Rpp1 encodes a ULP1-NBS-LRR protein that controls immunity to Phakopsora pachyrhizi in soybean

Kerry F. Pedley  
*United States Department of Agriculture*

Ajay K. Pandey  
*United States Department of Agriculture*

Amy Ruck  
*United States Department of Agriculture*

Lori M. Lincoln  
*United States Department of Agriculture*

Steven A. Whitham  
*Iowa State University, swhitham@iastate.edu*

Follow this and additional works at: [https://lib.dr.iastate.edu/plantpath_pubs](https://lib.dr.iastate.edu/plantpath_pubs)

Part of the [Agricultural Science Commons](https://lib.dr.iastate.edu/agriculturalscience), [Agriculture Commons](https://lib.dr.iastate.edu/agriculture), [Plant Breeding and Genetics Commons](https://lib.dr.iastate.edu/plantbreedingandgenetics), and the [Plant Pathology Commons](https://lib.dr.iastate.edu/plantpathology)

The complete bibliographic information for this item can be found at [https://lib.dr.iastate.edu/plantpath_pubs/260](https://lib.dr.iastate.edu/plantpath_pubs/260). For information on how to cite this item, please visit [http://lib.dr.iastate.edu/howtocite.html](http://lib.dr.iastate.edu/howtocite.html).
Rpp1 encodes a ULP1-NBS-LRR protein that controls immunity to Phakopsora pachyrhizi in soybean

Abstract
Phakopsora pachyrhizi is the causal agent of Asian soybean rust. Susceptible soybean plants infected by virulent isolates of P. pachyrhizi are characterized by tan-colored lesions and erumpent uredinia on the leaf surface. Germplasm screening and genetic analyses have led to the identification of seven loci, Rpp1 – Rpp7, that provide varying degrees of resistance to P. pachyrhizi (Rpp). Two genes, Rpp1 and Rpp1b, map to the same region on soybean chromosome 18. Rpp1 is unique among the Rpp genes in that it confers an immune response (IR) to avirulent P. pachyrhizi isolates. The IR is characterized by a lack of visible symptoms, whereas resistance provided by Rpp1b – Rpp7 results in red-brown foliar lesions. Rpp1 maps to a region spanning approximately 150 Kbp on chromosome 18 between markers Sct_187 and Sat_064 in L85-2378 (Rpp1), an isolate developed from Williams 82 and PI 200492 (Rpp1). To identify Rpp1, we constructed a bacterial artificial chromosome (BAC) library from soybean accession PI 200492. Sequencing of the Rpp1 locus identified three homologous nucleotide binding site-leucine rich repeat (NBS-LRR) candidate resistance genes between Sct_187 and Sat_064. Each candidate gene is also predicted to encode an N-terminal ubiquitin-like protease 1 (ULP1) domain. Co-silencing of the Rpp1 candidates abrogated the immune response in the Rpp1 resistant soybean accession PI 200492, indicating that Rpp1 is a ULP1-NBS-LRR protein and plays a key role in the IR.

Keywords
Plant immunity, soybean rust, sumoylation, integrated decoy

Disciplines
Agricultural Science | Agriculture | Plant Breeding and Genetics | Plant Pathology

Comments

Rights
Works produced by employees of the U.S. Government as part of their official duties are not copyrighted within the U.S. The content of this document is not copyrighted.

Authors
Kerry F. Pedley, Ajay K. Pandey, Amy Ruck, Lori M. Lincoln, Steven A. Whitham, and Michelle A. Graham

This article is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/plantpath_pubs/260
Rpp1 encodes a ULP1-NBS-LRR protein that controls immunity to Phakopsora pachyrhizi in soybean

Kerry F. Pedley,1,† Ajay K. Pandey,1,3 Amy Ruck,1 Lori M. Lincoln,2 Steven A. Whitham,3 and Michelle A. Graham2,†

1United States Department of Agriculture–Agricultural Research Service (USDA-ARS), Foreign Disease-Weed Science Research Unit, Ft. Detrick, MD 21702; 2USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA 50011; 3Iowa State University, Department of Plant Pathology and Microbiology, Ames, IA 50011

12 Current address of A. K. Pandey: National Agri-Food Biotechnology Institute, Sector 81, Mohali, India 140306

†Corresponding authors: K. F. Pedley; E-mail: Kerry.Pedley@ars.usda.gov and M. A. Graham; E-mail: Michelle.Graham@ars.usda.gov

Keywords: Plant immunity, soybean rust, sumoylation, integrated decoy
21 **Commercial Endorsement Disclaimer** – Mention of trade names or commercial
products in this publication is solely for the purpose of providing specific information
and does not imply recommendation or endorsement by the U.S. Department of
Agriculture.

26 **Equal Opportunity/Non-Discrimination Statement** – The U.S. Department of
Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis
of race, color, national origin, age, disability, and where applicable, sex, marital status,
familial status, parental status, religion, sexual orientation, genetic information, political
beliefs, reprisal, or because all or a part of an individual’s income is derived from any
public assistance program. (Not all prohibited bases apply to all programs.) Persons
with disabilities who require alternative means for communication of program
information (Braille, large print, audiotape, etc.) should contact USDA’s TARGET Center
at (202) 720-2600 (voice and TDD). To file a complaint of discrimination write to USDA,
Director, Office of Civil Rights, 1400 Independence Avenue, SW, Washington, DC 20250-
9410 or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal
opportunity provider and employer.
**ABSTRACT**

*Phakopsora pachyrhizi* is the causal agent of Asian soybean rust. Susceptible soybean plants infected by virulent isolates of *P. pachyrhizi* are characterized by tan-colored lesions and erumpent uredinia on the leaf surface. Germplasm screening and genetic analyses have led to the identification of seven loci, *Rpp1* – *Rpp7*, that provide varying degrees of resistance to *P. pachyrhizi* (*Rpp*). Two genes, *Rpp1* and *Rpp1b*, map to the same region on soybean chromosome 18. *Rpp1* is unique among the *Rpp* genes in that it confers an immune response (IR) to avirulent *P. pachyrhizi* isolates. The IR is characterized by a lack of visible symptoms, whereas resistance provided by *Rpp1b* – *Rpp7* results in red-brown foliar lesions. *Rpp1* maps to a region spanning approximately 150 Kb on chromosome 18 between markers Sct_187 and Sat_064 in L85-2378 (*Rpp1*), an isoline developed from Williams 82 and PI 200492 (*Rpp1*). To identify *Rpp1*, we constructed a bacterial artificial chromosome (BAC) library from soybean accession PI 200492. Sequencing of the *Rpp1* locus identified three homologous nucleotide binding site-leucine rich repeat (NBS-LRR) candidate resistance genes between Sct_187 and Sat_064. Each candidate gene is also predicted to encode an N-terminal ubiquitin-like protease 1 (ULP1) domain. Co-silencing of the *Rpp1* candidates abrogated the immune response in the *Rpp1* resistant soybean accession PI 200492, indicating that *Rpp1* is a ULP1-NBS-LRR protein and plays a key role in the IR.
INTRODUCTION

Plants are a potential source of sugars and other nutrients to microbes in the environment. Thus, it is not surprising that many bacteria and fungi have evolved mechanisms to overcome the preformed structural and chemical barriers plants use to protect these resources. As a result, plants have evolved additional systems for detecting would-be pathogens and mounting defenses to thwart infection. Collectively these systems and the underlying mechanisms are referred to as plant immunity.

It is now well-established that plants utilize two distinct, but overlapping immune responses (Jones and Dangl, 2006). The recognition of microbe- or pathogen-associated molecular patterns (MAMPs or PAMPs) activates MAMP- or PAMP-triggered immunity (MTI or PTI) (Jones and Dangl, 2006; Dodds and Rathjen, 2010). To counter MTI, pathogens deliver effector proteins into the plant cell that interfere with signaling and other downstream defense responses, resulting in effector-triggered susceptibility (ETS). Recognition or detection of microbial effectors by plant resistance proteins (R proteins) elicits effector-triggered immunity (ETI), which often culminates with a rapid, localized form of cell death known as the hypersensitive response (HR) (Jones and Dangl, 2006; Boller and Felix, 2009; Dodds and Rathjen, 2010). The HR functions to limit the ability of the pathogen to spread beyond the point of attack. Due to the dramatic and irreversible nature of the HR, it is essential that plants tightly regulate ETI in the absence of a bona fide infection.

The majority of plant R proteins contain nucleotide-binding site leucine-rich repeat (NBS-LRR) domains and are encoded by plant resistance genes (R genes) (Martin...
Many NBS-LRR proteins can be further divided into three functional classes based upon the presence of either Toll/interleukin-1 receptor (TIR), coiled coil (CC), or resistance to powdery mildew (RPW8) domains located at the amino terminus (Jones et al., 2016; Shao et al., 2016). NBS-LRR proteins are involved in pathogen detection with the TIR-NBS-LRR and CC-NBS-LLR classes operating through distinct, yet overlapping signaling pathways (Aarts et al., 1998; Elmore et al., 2011). Although many aspects of R protein function remain to be elucidated, there is evidence for both direct and indirect interactions of NBS-LRR proteins with microbial effectors, which initiate signaling cascades that trigger defense responses (Jones and Dangl, 2006; Dodds and Rathjen, 2010; Jones et al., 2016).

Soybean rust (SBR) is an aggressive foliar disease of soybean (Glycine max) and other legumes caused by the obligate biotrophic fungus, Phakopsora pachyrhizi. Three distinct reaction phenotypes occur on soybean in response to P. pachyrhizi infection (Bromfield, 1984), and the outcome of the interaction appears to follow the gene-for-gene resistance model (Flor, 1946). Susceptible soybean plants infected with virulent isolates of P. pachyrhizi develop tan-colored lesions (TAN phenotype) with abundant, sporulating uredinia on the leaf surface. Resistant plants are characterized by either a complete lack of visible symptoms, termed the immune response (IR), or by the appearance of reddish-brown lesions (RB reaction) in response to avirulent isolates. The IR effectively prevents reproduction of the pathogen and provides complete resistance (Miles et al., 2003). The RB reaction varies with respect to the pathogen’s ability to form.
uredinia and sporulate, and is considered incomplete resistance (Miles et al., 2003; Bonde et al., 2006).

