GENETIC CHARACTERIZATION OF A DOMINANT SUSCEPTIBLE *RPP1* ALLELE AND ANALYSIS OF QUANTITATIVE RESISTANCE TO ASIAN SOYBEAN RUST

BY

WEI WEI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Professor Matthew Hudson, Chair
Assistant Professor Steven Clough, USDA-ARS, Co-Chair
Professor Brian Diers
Assistant Professor Leslie Domier, USDA-ARS
Assistant Professor Tiffany Jamann

ABSTRACT

Asian soybean rust (ASR), the caused by fungal pathogen *Phakopsora pachyrhizi*, is a yield threatening disease in soybean growing countries (Hartman et al. 2012; Godoy et al. 2016). The yield loss can reach as high as 80% if the conditions are conducive, and no control measures are taken (Yorinori et al. 2005; Walker et al. 2011; Hartman, Miles, and Frederick 2007). Soybean natural resistance to *P. pachyrhizi* has been identified as single R gene-conferred, and seven *Rpp* (*Rpp1-Rpp7*; *Resistance to Phakopsora pachyrhizi 1-7*) gene loci have been identified and mapped (Hartman et al. 2012). Although multiple *Rpp* genes have been mapped, none have been cloned and validated, and the molecular mechanisms of *Rpp* gene-mediated resistance remain largely elusive. A genotype was identified that could suppress resistance from *Rpp1*, resulting in dominant susceptibility (Garcia et al. 2011). The second chapter of this dissertation describes an exploration of a possible molecular mechanism of this dominant susceptible allele of the *Rpp1* locus. To achieve that, genomic sequences of the *Rpp1* locus were cloned from the dominant susceptible line (DS), as well as two *Rpp1*-harboring soybean accessions PI561356 and PI594760B, revealing three NBS-LRR genes in each resistance line and one (DS-R) in the DS line. Identification of the DS-R gene was verified through virus-induced gene silencing that led to resistance being partially restored in the susceptible *Rpp1*/*DS-R* heterozygous genotypes. The possible involvement of DS-derived small RNA, a potential mechanism for dominant repression of *Rpp1* was explored. Rpp1 locus gene expression and small RNA abundances did not support this hypothesis. On the other hand, yeast two hybrid assays, and more than 6x higher expression of *DS-R* than *Rpp1* candidates, supported the hypothesis that an overly abundant DS-R protein could interact with and presumably disrupt the functional Rpp1 protein, probably leading to dominant susceptibility. In addition to exploring the DS mechanism, new information about the *Rpp1* gene that provides resistance to ASR, was also revealed by protein sequence comparison of candidate proteins across different soybean genotypes. The analyses indicate that the R2 gene (homologous to Glyma.18G281600 of Williams 82) within resistant genotypes

PI561356, PI594760B and PI200492 showed the highest probability of being the functional *Rpp1* gene.

The third chapter of this dissertation describes using genome-wide association (GWA) mapping to identify potential loci for soybean quantitative resistance to ASR. The phenotypic data was leaf disease severity scores from approximately 2,000 USDA soybean germplasm accessions in response to a mixture of four *P. pachyrhizi* isolates (Miles, Frederick, and Hartman 2006). The diverse soybean panel exhibited a wide distribution of disease scores, independent of *R* genes. Genotypic data was obtained from publicly available soybean 50K SNP arrays. Two different models were used for GWA: Mixed Linear Model (MLM) and FarmCPU. MLM did not identify any significant loci, but FarmCPU identified five significant loci. The five significant loci taken together explained 5% of the phenotypic variation and the model in total explained 13% of the phenotypic variation. Based on this study, the small effects of the significant loci detected suggested the complexity in the genetic basis of quantitative resistance to ASR, and the low heritability suggested that breeding for quantitative resistance to rust might not be practically effective. This data set was also used to look for possible new *Rpp* qualitative resistance genes by conducting GWA using the qualitative lesion type data of this soybean diversity panel. A logistic regression model was used for the binary phenotypic data. Five significant loci were detected on chromosome 6, chromosome 11 and chromosome 18, including the already mapped *Rpp3* and *Rpp1* loci. The locus on chromosome 11 and one of the three loci on chromosome 18, were not close to any known locus, so these markers could be associated with novel *Rpp* genes. Interestingly, among the two independent loci mapped to the *Rpp1* locus, one was not in the fine-mapped *Rpp1* locus, which suggested it could be a novel *Rpp* locus next to *Rpp1*. Additional evidence supporting this hypothesis was that the resistant allele of this proposed novel *Rpp* locus was not found in any of the previously characterized *Rpp1*-containing soybean accessions.

ACKNOWLEDGEMENTS

I would like to express my highest gratitude to my co-advisor Dr. Steven Clough for his patient advice and kind support during my entire Ph.D. study. Dr. Clough gave me tremendous encouragement and numerous ideas, especially when experiments were not progressing as well as I would have liked. He is an advisor and also a great friend in life. I would also like to thank my other co-advisor, Dr. Matthew Hudson, who provided me with abundant technical help and inspirations. I really value the time spent with Dr. Hudson and his professional and kind lab members. With them, I was able to expose myself to more diverse research fields, and at the same time, make more friends.

I acknowledge my thesis committee Dr. Leslie Domier, Dr. Brian Diers and Dr. Tiffany Jamann for their help, suggestions and criticisms on my thesis project. I really appreciated that Dr. Leslie Domier gave technical support for several key experiments in my project.

In addition to thesis committee, I would like to thank Dr. David Walker sincerely for constantly informing me with knowledge related to soybean rust and for being a good friend. I am thankful to Dr. Glen Hartman for offering me resources to do rust inoculations, which were really critical for my project. I also would like to thank Dr. Zlata Gvozdenov, Dr. Brian Freeman and Dr. Roger Innes for kindly providing helpful vectors and constructive suggestions.

I want to express my sincere gratitude to many co-workers, including David Neece, Nancy McCoppin, Theresa Herman, Gad Yousef and Chandra Paul. I would also like to thank Menghao Yu, Qiong Liu, Maria Malvino, Li'ang Yu, Jiayang Xie, Jing'en Xuan, Xiaoyue Zhang and Fan Zhu. Thanks to all of you for your help with research and companionship in life. Because of them, my life in UIUC has been more colorful.

I would like to say "thank you" and "I miss you" to my parents, grandparents and all the family members in China. You bring me the courage and solidify my

determination in obtaining the Ph.D. degree. Last, but most important, my deepest gratitude goes to my fiancé Xing Wu for his unconditional support and love.

TABLE OF CONTENTS

CHAPTER I LITERATURE REVIEW

Asian Soybean Rust

Disease history and economic importance

Asian Soybean Rust (ASR), caused by *Phakopsora pachyrhizi*, is a major disease in many soybean growing countries. The pathogen was first reported in Japan in 1902. Until the 1960's, the occurrence of disease was limited to tropical and subtropical countries in the Eastern Hemisphere, including China, India and Australia. However, in the 1960s the disease and pathogen were identified in Africa, and in the early 2000s, it was discovered in Paraguay and subsequently in Brazil and Argentina. ASR was found in the continental USA in 2004 (Hartman et al. 2012; Yorinori et al. 2005; Langenbach et al. 2016), starting in several southern-most soybean-producing states and later moving north, even into some north central states such as Illinois, Iowa and Nebraska over the course of the growing season (Hartman et al. 2012; Yorinori et al. 2005; Langenbach et al. 2016).

Yield losses due to ASR over 50% are commonly seen when environments are conducive, and losses could reach as high as 80% without control measures (Yorinori et al. 2005; Walker et al. 2011; Hartman, Miles, and Frederick 2007). In Brazil, about 4.6 million tons of soybean (1.22 billion dollars) were lost in 2003/2004 when farmers were not prepared for this disease and an epidemic hit the country. In recent years, the yield losses in Brazil largely decreased because of large-scale spraying of fungicides with several applications per season, which were estimated to cost 2.2 billion dollars (Godoy et al. 2016). In the USA, although the disease has not established itself in the major Midwest soybean belt, and no great yield losses have been reported, the USDA Economic Research Service still estimated that the net economic loss to range from $640 million to $1.3 billion depending on the severity and extent of the outbreak (https://plantpath.psu.edu/research/labs/ceal/research/soybean-rust/historial-risk-assessment/soybean-rust-risk-assessment).

