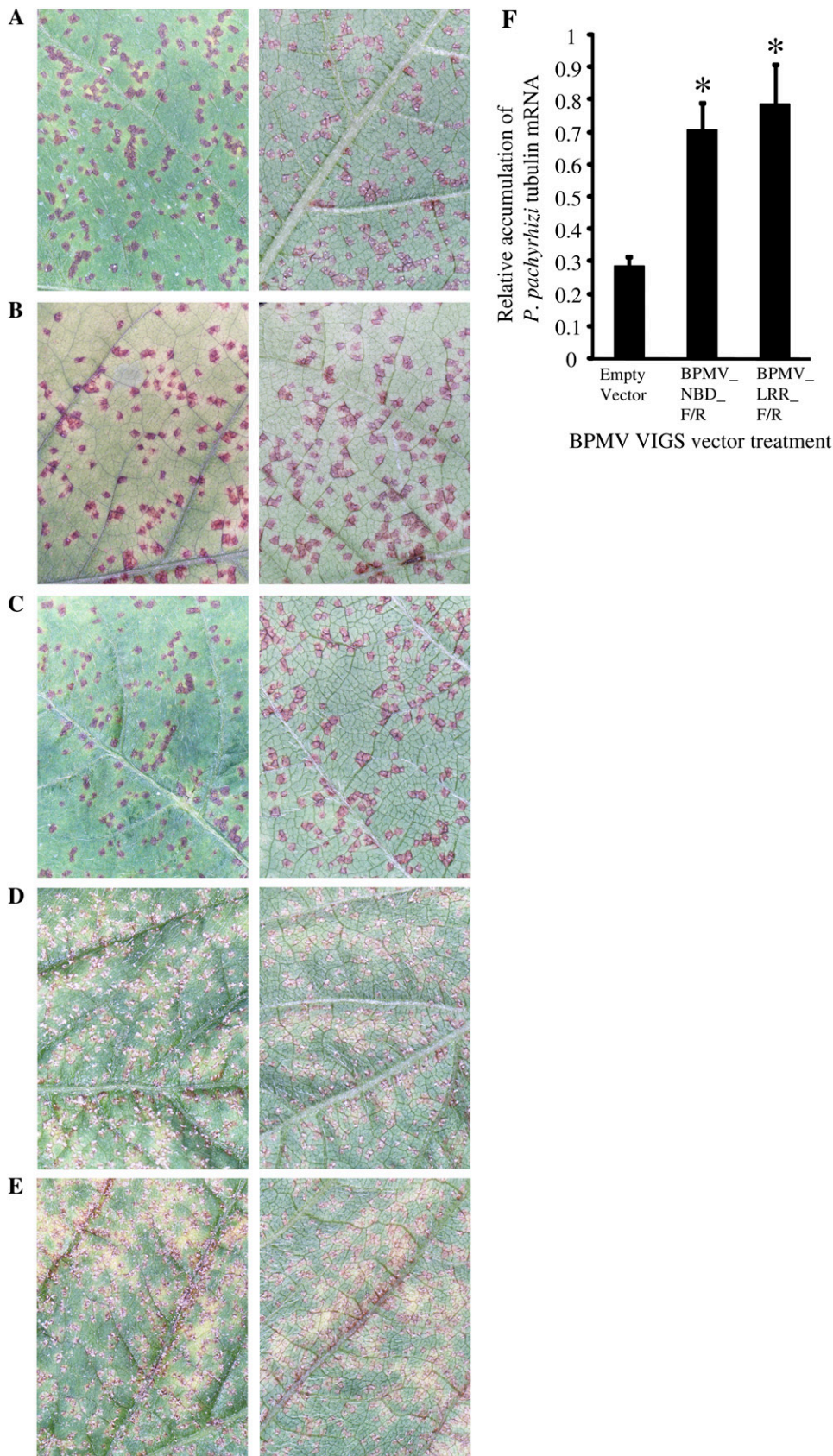


Figure 2. PI459025B response to ASR infection following VIGS. The images at left are the top (adaxial) leaf surfaces and those at right are the bottom (abaxial) leaf surfaces. A to E represent plants subjected to one of five pretreatments. The plant in A was not pretreated. The plant in B was mock inoculated with buffer and carborundum. The plant in C was infected with a BPMV vector lacking an insert (empty vector). The plant in D was inoculated with a BPMV construct carrying the *Rpp4* candidate BPMV_NBD_F/R insert (Supplemental Table S4). The plant in E was inoculated with a BPMV construct carrying the *Rpp4* candidate BPMV_LRR_F/R insert (Supplemental Table S4). All of these plants were subsequently infected with ASR and phenotyped. Red/brown lesions indicate a resistant interaction, and tan lesions indicate a susceptible interaction. F, Relative expression of *P. pachyrhizi* α -tubulin mRNA relative to soybean ubiquitin following BPMV and *P. pachyrhizi* inoculation. Each sample includes three biological and three technical replicates. The empty vector (C) sample has significantly lower (2–3-fold) expression of the fungal α -tubulin than either of the *Rpp4* candidate gene VIGS constructs. Asterisks designate $P < 0.01$ from a *t* test of *Rpp4*-silenced plants (D and E) compared with empty vector-silenced plants (C).

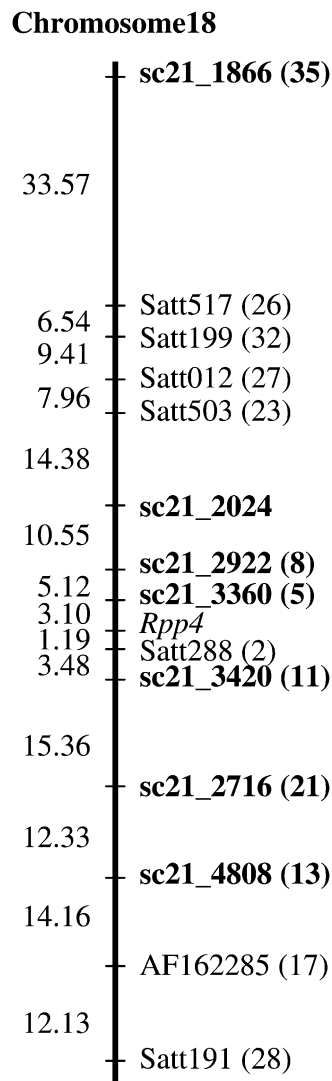


Figure 3. Genetic map of the PI459025B and BRS184 F2:3 population with newly developed markers. The genetic map shows chromosome 18 surrounding the *Rpp4* locus. New markers added to the map developed by Silva et al. (2008) include sc21_1866, sc21_2024, sc21_2716, sc21_2922, sc21_3420, sc21_3360, and sc21_4808. Recombination events (in relation to *Rpp4* phenotype) are indicated in parentheses to the right of each marker. Distance between markers (in cM) is shown on the left side of the map.