While the identity of the effectors or other determinants that contribute to the recognition of soybean rust isolates remain unknown, germplasm screening efforts have identified several sources of resistance to *P. pachyrhizi* (*Rpp*) that map to seven loci (*Rpp1 – Rpp7*; Hyten et al., 2007; Garcia et al., 2008; Silva et al., 2008; Hyten et al., 2009; Li et al., 2012; Childs et al., 2017). Two genes, *Rpp1* (Hyten et al., 2007) and *Rpp1b* (Chakraborty et al., 2009), map to an overlapping region on chromosome 18. *Rpp4* (Silva et al., 2008) and *Rpp6* (Li et al., 2012) also map to chromosome 18, but to loci distinct from *Rpp1*. The remaining genes, *Rpp2* (Silva et al., 2008), *Rpp3* (Hyten et al., 2009), the dominant and recessive alleles of *Rpp5* (Garcia et al., 2008), and *Rpp7* (Childs et al., 2017) map to chromosomes 16, 6, 3, and 19, respectively. With the exception of *Rpp1*, all of the known *Rpp* genes condition a RB reaction. *Rpp1*, but not *Rpp1b*, conditions an IR in response to avirulent isolates of *P. pachyrhizi*.

The soybean *Rpp* genes have yet to be cloned, but a candidate for *Rpp4* was identified in the resistant accession PI 459025B (*Rpp4*) using a combination of DNA sequencing, gene expression analysis, and virus-induced gene silencing (VIGS) (Meyer et al., 2009). Sequencing the *Rpp4* locus in the susceptible cultivar Williams 82 revealed the presence of three CC-NBS-LRR *R* genes with similarity to the RGC2 *R* gene family in lettuce. PI 459025B contains alleles of these genes not present in susceptible plants, and one allele (*Rpp4C4*) was predominantly expressed in resistant plants before and after challenge with an avirulent *P. pachyrhizi* isolate. Gene silencing resulted in the loss of...
resistance in PI 459025B, indicating that Rpp4 is a CC-NBS-LRR R gene within this gene cluster. To understand downstream genes contributing to Rpp4-mediated resistance, Morales et al. (2013) used microarray analyses to compare gene expression in Rpp4-silenced and control plants two weeks after infection.

In this study, we employed a similar approach to identify the source of resistance at the Rpp1 locus involved in the IR. We screened a BAC library developed from the Rpp1 resistant accession PI 200492 with a variety of markers targeting the Rpp1 locus mapped by Hyten et al. (2007). Sequencing of the BACs revealed the presence of three novel R gene candidates between markers Sct_187 and Sat_064, which define the Rpp1 locus. Co-silencing of the candidate genes using VIGS confirmed a role in resistance to P. pachyrhizi. RNA-seq of VIGS plants identified genes involved in the Rpp1-mediated defense network. This work shows that Rpp1 is an NBS-LRR protein with a novel Ubiquitin-like-specific protease 1 (ULP1) domain.

RESULTS

Identification of candidate Rpp1 genes

To identify candidate Rpp1 genes, a BAC library was constructed using DNA isolated from the resistant genotype (PI 200492). Markers defining the Rpp1 locus (Sct_187 and Sat_064) (Hyten et al., 2007) were used to identify the corresponding region in the Williams 82 genome sequence (Gm18:56,182,230 to 56,333,803). Primers for BAC library screening were developed from genes Glyma.18g280900 and Glyma.18g282700, which have relatively few homologs within the soybean genome.
(Supplemental Table S1). BAC library screening with these primers identified BACs PL_200492_2G03 and PL_200492_2E10 (Figure 1). Screening with BAC-end primers developed from PL_200492_2G03 and PL_200492_2E10 identified BACs PL_200492_1C06 and PL_200492_1A02. Paired end sequencing of 691, 402, 935 and 754 subclones from BACs PL_200492_1C06, PL_200492_2G03, PL_200492_1A02 and PL_200492_2E10, respectively, was used to generate a 324,316 bp contig (GenBank Accession MH590229) extending past the Rpp1 locus in both directions. BACs were assembled individually prior to complete contig assembly. BAC coverage ranged from 6.2X for PL_200492_1C06 to 10.0X for BAC PL_200492_1A02.

BLASTN comparisons against primary transcripts in the Williams 82 reference genome identified 31 orthologous genes in the PL 200492 Rpp1 contig (Glyma.18G280200-Glyma.18G282800, Glyma.18G283100, Glyma.18G283200, Glyma.U008900, Glyma.U008800). Of these, eight genes had homology to the NBS-LRR family of disease resistance genes. The genes are abbreviated as R1 – R8, corresponding to Glyma.18G280300, Glyma.18G280400, Glyma.18G281500, Glyma.18G281600, Glyma.18G281700, Glyma.18G283200, Glyma.U008800 and Glyma.U008900. Correspondence was determined by resistance gene order relative to the Williams 82 genome sequence. Based on multiple sequence alignments, the presence of conserved domains and the presence or absence of introns, the eight candidate R genes could be divided into two classes (Figure 2). Genes R1 – R5, contained NB-ARC (IPR002182), Winged helix-turn-helix (IPR011991), and LRR (IPR032675) domains distinct from those in R6, R7 and R8. R1 and R3 – R5 also contained a novel ULP1 protease domain.
R2 lacked the ULP1 protease domain and contained several in-frame stop codons, making it likely a pseudogene. R1 and R3 – R5, contained three predicted introns, while R2 contained a single predicted intron. R6 and R7 contained an additional coiled-coil domain and lacked introns entirely. R8 encoded a partial NBS domain and is also a likely pseudogene. Three R genes (R3, R4 and R5), were located between markers Sct_187 and Sat_064 (Figure 1), which define the Rpp1 locus (Hyten et al., 2007).

**VIGS of the Rpp1 candidate genes R3, R4 and R5 compromises immunity**

To determine if R3, R4 or R5 played a role in immunity, a VIGS construct was developed to silence their expression. If any of these genes were required for the IR, we expected visible lesions indicative of a compromised defense response to develop on silenced Rpp1-resistant plants challenged with an avirulent isolate of *P. pachyrhizi*. Since the genes shared 93 to 96% nucleotide identity, this prohibited the silencing of each gene individually. Therefore an identical 234 bp portion of the central NBS shared by all three candidate genes was used to generate a single VIGS construct for silencing. Aside from the viral symptoms associated with BPMV, propagation of the virus containing the Rpp1 insert (BPMV:Rpp1) did not result in any morphological abnormalities in susceptible or resistant plants, which were visually indistinguishable from control BPMV:GFP-infected plants. For VIGS experiments, Rpp1-silenced and control plants were challenged with *P. pachyrhizi* isolate LA04-1 4 weeks following the introduction of BPMV:Rpp1 by rub inoculation. The Rpp1-silenced plants exhibited visible RB lesions with limited uredinia formation 14 d after inoculation with isolate LA04-1 (Figure 3). By contrast, resistance was not affected in plants containing the BPMV:GFP recombinant
virus and no visible symptoms were observed. To confirm silencing, mRNA from three independent biological replicates was measured by qRT-PCR. The Rpp1 candidate gene transcript levels were reduced \(1.67 \pm 0.19\)-fold in BPMV:Rpp1-infected plants compared to BPMV:GFP-infected plants. To verify that Rpp1-mediated immunity was compromised in the silenced plants, the accumulation of \(P. pachyrhizi\) \(\alpha\)-tubulin transcript was used as a measurement of fungal growth. qRT-PCR performed with cDNA derived from three independent biological replicates from BPMV:GFP-infected plants did not produce a positive Ct value through 40 cycles. In contrast, three independent biological replicates of BPMV:Rpp1-infected plants yielded an average Ct value of \(36.05 \pm 0.56\), indicating the Rpp1-silencing allowed fungal accumulation to occur.

**R4 is the most highly expressed candidate gene in PI 200492, regardless of infection**

In order to determine which of the candidate R genes contribute to \(P. pachyrhizi\) resistance, a region of the NBS containing nucleotide differences that could be used to distinguish each of the genes was PCR amplified from cDNA, cloned and sequenced. Primers were designed from flanking regions that were absolutely conserved across all five genes in both PI 200492 and Williams 82 to reduce differences in amplification efficiency (Supplemental Figure S1). Prior to expression analyses, primers were tested on genomic DNA of PI 200492 and Williams 82 to determine amplification efficiency for each gene (Supplemental Table S2). In PI 200492 plants grown under standard conditions in the absence of the pathogen, 7.5, 0, 6.2, 65.3 and 21% of clones corresponded to \(R1 \sim R5\), respectively, suggesting R4 is the predominantly expressed gene in PI 200492 and confirming R2 as a pseudogene. In contrast, in susceptible
Williams 82 plants 2.6, 2.5, 2.7, 10.5 and 81.8 % of clones corresponded to R1 – R5, respectively, suggesting R5 is the predominantly expressed gene in Williams 82. In order to test how PI 200492 responded to P. pachyrhizi infection, we repeated the experiment collecting samples 24 and 72 hours after inoculation or mock-inoculation. Again, R4 was the predominantly expressed gene (50.7 to 63.7 % of clones), followed by R5 (14.9 to 18.4 % of clones). R1 and R4, were slightly induced by P. pachyrhizi inoculation at both time points, while R3 was repressed. R5 was slightly repressed by P. pachyrhizi inoculation at 24 hours, but induced by 72 hours.

**The Rpp1 ULP domain is a functional protease**

To determine if the ULP1 domain was functional, we used the ULP1 protein domains from R1, R3, R4, R5, Glyma.13G256800 and Glyma.15G058100 for BLASTP comparisons to all proteins in the *Arabidopsis* genome (TAIR version 10). Using an E-value cutoff of 10E-4, we identified seven predicted proteins (ULP1B, ELS1, ASP1, At4G33620, OST1 and OTS2). Sequence alignment of the ULP1 domains encoded by the Rpp1 candidate genes, the *Arabidopsis* ULP proteins and the two ULPs from *Saccharomyces cerevisiae* confirmed the presence of the catalytic triad residues (His-Asp-Cys, Supplemental Figure S2) required for protease activity. To test for functionality, the ULP1 domains of R3, R4 and R5 were used to complement *S. cerevisiae* strains that contained mutations in either of the two yeast small ubiquitin-like modifier (SUMO) protease genes, *Ulp1* or *Ulp2*. Ulp1 has several functions in yeast including the processing of Smt3 (SUMO) precursor peptides and the removal of Smt3 from post-translational conjugates (Li and Hochstrasser, 1999). Strains lacking a functional *Ulp1*
allele are not viable due to their inability to progress through the cell cycle, but \textit{ulp1} temperature sensitive (ts) alleles have been isolated (Li and Hochstrasser, 1999). A \textit{ulp1}-
ts allele permits growth at 30°C, but growth at 37°C is restricted. The second \textit{S. cerevisiae} SUMO protease, Ulp2, also functions to remove Smt3 from other proteins,
but is functionally distinct from Ulp1 (Li and Hochstrasser, 2000). Null Ulp2 mutants are
viable, but exhibit a pleiotropic phenotype that includes temperature-sensitive growth
(Li and Hochstrasser, 2000). We expressed the soybean ULP1 domains using a strong
galactose inducible promoter. Additionally, we used a truncated galactose-inducible
promoter for weaker expression, in case the expressed soybean gene fragment was
toxic in yeast. Control cells transformed with the empty vectors were unable to grow at
37°C on media containing glucose or galactose. The \textit{R5} ULP1 domain was able to restore
growth in the \textit{ulp1-ts} when expressed at high levels and the \textit{R4} ULP1 domain
complemented the \textit{ulp2} mutant cells at both high and low expression levels (Figure 4).
Conversely, expression of the \textit{R4} ULP1 domain encoding a mutation in the catalytic triad,
(C310S), failed to complement the \textit{ulp2} mutant (Supplemental Figure S3).