Disease cycle and symptoms

P. pachyrhizi has a host range of approximately 150 species in 53 genera, all restricted to the Fabaceae family (Hartman et al. 2012). The main hosts include soybean and related *Glycine* species, and also Kudzu, Cowpea and Pigeon Pea which could serve as epidemic bridges between soybean crops across seasons (Echeveste da Rosa 2015). A sexual cycle has yet to be identified, and urediniospores are the primary inoculum of soybean rust, which are light enough to be transported by air currents over long distances (Echeveste da Rosa 2015). After urediniospores land and germinate, direct penetration occurs directly through the cuticle and epidermal cell wall via appressoria (Echeveste da Rosa 2015). Primary and secondary hyphae grow extensively in the intercellular space and haustoria penetrate into cells without breaking cellular membranes. Approximately one week after penetration, lesions form on leaf surfaces and light brown uredinia producing urediniospores are found within the lesions. Urediniospores produced are carried by wind and continue as the secondary inoculum throughout the growing season (Echeveste da Rosa 2015). Lesions on soybean leaves may either be tan or red brown (RB) (Figure 1.1). The former is characterized by abundant uredinia and numerous urediniospores (Hartman, Miles, and Frederick 2007). The latter is characterized by a few uredinia and occasional sporulation, which is thought to be a resistance reaction (https://www.apsnet.org/edcenter/disandpath/fungalbasidio/pdlessons/Pages/SoybeanRust.aspx). Besides these two lesion types, there is another reaction type called immune (IM) that is characterized by no visible lesions and is considered as complete resistance (Hartman, Miles, and Frederick 2007).

Single gene resistance (R gene-conferred resistance)

Resistance to *P. pachyrhizi* has been commonly identified as single R gene-conferred, and specificity is seen to different pathogen isolates (Hartman et al. 2012). Seven major resistance loci (*Rpp1*-*Rpp7*) have been identified and mapped on the soybean genome. Three alleles of the *Rpp1* locus were mapped on chromosome 18 (Hyten et al. 2007; Ray et al. 2009a; Chakraborty et al. 2009; Kim et al. 2012). Based

on distinct reaction types to different isolates, they are classified as the original *Rpp1* (from PI 200492), *Rpp1-b* and *Rpp1-b* like' (from PI 594538A and PI 561356, respectively) and Rpp1-? (PI 587886 and PI 587880A). Among them, *Rpp1* and *Rpp1-b* have been overcome by Brazilian field isolates (Kim et al. 2012; Ribeiro et al. 2007) but *Rpp1* from PI 200492 was reported to be the most effective known resistance gene according to two screenings in the southern USA from 2006 to 2012 (Walker et al. 2014, 2011). *Rpp2* and *Rpp4* were mapped to chromosomes 16 and 18 from PI 230970 and PI 459025 (Silva et al. 2008), respectively. A recessive *rpp2* resistant allele has been identified and mapped (Garcia et al. 2008). *Rpp3* was mapped to chromosome 6 and another allele of resistance, *Rpp?* (Hyuuga) was also mapped to the same region as Rpp3 (Monteros et al. 2007; Hyten et al. 2009). There are at least 52 accessions in the USDA Soybean Germplasm collection with *Rpp3* resistance (Paul et al. 2021). For *Rpp5*, there are different alleles identified in the same interval on chromosome 3: a dominant allele from PI 200526, a recessive allele from PI 200456 and an incomplete dominant allele from PI 471904 (Garcia et al. 2008). *Rpp6,* which provides resistance to some isolates from Paraguay and US, was also mapped to chromosome 18, but confirmed to be at a different location from *Rpp1* and *Rpp4* (Li et al. 2012). Recently, a novel resistance locus was mapped to chromosome 19 in a Vietnam genotype PI 605823 and designated as *Rpp7* (Childs et al. 2018). Pyramiding of multiple resistance resources into one elite cultivar is a desirable strategy to increase the durability of R gene effectiveness. Several commercial cultivars developed with this strategy have become available in Brazil since 2009 (Godoy et al. 2016). R genes have also been identified in wild *Glycine* species and other legume species, for example *G. tomentella* (Hartman, Wang, and T. 1992; Singh and Nelson 2015) and pigeonpea (Kawashima et al. 2016).

Quantitative resistance

Compared to single R-gene conferred resistance, quantitative resistance is more likely to provide a broad spectrum of resistance to multiple pathogen isolates and is sometimes associated with enhancement in the pathogen-associated molecular

patterns (PAMPs) triggered immunity (PTI) process (Niks, Qi, and Marcel 2015). This form of resistance is more durable because it will not lose effectiveness due to the loss of a single effector in the fast-evolving pathogen populations. Since resistance is usually not complete, quantitative resistance is also called partial resistance. Quantitative resistance can be very effective by slowing down the progress of epidemic disease development, such as lowering infection frequency with a longer latent period, or smaller lesions and less spore production per uredinium. Soybean accessions with rate-reducing features have been reported but the quantitative trait loci (QTLs) underlying the partial resistance have yet to be identified (Mcebisi and Phinehas 2011). High tolerance to ASR is a desirable trait, which is defined as better yielding ability of soybean cultivars grown under severe rust stress (Mcebisi and Phinehas 2011). An obvious benefit of using ASR-tolerant cultivars is that they would not pose such a harsh selection pressure on rust populations as single gene resistance, which would minimize risk of pathogen adaptation (Langenbach et al. 2016).

Molecular mechanisms of plant resistance against ASR

Transcriptomic, proteomic and metabolic studies have revealed a plethora of critical signaling components in the effector triggered immunity (ETI) process of soybean response to *P. pachyrhizi*, including key regulators of defense, hormone signaling regulators, transcription factors and enzymes in secondary metabolism (Langenbach et al. 2016; van de Mortel et al. 2007; Schneider et al. 2011). Functional analysis using virus-induced gene silencing (VIGS) screened 140 candidate genes selected from a transcriptomic study about soybean *Rpp2*-mediated resistance and identified 11 genes that led to loss-of-resistance when silenced (Pandey et al. 2011). The genes showing an effect via VIGS were *GmEDS1*, *GmNPR1*, *GmPAD4*, *GmPAL1*, five predicted transcription factors, an O-methyl transferase, and a cytochrome P450 monooxygenase. Non-host resistance is conferred by a given plant species to all genetic variants of a given pathogen (Lipka et al. 2005). This type of resistance is considered more durable and targeting a broader range of pathogen races due to its trigger by recognition of general PAMPs or by several resistance genes corresponding

to multiple avirulence genes from a pathogen (Godoy et al. 2016). Thus, an abundance of research has been conducted on this topic, trying to discover new sources of genetic resistance that mostly would contribute to quantitative resistance (Langenbach et al. 2016). In *Arabidopsis*, a non-host of *P. pachyrhizi,* it was shown that mutation of the *pen1*, *pen2* and *pen3* genes could allow penetration of the fungus into plant epidermal cells and successfully invade the mesophyll cells, but without true colonization or uredia production. Jasmonic acid and salicylic acid signaling pathways were shown to play a role in resistance post invasion since deficient mutants in these two pathways exhibited increased haustoria formation (Loehrer et al. 2008; Bettgenhaeuser et al. 2014).

Plant NBS-LRR genes

Plant immune system

Based on molecular behaviors, the plant immune system could be classified into two branches: PAMP-triggered immunity (PTI) and effector-triggered immunity (ETI) (Jones and Dangl 2006). PTI is the first layer of defense that is initiated by plant pattern recognition receptors (PRRs) detecting PAMPs. The PAMPs are usually conserved and slowly evolving molecules harbored by a wide range of pathogens (Bent and Mackey 2007), such as bacterial flagellins, lipopolysaccharides and fungal chitin (Oakeley et al. 2004; Dow, Newman, and Roepenack 2000; Hanae et al. 2006). Successful pathogens release a set of proteins called “effectors” into host cells or apoplast to interfere with PTI and result in disease. Correspondingly, plants deploy resistance genes (R genes) to specifically recognize pathogen effectors and trigger strong resistance response such as hypersensitive response, which is defined as ETI. In ETI resistance, R genes often encode NBS-LRR proteins that contain nucleotide binding site (NBS) and leucine rich repeat (LRR) domains.

Structure and signaling

Most plant NBS-LRR type R genes are composed of at least three distinct domains joined by linker sequences: a variable amino-terminal domain (N-terminal), a central NBS domain, and a leucine rich repeat (LRR) domain (El Kasmi and Nishimura

2016). In some cases, four domains can be seen due to additional variable carboxyl-terminal domains. For instance, RRS1 has an extra putative nuclear localization signal and a WRKY transcription factor domain (McHale et al. 2006a; Lahaye 2002). Most NBS-LRR genes belong to two subclasses based on the N-terminal domain: whether it has a coiled-coil (CC) or a Toll and human interleukin receptor (TIR) domain. However, other N-terminal domains have also been reported, such as the RPW8 (resistance to powdery mildew 8) domain which was recently revealed to exist at low numbers but in many different plant species (Shao et al. 2016). CC and RPW8-containing NBS-LRR genes are found in both monocots and dicots while TIR is only present in dicots so far.