from resistant and susceptible lines that were infected with ASR or mock inoculated. In addition, genomic DNA of PI459025B and Wm82 was used as templates with primers *Rpp4*_NB_F/R, and the amplification products were cloned and sequenced to determine the amplification efficiency of the primers for all *Rpp4* candidate genes in both genotypes (Table I). A total of 96 clones were sequenced from each treatment × time point sample, resulting in 384 total clones per genotype. In resistant PI459025B, we detected 365 clones of *Rpp4C4* (PI459025B), two clones of *Rpp4C2* (PI459025B), and no clones of *Rpp4C1* (PI459025B),

Rpp4C3 (PI459025B), and *Rpp4C5* (PI459025B) across all time points and treatments. In the susceptible Wm82, we detected 337 clones of *Rpp4C3* (Wm82), eight clones of *Rpp4C2* (Wm82), and no clones of *Rpp4C1* (Wm82). While all genes could be detected in the genomic samples, the amplification efficiencies were not equal, making interpretation of the expression results more complex. Therefore, we used the observed amplification efficiencies to calculate the expected number of clones we would find for each gene if all genes were expressed equally (Table I). Based on these calculations, all of the genes should have been observed in the cDNA samples. However, only *Rpp4C4* (PI459025B) and *Rpp4C3* (Wm82) were detected in large numbers in the resistant and susceptible reactions, respectively. These results demonstrate that the differences in *Rpp4* candidate gene expression detected by semiquantitative RT-PCR in the resistant samples are due to *Rpp4C4* (PI459025B), making it the primary candidate for *Rpp4*-mediated resistance.

Evidence of Recombination in the *Rpp4* Locus

The structure of the *Rpp4* candidate genes can be examined to find evidence of duplication and recombination. *Rpp4C1*, *Rpp4C2*, *Rpp4C3*, and *RLG* (all Wm82) have 46, 47, 50, and 38 LRR motifs, respectively (Pfam E < 0.024; Finn et al., 2006; Fig. 1; Supplemental Fig. S3). We have isolated the amino acid sequence of each of the LRRs from these genes and aligned them using ClustalW (Thompson et al., 1994). LRRs that share greater than 80% amino acid identity are connected by a line as depicted in Supplemental Figure S3. Within each gene, the first 23 LRRs are unique and do not share greater than 80% amino acid identity with each other. Instead, LRRs located at the same position in each of the genes show the greatest amino acid identity. However, after LRR23, there is clear duplication and shuffling of LRR domains. For example, in *Rpp4C1*, *Rpp4C2*, and *Rpp4C3*, LRRs 24 to 31 have been duplicated to form LRRs 32 to 38 and 39 to 46. To see if recombination has also had a role in the evolution of these genes, we examined insertion/deletions (indels) across the entire length of *Rpp4C1* (Wm82), *Rpp4C2* (Wm82), and *Rpp4C3* (Wm82; Supplemental Table S5). If no recombination had occurred between the three genes, *Rpp4C2* (Wm82) and *Rpp4C3* (Wm82), which were duplicated more recently and share greater than 95% nucleotide identity, would share a majority of their indels along their entire length. If a single recombination event had occurred between *Rpp4C1* (Wm82) and *Rpp4C2* (Wm82) following the *Rpp4C2*/*Rpp4C3* duplication, *Rpp4C2* (Wm82) would share one stretch of indels with *Rpp4C1* (Wm82) and a second stretch of indels with *Rpp4C3* (Wm82). If multiple recombination events had occurred, all of the *Rpp4* candidate genes would share small patches of indels, as shown in Supplemental Table S5. Therefore, these LRR motif and indel data confirm that duplication and recombination events have occurred within the *Rpp4*

Table 1. Cloning of expression (cDNA) and genomic (DNA) PCR products of *Rpp4* candidate genes

Sample ^a	<i>Rpp4C1</i> ^b	<i>Rpp4C2</i> ^b	<i>Rpp4C3</i> ^b	<i>Rpp4C4</i> ^b	<i>Rpp4C5</i> ^b
PI459025B/ASR/12 hai	0	0	0	93	0
PI459025B/ASR/72 hai	0	2	0	90	0
PI459025B/Mock/12 hai	0	0	0	87	0
PI459025B/Mock/72 hai	0	0	0	95	0
PI459025B total ^c	0	2	0	365	0
Expected expression ^d	41	77	27	122	100
PI459025B genomic DNA	9	7	6	27	22
Wm82/ASR/12 hai	0	6	80		
Wm82/ASR/72 hai	0	1	88		
Wm82/Mock/12 hai	0	1	79		
Wm82/Mock/72 hai	0	0	90		
Wm82 total ^c	0	8	337		
Expected expression ^d	136	105	105		
Wm82 genomic DNA	35	27	27		

^aSamples refer to cDNA or genomic DNA. All RNA samples are named by genotype/treatment/time point, where genotype is PI459025B (resistant) or Wm82 (susceptible), treatment is either ASR infected or mock infected, and the time points are 12 or 72 hai. ^b*Rpp4C1* to *Rpp4C5* sequence numbers were determined by comparing the sequence from the cDNA reaction with the DNA sequences. *Rpp4C4* and *Rpp4C5* are not reported for Wm82 because they are not present in the Wm82 genome sequence. ^cTotal expression determined by summing all treatment/time point bulks for each genotype. These totals were compared with the expected expression for each *Rpp4* candidate gene. ^dExpected expression determined by taking the ratio of the genomic amplicons to the total number of cDNA sequences, assuming that efficiency of amplification is equivalent in DNA and cDNA.

candidate gene cluster of Wm82 and have also likely occurred at the *Rpp4* locus of PI459025B.