**Evolution of the \textit{Rpp1} locus across legumes**

To examine the evolution of the \textit{Rpp1} candidate genes across legumes, we took
advantage of the Legume Information System Genomic Context Viewer
(https://legumeinfo.org/lis_context_viewer/; Dash et al. (2016)). Since \textit{R3}, \textit{R4} and \textit{R5}
were located within the mapped \textit{Rpp1} locus and were closely related to \textit{R1} and \textit{R2}, we
focused our analyses on the region containing \textit{R1} through \textit{R5} (Gm18 from 56.10 to 56.30
MB, version Wm82.a2.v1). We identified the homeologous region on soybean Gm8,
representing a genome duplication event, and syntenic regions from *P. vulgaris* (common bean), *A. duranensis* and *A. ipaensis* (the diploid ancestors of cultivated peanut), *C. cajan* (Pigeon Pea), *L. angustifolius* (Narrow-leaved lupine), *L. japonicus* (Birdsfoot treefoil) and *M. truncatula* (Barrel Medic) (Figure 5). Colinearity was most conserved in legume species most closely related to soybean. We identified two genes, an ethylene responsive transcription factor and a homeodomain transcription factor, that flanked the *Rpp1* candidate genes *R3*, *R4* and *R5* on Gm18 and were conserved across all eight species and the homeologous region on Gm8. Only Gm18 contained R genes between these flanking genes. InterProScan analyses of all *R* proteins from across the entire region in all species confirmed that only the *R* genes in the Williams 82 Gm18 reference genome contained the novel ULP1 protease domain. Further, BLAST searches failed to identify any additional *R* proteins with ULP1 domains from the predicted proteins of any of these species. Taken together, this suggests that the presence of *R* genes in the *Rpp1* locus and the addition of the ULP1 protease domain is relatively new, following the separation of soybean and its closest sequenced relative, *P. vulgaris*.

Similarly, Meyer et al. (2009) found that *R* genes were absent from the region homeologous to the *Rpp4* Asian soybean rust resistant locus. Further, only one additional *Rpp4* homolog could be found elsewhere in the soybean genome.

In order to understand how the novel ULP1 protease domain was incorporated into the *Rpp1* candidate genes, we used BLASTN to compare the ULP1 domains from *R1*, *R3*, *R4* and *R5* to all predicted transcripts in the Williams 82 reference genome. This search identified best reciprocal matches to two homeologous genes Glyma.13G256800.
(E<0.0) and Glyma.15G058100 (E<10E-17), both lacking any R protein signatures (Figure 6). While the ULP1 domain in the Rpp1 candidates corresponded to a single exon, it corresponded to six exons in Glyma.13G256800 and Glyma.15G058100, which have 16 and 10 exons, respectively. Li et al. (2017) identified 13 ULP homologs in the soybean genome, each containing a minimum of four exons. The lack of introns in the ULP1 domain of the Rpp1 R genes suggests the ULP1 domain was likely inserted into an ancestral R gene at the Rpp1 locus by a retrotransposition event, likely within the first intron. Genes flanking a retrotransposon can be duplicated and transposed by readthrough transcription from the retroelement (Hoen et al., 2006). We took advantage of the soybean transposable element database, SoyTEdb (Du et al., 2010), to search for transposable elements within all R gene sequences of PI 200492 and Williams 82. We identified a Copia LTR retrotransposon (RLC_Gmr24_Gm10-Gm18, E=0) in intron 1 of R1 in both cultivars (Figure 6). The element immediately preceded the ULP1 domain of R1. Since the element is truncated relative to RLC-GMr24 (2035 of 2591 bases) and lacked typical long terminal repeats, it is likely no longer functional.

RNA-seq of Rpp1 Silenced Plants

Previous experiments silencing Rpp4 in PI 459025B resulted in the development of tan lesions with fully sporulating uredenia (Meyer et al., 2009). Microarray analyses of Rpp4-silenced plants revealed that genes normally induced during Rpp4-mediated resistance were repressed by Rpp4 silencing (Morales et al., 2013). In contrast, Rpp1 silencing resulted in plants with RB lesions and limited uredinia formation. However, inoculation of PI 200492 with compatible P. pachyrhizi isolates results in susceptible
TAN lesions (Miles et al., 2011). This suggests that silencing Rpp1 compromised immunity to P. pachyrhizi, but not resistance.

To evaluate the effect of Rpp1 silencing, the RNA used to confirm silencing was also used for RNA-seq analysis. By comparing gene expression between PI 200492 BPMV:Rpp1-infected plants and BMPV:GFP-infected plants, each infected with P. pachyrhizi isolate LA04–1, we hoped to identify downstream components of the Rpp1 signaling pathway responsible for resistance to P. pachyrhizi. Sequences for each of the samples are available from the National Center for Biotechnology Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra), BioProject Accession PRJNA479513.

Using a false discovery rate (FDR) < 0.001, we identified 1,211 genes induced by Rpp1 silencing and 2,566 genes repressed by Rpp1 silencing (Fold Change>2, Supplemental Table S3 and Fold change <-2, Supplemental Table S4). To better understand the biological processes affected by Rpp1 silencing, we used gene ontology (GO) terms to group differentially expressed genes by function. We identified 32 and 72 GO terms significantly (Corrected P<0.05) overrepresented among genes induced and repressed by Rpp1 silencing, respectively (Supplemental Table S5). Among the induced genes, we identified significant GO terms associated with defense (defense response, defense response to fungus, regulation of plant hypersensitive response, respiratory burst involved in defense response, flavonoid biosynthesis, positive regulation of flavonoid biosynthesis and phenylpropanoid biosynthesis), transport (amino acid, nitrate, proline, adenine, guanine and oligopeptide transport, amino acid import and phloem sucrose loading), metabolism (chlorophyll, lipoate, vitamin and sulfur amino
acid, oxidoreduction coenzyme and secondary metabolism and regulation of lipid metabolism) and biosynthesis (coenzyme, oxylipin, fat-soluble vitamin and sulfur compound biosynthesis). GO terms overrepresented among Rpp1-silencing repressed genes included those associated with photosynthesis (photosynthesis, light reaction, light harvesting, response to far red and red light, electron transport and others), growth (growth, cell tip growth, multidimensional growth, regulation of cell size and meristem growth), defense response (incompatible interaction), abiotic stress responses (response to cold, dessication, sucrose and temperature stimulus) and biosynthesis (biosynthesis of chlorophyll, brassinosteroid, lignin, cutin, carbohydrate and lipids). This suggests that Rpp1 silencing repressed photosynthetic processes, abiotic stress responses and growth while inducing defense, similar to resistant responses governed by Rpp2 (van de Mortel et al., 2007), Rpp3 (Schneider et al., 2011) and Rpp4 (Morales et al., 2013). Morales et al. (2013), identified 54 probes unique to resistance responses governed by Rpp2, Rpp3 and Rpp4. Using the SoyBase Gene Model Correspondence Lookup (https://www.soybase.org/correspondence/), these correspond to 42 genes in the current genome assembly. Of these, 11 were also differentially expressed in response to Rpp1 silencing. Morales et al. (2013) also used microarray analysis to identify genes differentially expressed in response to Rpp4 silencing. Of the 260 genes differentially expressed in response to Rpp4 silencing, 97 were also differentially expressed in response to Rpp1 silencing. Of these, 89 genes (92%) were repressed by
344  *Rpp4*-silencing. In contrast, 62 genes (70%) were induced by *Rpp1* silencing, again
345  suggesting that *Rpp1* silencing compromised the IR but not defense.
346  
347  To identify the transcription factors responding to *Rpp1* silencing, we took
348  advantage of the SoyDB Transcription Factor database (Wang et al., 2010). We identified
349  348 differentially expressed transcription factors, representing 39 transcription factor
350  families (Supplemental Table S6). Of these, WRKY, NAC and PLATZ transcription factors
351  were significantly over represented among genes differentially expressed in response to
352  *Rpp1* silencing (Corrected-P <0.01). Of the 43 WRKY transcription factors identified, 12
353  were repressed (log2FC from -1.73 to -3.32) by *Rpp1* silencing, including homologs of
354  AtWRKYs 6, 15, 40, 50, 51 and 70. Induced WRKYs (log2FC from 1.39 to 6.94) included
355  homologs of AtWRKYs 6, 23, 28, 30, 33, 40, 41, 42, 57, 65, 72 and 75. WRKYs are
356  associated with abiotic stress tolerance (WRKYs 6, 30, 33, 40,41, 42, 57, 72, 75) and
357  biotic stress responses (WRKYs 6, 28, 33, 40, 72) (reviewed by (Bakshi and Oelmüller,
358  2014; Phukan et al., 2016)). Glyma.13G310100, a homolog of AtWRKY6, was identified
359  by (Pandey et al., 2011) as required for *Rpp2*-mediated resistance to ASR.
360  Glyma.13G310100 was induced by *Rpp1* silencing with a log2FC of 3.46.

361  **DISCUSSION**

362  The BPMV-based VIGS system has proven to be a powerful tool for analyzing
363  soybean genes involved in both the recognition event and downstream signaling that
364  occurs during the resistance response to *P. pachyrhizi*. Examples of this include the
365  identification of *Rpp4* (Meyer et al., 2009; Morales et al., 2013) and the functional
analysis of genes involved in $Rpp1$- and $Rpp2$-mediated responses (Pandey et al., 2011; Cooper et al., 2013). In this study, we employed VIGS to test candidate NBS-LRR genes identified through sequencing BACs spanning the $Rpp1$ locus. The results presented here demonstrate that the soybean resistance gene responsible for the IR to $P. pachyrhizi$ encodes an NBS-LRR protein with an N-terminal ULP1 domain belonging to the C48 peptidase family.