What role does each NBS-LRR type R protein domain play when signaling for resistance? The LRR is typically 20-30 amino acids in length and usually forms tandem repeats with numbers ranging from 8 to 25 in *Arabidopsis*. The LRR contains a series of beta-sheets that form a horseshoe-like concave face, which is proposed to function in protein-protein interaction (McHale et al. 2006). In many studies, the LRR domain is reported to be involved in pathogen effector binding specificity (DeYoung and Innes 2006; Zhou et al. 2006). At the same time, several lines of evidence also indicate that the LRR domain may have a regulatory function, participating in intramolecular interactions with the NBS domain (McHale et al. 2006; DeYoung and Innes 2006). The function and structure of NBS domains were partially revealed by the crystal structure of the NBS domain the human Apaf-1 and the *Caenorhabditis elegans* CED-4. Being highly conserved between animals and plants, the NBS domain consists of four subdomains that form a pocket-like 3D structure that was shown to bind a nucleotide (ATP or ADP) (McHale et al. 2006; DeYoung and Innes 2006). Supported by an array of evidence, a model has been proposed to explain the R protein activation, during which the NBS domain acts like a molecular switch for conformational changes (Takken, Albrecht, and Tameling 2006; Dodds and Rathjen 2010). With the absence of an effector trigger, NBS-LRR proteins are in an ADP-bound “OFF” state. Maintenance of this state may require the negative regulation by either an attached protein, or the LRR domain that might autoinhibit through intramolecular interaction (Tao et al. 2000; McHale et al. 2006). When an effector protein is detected, a conformation change is

induced that allows the change of ADP to ATP, and subsequently a second conformational change occurs at the N-terminal region for downstream signaling. This state is called an ATP-bound "ON" state. Gradually the signal reduces and the R proteins go back to the resting state (Figure 1.2).

A collection of evidence suggests that the N-terminal domain can be involved in downstream signaling (DeYoung and Innes 2006). The downstream signaling pathways of TIR-NBS-LRR and CC-NBS-LRR seem to be distinct from each other in *Arabidopsis*, as a TIR-type is dependent on the Enhanced Disease Susceptibility 1 (EDS1) protein whereas a CC-type on Non-race-specific Disease Resistance 1 (NDR1) (Centre et al. 1998; McHale et al. 2006). Another important characteristic of the TIR domain is that it may form homodimers (or higher-order oligomers), which are required for downstream signaling, as revealed by a series of studies (Takken and Goverse 2012; Bernoux et al. 2011; Nishimura et al. 2017). Meanwhile, the CC domain was also reported to self-associate during signaling in NBS-LRR proteins from a wide range of plant species (Casey et al. 2016).

Mechanisms of effector recognition

Interaction between NBS-LRR genes and pathogen effectors involves direct and indirect detection. Using *in vivo* protein interaction assays, RRS1 protein from Arabidopsis was shown to bind with bacterial wilt pathogen effector PopP2 (Deslandes et al. 2005). The resistance proteins L5, L6 and L7 was shown to recognize the flax rust AvrL567 effector directly (DeYoung and Innes 2006). However, more evidence supports that indirect recognition exists abundantly in nature (DeYoung and Innes 2006; Jones and Dangl 2006; Chisholm et al. 2006). When released into plant cells, the effector protein AvrPphB from *Pseudomonas syringae* targets and cleaves host protein PBS1, which was perceived by R protein RPS5 and triggers downstream signaling (Swiderski and Innes 2001). Another well-known example are three different *P. syringae* effector proteins AvrRPM1, AvrB and AvrRpt2, which all interact with a host plasma-membrane-associated protein, RIN4. Either through phosphorylation or cleavage, the change in RIN4 activates two different R proteins; RPM1 (for AvrRPM1 and AvrB) and RPS2 (for

AvrRpr2). These findings contribute to the development of the Guard Model that the NBS-LRR proteins monitor the status of host proteins targeted by pathogen effectors (McHale et al. 2006). This indirect-detection hypothesis was proposed to explain the vast number of effectors corresponding to a relatively limited number of plant R proteins since R proteins could work as the one-to-multiple guards but not the direct one-to-one receptors (McHale et al. 2006b; DeYoung and Innes 2006; Jones and Dangl 2006; Chisholm et al. 2006). This could be viewed as an expansion of the Flor's gene-for-gene theory that describes a pairwise relationship between a pathogen effector and a host resistance protein (Flor 1971). With more extensive research into R gene and effector interactions, several variants of the Guard Hypothesis have been proposed. The Decoy Model differs from the Guard Model that in addition to having one target in the plant cell, effectors can have multiple mimic targets. One of them guarded by R protein is a decoy that leads to resistance and the others contribute to susceptibility (van der Hoorn and Kamoun 2008). These indirect interaction models raise the possibility of engineering the specificity of a certain host target rather than the large-sized R protein into recognizing other effectors, therefore conferring resistance to a new pathogen (Kim et al. 2016).

NBS-LRR protein oligomerization

In the past two decades, multiple evidence generated by genetic and biochemical studies suggested NBS-LRR proteins might signal through oligomerization. Using transient expression in *Nicotiana benthamiana* followed by co-immunoprecipitation, the tobacco N protein was shown to oligomerize in the presence of the elicitor (Mestre and Baulcombe 2006). The oligomerization was shown to be an early event and required for full activation of N-mediated resistance to *Tobacco mosaic virus*. Furthermore, oligomerization was abolished by a mutation in the P-loop motif of the NB-ARC domain, suggesting the NBS domain was responsible for oligomerization. This importance of P-loop in oligomerization was also supported by another R-Avr system, *A. thaliana* RPP1 protein – effector ATR1 of *Hyaloperonospora arabidopsidis* (Schreiber et al. 2016). More interestingly, in this study, multiple domains including TIR and LRR were shown to self-associate, indicating multiple interaction interfaces could exist in the RPP1

oligomers. The TIR domain self-association was likely facilitated by the NB-ARC domain self-association and correlated with cell death induction of this R protein. Utilizing cryo-electron microscopy, the structure of tetramers formed by ROQ1 protein (a TIR-NBS-LRR protein) was elucidated, and called “resistosome” (Martin et al. 2020). The structure not only revealed the oligomerization interfaces in the NB-ARC domain, but also provide an explanation for how the oligomerization of NB-ARC could bring the TIR domain in close proximity for downstream immune signaling. Compared to TIR domain, self-association between CC-NBS-LRR was observed without the presence of effectors (Ade et al. 2007; Casey et al. 2016). The cryo-EM revealed the structure of the wheel-like pentameric resistosome formed by a CC-NBS-LRR ZAR1, an associated protein kinase RKS1 and its guardee protein PBL2 (Wang, Hu, et al. 2019). Besides the NB-ARC domain that mediated the oligomerization, the CC domain also contributed to the oligomerization structurally and a funnel-shaped structure of CC resulted from the oligomerization was found to be required for downstream signaling.

Working in pairs

With the ongoing research towards the NBS-LRR protein signaling mechanisms, Flor’s gene-for-gene theory (Flor 1971) has been extended: the interaction between pathogen effector and R proteins might require two NBS-LRRs working in pairs. In a 2010 review (Eitas and Dangl 2010), seven NBS-LRR pairs were listed as dimerizing, which may suggest that this mechanism is not rare. In rice, the RGA4 and RGA5 proteins were shown to work coordinately and sufficiently to confer resistance to the rice blast fungus (Césari et al. 2014). Revealed by co-immunoprecipitation, these two proteins could form both homo- and hetero-complexes *in planta*. More interestingly, structural and experimental evidence showed that each of them plays distinct roles during resistance signaling: RGA4 could mediate cell death whereas RGA5 contains an additional RATX1 domain and may function as an effector detector and a signaling repressor (Césari et al. 2014). Harboring an extra domain in one of the NBS-LRR pairs was observed in many cases and this domain was proposed to be a decoy fused into the R protein during evolution (Cesari et al. 2014). Despite that, Pikm1-TS and Pikm2-

TS, a NBS-LRR pair lacking TIR or CC domains and an extra integrated decoy domain were identified in rice to confer resistance to rice blast fungus (Ashikawa et al. 2008). But how do the two proteins form hetero-dimers? Do they have different functions? The working mechanisms still remain unknown.