The *Rpp4* Locus Is Duplicated in the Soybean Genome

In order to examine the effect of whole genome duplication events (Shoemaker et al., 1996; Schlueter et al., 2004) on the evolution of the *Rpp4* candidate genes in soybean, the region homologous to the *Rpp4* locus was identified. Scaffold_21 (Supplemental Fig. S5), which includes the *Rpp4* locus, was compared with the 7X genome assembly (version Glyma0; United States Department of Energy Joint Genome Institute, 2008) using BLASTN (Altschul et al., 1997). Scaffold_75 shared the greatest synteny with scaffold_21 (Supplemental Fig. S5). BLASTN analyses against the sequences of previously mapped markers placed scaffold_75 on soybean chromosome 9 (LG K). Comparison of scaffold_21, scaffold_75, and our BAC sequence data revealed an approximately 162-kb deletion in scaffold_75 relative to our sequence and scaffold_21. Clear evidence of synteny could be observed in the regions surrounding the deletion (Supplemental Fig. S5). Genes encompassed by the deleted region include the *Rpp4* candidate genes (Q9ZT68), one gene prediction with homology to an uncharacterized protein (Q01HF9), one gene prediction with homology to a cellulase protein (Q43105), and seven gene predictions with homology to repetitive elements (Q9SIM3, Q6UUN7, Q5MG92, Q9C739, Q9M2D1, Q2R3T4, and Q9ZRJ0; Supplemental Figs. S2 and S5). Since *RLG* shows significant nucleotide identity to the

candidate *Rpp4* genes, scaffold_48 (chromosome 1, LG D1A) was also compared with scaffold_21. No evidence of synteny between the two regions was detected.

DISCUSSION

The *Rpp4* ASR Resistance Gene Is a Member of the CC-NBS-LRR Family of Disease Resistance Genes

Using molecular markers, we developed a BAC contig corresponding to the *Rpp4* locus in the susceptible cv Wm82. Sequencing of two BACs within this region identified three candidate disease resistance genes belonging to the CC-NBS-LRR family of disease resistance genes. Sequencing of genomic DNA from the resistant (PI459025B) and susceptible (Wm82) genotypes confirmed the presence of five and three candidate genes in these genotypes, respectively. We used VIGS to demonstrate that silencing of *Rpp4* diminished resistance in PI459025B, confirming that one of the *Rpp4* candidate genes is responsible for resistance. Markers developed from the *Rpp4* contig defined the region responsible for *Rpp4*-mediated resistance. Only *Rpp4C2* (PI459025B) and likely *Rpp4C4* (PI459025B) and *Rpp4C5* (PI459025B) mapped within this region. Sequence analyses of the RT-PCR products from the resistant (PI459025B) and susceptible (Wm82) genotypes demonstrated that *Rpp4C4* (PI459025B) is almost exclusively expressed in the resistant line, while the other gene sequences were almost undetected. Based on our analyses and the added knowledge that *Rpp4C4* (PI459025B) is not present in Wm82, *Rpp4C4*

(PI459025B) is the primary candidate gene for *Rpp4*-mediated resistance to ASR.

The *Rpp4* Candidate Resistance Genes Show Greatest Similarity to the RGC2 Family of Disease Resistance Genes from Lettuce

The RGC2 family contains the largest genes to be characterized in the NBS-LRR family to date. Each gene spans 15 to 25 kb, and the estimated number of RGC2 genes in the lettuce genome varies from 14 to 40 copies depending on the genotype (Meyers et al., 1998; Kuang et al., 2004). One member of the RGC2 gene family, *Dm3*, confers resistance to downy mildew (*Bremia lactucae*; Meyers et al., 1998; Shen et al., 2002), while other family members are responsible for up to eight different *Dm* specificities and root aphid resistance (Crute and Dunn, 1980; Hulbert and Michelmore, 1985; Farrara et al., 1987; Maisonneuve et al., 1994). All of the individual specificities confer resistance to downy mildew in a gene-for-gene manner; however, only single specificities are usually identified in individual genotypes. The phenotypes associated with the expression of different RGC2 family members vary. *Dm3* and *Dm14* condition an immune response where resistance occurs in cells initially infected by the pathogen, resulting in a reaction that is not detectable macroscopically. For *Dm16*, the response to *B. lactucae* is delayed until the haustoria penetrate cells beyond the initial infection, resulting in a visible yet still incompatible interaction (Wroblewski et al., 2007). The delayed reaction conditioned by *Dm16* is much like the reaction observed for *Rpp4*-mediated resistance to ASR.

Microscopic and Microarray Evidence Supports the Involvement of *Rpp4* Candidate Genes in *Rpp4*-Mediated Resistance

Thus far, in depth microscopic analyses of ASR resistance reactions have only been reported for *Rpp2* (Hoppe and Koch, 1989). Microscopic analysis of the reddish-brown phenotype conferred by *Rpp2* suggests that the early life cycle of ASR is similar in both susceptible and *Rpp2*-resistant soybean accessions. The similarity in development concludes by 2 d after inoculation, after which localized host cell death occurs at the time of mycelial development in the *Rpp2* genotype (Hoppe and Koch, 1989). These observations are consistent with the microarray experiments reported by van de Mortel et al. (2007). Resistant and susceptible plants both induce early expression of basal defense pathways. After 24 h, gene expression returns to normal levels. A second round of differential gene expression occurs at approximately 72 h and is observed first at a higher magnitude in the resistant genotype. The second defense response in the resistant genotype (72 hai) correlates with the time at which haustoria are expected to be rapidly developing (van de Mortel et al., 2007). A similar biphasic response is

observed in *Rpp4*-mediated resistance (M. van de Mortel, unpublished data). Since the resistance response coincides with haustoria development, it is likely that the *Rpp4* gene product detects the presence of an ASR effector secreted at this time. The predicted cytoplasmic localization of *Rpp4C4* (PI459025B) is compatible with the hypothesis that the corresponding ASR effector may be produced in the haustoria and secreted into the host cells where recognition would occur. This model for *Rpp4* function is consistent with the expression and secretion of AvrL5, AvrL6, or AvrL7 of flax rust and subsequent recognition by the intercellular resistance proteins L5, L6, or L7 of flax (*Linum usitatissimum*; Catanzariti et al., 2006).

What Causes Susceptibility to ASR?

By examining the indel patterns between the *Rpp4* candidate genes, we see clear evidence of the evolutionary forces acting on the *Rpp4* locus. Differences in gene number between Wm82 and PI459025B are likely due to duplication or unequal recombination. Furthermore, the observed pattern of indel swapping provides evidence of intragenic recombination. All of these phenomena have previously been reported in other disease resistance gene clusters (Baumgarten et al., 2003; Meyers et al., 2003). Given the similarity of all *Rpp4* candidate genes between genotypes, it is possible that small amino acid differences may play a key role in resistance. For example, a single amino acid change in the LRR of the rice blast resistance protein Pi-ta results in susceptibility (Bryan et al., 2000), and six amino acid changes in the LRR of the P and P2 resistance proteins of flax determine specificity differences (Dodds et al., 2001).