Based on our previous VIGS work with $Rpp4$ (Meyer et al., 2009) and downstream $Rpp2$ signaling components (Pandey et al., 2011), we expected that silencing $Rpp1$ would yield a TAN phenotype with abundant uredinia, indicating complete susceptibility. However, our results suggest $Rpp1$ silencing altered the IR, but not defense. There are three possible explanations for this unexpected phenotype. First, although silencing of the $Rpp1$ candidate genes greatly reduced the levels of $Rpp1$ candidate transcripts, VIGS is not absolute. Reduced levels of $Rpp1$, or components of the $Rpp1$ signaling pathway, may condition a RB reaction rather than the IR. Expression levels of $CcRpp1$, a $P. pachyrhizi$ resistance gene obtained from Cajanus cajan (pigeonpea), condition different degrees of resistance (Kawashima et al., 2016). Plants homozygous for $CcRpp1$ were immune to $P. pachyrhizi$ and displayed no visible symptoms. However, hemizygous plants displayed a RB type resistance. Expression analysis of the $CcRpp1$ transgene revealed greater expression in homozygous plants, suggesting that expression levels influenced the efficacy of the transgene (Kawashima et al., 2016). Although $CcRpp1$ is not orthologus to $GmRpp1$, it demonstrates that
expression levels of resistance genes can have a profound impact on the phenotype of resistance to soybean rust.

A second explanation for the observed RB phenotype in Rpp1-silenced plants is that an additional gene may condition resistance to *P. pachyrhizi* in PI 200492. McClean and Byth (1980) were the first to demonstrate that Rpp1 segregated as a single locus in PI 200492. Similarly, Hyten et al. (2007) mapped Rpp1 as a single locus in L85-2378 (Rpp1), an isoline developed from Williams 82 and PI 200492 (Bernard et al., 1991).

Silencing of the candidate Rpp1 genes in L85-2378 also compromised the IR (data not shown), suggesting no R genes outside the Rpp1 locus contribute to rust resistance. However, the Rpp1 locus does contain multiple tightly linked resistance genes.

Chakraborty et al. (2009) mapped Rpp1b RB resistance in PI 594538A to a region spanning the Rpp1 locus and recently Sahoo et al. (2017) mapped the Rps12 (*Resistance to Phytophthora sojae* 12) gene in PI 399036 to a 368 kb region encompassing the Rpp1 locus. Since silencing R3, R4 and R5 compromised the IR, but not defense to rust isolate LA04-1, a second R gene in the Rpp1 locus would have to confer RB resistance to LA04-1 and not be silenced by our VIGS construct. This leaves R6 and R7 and additional R genes near the Rpp1 locus not spanned by our BAC contig, as the most likely candidates for an additional rust resistance gene. VIGS constructs developed from identical regions in R6 and R7 failed to compromise resistance, even when co-silenced with R3, R4 and R5 (data not shown).

A third explanation is that Rpp1 confers a completely novel phenotype. This hypothesis is supported by the findings of Cooper et al. (2013) who reported disruption
of Rpp1-mediated immunity following silencing of five different genes implicated in
Rpp1-mediated defense. While silencing resulted in small lesions associated with a
hypersensitive response and cell death, no sign of rust sporulation was detected,
mirroring the Rpp1-silencing phenotype. Further, PI 200492 has been tested with over
30 distinct P. pachyrhizi isolates representing geographical and temporal variation and
only the IR or TAN phenotypes are observed (data not shown). Taken together these
data suggest that no additional genes condition RB resistance to P. pachyrhizi in the
Rpp1 locus, but instead suggest silencing of Rpp1 results in a novel phenotype. Future
experiments involving Rpp1 candidate transgene expression in a susceptible soybean
background will hopefully shed light on these possibilities, demonstrating whether a
single gene is sufficient to confer resistance and if so, whether the level of transgene
expression influences the resistance phenotype. Furthermore, complementation with
Rpp1 candidate transgenes with inactive ULP1 domains should also provide evidence
regarding the role of this domain in recognition and/or signaling.

The unusual structure of Rpp1 raises several questions regarding how it
functions in the resistance response. Specifically, does the ULP1 domain play a direct
role in either recognition of the pathogen, downstream signaling following recognition,
or potentially both events? Like other biotrophic plant-pathogenic fungi, P. pachyrhizi
delivers immunity-suppressing effector proteins inside the plant cell via haustoria during
the infection process to promote virulence. Little is known regarding the effectors
deployed by P. pachyrhizi, but several transcriptomic studies have identified potential
candidates that are likely transferred into the host (Link et al., 2014; Kunjeti et al., 2016;
A small subset of these candidates have also been shown to promote virulence (Kunjeti et al., 2016), suppress the HR (de Carvalho et al., 2017), or suppress plant immunity (Qi et al., 2016; Qi et al., 2018). Although a cognate effector protein that triggers \textit{Rpp1}-mediated resistance has yet to be discovered, the \textit{Rpp1} ULP1 domain may serve as useful tool for the identification of at least one.

Plant NBS-LRR proteins have evolved to detect pathogen effectors inside the host cell, either by direct or indirect recognition. In direct recognition, an effector directly binds to the NBS-LRR protein, which triggers a response by the host cell. In indirect recognition, the NBS-LRR protein monitors one or more host proteins acted upon by the effector. If a host protein targeted by the effector plays a role in immunity it is called a “guardee” (Dangl and Jones, 2001), and in cases where the host protein mimics the authentic host protein targeted by the effector it is referred to as a “decoy” (van der Hoorn and Kamoun, 2008). Recently, a new model that is an amalgam between direct and indirect recognition has been put forth. This new model, known as the “integrated decoy” model, attempts to explain why some NBS-LRR proteins contain additional domains (Cesari et al., 2014). The authors of this model propose that some host proteins targeted by effectors have been incorporated through evolution into NBS-LRR proteins, allowing for direct recognition of effector proteins.

Although originally thought to be rather anomalous, the presence of additional integrated domains within NBS-LRR proteins appears to be widespread throughout all plant lineages (Cesari et al., 2013; Kroj et al., 2016; Sarris et al., 2016). In their search of 40 publically available plant predicted proteomes, Sarris et al. (2016) identified 61
distinct Pfam domains that occur in at least two different plant families. This group includes the ULP1 protease domain (called a Peptidase_C48 domain by Sarris et al. (2016)), which was found in NBS-LRR proteins in soybean, *Fragaria vesca*, and *Zea mays*. With the exception of soybean, presented in this study, these genes have not been demonstrated to play a role in defense. Nevertheless, it has been suggested that the mere presence of the same integrated domain in different plant species is an indication that these hosts may be attacked by pathogens that attempt to manipulate similar host proteins (Malik and Van der Hoorn, 2016). This hypothesis is consistent with the findings of Mukhtar et al. (2011) and Wessling et al. (2014) that demonstrate effectors from diverse pathogens may target common host proteins. We hypothesize that the ULP1 domain of Rpp1 functions as an integrated decoy and that one or more components of the sumoylation machinery may be targeted by *P. pachyrhizi* effectors.

The Rpp1 ULP1 domain most closely resembles proteases involved in sumoylation. Sumoylation is a rapid and reversible post-translational modification conserved in all eukaryotic organisms involving the conjugation of a SUMO peptide to a protein. The presence or absence of SUMO on a particular protein can alter its stability, interaction with other proteins, localization, or enzymatic activity, and there is an increasing body of evidence showing that sumoylation plays a key role in plant immunity (Verma et al., 2017). Among the first studies implicating sumoylation in plant-pathogen interactions was the discovery that XopD, a *Xanthomonas campestris pv. vesicatoria* effector protein, which is a SUMO protease with plant-specific SUMO substrate specificity (Hotson et al., 2003). During infection XopD is delivered inside the plant cell.
via the bacterial type III secretion system where it mimics plant SUMO isopeptidases.

This finding demonstrates that pathogens likely target SUMO-dependent processes in their attempts to thwart host defenses.

Forward and reverse genetic approaches have also identified components of the SUMO machinery that participate in plant defense responses (Lee et al., 2007; van den Burg and Takken, 2010; Bailey et al., 2016; Castaño-Miquel et al., 2017; Gou et al., 2017). The siz1 E3 SUMO ligase mutant in *Arabidopsis*, exhibits constitutive systemic-acquired resistance (SAR) with elevated levels of salicylic acid (SA) and increased pathogenesis-related (PR) genes (Lee et al., 2007). Interestingly, downregulation of soybean GmSIZ1a and GmSIZ1b using RNA interference (RNAi)-mediated gene silencing did not significantly alter SA levels (Cai et al., 2017), leaving open the possibility that other E3 SUMO ligases may play a role in SAR in soybean. NPR1, a central regulator of SAR, was recently shown to be sumoylated upon immune induction by salicylic acid. Sumoylation of NPR1 altered its association with WRKY transcriptional repressors to TGA transcriptional activators (Saleh et al., 2015). Sumoylation of specific transcription factors, including WRKYs, may also serve to suppress the expression of early defense genes which are then rapidly activated through mitogen-activated protein kinase (MAPK) signaling cascades in response to MTI or ETI (van den Burg and Takken, 2010). Together, SUMO- and MAPK-signaling may provide dynamic control over plant immune responses (van den Burg and Takken, 2010). Significantly, silencing of *GmNPR1* and *GmWRKYs* 36, 40, and 45 were shown to compromise *Rpp2*-mediated resistance (Pandey et al., 2011).
Since independently evolved effectors from diverse pathogens target common host proteins, or “hub proteins” (Mukhtar et al., 2011; Wessling et al., 2014), it is reasonable to assume that enzymes involved in sumoylation, including the SUMO proteases that remove SUMO peptides from other proteins, are ideal targets for effectors that compromise MTI or ETI. Although we are unaware of any SUMO proteases that are targeted by plant pathogens, several fungal and oomycete effectors have been shown to exhibit protease inhibition (Tian et al., 2007; van Esse et al., 2008; Pretsch et al., 2013). RTP1 from the rust fungi *Uromyces fabae* and *U. striatus* were the first fungal proteins shown to be specifically expressed in haustoria and transferred to the host cytoplasm during infection (Kemen et al., 2005). More recently the C-terminal domain of RTP1 was shown to have similarity to the C-terminal domain of cysteine protease inhibitors (Pretsch et al., 2013). RTP1 homologs have now been identified in several rust fungi, including two from *P. pachyrhizi* (Pretsch et al., 2013). The putative targets of PpRTP1 and PpRTP2 are not known, but this indicates that *P. pachyrhizi* deploys protease inhibitors as part of its effector repertoire.