Genome organization and evolution

The NBS-LRR gene family is one of the largest gene families in plants. About 150 NBS-LRR-encoding genes were identified in *Arabidopsis*, over 400 in *Oryza sativa*, from 400 to 500 in *Medicago truncatula*, and about 320 in soybean (McHale et al. 2006; Meyers et al. 2003; Ameline-Torregrosa et al. 2007; Kang et al. 2012). This group of genes is thought to have an ancient origin. The two domains, NBS and LRR, existed even before the prokaryotes and eukaryotes diverged and their fusion was detected in land plant lineages (Yue et al. 2012). NBS-LRR encoding genes are commonly observed as clusters in plant genomes, revealed by analyses of a wide range of plant species (McHale et al. 2006; Meyers et al. 2003; Ameline-Torregrosa et al. 2007; Kang et al. 2012). It is also noted that NBS-LRR sequences often locate in clusters, which was proposed to be a consequence of tandem duplications (Leister 2004). Segmental genomic duplications or rearrangement may copy or translocate the NBS-LRR genes to more distant genomic regions, as the similarity between two remote clusters was also seen (Meyers et al. 2003; Meyers, Kaushik, and Nandety 2005). Surveying of recently duplicated regions in the soybean genome revealed that segmental genomic duplications also caused functional redundancy of resistance genes (Kang et al. 2012). Recombination is considered another driving force for NBS-LRR gene evolution. Intragenic crossing-overs could lead to domain swaps between different alleles and sometimes between different NBS-LRR genes in a cluster, while unequal intergenic crossing-over results in R gene copy number variation between clusters (Leister 2004). Having NBS-LRR genes present in clusters would facilitate random rearrangements of the homologous DNA regions, leading to the generation of new R genes.

Figure 1.1. Reaction types of rust infection. Red brown lesion (RB) on the left and TAN lesion on the right. Photos were taken 14 days after inoculation.

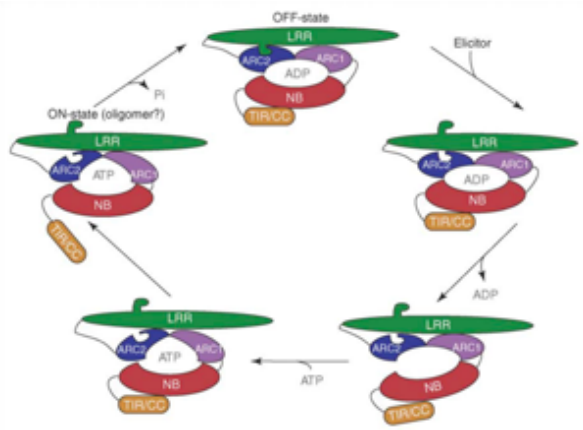


Figure 1.2. The conformation changes during NBS-LRR protein activation by pathogen effectors. From Takken, Albrecht, and Tameling 2006.

CHAPTER II. Molecular cloning of the *Rpp1* locus and characterization of the potential mechanism underlying a dominant susceptible *Rpp1* allele

Introduction

Asian soybean rust (ASR), caused by fungal pathogen *Phakopsora pachyrhizi*, is a yield threatening disease in many soybean growing countries (Hartman et al. 2012; Godoy et al. 2016). New World soybean rust, caused by *P. meibomiae*, rarely causes widespread crop losses. The yield loss of ASR can reach as high as 80% if conditions are conducive and no control measures are taken (Yorinori et al. 2005; Walker et al. 2011; Hartman, Miles, and Frederick 2007). Fungicides, together with soybean genetic resistance, are used to control this disease. After being inoculated with *P. pachyrhizi*, soybean leaves can exhibit immune (IM), red-brown (RB) or tan lesions (TAN). The former two are considered as resistance and found to be controlled by single genes. To date, seven major single-gene resistance loci have been identified and mapped (*Rpp1-Rpp7*) (Godoy et al. 2016; Childs et al. 2018). Based on leaf reactions to isolates, the *Rpp1* locus contains at least three different alleles: *Rpp1*, *Rpp1b* and several *Rpp1-?* Alleles (Hyten et al. 2007; Paul and Hartman 2009; Kim et al. 2012; Ray et al. 2009a). Among them, the *Rpp1* allele from PI200492 was assessed as most effective in the southern USA, but not in other global regions. *Rpp1-b* has been generally effective in the eastern hemisphere and in South America, and *Rpp1-?* From PI587880A was the most effective in Brazil (Walker et al. 2011, 2014; Godoy et al. 2016; Kato 2017). Fine-mapping of the *Rpp1* from PI200492 and *Rpp1-?* Allele from PI561356 located the *Rpp1* gene to overlapping 150 kb and 100 kb intervals on soybean chromosome 18, respectively (Kim et al. 2012; Hyten et al. 2007). Recently, the genomic region of the *Rpp1* locus from PI200492 was molecularly cloned (Pedley et al. 2018). Three homologous NBS-LRR family genes in a cluster were selected as the best candidates using virus-induced gene silencing (VIGS). None of the other *Rpp1* alleles have had their genomic sequence cloned, although they might bring more economic benefits

worldwide. The causal R gene controlling *Rpp1* mediated resistance has yet to be confirmed using transgenic plants.

The largest class of R genes consists of NBS-LRR proteins. These NBS-LRR proteins could mediate the recognition of specialized pathogen effector proteins and trigger intense defense reactions (Jones and Dangl 2006). Flor's classic gene-for-gene model proposed the incompatibility (resistance) resulted by the interaction of a dominant R gene and a corresponding avirulence gene. Accordingly, most R genes identified work as dominant or semi-dominant genes (Hammond-Kosack and Jones 1997). However, exceptions exist and emerging cases were reported about recessive R genes. It was also found that recessive R genes might have distinct mechanisms from the dominant R genes. For example, mutation-induced recessive alleles of the barley *Mlo* locus exhibited broad spectrum resistance to *Blumeria* (*Erysiphe) graminis* f. sp. Hordei, the barley powdery mildew pathogen (Büschges et al. 1997). Research on the molecular function of the wild-type *Mlo* gene suggested that it could work as a broad negative regulator of defense response and cell death (Piffanelli et al. 2002). Another extensively studied example is a natural recessive resistance gene from pepper, *eIF4E* that conferred resistance to potato virus Y (Ruffel et al. 2002). The eIF4E protein was be required for replication cycle of the virus (Duprat et al. 2002).

Most soybean *Rpp* genes behave as dominant genes, while incomplete dominant or recessive alleles were reported for some. The *Rpp1* allele from PI587880A generated an immune phenotype if homozygous for *Rpp1,* whereas RB lesions were seen in heterozygous *Rpp1*/*rpp1*, indicative of incomplete dominance (Ray et al. 2009). The *Rpp2* locus had a recessive resistant allele from PI224270 and *Rpp5* locus had a *rpp5* allele from PI200456 (Garcia et al. 2008). Garcia (2011) reported a dominant susceptible *Rpp1* allele from cultivar TMG06_0011 that could invert the dominance of resistant *Rpp1* alleles into recessiveness. When TMG06_0011 was crossed with PI594760B or PI561356 containing dominant *Rpp1*, the F1 progeny showed complete susceptibility and the F2 progenies showed 1:3 (R:S) segregating patterns. The inversion of dominance was apparently specific to the *Rpp1* locus because no inversion was observed when crossed with *Rpp2*, *Rpp4* and *Rpp5*-containing genotypes.

The switch between dominance and recessiveness as seen in TMG06_0011 is uncommon, and it is therefore worthwhile to explore the underlying mechanisms. It could potentially extend our knowledge regarding the biological behaviors of R genes or regulatory process of R genes. More understanding of R genes would eventually contribute to guiding the incorporation of these resistant resources into plant breeding programs. Two major hypotheses on the mechanism of dominant susceptibility have been proposed for this study. The first hypothesis is transcript-level suppression. More specifically, we hypothesize the expression of resistant *Rpp1* gene could be suppressed by small RNAs generated from TMG06_0011 *Rpp1* locus, as small RNAs were reported to extensively regulate NBS-LRR genes (Fei et al. 2016). The second hypothesis is protein-level suppression. In this case, the Rpp1 protein produced by TMG06_0011 could be a dominant negative protein that interferes with the functional Rpp1 protein. Protein-level suppression would presumably be through protein-protein interactions as studies have shown that NBS-LRR proteins can dimerize or oligomerize (McHale et al. 2006; DeYoung and Innes 2006; El Kasmi and Nishimura 2016). Wu (2017) conducted whole genome sequencing, RNA sequencing and small RNA sequencing on PI594760B, PI561356 and TMG06_0011. However, due to the repetitive feature of NBS-LRR gene clusters, genomic sequences of the *Rpp1* locus were not well-assembled and no explicit evidence supporting these two hypotheses was found in that study. To further evaluate these two hypotheses, this chapter aimed to (1) achieve better genomic sequences of the *Rpp1* locus from resistant PI594760B and PI561356, as well as from the dominant susceptible TMG06_0011; (2) use the more accurate sequences, select candidate genes for dominant susceptibility and compare the expression of candidate genes within the *Rpp1* loci on genomic and transcriptomic levels; (3) conduct genetic and biochemical assays to evaluate possible interactions involving candidate Rpp1 proteins.