Why Is Resistance to ASR So Rare?

One of the challenges for controlling ASR outbreaks is the unique ability of *P. pachyrhizi* to infect a broad range of legume species within the Fabaceae. Bromfield (1984) identified 95 legume species, from over 42 genera, which could serve as hosts for overwintering and inoculum accumulation. More recently, new host species from 25 genera were identified in greenhouse evaluations, including 12 genera that had not been reported previously (Slaminko et al., 2008). While some sources of resistance have been detected in other legumes, resistance to ASR in soybeans is relatively rare. In order to identify new sources of resistance in soybean, Miles et al. (2006) evaluated the entire U.S. germplasm collection (16,000 accessions) against a mixture of five *P. pachyrhizi* isolates. After two rounds of evaluation, only 805 accessions were identified with even partial tolerance or resistance reactions to *P. pachyrhizi*, which correlates to less than 5% of the U.S. germplasm collection. The low frequency of the *Rpp4* candidate genes in the soybean genome may account for the lack of resistance to ASR in soybean accessions. BLASTN (Altschul et al., 1997) analyses of

the *Rpp4* candidate genes from Wm82 revealed that only four genes (*Rpp4C1*, *Rpp4C2*, *Rpp4C3*, and *RLG*) with significant homology were present in the soybean genome (also Wm82).

One possibility for the rarity of the *Rpp4* candidate genes is that the cost of maintaining *Rpp4* when no pathogen is present may be greater than the benefit of resistance during pathogen attack. For example, the *RPM1* resistance gene from *Arabidopsis* is also a member of the CC-NBS-LRR family (Pan et al., 2000) and confers resistance to *Pseudomonas syringae*. Susceptible plants completely lack the *RPM1* allele (Grant et al., 1995); however, evolutionary analyses of flanking regions suggest that the resistant and susceptible alleles have existed for over 9 million years (Stahl et al., 1999). Tian et al. (2003) created four independent pairs of *Arabidopsis* lines that were genetically identical except for the presence or absence of *RPM1*. Field trials were used to test plant fitness in the absence of *P. syringae*. The four lines containing *RPM1* had 9% fewer seeds per plant than those missing *RPM1*. Therefore, the cost of maintaining *RPM1* in a pathogen-free environment reduced yield. If the *Rpp4* candidate genes have a similar effect on yield, these genes would have been selected against, especially if selection in China (the geographic origin of soybean), the United States, and South America occurred without a threat from ASR.

The loss of genes by selection may explain the rarity of ASR resistance in soybean and other legumes. Soybean has undergone two whole genome duplication events in its evolutionary history (Shoemaker et al., 1996; Schlueter et al., 2004), with the most recent event reported as occurring 14.5 million years ago (Schlueter et al., 2004) or 3.5 to 5 million years ago (Blanc and Wolfe, 2004). It is uncertain whether this event was a whole genome duplication event (autopolyploidy) or whether two related genomes combined (allopolyploidy; Schlueter et al., 2004; Straub et al., 2006). Examination of the *Rpp4* duplicated region on chromosome 9 (LG K) revealed a 162-kb gap relative to the *Rpp4* locus. The gap spans the region containing the *Rpp4* candidate genes. If autopolyploidy had occurred, the two *Rpp4* loci could undergo unequal recombination, leading to the loss of the *Rpp4* candidate genes in one locus but not the other. However, if allopolyploidy occurred, when the two genomes came together it is possible that one had the *Rpp4* chromosome 18 (LG G) while the other had the *Rpp4* chromosome 9 (LG K) and no *Rpp4* resistance genes. If this were the case, there are likely ancestral legume species that completely lack the *Rpp4* candidate resistance genes in their genomes. This may help explain *P. pachyrhizi*'s broad host range among legumes.

Future Directions

The identification of the gene responsible for *Rpp4*-mediated resistance is important not only for its direct

agronomic impact but also for evolutionary studies of *R* genes in genomes with polyploid histories. Given the rarity of ASR resistance genes in soybean, we will use the candidate genes identified in this project to detect novel sources of resistance to ASR in the broader legume family. These genes will be vital for building an arsenal against ASR in the commercial soybean germplasm. Identification of a candidate gene controlling *Rpp4*-mediated ASR resistance would not have been possible without the use of forward (map-based cloning) and reverse (VIGS) genetic approaches. In the future, the well-developed genetic map and the recently released soybean genome (United States Department of Energy Joint Genome Institute, 2008) will aid in the identification of candidate disease resistance gene sequences for other agronomically important diseases. VIGS technology will enhance these advancements by allowing researchers an efficient and cost-effective method for testing candidate gene function.

MATERIALS AND METHODS

BAC Contig Development

Molecular marker Satt288, which is linked to the *Rpp4* locus (Silva et al., 2008), was used to identify BAC clones from the Iowa State University/United States Department of Agriculture-Agricultural Research Service soybean (*Glycine max*) Wm82 (GM_WBa) library (Marek and Shoemaker, 1997) by PCR amplification (Supplemental Fig. S1). Marker Sat_143 was also used to screen the BAC library and anchor the *Rpp4* contig. End sequences from identified BACs were used to design primers (Supplemental Table S1) to extend and join the contigs. BAC clones from the University of Missouri-Columbia Wm82 (GM_WBb) library (Wu et al., 2008) were used to fill gaps when possible by comparison of publicly available BAC-end sequences using BLASTN (Altschul et al., 1997) and visualization of BLASTN output for 100% identity. BAC-end primers were designed using Oligo 6.8 (Molecular Biology Insights). PCR reagents and conditions are listed in Supplemental Protocol S1. PCR products were visualized by gel electrophoresis. BAC-end sequences were compared with the UniProt protein database (Apweiler et al., 2004) using BLASTX (Altschul et al., 1997).