An additional inference that can be drawn from the integrated decoy model is that incorporated proteins may eventually lose their biochemical activity while retaining effector-binding properties (Maqbool et al., 2015). Analysis of protein kinases fused to NBS-LRR proteins, the most common class of integrated domain, revealed that most kinases identified as integrated domains are potentially catalytically active (Sarris et al., 2016). This suggests that integrated domains originating from proteins with enzymatic activity may not act merely as decoys; rather, such domains may retain their original...
function and serve as integrated sensors (Sarris et al., 2016). Our analysis of the genes at the Rpp1 locus is consistent with this assertion as the predicted proteins encoded by these genes all retain the conserved protease catalytic residues and expression of the ULP1 domain of R4 and R5 were shown to complement the yeast ULP mutants.

Further studies are needed to determine if the enzymatic activity of Rpp1 is required for the IR. If the protease activity is required, and assuming that disruption of the activity does not alter its interaction with a putative effector, then this would be an indication that the ULP1 domain functions in signal transduction. In this scenario, the ULP1 domain would become activated following the recognition event which would lead to the desumoylation of one or more downstream proteins leading to the IR. If the protease activity turns out to be dispensable for resistance, then the ULP1 domain of Rpp1 may have a function independent of resistance. The fact that the Williams 82 susceptible soybean line retains the ULP1 domains at this locus would argue that there may be a selective advantage to retaining this activity.

In previous experiments, we used microarray analysis to compare gene expression in control and Rpp4-silenced plants (Meyer et al., 2009). Unlike Rpp1, Rpp4 contained no predicted integrated domains. Of the 264 differentially expressed genes responding to Rpp4 silencing, 7.2% (19 of 264) were induced. In contrast, 32.1% (1,211 of 3,777) of differentially expressed genes were induced in response to Rpp1-silencing, suggesting the ULP1 domain found in R1, R3, R4 and R5 could impact gene expression. The Arabidopsis genome contains eight predicted ULPs: OTS1/ULP1D (At1g60220), OTS2/ULP1C (At1g10570), ELS1/ULP1A (At3g06910), ESD4 (At4g15880),
ASP1/SPF1/ULP2B (At1g09730), ULP1B (At4g00690), SPF2/ULP2A (At4g33620) and

At3g48480 (Reeves et al., 2002; Conti et al., 2008; Novatchkova et al., 2012; Kong et al.,

2017). OTS1, OTS2, ESD4, ELS1 and ASP1 regulate flowering time in Arabidopsis (Reeves
et al., 2002; Conti et al., 2008; Novatchkova et al., 2012; Kong et al., 2017). OTS1 and

OTS2 regulate responses to osmotic stress, light-induced signaling and SA-mediated

signaling (Sadanandom et al., 2015; Castro et al., 2016). In addition, OTS1 is involved in

transcriptional gene silencing and tolerance to high copper levels (Liu et al., 2017a; Zhan
et al., 2018). ASP1, which is most closely related to the ULP1 domains present in R1, R3,

R4, and R5, regulates flowering time, fertility and ABA signaling during seedling
development (Kong et al., 2017; Liu et al., 2017b; Wang et al., 2018). Of the eight

Arabidopsis ULPs identified, only an OTS1 and OTS2 double mutant (ulp1c/ulp1d) has

been characterized using whole genome expression analyses (Castro et al., 2016). In

ulp1c/ulp1d, 112 genes were differentially expressed relative to wild-type controls.

Genes associated with abiotic and biotic stress responses were significantly

overrepresented. Nineteen of these genes (corresponding to 41 soybean best BLAST

orthologs) were also differentially expressed in response to Rpp1-silencing. However

none of the 41 soybean orthologs were differentially expressed in response to Rpp4-
silencing, again suggesting a role for the ULP1 domain in R1, R3, R4 and R5. Included

within the 19 genes common to ulp1c/ulp1d and Rpp1-silencing were orthologs of

Arabidopsis flowering time genes (Flowering locus T (AtFT, Glyma.08G363100 and

Glyma.16G150700 ) and AtFD (Glyma.04G022100), AtPSK5 (Glyma.09G277600 and

Glyma.18G21270) and AtWRKY28 (Glyma.02G285900, Glyma.05G127600 and
AtPKS5 phosphorylates and interacts with NPR1 to regulate expression of defense-related WRKYS WRKY33 and WRKY62 (Xie et al., 2010). AtWRKY28 enhances tolerance to abiotic and biotic stress (Babitha et al., 2013; Chen et al., 2013). This suggests the ULP1 domain can modulate defense responses and helps explain the novel phenotype of Rpp1-silenced plants.

Using a bioinformatic approach, Sarris et al. (2016) identified R proteins with integrated domains in 40 different plant species. Interestingly, integration of the same domain could be found in distinct plant lineages, indicating independent insertion events. The peptidase C48 domain, corresponding to the ULP1 domain, was reported in soybean, wild strawberry (Fragaria vesca, mrna24089.1-v1.0) and maize (Zea mays, GRMZM2G033519_P01). In strawberry, the peptidase C48 domain was located downstream of the LRR domain. In both soybean and maize, the peptidase C48 domain was located upstream of the NBS. In maize, the peptidase C48 domain of GRMZM2G033519_P01 contains multiple introns and exons. In contrast, the peptidase C48 domain corresponds to a single exon in soybean. This confirms independent integration of this domain in these three species. Given that all other ULP genes in the soybean genome contain at least 4 exons (Li et al., 2017) and the presence of Copia element immediately preceeding the ULP1 domain in R1, it seems likely the ULP1 domain was inserted into the first intron of an ancestral gene at the Rpp1 locus by readthrough transcription from a retroelement. Hoen et al. (2006) demonstrated that a family of ULP-like genes (97 genes and psuedogenes) had expanded in the genome of Arabidopsis through the action of Mutator-like transposable elements. Further,
transduplicated genes were under selective pressure to maintain their function. ULP domains have also been identified in transposable elements from diverse species including grape (Benjak et al., 2008), melon and rice (van Leeuwen et al., 2007) and from the soybean oomycete pathogens *Phytophthora sojae* and *P. ramorum* (Kojima and Jurka, 2011).

The identification of Rpp1 as a novel ULP1-NBS-LRR advances our understanding of the molecular basis of resistance employed by soybean against *P. pachyrhizi* and has allowed us to formulate new hypotheses to further explore this important pathosystem. For example, although many putative *P. pachyrhizi* effectors have been identified, little is known concerning their targets within the cell. The Rpp1 ULP1 domain may represent a key target of *P. pachyrhizi* effectors, which would provide additional support for the sumoylation pathway as an essential component of immunity towards microbial pathogens. Further investigation into the role of the Rpp1 ULP1 domain is an important future goal that should provide additional insight into the soybean defense mechanisms and *P. pachyrhizi* virulence determinants.

**MATERIALS AND METHODS**

**Plant materials**

The resistant soybean accession PI 200492 (Rpp1; McLean and Byth, 1980) and the susceptible genotype Williams 82 (Hyten et al., 2007) were used in this study. Seeds were germinated in a growth chamber at 20°C with a 16 h photoperiod. Plants
were fertilized with Peters Professional® 20-20-20 General Purpose (Everris NA Inc. Dublin, OH) at 3 weeks following germination.

Sequencing of the Rpp1 locus

A bacterial artificial chromosome (BAC) library of the resistant soybean accession PI 200492 (Rpp1) was constructed from high molecular weight genomic DNA isolated from 10 g of young leaf tissue by Bio S&T Inc. (Montreal, Canada). The genomic DNA was partially digested with HindIII, cloned into pIndigo BAC (HindIII) (Epicentre Inc., Madison, WI, USA), and transformed into E. coli strain DH10B (Invitrogen, Canada). Estimated genome coverage was 10X with an average insert size of 130 Kb. The library was screened by PCR according to the manufacturer’s recommendations with primers designed to amplify DNA sequences located at the Rpp1 locus (Supplemental Table S1).

Four BACs (PI _200492_1C06, PI _200492_2G03, PI _200492_1A02 and PI _200492_2E10) were identified and used to construct shotgun libraries. Briefly, DNA was extracted from E. coli using the Qiagen Large-Construct Kit (Qiagen, Valencia, CA, USA). The recovered DNA was sheared and cloned using the TOPO Shotgun Subcloning Kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA). DNA sequencing of subclones from BACs PI _200492_1A02 and PI _200492_2E10 was performed by Eurofins MWG Operon LLC (Huntsville, AL, USA). DNA sequencing of subclones from BACs PI _200492_1C06 and PI _200492_2G03 was performed in house at the USDA-ARS-CICGRU. All sequencing was performed using an Applied Biosystems 3730 DNA Analyzer with a 96-capillary array. Sequences were trimmed and assembled using Sequencher version 5.4 with the default parameters with the exception of a minimum match.
In order to maximize read lengths and improve assembly accuracy, forward and reverse reads for the same subclone were preassembled prior to BAC assembly. Each BAC was assembled individually, prior to assembly of all BACs. The Williams 82 reference genome (version Williams82.a2.v1, Schmutz et al., 2010) was used to orient contigs when necessary to fill gaps. Gaps were filled using polymerase chain reaction (PCR) to amplify from the BAC in question. PCR products were cloned and sequenced at the USDA-ARS-CICGRU facility using Hi-Fi Platinum Taq DNA polymerase (Invitrogen, no. 10342-053), TA cloning kit (Invitrogen, no. K4560-01), and One Shot TOP10 Electrocompetent E. coli (Invitrogen, no. C404052).

Annotation of the Rpp1 locus in PI 200492

The sequence from the Rpp1 contig from PI 200492 (324,316 bp, GenBank Accession MH590229) was divided into 2000 bp intervals which were compared to the UniProt (Apweiler et al., 2004) and TAIR (The Arabidopsis Information Resource version 10, www.arabidopsis.org) protein databases using BLASTX (Altschul et al., 1997) to identify potential protein coding sequences. In addition, BLASTN (Altschul et al., 1997) was used to compare against predicted genes and transcripts from the Williams 82 reference genome (Schmutz et al., 2010). FGENESH (Solovyev et al., 2006) and the NetPlantGene2 Server (Hebsgaard et al., 1996) analyses were used to predict exon positions and splice sites within candidate genes of interest. InterProScan (Jones et al., 2014) was used to identify conserved domains within predicted protein sequences of...
interest. BLASTN (Altschul et al., 1997) against the SoyTE database (Du et al., 2010) was used to search for transposable elements within the \textit{Rpp1} candidate genes.