Materials and methods

For convenience, the dominant susceptible TMG06_0011 will be represented by "DS", the resistant PI594760B will be called "B" and the resistant PI561356 will be called "J".

Plant growth and rust inoculation

Soybean plants were grown in an environmentally controlled plant growth chamber under a 16-h photoperiod at 22°C. At the V3 stage, fully-expanded trifoliates were detached and leaves were cut at about 0.5 cm from the base removing the petioles, using a clean straight edge razor blade. Leaves were put on wet paper towels in sterile plastic boxes for inoculation and incubation. Rust urediniospores were collected from fresh sporulating uredinia. Inoculum was prepared in distilled water with 0.01% Tween20 and vortexed for 30s at full speed to break spore clusters. Spore suspensions were adjusted to approximately 5,000 per ml with the aid of a hemocytometer. Rust inoculation was performed in two ways. For spray inoculation, an atomizer was used to spray evenly the spore suspensions onto the underside of leaf surfaces. For drop inoculation, 15 μL of a spore suspension was deposited between veins on the underside of leaf surfaces. Plastic boxes containing the leaves were placed in the dark for 24h and then placed under a 12/12h light cycle at 22°C to provide conducive conditions for rust spore germination and infection. Symptoms were recorded 14 days post inoculation (dpi). Rust isolates "MAL2019" and "BC09-1" were obtained from the USDA-ARS scientists Dr. Glen Hartman and David Walker, respectively, both stationed at the University of Illinois, Urbana.

Fosmid library constructions

High molecular weight genomic DNA samples were extracted from young leaves for all three genotypes according to a modified cetyl trimethylammonium bromide (CTAB) method (Porebski, Bailey, and Baum 1997). The DNA was sheared to approximately 40 kb by repeated pipetting through 200 μL tips or passing through a gauge needle. Pulse field gel electrophoresis using the CHEF-DRII system (Bio-Rad)

was performed to check sizes of the sheared DNA. Fosmid libraries were constructed using the CopyControl HTP Fosmid Library Production Kit (Lucigen). More specifically, the DNAs were end-repaired, purified and ligated into vector pCC2FOS. The pCC2FOS-based fosmid constructs were packaged into lambda phages and the packaged phages were incubated with *E. coli* strain EPI300-T1^{R} to transfect the fosmid clones into bacteria. Usually a total of 10 ml of the *E. coli* transfection culture would be obtained after this step. Around 300 μL of the *E. coli* transfection culture was spread onto each 150 mm x 15 mm LB plate (chloramphenicol at 12.5 μg/ml), resulting in about 30-40 plates. Plates were incubated at 37°C for 16h. Bacteria on plates (ideally, 3,000 colony-forming unit (CFU)/plate) were collected as primary pools by adding 2 ml LB to each plate, scraping colonies with a sterilized spreader, transferring to a 2 ml tube, and adding glycerol to 20% final concentration. Bacteria from multiple plates could be pooled to make each primary pool generated from about 3,000 colonies. Each tube was vortexed and stored at -80°C.

Fosmid libraries screening

A PCR-based screening strategy was utilized (Figure 2.1). Starting with the tubes of primary libraries (described in the previous paragraph), PCR was used to identify fosmids carrying the DNA of interest. Template DNA for PCR was the supernatant obtained by adding 1-2 μL from a tube of a bacterial pool to 20 μL H_2O, heated at 95°C for 10min and centrifuged for 1min. Phusion High-Fidelity DNA Polymerase (NEB) was used for the screening. One 20 μL PCR reaction was composed of Phusion HF buffer, 10 mM dNTPs, 10 μM primers, 1μL template DNA and 0.2 μL DNA polymerase. PCR thermocycling conditions contained an initial denaturation of 98°C for 30s, followed by 35 cycles of 98°C 10s, denaturing 30s and 72°C for 20s, and final extension at 72°C for 5 min. Once a positive signal was detected, the corresponding bacteria pool was diluted into secondary bacteria pools (Figure 2.1). The screening process was repeated until a positive signal could be identified from plates that contained about 100 colonies. The bacteria on the plate were collected, diluted and cultured on three new plates, with about 100 colonies on each. Each of these final colonies was picked and transferred to

a single well in 96-well culture plates and cultured with 150 μL LB (12.5 μg/ml chloramphenicol). PCR was simplified for final identification of the targeted fosmid clone by pooling bacterial culture across rows. The positive fosmid clones were extracted from *E. coli* using the QIAprep Spin Miniprep Kit (Qiagen), and linearized prior to sequencing using the restriction enzyme *Not*I.

Rpp1 locus sequence assembly and annotation

Identified fosmids were sequenced on both an Oxford Nanopore and an Illumina MiSeq platform. Raw Nanopore reads were assembled into contigs and refined with Canu v1.7 and Nanopolish v0.9.0. Five iterations of corrections using Illumina reads were performed using Pilon v1.22. For fosmid clones without Nanopore reads, SPAdes v3.14 was used for assembling Illumina MiSeq reads, with the option "–careful" on. Polished contigs, each of which represented a fosmid-inserted DNA, were assembled into larger contigs with the built-in assembler in Geneious v11.1.5. Coding regions of genes were predicted using Augustus v.3.2.3 combining the extrinsic information from RNA-Seq data. Genes were marked with the best-hit genes of the Williams 82 a2. V1 reference genome based on BLASTN results.

To better evaluate the structural variation between the *Rpp1* loci, pairwise alignments were conducted between the three genotypes of this study using the long sequence aligner Mummer/4 for the *Rpp1* locus. Multiple alignments for predicted protein sequences and DNA sequences were performed with ClustalW. Sequences of PI200492 were retrieved from the Genebank database (Pedley et al. 2018).

Virus-induced gene silencing of the dominant susceptible gene

- Plant materials

The B genotype was crossed to DS to obtain F_1 progeny. Multiple F_1 progenies were planted, and the seeds collected and bulked as F_2 plants. Since F_2 plants were segregating at the *Rpp1* locus, a set of co-dominant DNA markers were used to select heterozygous plants for VIGS (Appendix A). Electrophoresis using 3% metaphor agarose gel was utilized to separate bands of small size difference. Plants were grown

under a 16h/8h light/dark cycle at 24°C until unifoliates were fully expanded. Two days before virus rub-inoculation, plants were transferred to 20°C to adapt to conditions more favorable for this soybean VIGS system.

- Constructs

DNA fragments designed to target the 5' UTR of the dominant susceptible gene transcript were cloned from fosmid clones using PCR. DNA fragments targeting the 5' ORF as well as the NB and LRR domains were amplified with two long primers directly without a template since the amplicons were shorter than 100 bp (Appendix A). The 5' end of the forward and reverse primers contained the restriction enzyme sites *Bam*HI and *Xho*I respectively. Double digestion was performed with these two enzymes in CutSmart buffer (New England BioLabs) for both the DNA fragments and the bean pod mottle virus (BPMV) vector pBPMV-IA-V2 (Zhang et al. 2010). Digested fragments were ligated with digested vectors in the presence of T4 ligase (NEB) and subsequently transformed into NEB Turbo competent *E. coli*. Plasmids were propagated by growing plasmids-containing *E. coli* in LB media and extracted with the QIAprep Spin Miniprep Kit. The cloning procedures were similar for most work in this thesis.

- Virus inoculation

20 μg recombined pBPMV-IA-V2 vectors were co-inoculated with 20 μg pBPMV-IA-R1M vector and 10 μg pCambia1305:P19 in a volume no more than 100 μL onto B genotype leaves via a biolistic method using gold microcarriers (Zhang, Whitham, and Hill 2013). After two weeks, symptomatic leaves were collected and used as inocula via rub-inoculation. Five 'inoculum leaves' were ground with a mortar and pestle, and homogenized in 25 ml 10 mM sodium phosphate buffer (pH 7). Plant unifoliate leaves, after applying a small amount of carborundum powder, were rubbed using a cloth soaked with sap from viral inoculum plants. At 20 min post inoculation (leaf surface became dry), the remaining carborundum was washed off with ddH_2O. Plants were placed in 80% relative humidity at 20°C, 16h/8h cycle and fertilized twice a week. pBPMV-IA-V2 vector containing a fragment of phytoene desaturase (PDS) gene was used as the positive control to indicate gene silencing if bleaching is observed.