Sequencing of *Rpp4* Candidate BACs

BACs GM_WBb0070A12 and GM_WBb176I01 were selected for complete sequencing due to the presence of *R* genes in the BAC-end sequences and estimated coverage of the detected *R* gene region. The BAC DNA was subcloned and sequenced using the manufacturer's recommendations and the following kits and supplies: Large-Construct Kit (Qiagen, no. 12462), TOPO Shotgun Subcloning Kit (Invitrogen, no. K7000-01), One Shot TOP10 Chemically Competent *Escherichia coli* (Invitrogen, no. C404003), miniprep solutions (Qiagen, P1 [no. 19051], P2 [no. 19052], and P3 [no. 19053]), 96-well unifilters and uniplates (Whatman, nos. 7770-0062 and 7701-1750), and ABI Big Dye version 3.1 chemistry protocol and Hi-Di formamide (Applied Biosystems, nos. 4311320 and 4337457). Sequencing was performed using an Applied Biosystems 3730 DNA Analyzer with a 96-capillary array. Sequences were trimmed and assembled using Sequencher version 4.7 default parameters with the exception of a minimum match percentage of 100% (Gene Codes Corporation). In order to maximize read lengths, forward and reverse reads for the same clone were preassembled prior to complete assembly.

Gene Prediction of Sequenced BACs

The sequenced contig developed from BACs GM_WBb0070A12 and GM_WBb0176I01 (208,603 bp; GenBank accession nos. FJ225394 and FJ225395) was divided into 2,000-bp pieces and compared with the UniProt

protein database (Apweiler et al., 2004) using BLASTX (Altschul et al., 1997) to identify genic regions (Supplemental Fig. S2). In addition, FGENESH (Salamov and Solovyev, 2000) was utilized to predict coding regions within the entire region. The NetPlantGene Server (Hebsgaard et al., 1996), BLASTX comparisons with UniProt, and BLASTN and TBLASTX comparisons with the Dana Farber Soybean Gene Index (Lee et al., 2005) were used to predict and confirm exon positions and splice sites. The three candidate *R* genes identified were named *Rpp4C1*, *Rpp4C2*, and *Rpp4C3* (Wm82; Fig. 1). Since correspondences between alleles of the same gene could not be confirmed, the genotype of each allele is indicated following the gene name. To determine the abundance of genes with homology to the *Rpp4* candidate genes in the soybean genome, the nucleotide sequences of the three candidate genes were compared with the soybean whole genome assembly (version Glyma0; United States Department of Energy Joint Genome Institute, 2008) using MEGA-BLAST ($E < 10^{-4}$). One additional gene was identified and designated *RLG* (Wm82).

Amplification of *Rpp4* Candidate Genes from the Resistant Genotype (PI459025B)

An alignment of the coding sequences of the three *Rpp4* candidate genes and *RLG* from Wm82 was made using ClustalW with the default settings (Thompson et al., 1994). Oligo 6.8 (Molecular Biology Insights) was used to design conserved primers that would amplify all four Wm82 genes (three *Rpp4* candidate genes and *RLG*) and would be used to amplify similar genes from PI459025B (*Rpp4*_NB_F/R; Supplemental Table S3). PCR, cloning, and sequencing were performed using the following reagents: Hi-Fi Platinum Taq DNA polymerase (Invitrogen, no. 10342-053), TA cloning kit (Invitrogen, no. K4500-01), and One Shot TOP10 Chemically Competent *E. coli* (Invitrogen, no. C404003). Additional details are provided in Supplemental Protocol S1.

Functional Analyses of the *Rpp4* Candidate Genes via VIGS

Two primer pairs (BPMV_NBD_F/R and BPMV_LRR_F/R; Supplemental Table S2) were used for PCR amplification using BAC DNA (GM_WBb0070A12 and GM_WBb0176I01) as template. The PCR products were directionally cloned into RNA2 of the BPMV VIGS vector (Zhang et al., 2009) that we adapted for high-throughput cloning by modification with the topoisomerase enzyme (Invitrogen). The orientation and identity of the VIGS inserts were confirmed by sequencing using a vector-specific forward primer (1548F, 5'-CAAGAGAAAGATTATTGGAGGGA-3'). To generate inoculum for the VIGS experiments, BPMV RNA1 and RNA2 DNA clones were used to bombard Wm82 leaves at 14 d after sowing, as described previously (Zhang et al., 2009). BPMV-infected leaf tissue was collected at 3 weeks after bombardment, lyophilized, and stored at -20°C . Plants of the resistant genotype (PI459025B) were germinated in a growth chamber at the Foreign Disease-Weed Science Research Unit at Fort Detrick, Maryland. Two weeks following germination, plants were dusted with carborundum and the first true leaves were rub inoculated with the lyophilized leaf tissue inoculum corresponding to the BPMV_NBD_F/R or BPMV_LRR_F/R construct. Six plants from PI459025B were infected with each construct. Three weeks following BPMV infection, plants were transferred to the U.S. Department of Agriculture Plant Pathogen BSL-3 containment facility at Fort Detrick (Melching et al., 1983), inoculated with *Phakopsora pachyrhizi* isolate LA04-1, and placed in a dew chamber overnight. Plants were then moved to a greenhouse and evaluated for a resistant (red/brown lesions) or susceptible (tan lesions) phenotype 2 weeks after inoculation with *P. pachyrhizi*. In addition to the BPMV vector-inoculated plants, controls included mock-inoculated plants (the same experimental conditions as the VIGS-treated plants but rub inoculated with buffer instead of the BPMV inoculum), plants that were inoculated with a BPMV vector lacking an insert, and plants that were not treated prior to inoculation with *P. pachyrhizi*. All of the control plants were infected with *P. pachyrhizi* as described for the experimental plants. Three independent replicates of the experiment were performed with similar results.

Taqman RT-PCR Analysis of *Rpp4C* mRNA Expression and *P. pachyrhizi* Growth in *Rpp4C*-Silenced Plants

RNA was extracted from the *Rpp4C* NBD- and LRR-silenced plants using the Qiagen Plant RNeasy kit and subsequently DNase treated. *Rpp4C* cluster

expression was assessed in control and VIGS-treated plants using the primers *Rpp4TMF* (5'-GTTTGCTTCAAGGGGTCCACA-3') and *Rpp4TMR* (5'-AAC-ATCCCGCACAAATGTCATGC-3') and the probe *Rpp4TMP* (5'-TGGTGGGAAGTCTCTCTCATGACCGCCT-3'). The probe was modified with 6-carboxy fluorescein at the 5' end and with Blackhole Quencher I at the 3' end (Integrated DNA Technologies). *P. pachyrhizi* growth was assessed in control and VIGS-treated plants by quantifying the constitutively expressed *ASR* α -tubulin gene by Taqman qRT-PCR as described (van de Mortel et al., 2007). Three biological and three technical replicates were included in the analysis. The iScript One-Step RT-PCR kit for probes (Bio-Rad) was used according to the manufacturer's protocol with 25 ng of total RNA, 300 nM final concentration of primers, and 150 nM probe in the following RT-PCR program: cDNA synthesis for 10 min at 50°C , iScript reverse transcription inactivation for 5 min at 95°C , PCR cycling at 95°C for 10 s, and data collection for 30 s at the extension temperature of 60°C for 45 cycles. *Rpp4C* and *ASR* α -tubulin signals were normalized to the soybean *ubiquitin3* gene (GenBank accession no. D28123), which is not differentially expressed in response to ASR (van de Mortel et al., 2007).