### Evolutionary analyses of the \textit{Rpp1} locus across legumes

To examine the evolution of the \textit{Rpp1} locus across sequenced legume species, we took advantage of the Legume Information System Genomic Context Viewer ([https://legumeinfo.org/lis_context_viewer/](https://legumeinfo.org/lis_context_viewer/), (Dash et al., 2016)). We used the sequence from the \textit{Rpp1} locus in PI_200492 to identify the corresponding region in the Williams 82 reference genome (Gm18 from 56.10 to 56.30 MB, version Wm82.a2.v1, (Schmutz et al., 2010)), the homeologous region on Gm08 (22.28 to 23.15 MB) of soybean, and syntenic regions from \textit{Phaseolus vulgaris} (Pv08: 1.65 to 1.83 MB, genome version 2.0, (Schmutz et al., 2014)), \textit{Arachis duranensis} (A04: 120.49 to 121.05 MB, genome version 1.0, (Bertioli et al., 2015)), \textit{Arachis ipaensis} (B04: 130.67 to 131.29 MB, genome version 1.0, (Bertioli et al., 2015)), \textit{Cajanus cajan} (LG07: 16.70 to 16.79 MB, genome version 1.0, (Varshney et al., 2012)), \textit{Lupinus angustifolius} (NLL08: 1.99 to 2.16 MB, genome version 2.5, (Sato et al., 2008)) and \textit{Medicago truncatula} (Chr07: 5.93 to 6.60 MB, genome version 4.0, (Tang et al., 2014)). To determine if R proteins with ULP1 domains were present in any of these species, we divided each of the predicted protein sequences of R1, R3, R4 and R5 into two sections, one containing the ULP1 domain and the other containing the remaining protein sequence. BLASTP (Altschul et al., 1997) was used to compare each of these sections to all predicted proteins from soybean, \textit{P. vulgaris}, \textit{A. duranensis}, \textit{A. ipaensis}, \textit{C. cajan}, \textit{L. angustifolius}, \textit{L. japonicus} and \textit{M. truncatula}.
truncatula. BLAST reports of the two regions were compared to identify any putative R proteins with a ULP1 domain in any species.

**P. pachyrhizi inoculations**

P. pachyrhizi inoculum was prepared as previously described (Kendrick et al., 2011) using isolate LA04-1 which elicits an IR with soybean PI 200492 (*Rpp1*), but is fully pathogenic on the susceptible Williams 82 (Ray et al., 2009). Spore concentrations were measured with a hemacytometer and adjusted to approximately 5.0 × 10⁴ spores ml⁻¹ in a solution of 0.01% Tween 20 in sterile distilled water. Inoculum was applied with an atomizer until leaves were saturated. Plants were incubated in a 20°C dew chamber overnight then transferred to a 25°C greenhouse.

**VIGS of Rpp1 candidate genes**

A DNA fragment representing an identical 234 bp region of *R3, R4,* and *R5* was synthesized from overlapping oligonucleotides KP863, KP864, KP865, and KP866 (Supplemental Table S1). Briefly, 2 μM of each oligonucleotide was reacted in PCR buffer with 0.2 mM dNTPs and 1 unit Taq polymerase with an initial denaturation step at 94°C for 2 min followed by 5 PCR cycles (94°C for 1 min, 50°C for 30 s, and 72°C for 1 min) followed by 2 min extension at 72°C. One μl of the reaction was used as template for PCR under the same conditions with 20 cycles using oligonucleotides KP867 and KP868 to generate *BamHI* and *KpnI* sites for directional cloning into pBPMV-R2.

Orientation of the insert was confirmed by sequencing using a vector-specific forward primer 1548F (Zhang et al., 2009). To generate inoculum for VIGS experiments, BPMV RNA1 (pBPMV-IA-R1M) and either the *Rpp1* experimental (pBPMV-*Rpp1*) or a green
fluorescent protein (GFP) control (pBPMV-GFP) plasmids were co-inoculated by via particle bombardment on leaves of Williams 82 plants at 14 days after sowing, as previously described (Zhang et al., 2009; 2010; Whitham et al., 2016). Symptomatic BPMV-infected leaf tissue was collected at 3 to 5 weeks after bombardment, lyophilized, and stored at −20°C.

Two weeks after germination, primary leaves of PI 200492 (Rpp1) plants were dusted with Carborundum and rub-inoculated with lyophilized BPMV-infected leaf tissue that was ground to a powder and suspended in 50 mM potassium phosphate buffer, pH 7.0. Plants were transferred to the United States Department of Agriculture BSL-3 plant pathogen containment facility at Fort Detrick (Melching et al., 1983) approximately 5 weeks after BPMV inoculation. BPMV-infected plants were inoculated with P. pachyrhizi and assessed for soybean rust symptoms 2 weeks after inoculation. For each experiment, six plants inoculated with each BPMV construct were used. Three independent replicates of the experiment were performed.

Assessment of gene silencing and fungal growth in VIGS plants

In order to assess Rpp1 candidate gene silencing, 1 cm diameter leaf disks were removed from the fourth trifoliate of BPMV-infected plants 2 weeks after inoculation with P. pachyrhizi. Collected tissue was immediately frozen in liquid nitrogen and stored at −80°C. Leaf tissue was ground in liquid nitrogen and RNA was extracted using the Qiagen Plant RNeasy kit (Qiagen, Valencia, CA, U.S.A.) and treated with Turbo DNase (Ambion, Thermo Fisher Scientific, Waltham, MA, U.S.A.). First-strand cDNA synthesis was performed with the Transcriptor First Strand cDNA synthesis kit (Roche,
Indianapolis, IN, U.S.A.). qRT-PCR was performed on cDNA derived from experimental and control plants using Rpp1 candidate-specific oligonucleotide primers KP920 and KP921 and probe KP922 (Supplemental Table S1) modified with 6-carboxy fluorescein at the 5’ end and with Blackhole Quencher I at the 3’ end (Integrated DNA Technologies, Coralville, IA, U.S.A.) using a 120 s denaturation step at 95°C s followed by 40 PCR cycles (95°C for 15 s and 68°C for 60 s). Fungal growth was assessed by measuring the constitutively expressed P. pachyrhizi α-tubulin gene by qRT-PCR using primers KP1045 and KP1046 (Supplemental Table S1) using a 90 s denaturation step followed by 40 reaction cycles (95°C for 15 s, 60°C for 30 s, and 72°C for 30 s). Soybean Rpp1 and P. pachyrhizi α-tubulin transcript levels were normalized to the soybean ubiquitin-3 gene (GenBank accession no. D28123) using a 120 s denaturation at 95°C followed by 40 reaction cycles (95°C for 15 s, 52°C for 30 s, and 72°C for 30 s). The ubiquitin-3 gene is not differentially expressed during P. pachyrhizi infection (van de Mortel et al., 2007). Three experimental replicates were performed for both qRT-PCR experiments using three independent biological replicates.

**Gene expression of Rpp1 candidate genes**

To measure the expression of Rpp1 candidate genes R1 through R5, we used ClustalW (Larkin et al., 2007) to align predicted RNA transcripts from PI 200492 and Williams 82. We identified an ~300bp region within the NBS domain that was highly conserved across all R genes in the Rpp1 locus from PI 200492 and Williams 82 (Supplemental Figure S1). Primers (KP1041 and KP1042, Supplemental Table S1) were designed from regions identical across all genes. Amplified products varied in size and...
contained unique sequence differences that could be used to identify the corresponding genes. To measure amplification efficiency for each of the genes, the primers were tested on genomic DNA of PI 200492 and Williams 82. Reaction conditions included a 90 s denaturation step at 94°C followed by 35 PCR cycles (94°C for 30 s, 54°C for 30 s, and 72°C for 60 s) and a 72°C extension for 120 s. PCR amplicons were cloned into pCR2.1-TOPO (Invitrogen, Thermo Fisher Scientific, Waltham, MA), and transformed into TOP10 Electrocomp *E. coli* (Invitrogen, Thermo Fisher Scientific, Waltham, MA). Cloned DNA was prepared and sequenced at the USDA-ARS-CICGRU as described above. Sequences were imported into Sequencher version 5.4 (Gene Codes Corporation, Ann Arbor, MI, USA) and vector sequences and low quality sequences were trimmed. Sequences were aligned by clone pair and consensus sequences were compared to the *Rpp1* locus in PI 200492 and Williams 82 using BLASTN (Altschul et al., 1997) to remove non-target genes. Gene-specific nucleotide differences were used to assign subclones to specific genes (Supplemental Figure S1 and Table S2). Paired sequences of 100+ genomic clones from each genotype were used to determine an application efficiency correction factor for *Rpp1* R1 – R5 in each genotype. Individual gene representation was determined by dividing the number of clones assigned to each gene by the total number of assigned clones overall. An amplification efficiency correction factor was then determined for each gene in each genotype by dividing expected representation by observed representation. Since the primers matched all genes perfectly, and there were five genes in each genotype, expected representation was 0.2 for all genes.
To measure baseline \textit{Rpp1} \textit{R1} – \textit{R5} gene expression in resistant and susceptible plants, the second trifoliate leaves were collected from non-inoculated PI 200492 and Williams 82. To measure \textit{Rpp1} candidate gene expression in response to \textit{P. pachyrhizi}, tissue was collected from mock-inoculated and infected plants 24 and 72 hours after inoculation. Collected tissue was immediately frozen in liquid nitrogen and stored at \(-80^\circ\text{C}\). Leaf tissue was ground in liquid nitrogen and RNA was extracted using the Qiagen Plant RNeasy kit (Qiagen, Valencia, CA, U.S.A.) and treated with Turbo DNase (Ambion, Thermo Fisher Scientific, Waltham, MA, U.S.A.). First-strand cDNA synthesis was performed with the Transcriptor First Strand cDNA synthesis kit (Roche, Indianapolis, IN, U.S.A.). cDNAs were amplified with PCR primers KP1041 and KP1042, cloned, sequence and analyzed as described above. To correct expression for differences in amplification efficiency, the number of observed clones for a given cDNA was multiplied by the correction factor determined for each given gene above.