- Measurement of gene expression after silencing

When the top fully-expanded leaves of the positive control plants showed obvious bleaching phenotypes (usually about three weeks post virus inoculation), the top fully-expanded leaves of the treatment plants were collected for analysis. Three to five biological replicates from different BxDS F_1 plants were used for each treatment. RNA extraction from leaves was conducted using Trizol reagent (Thermo Fisher), followed by chloroform clean-up and isopropanol precipitation (Calla et al. 2014). DNA was removed by TURBO DNase (Thermo Fisher), after which RNA purification was conducted using the phenol/chloroform method (Green and Sambrook 2017). Reverse transcription was performed with SuperScript IV Reverse Transcriptase (Thermo Fisher), using oligo dT as the primer. Gene expressions of the dominant susceptible candidate gene in each treatment were measured using quantitative PCR with Brilliant III Ultra-Fast SYBR Green QPCR Kit (Agilent Technologies) on a Mx3000P Real-Time PCR system and using primers listed in Appendix A. The *Cons15* gene was used as a reference (Libault et al. 2008). The delta-Ct method was used to transform raw Ct values into relative gene expression compared to the reference gene (Rao et al. 2013).

- Assessment of rust resistance after gene silencing

Disease severity after gene silencing was assessed for the selected pBPMV construct and empty vector treatment. Since BxDS F_1 plants were hard to obtain due to low crossing rate of soybean flowers, BxDS F_2 plants were used in the phenotype assessment with a large biological replicate number. About three weeks post virus rub-inoculation, the top fully-expanded leaves of the treatment plants were detached and inoculated with rust using the drop-inoculation method. Six technical replicates were performed on each leaf, 10 biological replicates were used for each treatment and the whole experiment was repeated once. Several indexes of disease severity were recorded 14 days post rust inoculation, including the reaction types, sporulating uredinium number per lesion and relative spore number per lesion. For evaluating relative spore number, six leaf discs covering inoculation sites were obtained with a cork borer and placed in a 2 ml Eppendorf tube with 500 μL ddH_2O with 0.01% Tween. The mixture was vortexed for 30s to detach urediniospore clusters from leaves and to

homogenize spore suspensions. The concentration of spore suspensions was estimated using a hemocytometer, divided by 6 and recorded as the relative spore number per lesion.

Statistical analyses were conducted using SAS OnDemand for Academics (https://welcome.oda.sas.com/login). The assumptions of normal distribution and homogeneity of errors was tested. Two-way ANOVA including treatment and batch as factors was performed for datasets that met the assumptions. Friedman's Two-way nonparametric ANOVA was performed for datasets that did not meet the assumptions.

Measurement of gene expression and small RNA expression of the NBS-LRR genes in the Rpp1 locus

To measure the expression of candidate genes, RNA-Seq reads were aligned to the *Rpp1* locus with STAR v2.7 (Dobin et al. 2013). To normalize expression of genes with different length, reads per kilobase transcript, per million mapped reads (RPKM) were calculated. Small RNA-Seq data targeting clusters were predicted and small RNA abundance was counted using Shortstack v3.8.5 (Axtell 2013), in the unit of reads per million mapped reads (RPM).

The F1 plants (BxDS) were generated by crossing B and DS. A specific PCR marker was designed to check the heterogeneity of the F1 (Appendix A). Detached trifoliates from B, DS and BxDS plants were inoculated with rust isolate MAL2019 by spraying spore suspension. The concentration for spore suspension was estimated with a hemocytometer and adjusted to approximately 10^4 spores/ml. Samples were collected at 0h and 48h post inoculation. Four biological replicates were obtained for each treatment. RNA extraction was done using Trizol reagents. qRT-PCR was conducted using Brilliant III Ultra-Fast QPCR Master mix on a Mx3000P Real-Time PCR system. For details of RNA extraction and qRT-PCR, see previous section. *Cons15* gene was used as reference gene.

One-way ANOVA of genes was performed for each time point and genotype separately. Means were computed and compared pairwisely. The cutoff for significant difference was alpha=0.05, with multiple testing adjusted using Tukey's method.

Yeast-two-hybrid (Y2H) assays

The Y2H assays were conducted using the Matchmaker Gold Yeast Two-Hybrid System (Takara). Coding regions of the BR1, BR2, BR3 and DSR genes were PCR amplified using Phusion DNA polymerase and the same system described previously. Each of BR CDS sequences were cloned in the pGADT7 vector to produce the activation domain (AD)-fused proteins and the DSR coding sequences were cloned into pGBKT7 vector to produce the DNA-binding domain (BD)-fused proteins. The Ulp1, NB-ARC and LRR domains of each BR gene were cloned separately and fused with AD, and domains of DSR gene were cloned and fused with BD. Primers for all the PCR reactions mentioned here can be found in Appendix A. Each AD construct was co-transformed with the BD construct into the yeast strain AH109 (kindly provided by Dr. Brian Freeman, UIUC) using a Lithium acetate-mediated method following the yeast co-transformation protocol (Paiano et al. 2019). After transformation, transformants verified by colony PCR were evaluated for their viability on different selection media using a 10-fold serial dilution spot-test assay. Briefly, yeast transformants were cultured overnight in synthetic defined (SD) liquid medium deficient in essential amino acids. The cultures were adjusted to OD_{600}=0.2 and further serially diluted 10-fold five times. 15 μL of each diluted culture were plated onto two types of SD selective medium: one lacking leucine and tryptophan, and one lacking leucine, tryptophan and histidine. Results were recorded 3 days post plating.

Results

Validating the dominant susceptible phenotype

PI594760B (B) and TMG06_0011 (DS) were verified to be homozygous at the *Rpp1* locus, and generated RB (resistant) and TAN lesions (susceptible) respectively when inoculated with the avrRpp1b-carrying rust isolate MAL2019 (Figure 2.2). Similar to the previously reported study on the dominant susceptible behavior of DS (Garcia et al. 2011), the F1 progeny from B and DS crosses generated obvious TAN lesions with

abundant sporulation, confirming the dominant susceptibility of the *Rpp1* allele in DS genotype. We saw similar responses when looking at PI200492 that harbored a different *Rpp1* allele. Inoculation of PI200492 with the avrRpp1-carrying rust isolate BC09-1 produced an immune phenotype (Figure 2.2), with small red dots observed under 35X magnification. Inoculating isolate BC09-1 on DS, or the F1 progeny from DS crossed with PI200492, displayed TAN lesions, again showing that the susceptible DS *Rpp1* allele is dominant over multiple resistance *Rpp1* alleles.

Fosmid library screening and assembly of fosmid clones

A total of 30 – 50 primary fosmid library pools were constructed for each genotype B, J, or DS, with the goal of having 3,000 to 4,000 fosmid clones per pool. Screening was based on the *Rpp1-?* Locus mapped by Kim et al. (2012) in the J genotype. Primers specific to that corresponding region in B, J and DS were designed to screen each of the libraries (Table 2.1), with the help of the William 82 reference genome and draft genomes of B, J and DS assembled by Wu et al. (2017). The screening identified a total of 13 fosmids containing regions of the *Rpp1* locus for sequence analysis: five from B, four from J, and four from DS.

Of the 13 fosmid clones, eleven were selected for sequencing with the Oxford Nanopore and generated from 73,323 to 155,463 reads per library with mean read lengths from 6,136 to 8,622 nt (Table 2.2). Fosmid inserts are expected to be around 40 kb, and Oxford Nanopore can generate very long contiguous sequences, therefore only read lengths of 30-50 kb were used for assemblies, and the libraries contained from 238 to 4,239 reads of that size range. Additionally, all thirteen fosmid clones were sequenced on an Illumina MiSeq platform and generated 27,037 to 58,035 paired-end reads with a length of 250 nt. Illumina MiSeq reads were used to correct errors in the 11 Nanopore sequenced fosmids, or for assembly by themselves. As a result, each fosmid clone was successfully assembled into single complete sequences ranging between 30 to 45 kb (Table 2.2). One exception is the clone “DS32” that was only 19 kb in length. This clone was found to be chimeric so only part of it was from the *Rpp1* region.

Subsequently, these contigs were further assembled into larger contigs to reveal the *Rpp1* locus for each genotype.