SSR Marker Development and Screening

Six SSRs were identified within and immediately surrounding the *Rpp4* candidate gene cluster, including two SSR markers defining the genes with the Wm82 *Rpp4* cluster (Supplemental Table S4). Forty-eight additional SSRs were identified in genome scaffold_21 (version Glyma0; United States Department of Energy Joint Genome Institute, 2008) surrounding the cluster. Primers surrounding the SSRs were designed using Oligo 6.8 (Molecular Biology Insights; Supplemental Table S4) and screened on a panel of 10 germplasm sources: PI459025B, BRS184, PI230970, BSR101, PI437654, Noirl, Minsoy, *Glycine soja* (PI468916), *G. max* (A81-356022), and Wm82. PCR conditions are provided in Supplemental Protocol S1. Six primer pairs, polymorphic between PI459025B (*Rpp4*) and BRS184 (*rpp4*), were mapped in an F2:3 population as described previously (Silva et al., 2008).

Expression Analyses of the *Rpp4* Candidate Genes in Resistant and Susceptible Interactions with ASR

In order to design a single pair of primers that could amplify all of the *Rpp4* candidate genes (*Rpp4F*/R; Supplemental Table S3), an alignment of the coding sequences of the three *Rpp4* candidate genes and *Rpp4L* was made using ClustalW (Thompson et al., 1994). Primers designed for a soybean tubulin gene (Dana Farber Soybean Gene Index no. TC204178; Supplemental Table S3; Graham et al., 2002; O'Rourke et al., 2007) were used to normalize RNA levels (Supplemental Table S3). Primers were designed so that amplicons spanned an intron and could control for DNA contamination.

RNA was extracted from susceptible (Wm82) and resistant (PI459025B) infected and mock-infected plants grown at the U.S. Department of Agriculture containment facility at Fort Detrick. Three leaflets of the second trifoliate leaf of two plants (six leaflets total) were collected at 12, 24, 72, and 216 h. Leaves were immediately frozen in liquid nitrogen and stored at -80°C . Leaf tissue was ground in liquid nitrogen, and RNA was extracted using 1 mL of Tri Reagent (no. TR118; Molecular Research Center) according to the manufacturer's protocols. RNA samples were stored as pellets at -80°C and shipped to Iowa State University, where the RNA samples were resuspended in 50 μL of Nuclease-free water (Applied Biosystems, no. AM9937). Total RNA samples were DNase-d using TURBO DNA-free (Applied Biosystems, no. AM1907) according to the manufacturer's directions.

Expression analyses were performed by qRT-PCR using the *Rpp4F*/R primers that amplify all *Rpp4* candidate genes from both genotypes (Table I). Invitrogen's SuperScriptIII Platinum SYBR Green One-Step qRT-PCR Kit (no. 11736-051) was used for 50- μL reactions with 30 ng of total RNA for sample reactions following the manufacturer's instructions. Cycling conditions are provided in Supplemental Protocol S1. The PCRs were run in a Stratagene Mx3000P followed by a dissociation curve, taking a fluorescent measurement at every degree between 55°C and 95°C . The fold change was calculated from the differences in threshold cycle (C_t) using the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001). Each sample was also run in triplicate and normalized against tubulin amplification to ensure that the differential expression was not due to differing amounts of initial RNA template added to each sample. RT reactions to make cDNA were made using RETROscript (Applied Biosystems, no. AM1710) according to the manufacturer's instructions. PCR components and conditions are provided in Supplemental Protocol S1. RT-PCR products

were cloned, and 96 colonies were sequenced from each time point/treatment bulk sample. Sequenced products were compared with the genomic sequences described previously to determine which gene copy was being expressed in each sample. Based on a Moloney murine leukemia virus enzyme error rate of one error per 3,500 bases (provided by Ambion technical support), RT sequences were considered a match at 99% identity, provided that the base change did not match the genomic sequence of any of the *Rpp4* candidate genes. *P. pachyrhizi* growth was assessed for all time points as described previously for VIGS samples.

Protein Domain Analyses of the *Rpp4* Candidate Genes

Searches against the Pfam database (Finn et al., 2006) were used to identify conserved domains within the three *Rpp4* candidate genes and *RLG* from Wm82 (Fig. 1; Supplemental Fig. S3). In addition, the program COILS (Lupas et al., 1991) was used to predict a CC domain. LRRs from each gene were combined in a single fasta file and compared utilizing ClustalW (Thompson et al., 1994; Supplemental Fig. S3).

Analyses of Indels in the *Rpp4* Candidate Genes

ClustalW (default settings) was used to align the predicted gene sequences for *Rpp4C1*, *Rpp4C2*, and *Rpp4C3* (all Wm82), including 2,000 bases upstream and downstream of each gene (Supplemental Table S5). The sequence alignment was analyzed to identify conserved indel sites shared by two of the three genes. The identified indels were scored for each gene, and the locations were recorded relative to *Rpp4C1* (Wm82).

Alignment of the *Rpp4* Contig with the Soybean Whole Genome Assembly and Identification of a *Rpp4* Homologous Region

The sequenced BACs and BAC-end sequences from the *Rpp4* contig were used to identify the soybean genome scaffold (version Glyma0; United States Department of Energy Joint Genome Institute, 2008) corresponding to the *Rpp4* locus using BLASTN (Altschul et al., 1997). Scaffold_21 was screened for the presence of additional candidate resistance genes by dividing the region between markers Sat_143 and A885_1 on scaffold_21 (approximately 1.065 Mb) into 2,000-bp pieces and comparing them with the UniProt protein database (Apweiler et al., 2004) using BLASTX (Altschul et al., 1997). To determine if the *Rpp4* locus was duplicated in the soybean genome, the same 2,000-bp intervals of scaffold_21 were blasted (BLASTN; Altschul et al., 1997) against the entire genome assembly (version Glyma0; United States Department of Energy Joint Genome Institute, 2008). The first hit identified was the sequenced region itself (scaffold_21), and the second best hit was presumed to represent a possibly duplicated region. The programs WebACT (Abbott et al., 2005) and JDotter (Brodie et al., 2004) were used to visualize the sequence identity between the duplicated regions (Supplemental Fig. S5).