**Experiments in yeast**

Yeast strains TSA566 (MAT\(a\); ura3\(\Delta 0\); leu2\(\Delta 0\); his3\(\Delta 1\); met15\(\Delta 0\); ulp1-333:kanMX) and Y21424 (MAT\(a\)/MAT\(a\); ura3\(\Delta 0\)/ura3\(\Delta 0\); leu2\(\Delta 0\)/leu2\(\Delta 0\); his3\(\Delta 1\)/his3\(\Delta 1\); met15\(\Delta 0\)/MET15; LYS2/lys2\(\Delta 0\); YIL031w/YIL031w::kanMX4) were obtained from the European \textit{Saccharomyces cerevisiae} Archive for Functional Analysis (Euroscarf; www.euroscar.de). A haploid derivative of Y21424, AR01, was obtained via random spore analysis following standard protocols (Treco and Winston, 2008) and screened for lysine and methionine auxotrophic traits and the \textit{Ulp2} deletion phenotype. Mutations in TSA566 (\textit{ulp1}) and AR01 (\textit{ulp2}) cause temperature sensitivity of growth. For
complementation studies TSA566 and AR01 were transformed with the ULP1 domains of R3, R4 or R5 driven by the galactose-inducible promoter $P_{GAL1}$ or its weaker derivative $P_{GALS}$ in the CEN6/URA3-based plasmids p416GAL1 and p416GALS, respectively (Mumberg et al., 1994). In addition, an inactive version of the R4 ULP1 domain containing a mutation in the catalytic triad (C310S) was also expressed. Plasmids p416GAL1 and p416GALS were obtained from the American Type Culture Collection (ATCC; www.atcc.org). ORFs of the native R3, R4 and R5 ULP1 domains and mutant R4 ULP1 domain, each modified to include an N-terminal myc tag were synthesized by GenScript (Piscataway, NJ) and excised from pUC57-Simple as XbaI/SalI fragments and cloned into XbaI/SalI digested p416GAL1 and p416GALS vectors. Transformed yeast cells were recovered on SD (−ura) with glucose as the sole carbon source at 30°C. To test for complementation, cells were grown overnight at 30°C in liquid SD (−ura) glucose media and streaked onto SD (−ura) glucose or SD (−ura) galactose/rafinose media followed by incubation at 30°C or 37°C for 3 days.

**RNA-seq analyses of VIGS plants**

The six RNA samples used to confirm VIGS silencing were also used for RNA-seq analyses. Prior to RNA-seq, samples were purified and concentrated using the Qiagen RNeasy MinuElute Cleanup Kit (74204, Qiagen, Germantown, MD). The Agilent 2100 Bioanalyzer™ (Agilent, Santa Clara, CA) was used to confirm an RNA integrity number (RIN) greater than seven. RNA-seq library preparation and 150 bp single end sequencing was performed at the Iowa State University DNA facility, using the Illumina HiSeq 2500 platform (Illumina, San Diego, CA). Following sequencing, reads from individual libraries
were trimmed to remove adaptor sequences, sequencing artifacts and low quality sequences as described by Atwood et al. (Atwood et al., 2014). Tophat version 2.1.1 (Trapnell et al., 2009) was used to align reads to the Williams 82 reference genome sequence (Wm82.a2.v1, (Schmutz et al., 2010)). Samtools (Li et al., 2009) was used to remove unreliably mapped reads. The resulting mapping files were imported in the statistical program R (R Core Team, 2014) using Rsamtools (Morgan and Pages, 2013).

Gene correspondences were made using the Williams82.a2.v1 gene feature file, imported using rtracklayer (Lawrence et al., 2009). Mapped reads per gene per sample were counted using Genomic Alignments (Lawrence et al., 2013). The R graphics program ggplot2 (Wickham, 2009) was used to compare and visualize counts for sample replicates for technical reproducibility. One vector control sample was removed from further analysis.

Prior to data normalization and statistical analysis of differentially expressed genes, counts assigned $R3$, $R4$ and $R5$ were removed, as they were likely of viral origin. However, the high counts did confirm presence of the virus in silenced plants. All genes with counts per million <1 across all biological replicates were removed from further analysis. EdgeR (Robinson et al., 2010) was used for single factor, pairwise comparisons to calculate normalization factors and identify differentially expressed genes (DEGs) using a false discovery rate (FDR < .001). DEGs were annotated using the SoyBase Genome Annotation Report page (https://www.soybase.org/genomeannotation/) which provided best A. thaliana homologs and inferred gene ontology (GO) information (The Arabidopsis Information Resource [TAIR] version 10, www.arabidopsis.org). Significantly
(corrected P-value <0.05) overrepresented biological process GO terms were identified using the SoyBase GO Term Enrichment tool (https://www.soybase.org/goslimgraphic_v2/dashboard.php). Differentially expressed transcription factors were identified using The SoyDB transcription factor database (Wang et al., 2010).

ACKNOWLEDGEMENTS

We thank Andrea Luquette and Will Lane for their technical assistance. This research was financed by the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) Projects 8044-22000-046-00D and 5030-21220-005-00D, the Iowa Soybean Association, the North Central Soybean Research Program, the United Soybean Board and the National Science Foundation (award number 0820642).

AUTHOR CONTRIBUTIONS

K. Pedley, S. Whitham, and M. Graham designed the research; K. Pedley, A. Pandey, A. Ruck, L. Lincoln, and M. Graham planned and performed the experiments; all authors analyzed the data; K. Pedley and M. Graham wrote the manuscript; all authors read, edited, and approved the final manuscript.
**FIGURE LEGENDS**

**Figure 1.** Schematic diagram of the *Rpp1* locus. Four bacterial artificial chromosomes (BACs; grey boxes) spanning the *Rpp1* locus on chromosome 18 were identified from a library constructed from soybean accession PI 200492 (*Rpp1*) DNA. Positions of the physical markers used to map *Rpp1* (Hyten et al., 2007), Sct_187 and Sat_046, are illustrated at top along with targets of the primers used to screen the library. The blue box indicates the location of the *Rpp1* locus. Primers were designed based on the Williams 82 soybean genome sequence or PI_200492 BAC end sequences. The location and orientation of the eight candidate *R* genes is depicted at bottom (green arrow heads).

**Figure 2.** Structure of candidate R proteins in the *Rpp1* locus in PI 200492. Comparisons to predicted genes and transcripts in the *Rpp1* locus from Williams 82 was used to predict *R* gene structure in the *Rpp1* locus from PI 200492. The programs FGENESH (Solovyev et al., 2006) and the NetPlantGene2 Server (Hebsgaard et al., 1996) were used to confirm the positions of exons and splice sites within candidate *R* genes. InterProScan (Jones et al., 2014) was used to identify conserved domains within R proteins. These analyses divided the R proteins from the *Rpp1* locus into two distinct classes based on motif differences and the presence or absence of introns in the corresponding genes. Intron lengths (bp) are as follows: \( R1 \) (4,485, 328 and 661), \( R2 \) (635), \( R3 \) (580, 329 and 712), \( R4 \) (2,310, 332 and 674) and \( R5 \) (1,153, 330 and 656).
**Figure 3.** Loss of *Rpp1*-mediated resistance in plants following VIGS of *Rpp1* candidate genes. Plants of soybean accession PI 200492 (*Rpp1*) were infected with *Bean pod mottle virus* (BPMV) carrying either a 234 bp fragment corresponding to an identical region in *R3, R4,* and *R5* or a portion of the green fluorescent protein gene (control). BPMV-infected plants were challenged with *P. pachyrhizi* isolate LA04-1 approximately 4 weeks after BPMV inoculation and symptoms were photographed 2 weeks later. *Rpp1*-silenced plants show typical soybean rust symptoms indicating a loss of immunity whereas control plants are immune. Top row images show the adaxial leaf surface and bottom row images show the abaxial leaf surface. Susceptible Williams 82 soybean plants (not infected with BPMV) were included to serve as a susceptible control. While *Rpp1*-silenced plants have RB lesions with few fungal spores, Williams 82 plants have tan lesions with fully sporulating uredinia.

**Figure 4.** Complementation of yeast mutants by *Rpp1* candidate gene ULP1 domains. The ULP1 domains of *R3, R4,* or *R5* were expressed from either a strong (Gal1) or weak (GalS) promoter in yeast mutants carrying either a temperature-sensitive (ts) allele of *ULP1* or lacking the *ULP2* gene. Both yeast mutants are normally impaired at 37°C. The *Rpp1 R5* ULP1 domain was able to partially restore growth in the *ulp1 ts* mutant cells when expressed at high levels. In *ulp2* mutant cells, the *R4* ULP1 domain was able to complement the mutation at both high and low expression levels. Yeast cells were grown for 3 days and photographed.
Figure 5. Evolution of the \textit{Rpp1} locus across legumes reveals the novel R proteins unique to soybean. The Legume Information System Genomic Context Viewer (https://legumeinfo.org/lis_context_viewer/, Dash et al. (2016)) was used to examine the region corresponding to \textit{R1} through \textit{R5} in the \textit{Rpp1} locus (Gm18 from 56.10 to 56.30 MB, version Wm82.a2.v1) across a broad range of legumes. The location and orientation of predicted genes is depicted by arrow heads. The color of the arrow head indicates sequence homology, a full description is provided in the box below the figure. To ease visualization, genes conserved across two or more species are aligned vertically.

Figure 6. Characterization of the ULP1 domains in the \textit{Rpp1} candidate genes. BLASTN comparing the ULP1 domains from \textit{R1, R3, R4} and \textit{R5} to all predicted transcripts in the Williams 82 reference genome identified best matches to homeologous genes Glyma.13G256800 (E<0.0) and Glyma.15G058100 (E<10E-17). While the ULP1 domain is encoded by a single exon in the \textit{Rpp1} candidate genes, it corresponds to six exons in Glyma.13G256800 and Glyma.15G058100. Lack of introns in the ULP domains of the \textit{Rpp1} candidate genes suggests the ULP1 domain was inserted into an ancestral \textit{R} gene by a retrotransposition event between exons 1 and 2. BLASTN analyses identified a Copia retroelement in intron 1 of \textit{R1}, immediately preceeding the ULP1 domain.
**Supplemental Tables**

**Supplemental Table S1.** Oligonucleotides used in this study.

**Supplemental Table S2.** Cloning of genomic (DNA) and expression (cDNA) PCR products from candidate R genes in the Rpp1 locus in PI 200492 (Rpp1) and Willams 82.

**Supplemental Table S3.** Genes induced by Rpp1-silencing relative to vector control plants.

**Supplemental Table S4.** Genes repressed by GmRpp1-silencing relative to vector control plants.

**Supplemental Table S5.** Gene Ontology (GO) terms significantly (Corrected P<.05) overrepresented among differentially expressed genes induced or repressed by Rpp1 silencing.

**Supplemental Table S6.** Transcription factors differentially expressed in response to GmRpp1 silencing.
Supplemental Figures

Supplemental Figure S1. Portion of NBS used for PCR, cloning and sequencing R gene products from the Rpp1 locus in PI 200492 and Williams 82. Alignment was generated using CLUSTAL (Larkin et al., 2007) and bases are colored to indicate nucleotide identity or conservation. The location of primers KP1041 and KP1042 are indicated by black arrows.