Genome sequence comparisons of the Rpp1 locus in DS to B and J

The assembled *Rpp1* regions of B, J and DS were 124, 122 and 90 kb, respectively (Figure 2.3a). All covered a major part of the *Rpp1* region mapped by Kim et al. (2012) that spanned from Glyma.18G281500 to Glyma.18G282500 on the Williams 82 genome. B and J were very similar in terms of gene structure, and both had 19 predicted genes. Using BLASTN, 15 had orthologs identified in the Williams 82 genome (Glyma.18G281100 – Glyma.18G281600, Glyma.18G281800, Glyma.18G282000 – Glyma.18G282300, Glyma.08G257400 and Glyma.17G256100). Three genes hit NBS-LRR protein-encoding genes: one to Glyma.18G281500 and two to Glyma.18G281600. By order, they were named as *B-R1*, *B-R2*, *B-R3 and J-R1, J-R2 and J-R3*. The two genes flanking the R1s were partial orthologs of Glyma.08G257400, a gene from the homoeologous region on chromosome 8. Both of these predicted flanking genes also exist in the Williams 82 reference sequence, but were not annotated as genes in the current annotation at the time of writing. Between R1 and R2, the sequence homologous to Glyma.17G256100, and two genes with no matches, were not found at the corresponding position in the Williams 82 genome. The gene matching Glyma.17G256100 harbored a reverse transcriptase domain (IPR000477) and an endonuclease/exonuclease/phosphatase domain (IPR005135).

DS differed from B and J in that it had only 10 predicted genes (Figure 2.3a) and all of them had orthologs in Williams 82 (Glyma.18G281300, Glyma.18G281400, Glyma.18G281600, Glyma.18G281800, Glyma.18G282000 – Glyma.18G282400 and Glyma.08G257400). Only one of these 10 genes hit an NBS-LRR protein-encoding gene, Glyma.18G281600, and was named as *DS-R*.

To conserve resources to minimize redundancies, some of the subsequent work only focused on B, even though fosmid libraries were made and sequenced for both *Rpp1* resistant genotypes B and J. The two ends of the B and DS *Rpp1* locus had high similarity from Glyma.18G281300 to Glyma.18G281400, and from Glyma.18G281800 to

Glyma.18G282300 (Figure 2.3b). The middle part (where the NBS-LRR genes were located) were less similar due to different numbers of NBS-LRR genes. However, all of R1, R2 and R3 in B aligned to DS-R, showing that the four genes are homologous. The intergenic regions of the Rpp1 region corresponding to Glyma18G1400 – R1 and R3 – Glyma.18G281800 in B were aligned to the downstream and upstream regions of DS-R (Figure 2.3b), but the corresponding intergenic regions of R1-R2 and R2-R3 in B were largely missing in DS.

The presence of the DS *rpp1* allele in other soybean lines

The DS genotype is a Brazilian breeding line generated from cultivars Embrapa48 and IAC-12 (Garcia et al. 2011). We investigated whether this allele was inherited from any parent or generated during the breeding process. It turned out that Embrapa48 also harbors the same dominant susceptible allele (results not shown). Furthermore, this allele was searched for in around 50 soybean accessions using short-read alignment with resequencing data from public database (NCBI BioProject ID: PRJNA289660 and PRJNA384190). The searches found that this allele might exist in three soybean accessions: PI548364, PI548452 and PI578495. More interestingly, this allele showed more than 99.9% similarity with the *Rpp1* locus of a *Glycine soja* accession, PI549046, with the well-assembled genomic sequence obtained from Liu et al. (2020).

Comparisons of NBS-LRR protein sequences within the Rpp1 region

Predicted protein sequences of NBS-LRR genes in the *Rpp1* regions of B, J and DS were searched for protein domains, revealing that all of them contained Ulp1 protease (pfam02902), NB-ARC (pfam00931), and LRR domains (Pfam13855) (Figure 2.4a). These three domains were also identified in corresponding proteins in PI200492 that harbored a different resistant *Rpp1* allele (Pedley et al. 2018), as well as in susceptible Williams 82 (Figure 2.2). Multiple alignments of protein sequences from these genotypes showed mean identities greater than 80% between all the proteins. One strikingly variant region was the 40 amino acid sequence in the NB-ARC domain

that harbored variable numbers of glutamic acid repeats (Figure 2.4a; Figure 2.4b). Noticeably, the DS-R and all the R3 proteins had only three to five glutamic acid repeats, while the R1 and R2 proteins of resistant genotypes had from 8 to 25 glutamic acids in a row (Figure 2.4b). Another region exhibiting high sequence variability was the region from 1000 to 1200 amino acids. Variable numbers of LRRs were identified, although some were identified with insignificant e-values (>0.01) in the Pfam database.

Shown by the neighbor-joining phylogenetic tree (Figure 2.4), the DS-R protein was closest to the J-R3 and PI200492-R3, followed by the Williams 82-R2 and B-R3 (Figure 2.4c). All the R1 sequences were greater than 99% identical, and formed a distinct cluster in the phylogenetic tree. The R2 genes from all genotypes also formed a cluster, but the R2 from the resistant genotypes B, J and PI200492 were much more similar to each other than to that from Williams 82. The R3 genes were the most variable. J-R3 and PI200492-R3 were about 99% identical but more distant from B-R3 (~85%) and Williams 82-R3 (~85%). Interestingly, J and PI200492 harbored different *Rpp1* alleles but their R1, R2 and R3 proteins were all highly similar, with identities for each pair greater than 97% (Figure 2.4c). J-R1 and PI200492-R1 only displayed one single nucleotide polymorphism that was not within any domains (Table 2.3). J-R3 and PI200492-R3 displayed three differences, but none of them were located within functional domains either. Polymorphisms between J R2 and PI200492 R2 included one located in the Ulp1 domain, three in the NB-ARC domain and two in the LRR domain. The most obvious variation was the previously mentioned glutamic acid-repeat sequence, going from amino acid residue 622 to 662.

VIGS construct design and specificity verification

We used VIGS to determine if the NBS-LRR gene within the DS *Rpp1* locus is responsible for the dominant susceptible phenotype. Based on the results of multiple alignment, four regions in the *DS-R* transcript were identified as showing the most nucleotide variations between the *DS-R* transcript and the three B-R transcripts. These variations were located at different positions of the *DS-R* coding region (Figure 2.5): one in the 5' UTR region (DSR-1), one in the Ulp1 domain-encoding region (DSR-2), one

near the NB domain (DSR-3) and one near the LRR domain (DSR-4), with 208, 102, 92 and 84 bp in length, respectively. These regions were inserted into a BPMV-based VIGS system to determine if any could reduce expression of *DS-R*. Measuring DS-R expression levels after infecting these virus-based constructs, pBPMV:DSR-1 reduced the gene expression by 50% on average, pBPMV:DSR-2 by 45%, pBPMV:DSR-3 by 25% and pBPMV:DSR-4 by 35%. Being the most effective, pBPMV:DSR-1 was selected for further VIGS studies.

Since nonspecific targeting is a major concern when utilizing a VIGS system, multiple alignments were performed to compare sequences of the targeted 5'UTR of *DS-R* and three *B-R*s. The three *B-R*s displayed 65%, 54% and 65% similarities with DSR, respectively, and are unlikely to cause unintended silencing of B R genes (Figure 2.6a). As another verification, B with pBPMV:DSR-1 was inoculated with rust spores, and the leaves presented RB lesions with no sporulation (Figure 2.6c), showing that the B *Rpp1* gene was fully effective, supporting that the pBPMV:DSR-1 construct was specific to the *DS-R* gene.

VIGS of the DS-R gene reduced suppression of resistance in B x DS heterozygous plants

B x DS F2 lines with a heterozygous *Rpp1* locus were selected to conduct VIGS experiments. The positive control pBPMV:PDS VIGS vector induced obvious bleaching on top leaves at three weeks post inoculation, indicating that the BPMV-based VIGS system was working (Figure 2.6b). Presence of either the empty pBPMV VIGS vector (EV), or the pBPMV:DSR-1 VIGS construct, resulted in mild viral symptoms on top leaves reflective of viral systemic spread (Figure 2.6b). At three weeks post-inoculation, newly-expanded leaves that were healthy but with very mild viral symptoms (slight yellowing), were used for subsequent rust inoculations. TAN lesions with a range of sporulation levels were observed for both the EV and pBPMV:DSR-1 treatments (Figure 2.7a). However, leaves treated with pBPMV:DSR-1 presented significantly lower uredinium number per lesion and relative spore number per lesion than those treated with EV, with p-values at 0.0001 and 0.0002, respectively (Figure 2.7b,c). The median

of uredinium number per lesions decreased by 20% due to the DSR-1 VIGS construct, and the medium of spore number per lesion decreased by 50%, both of which support that the suppression of *Rpp1*-mediated resistance was being released after silencing of *DS-R* in the B x DS *Rpp1* heterozygous plants. Additionally, some RB-like lesions (with sporulation) were also observed occasionally on pBPMV:DSR-1 treated plants (Figure 2.7a-6). Overall, the partial restoration of resistance supports the hypothesis that the *DS-R* gene is responsible for the dominant susceptible phenotype.