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers FJ225394 and FJ225395.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. The Wm82 *Rpp4* BAC contig.

Supplemental Figure S2. Annotation of the sequenced BACs M70A12 and M176101.

Supplemental Figure S3. Comparison of protein motifs in the three *Rpp4* candidate genes (*Rpp4C1*–*Rpp4C3* [Wm82]) and a homolog (*RLG* [Wm82]) identified from the soybean genome sequence.

Supplemental Figure S4. Partial sequence alignment of the *Rpp4* candidate genes from Wm82 and PI459025B.

Supplemental Figure S5. Identification of a duplicated region on chromosome 9 syntenic to the *Rpp4* locus.

Supplemental Table S1. Primers designed from BAC-end sequences to develop the Wm82 *Rpp4* BAC contig.

Supplemental Table S2. Primers used to make VIGS constructs.

Supplemental Table S3. Primers used for expression analyses and cloning of the *Rpp4* candidate genes in Wm82 and PI459025B.

Supplemental Table S4. SSR markers developed for the *Rpp4* locus and the surrounding region from the soybean genome scaffold_21.

Supplemental Table S5. Recombination activity identified using indels from the Wm82 *Rpp4* candidate genes.

Supplemental Protocol S1. PCR and conditions.

ACKNOWLEDGMENTS

We acknowledge the contributions made by Lori Lincoln, Sehiza Grosic, Jaime Dittman, Greg Peiffer, Jamie O'Rourke, and Heather Davidson.

Received December 17, 2008; accepted February 24, 2009; published February 27, 2009.

LITERATURE CITED

- Abbott JC, Aanensen DM, Rutherford K, Butcher S, Spratt BG (2005) WebACT: an online companion for the Artemis Comparison Tool. *Bioinformatics* **21**: 3665–3666
- Altschul SE, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402
- Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**: D115–D119
- Baumgarten A, Cannon S, Spangler R, May G (2003) Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics* **165**: 309–319
- Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**: 1667–1668
- Bonde MR, Nester SE, Austin CN, Stone CL, Frederick RD, Hartman GL, Miles MR (2006) Evaluation of virulence of *Phakopsora pachyrhizi* and *P. meibomia* isolates. *Plant Dis* **90**: 708–716
- Brodie R, Roper RL, Upton C (2004) JDotter: a Java interface to multiple dotplots generated by dotter. *Bioinformatics* **20**: 279–281
- Bromfield KR (1984) Soybean Rust, Monograph No. 11. American Phytopathological Society, St. Paul
- Bromfield KR, Hartwig EE (1980) Resistance to soybean rust and mode of inheritance. *Crop Sci* **20**: 254–255
- Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R, Valent B (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *Plant Cell* **12**: 2033–2046
- Catanzariti A, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis JG (2006) Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**: 243–256
- Cheng YW, Chan KL (1968) The breeding of rust resistant soybean Tainung 3. *J Taiwan Agr Res* **17**: 30–34
- Crute IR, Dunn JA (1980) An association between resistance to root aphid (*Pemphigus bursarius* L.) and downy mildew (*Bremia lactucae* Regel) in lettuce. *Euphytica* **29**: 483–488
- Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. *Nature* **411**: 826–833
- Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JDG (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell* **10**: 1915–1926
- Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JD (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* **84**: 451–459
- Dodds PN, Lawrence GJ, Ellis JG (2001) Six amino acid changes confined to the leucine-rich repeat β -strand/ β -turn motif determine the difference between the *P* and *P2* rust resistance specificities in flax. *Plant Cell* **13**: 163–178