Supplemental Figure S2. Sequence alignment of the ULP1 domains of candidate R proteins from the Rpp1 locus in PI 200492 and Williams 82 with ULP proteases from Arabidopsis thaliana and Saccharomyces cerevisiae. Sequences were aligned using CLUSTAL (Larkin et al., 2007) and shaded using the CLUSTALX color scheme. Identical amino acids in the catalytic triad (His-Asp-Cys) are indicated with an asterisk.

Supplemental Figure S3. Complementation of yeast ulp2 mutant with the R4 ULP1 domain. The R4 ULP1 domain and a mutant version (C310S) of R4 were expressed using either a strong (Gal1) or weak (GalS) promoter in yeast cells lacking the ULP2 gene. Growth of ulp2 mutants are impaired at 37°C. The wild-type R4 ULP1 domain was able to restore growth in the ulp2 mutant cells when expressed at high levels or low levels. The R4 (C310S) domain was unable to complement the mutation. Yeast cells were grown for 3 days and photographed.
Literature Cited


Pedley, *Molecular Plant-Microbe Interactions*


component of a set of federated data resources for the legume family. Nucleic Acids Res. 44:D1181–D1188.


Hane, J.K., Ming, Y., Kamphuis, L.G., Nelson, M.N., Garg, G., Atkins, C.A., Bayer, P.E.,
Bravo, A., Bringans, S., Cannon, S., Edwards, D., Foley, R., Gao, L., Harrison, M.J.,
Huang, W., Hurgobin, B., Li, S., Liu, C., McGrath, A., Morahan, G., Murray, J.,
for lupin (Lupinus angustifolius), an emerging health food: insights into plant–
Hebsgaard, S.M., Korning, P.G., Tolstrup, N., Engelbrecht, J., Rouze, P., and Brunak, S.
1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local
2006. Transposon-mediated expansion and diversification of a family of *ULP*-like
Bulked segregant analysis using the GoldenGate assay to locate the locus that
and Cregan, P.B. 2007. Map location of the *Rpp1* locus that confers resistance to


mediated innate immunity in *Arabidopsis* is regulated by SIZ1 SUMO E3 ligase.


responses by predicted secreted proteins of the soybean rust pathogen


Foundation for Statistical Computing, Vienna, Austria.

Ray, J.D., Morel, W., Smith, J.R., Frederick, R.D., and Miles, M.R. 2009. Genetics and

mapping of adult plant rust resistance in soybean PI 587886 and PI 587880A.


Reeves, P.H., Murtas, G., Dash, S., and Coupland, G. 2002. *Early in short days 4*, a

mutation in Arabidopsis that causes early flowering and reduces the mRNA

abundance of the floral repressor FLC. Development 129:5349-5361.


for differential expression analysis of digital gene expression data. Bioinformatics

26:139-140.

Sadanandom, A., Ádám, É., Orosa, B., Viczián, A., Klose, C., Zhang, C., Josse, E.-M.,

Kozma-Bognár, L., and Nagy, F. 2015. SUMOylation of phytochrome-B negatively


Sahoo, D.K., Abeysekara, N.S., Cianzio, S.R., Robertson, A.E., and Bhattacharyya, M.K.

2017. A novel *Phytophthora sojae* resistance *Rps12* gene mapped to a genomic

region that contains several *Rps* genes. PloS One 12:e0169950.

Saleh, A., Withers, J., Mohan, R., Marques, J., Gu, Y., Yan, S., Zavaliev, R., Nomoto, M.,

Tada, Y., and Dong, X. 2015. Posttranslational modifications of the master...


M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B.,
Shinozaki, K., Nguyen, H., Wing, R., Cregan, P., Specht, J., Grimwood, J., Rokhsar,
D., Stacey, G., Shoemaker, R., and Jackson, S. 2010. Genome sequence of the

Schneider, K.T., Van de Mortel, M., Bancroft, T.J., Braun, E., Nettleton, D., Nelson, R.T.,
gene expression changes elicited by Phakopsora pachyrhizi in soybean correlate

2016. Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich
repeat genes reveal three anciently diverged classes with distinct evolutionary

Molecular mapping of two loci that confer resistance to Asian rust in soybean.

Solovyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. 2006. Automatic annotation of
eukaryotic genes, pseudogenes and promoters. Genome Biol. 7:S10.

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L.,
Childs, K.L., Yandell, M., Gundlach, H., Mayer, K.F., Schwartz, D.C., and Town,


Figure 1. Schematic diagram of the *Rpp1* locus. Four bacterial artificial chromosomes (BACs; grey boxes) spanning the *Rpp1* locus on chromosome 18 were identified from a library constructed from soybean accession PI 200492 (*Rpp1*) DNA. Positions of the physical markers used to map *Rpp1* (Hyten et al., 2007), Sct_187 and Sat_046, are illustrated at top along with targets of the primers used to screen the library. The blue box indicates the location of the *Rpp1* locus. Primers were designed based on the Williams 82 soybean genome sequence or PI_200492 BAC end sequences. The location and orientation of the eight candidate *R* genes is depicted at bottom (green arrow heads).
Figure 2. Structure of candidate R proteins in the Rpp1 locus in PI 200492. Comparisons to predicted genes and transcripts in the Rpp1 locus from Williams 82 was used to predict R gene structure in the Rpp1 locus from PI 200492. The programs FGENESH (Solovyev et al., 2006) and the NetPlantGene2 Server (Hebsgaard et al., 1996) were used to confirm the positions of exons and splice sites within candidate R genes. InterProScan (Jones et al., 2014) was used to identify conserved domains within R proteins. These analyses divided the R proteins from the Rpp1 locus into two distinct classes based on motif differences and the presence or absence of introns in the corresponding genes. Intron lengths (bp) are as follows: R1 (4,485, 328 and 661), R2 (635), R3 (580, 329 and 712), R4 (2,310, 332 and 674) and R5 (1,153, 330 and 656).

254x190mm (72 x 72 DPI)
Figure 3. Loss of Rpp1-mediated resistance in plants following VIGS of Rpp1 candidate genes. Plants of soybean accession PI 200492 (Rpp1) were infected with Bean pod mottle virus (BPMV) carrying either a 234 bp fragment corresponding to an identical region in R3, R4, and R5 or a portion of the green fluorescent protein gene (control). BPMV-infected plants were challenged with P. pachyrhizi isolate LA04-1 approximately 4 weeks after BPMV inoculation and symptoms were photographed 2 weeks later. Rpp1-silenced plants show typical soybean rust symptoms indicating a loss of immunity whereas control plants are immune. Top row images show the adaxial leaf surface and bottom row images show the abaxial leaf surface. Susceptible Williams 82 soybean plants (not infected with BPMV) were included to serve as a susceptible control. While Rpp1-silenced plants have RB lesions with few fungal spores, Williams 82 plants have tan lesions with fully sporulating uredinia.

254x190mm (72 x 72 DPI)
Figure 4. Complementation of yeast mutants by Rpp1 candidate gene ULP1 domains. The ULP1 domains of R3, R4, or R5 were expressed from either a strong (Gal1) or weak (GalS) promoter in yeast mutants carrying either a temperature-sensitive (ts) allele of ULP1 or lacking the ULP2 gene. Both yeast mutants are normally impaired at 37°C. The Rpp1 R5 ULP1 domain was able to partially restore growth in the ulp1 ts mutant cells when expressed at high levels. In ulp2 mutant cells, the R4 ULP1 domain was able to complement the mutation at both high and low expression levels. Yeast cells were grown for 3 days and photographed.

254x190mm (72 x 72 DPI)
Figure 5. Evolution of the Rpp1 locus across legumes reveals the novel R proteins unique to soybean. The Legume Information System Genomic Context Viewer (https://legumeinfo.org/lis_context_viewer/, Dash et al. (2016)) was used to examine the region corresponding to R1 through R5 in the Rpp1 locus (Gm18 from 56.10 to 56.30 MB, version Wm82.a2.v1) across a broad range of legumes. The location and orientation of predicted genes is depicted by arrow heads. The color of the arrow head indicates sequence homology, a full description is provided in the box below the figure. To ease visualization, genes conserved across two or more species are aligned vertically.
Figure 6. Characterization of the ULP1 domains in the *Rpp1* candidate genes. BLASTN comparing the ULP1 domains from *R1*, *R3*, *R4* and *R5* to all predicted transcripts in the Williams 82 reference genome identified best matches to homeologous genes Glyma.13G256800 (*E*<0.0) and Glyma.15G058100 (*E*<10E-17). While the ULP1 domain is encoded by a single exon in the *Rpp1* candidate genes, it corresponds to six exons in Glyma.13G256800 and Glyma.15G058100. Lack of introns in the ULP domains of the *Rpp1* candidate genes suggests the ULP1 domain was inserted into an ancestral *R* gene by a retrotransposition event between exons 1 and 2. BLASTN analyses identified a Copia retroelement in intron 1 of *R1*, immediately preceding the ULP1 domain.

254x190mm (72 x 72 DPI)
Supplemental Figure S1. Portion of NBS used for PCR, cloning and sequencing R gene products from the Rpp1 locus in PI 200492 and Williams 82. Alignment was generated using CLUSTAL (Larkin et al. 2007) and bases are colored to indicate nucleotide identity or conservation. The location of primers KP1041 and KP1042 are indicated by black arrows.

254x190mm (72 x 72 DPI)
Supplemental Figure S2. Sequence alignment of the ULP1 domains of candidate R proteins from the Rpp1 locus in PI 200492 and Williams 82 with ULP proteases from Arabidopsis thaliana and Saccharomyces cerevisiae. Sequences were aligned using CLUSTAL (Larkin et al. 2007) and shaded using the CLUSTALX color scheme. Identical amino acids in the catalytic triad (His-Asp-Cys) are indicated with an asterisk.

254x190mm (72 x 72 DPI)
Supplemental Figure S3. Complementation of yeast ulp2 mutant with the R4 ULP domain. The R4 ULP domain and a mutant version (C310S) of R4 were expressed from either a strong (Gal1) or weak (GalS) promoter in yeast cells lacking the ULP2 gene. Growth of ulp2 mutants are impaired at 37°C. The wild-type R4 ULP domain was able to restore growth in the ulp2 mutant cells when expressed at high levels or low levels. The R4 (C310S) domain was unable to complement the mutation. Yeast cells were grown for 3 days and photographed.