- Farrara BF, Illott TW, Michelmore RW (1987) Genetic analysis of factors for resistance to downy mildew (*Bremia lactucae*) in species of lettuce (*Lactuca sativa* and *L. serriola*). *Plant Pathol* 36: 499–514
- Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34: D247–D251
- Gao H, Bhattacharyya MK (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
- García A, Calvo ES, de Souza Kiihl RA, Harada A, Hiromoto DM, Vieira LG (2008) Molecular mapping of soybean rust (*Phakopsora pachyrhizi*) resistance genes: discovery of a novel locus and alleles. *Theor Appl Genet* 117: 545–553
- Graham MA, Marek LF, Shoemaker RC (2002) Organization, expression and evolution of a disease resistance gene cluster in soybean. *Genetics* 162: 1961–1977
- Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1995) Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. *Science* 269: 843–846
- Hartwig EE (1986) Identification of a fourth major gene conferring resistance to soybean rust. *Crop Sci* 26: 1135–1136
- Hartwig EE, Bromfield KR (1983) Relationships among three genes conferring specific resistance to rust in soybeans. *Crop Sci* 23: 237–239
- Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996) Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Res* 24: 3439–3452
- Hidayat OO, Somaatmadja S (1977) Screening of soybean breeding lines for resistance to soybean rust (*Phakopsora pachyrhizi* Sydow). *Soybean Rust Newsl* 1: 9–22
- Hoppe HH, Koch E (1989) Defense reactions in host and nonhost plants against the soybean rust fungus (*Phakopsora pachyrhizi* Syd.). *J Phytopathol* 125: 77–88
- Hulbert SH, Bennetzen JL (1991) Recombination at the *Rp1* locus of maize. *Mol Gen Genet* 226: 377–382
- Hulbert SH, Michelmore RW (1985) Linkage analysis of genes for resistance to downy mildew (*Bremia lactucae*) in lettuce (*Lactuca sativa*). *Theor Appl Genet* 70: 520–528
- Jones DA, Jones JDG (1997) The role of leucine-rich repeat proteins in plant defenses. *Adv Bot Res* 24: 89–167
- Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JD (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science* 266: 789–793
- Jones JD, Dangl JL (2006) The plant immune system. *Nature* 444: 323–329
- Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. *Proc Natl Acad Sci USA* 93: 11746–11750
- Kuang HH, Woo SS, Meyers BC, Nevo E, Michelmore RW (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* 16: 2870–2894
- Lee Y, Tsai J, Sunkara S, Karamycheva S, Perteu G, Sultana R, Antonescu V, Chan A, Cheung F, Quackenbush J (2005) The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res* 33: D71–D74
- Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet* 20: 116–122
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_T} method. *Methods* 25: 402–408
- Lupas A, Van Dyke M, Stock J (1991) Predicted coiled coils from protein sequences. *Science* 252: 1162–1164
- Maisonneuve B, Bellec Y, Anderson P, Michelmore RW (1994) Rapid mapping of two genes for resistance to downy mildew from *Lactuca serriola* to existing clusters of resistance genes. *Theor Appl Genet* 89: 96–104
- Marek LF, Shoemaker RC (1997) BAC contig development by fingerprint analysis in soybean. *Genome* 40: 420–427
- McLean RJ, Byth DE (1980) Inheritance of resistance to rust (*Phakopsora pachyrhizi*) in soybean. *Aust J Agric Res* 31: 951–956
- Melching JS, Bromfield KR, Kingsolver CH (1983) The plant pathogen containment facility at Frederick, Maryland. *Plant Dis* 67: 717–722
- Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 10: 1817–1832
- Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. *Curr Opin Plant Biol* 8: 129–134
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15: 809–834
- Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res* 8: 1113–1130
- Miles M, Frederick R, Hartman G (2006) Evaluation of soybean germplasm for resistance to *Phakopsora pachyrhizi*. *Plant Health Prog* 10: doi/10.1094/PHP-2006-0104-01-RS
- Ogle HJ, Byth DE, McLean R (1979) Effect of rust (*Phakopsora pachyrhizi*) on soybean yield and quality in southeastern Queensland. *Aust J Agric Res* 30: 883–893
- O'Rourke JA, Charlson DV, Gonzalez DO, Vodkin LO, Graham MA, Cianzio SR, Grusak MA, Shoemaker RC (2007) Microarray analysis of iron deficiency chlorosis in near-isogenic soybean lines. *BMC Genomics* 8: 476
- Pan Q, Liu YS, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D, Fluhr R (2000) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons tomato and Arabidopsis. *Genetics* 155: 309–322
- Patil VS, Wuike RV, Thakare CS, Chirame BB (1997) Viability of uredospores of *Phakopsora pachyrhizi* Syd. at different storage conditions. *J Maharashtra Agric Univ* 22: 260–261
- Rairdan GJ, Collier SM, Sacco MA, Baldwin TT, Boettlich T, Moffett P (2008) The coiled-coil and nucleotide binding domains of the potato Rx disease resistance protein function in pathogen recognition and signaling. *Plant Cell* 20: 739–751
- Richter TE, Ronald PC (2000) The evolution of disease resistance genes. *Plant Mol Biol* 42: 195–204
- Salamov AA, Solov'yev VV (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res* 10: 516–522
- Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47: 868–876
- Schneider RW, Hollier CA, Whitam HK, Palm ME, McKemy JM, Hernandez JR, Levy L, DeVries-Paterson R (2005) First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. *Plant Dis* 89: 774
- Shen KA, Chin DB, Arroyo-García R, Ochoa OE, Lavelle DO, Wroblewski T, Meyers BC, Michelmore RW (2002) *Dm3* is one member of a large constitutively expressed family of nucleotide binding site-leucine-rich repeat encoding genes. *Mol Plant Microbe Interact* 15: 251–261
- Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* 144: 329–338
- Silva DCG, Yamanaka N, Brogin RL, Arias CAA, Nepomuceno AL, Di Mauro AO, Pereira SS, Nogueira LM, Passianotto ALL, Abdelnoor RV (2008) Molecular mapping of two loci that confer resistance to Asian rust in soybean. *Theor Appl Genet* 117: 57–63
- Singh BB, Thapliyal PN (1977) Breeding for resistance to soybean rust in India. In RE Ford, JB Sinclair, eds, *Rust of Soybean: The Problem and Research Needs*. Report of a workshop held in Manila, The Philippines. University of Illinois, Urbana-Champaign, IL, pp 62–65
- Slaminko TL, Miles MR, Bonde MR, Hartman GL (2008) New legume hosts of *Phakopsora pachyrhizi* based on greenhouse evaluations. *Plant Dis* 92: 767–771
- Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald PC (1997) Evolution of the rice *Xa21* disease resistance gene family. *Plant Cell* 9: 1279–1287
- Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of Arabidopsis. *Nature* 400: 667–671
- Straub SCK, Pfeil BE, Doyle JJ (2006) Testing the polyploid past of soybean using a low-copy nuclear gene: is *Glycine* (Fabaceae: Papilionoideae) an auto- or allopolyploid? *Mol Phylogenet Evol* 39: 580–584
- Sun X, Cao Y, Yang Z, Xu C, Li X, Wang S, Zhang Q (2004) *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. *Plant J* 37: 517–527

- Tameling WIL, Elzinga SDJ, Darmin PS, Vossen JH, Takken FLW, Haring MA, Cornelissen BJC** (2002) The tomato R gene products I-2 and Mi-1 are functional ATP binding proteins with ATPase activity. *Plant Cell* **14**: 2929–2939
- Tao Y, Xie Z, Chen W, Glazebrook J, Chang HS, Han B, Zhu T, Zou G, Katagiri F** (2003) Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell* **15**: 317–330
- Thompson JD, Higgins DG, Gibson TJ** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680
- Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J** (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**: 74–77
- United States Department of Energy Joint Genome Institute** (2008) Soybean Genome Project Version Glyma0. <http://www.phytozome.net/soybean> (January 1, 2008)
- van de Mortel M, Recknor JC, Graham MA, Nettleton D, Dittman JD, Nelson RT, Godoy CV, Abdelnoor RV, Almeida AMR, Baum TJ** (2007) Distinct biphasic mRNA changes in response to Asian soybean rust infection. *Mol Plant Microbe Interact* **20**: 887–899
- Wroblewski T, Piskurewicz U, Tomczak A, Ochoa O, Michelmore RW** (2007) Silencing of the major family of NBS-LRR-encoding genes in lettuce results in the loss of multiple resistance specificities. *Plant J* **51**: 803–818
- Wu X, Zhong G, Findley SD, Cregan P, Stacey G, Nguyen HT** (2008) Genetic marker anchoring by six-dimensional pools for development of a soybean physical map. *BMC Genomics* **9**: 28
- Zhang C, Yang C, Whitham SA, Hill JH** (2009) Development and use of an efficient DNA-based viral gene silencing vector for soybean. *Mol Plant Microbe Interact* **22**: 123–131