VIGS of the DS-R gene also partially restored the resistance in PI200492 X DS heterozygous Rpp1 plants

As shown above, the DS genotype could exert dominant susceptibility on both resistance from the *Rpp1* (PI220492) and *Rpp1-?* (B and J) loci. Therefore, the VIGS treatment was also performed in the PI200492 x DS plants as an additional evaluation of *DS-R* as the cause of resistance suppression. As in the VIGS experiment above for the B x DS *Rpp1* heterozygotes, TAN lesions were observed (Figure 2.8a) but occasional dark lesions could be observed for leaves treated with the VIGS construct pBPMV:DSR-1 (Figure 2.8a 6-8). pBPMV:DSR-1 treated leaves displayed significantly lower levels of sporulation than the EV control treatment (P=0.01) (Figure 2.8b). This result showed that silencing of *DS-R* could also partially restore the resistance in the PI200492 x DS *Rpp1* heterozygous plants, supporting the *DS-R* gene as the probable causal gene for the dominant susceptibility phenotype.

Gene expression and small RNA abundances of the NBS-LRR genes within the Rpp1 locus

- ***Gene expression by qRT-PCR***

To investigate the mechanism of dominant susceptibility, gene expressions of *DS-R*, *B-R1*, *B-R2* and *B-R3* were measured under various conditions, including in the homozygous parents and heterozygous F1 B x DS plants, as well as with or without rust infection. Expression of *DS-R* was significantly higher (6-256 x) than any of the three R genes in B, under all conditions (Figure 2.9a). At 0 hours post inoculation(hpi), *DS-R*

was expressed 6.3-fold higher than the second highest gene in F1 lines (*B-R1*), and at 48 hpi, 14.2-fold higher than the second highest gene (*B-R2)*. As for the expression levels of *Rpp1* gene candidates in B, *B-R1* was the most abundant at 0 hpi while *B-R2* was the most abundant at 48 hpi, in both B and the B x DS F1 lines. The differences between B-R1 and B-R2 were not statistically significant but they were significantly higher than B-R3 in all conditions, with ratios ranging from 4- to 16-fold (statistical test data not shown). Slight decreases in expression levels were observed (less than 2-fold) from homozygous to heterozygous genotypes for every *B-R*, but none of the decreases were statistically significant (Figure 2.9a).

- ***Gene expression by RNA-Seq***

Gene expression of B and DS lines was also evaluated via RNA-Seq, which included more time points post inoculation. The expression levels of the DS-R gene were consistently higher than all the *Rpp1* gene candidates from B at each time point, with the smallest difference being about 4-fold and largest difference more than 100-fold (Figure 2.9b). The gene expression of B-R genes fluctuated across six time points but B-R2 was the most abundant at all time points. The B-R3 genes had the lowest expression, consistent with qRT-PCR results. The RNA-Seq result further validated the significantly higher expression of the DS-R gene compared to any of the candidate *Rpp1* genes.

- ***Small RNA abundances***

A reasonable hypothesis to explain the ability of the DS genotype to reverse the dominant inheritance of resistance in *Rpp1*-containing plants, would be RNAi via small RNAs (sRNAs). Therefore, sRNA-Seq was performed on B and DS, and sequence reads were both aligned to the B genotype. sRNA loci were predicted, and sRNA abundances were estimated and compared (Figure 2.10a). The B and DS genotypes both generated sRNAs targeting *B-R* genes at 0, 24, 48 and 72 hpi. However, their levels were low (all RPMs less than 30), and there were no striking differences in abundance between them, except that at 48 hpi (Figure 2.10a), DS generated much more small RNA targeting *B-R2* than the B genotype (6.7 RPM vs 0 RPM). However, comparisons were also performed between sRNA data from the J and DS genotypes,

and the difference at 48 hpi was not observed (Figure 2.10b). Hence, no strong evidence was discovered to support the role of sRNA or any other transcriptional suppression on resistant *Rpp1* gene by DS.

Interaction between DS-R and B-R proteins in yeast

The results of the sRNA and VIGS analyses above suggest that suppression of the *Rpp1*-mediated resistance by DS-R might be occurring at the protein-level. Therefore, we investigated whether the DS-R protein could interact with any of the B-R proteins via the Y2H assay. The growth of yeast transformants that could overexpress different combinations of proteins was evaluated on selective media. As positive controls, all of the transformants were able to grow on SD media lacking leucine (Leu) and tryptophan (Trp) (Figure 2.11a), indicative of successful co-transformation of pGBKT7 vector and pGADT7 vector that harbored a TRP1 nutritional selection marker and a LEU2 marker, respectively. Yeast transformants that co-expressed the DS-R protein and the B-R2, or DS-R and the B-R3 were able to grow on SD media lacking Leu, Trp and Histidine, while none of the negative controls grew significantly (Figure 2.11a). As serial dilutions from the same original yeast cell concentration provided a semi-quantitative measurement, yeast co-expressing DS-R and the B-R2 was found to grow stronger than DS-R plus B-R3. However, yeast transformants that co-expressed the DS-R and the B-R1 did not grow significantly compared to the negative control. Based on the evidence above, the DS-R protein can interact with the B-R2 and B-R3 proteins in yeast.

To test the interaction between individual protein domains, the Ulp1, NB-ARC and LRR domains of B-R proteins were isolated and fused with the AD domain, and those of the DS-R protein were fused with the DBD domain of the Y2H system. The growth of yeast transformants that could overexpress different combinations of proteins was evaluated on SD media lacking Leu, Trp and His (Figure 2.11b). Vigorous growth was detected of yeast transformants co-expressing the NB-ARC domain of DS-R and NB-ARC from B-R1, B-R2 or B-R3. None of the negative controls showed significant growth. Neither the Ulp1 nor LRR domain showed strong growth compared to the

negative controls. This result suggested that an interaction between the DS-R protein and the B-R2 or B-R3 proteins might be mediated by the NB-ARC domain.

Discussion

This chapter summarizes our attempts to find the most probable explanation for the dominance of the susceptible *Rpp1* allele from soybean line TMG06_0011 (DS) reported by Garcia et al. (2011). The primary key to molecular insight was to sequence the *Rpp1* regions in DS as well as two Rpp1 resistant genotypes. Also of use, was the *Rpp1* locus sequenced and published from another resistant genotype, PI200492 (Pedley et al. 2018). To successfully sequence the *Rpp1* regions, we used mapping data from Kim et al. (2012) to clone approximately 100 kb of the *Rpp1* locus from the *Rpp1*-harboring soybean genotypes PI594760B (B) and PI561356 (J), as well as from DS. The well-assembled sequences revealed three NBS-LRR genes in the two resistance lines (similar to that found in PI200492), but only one NBS-LRR (DS-R) in the DS line. Thus, as in the PI200492 study (Pedley et al. 2018), the three NBS-LRR genes (R1-R3 corresponding to Glyma.18G281500, Glyma.18G281600 and Glyma.18G281700 in Wm82.a2.v1) were considered as *Rpp1* candidate genes. The dominant susceptibility was thought to be caused by a dominant susceptible *Rpp1* allele based on the 3:1 (S:R) segregation ratio of F_2 population (Garcia et al. 2011). With this evidence, the possibility could not be ruled out that a different but closely linked gene was causing the susceptibility; for example, a gene encoding a negative regulator of defense. However, the alignments of the *Rpp1* locus between the two resistant genotypes and the DS genotype revealed high similarity in other genes in the region but substantial differences concerning the NBS-LRR gene cluster. This finding led us to mainly focus on the *DS-R* gene for exploring the mechanism of dominant susceptibility.

We detected the presence of this DS *rpp1* allele in other soybean accessions at a low frequency and also found it in soybean's wild relative, a *G. soja* line. Thus, the local rearrangement of NBS-LRR genes was not, as previously thought, a mutation during breeding of this Brazilian cultivar, but must have occurred long ago. NBS-LRR

clusters in the soybean genome exhibited higher frequency of copy number variations than other genomic regions (McHale et al. 2012), and gain or loss of NBS-LRR genes in a cluster was proposed to be the result of unequal crossing-over (McHale et al. 2006b; Leister 2004). Unequal crossing-over could give rise to new R genes by recombining between diverged haplotypes (McHale et al. 2012). This dominant susceptible allele is also likely to be the product of this mechanism. However, the fact that it appears to have persisted in the population since before soybean domestication makes it possible that there is a reason that this allele has been subject to positive selection. For example, perhaps this gene may provide or modulate resistance to another pathogen.

With the assembled and annotated *Rpp1* loci, cross comparisons were made of protein sequences between the DS-R protein and Rpp1 candidate proteins. The DS-R protein was shown to be highly homologous to all Rpp1 candidate proteins, with slightly higher similarity to R3 than R2 and R1. Although there were many similarities, multiple polymorphisms were identified throughout the whole sequence. The most intriguing variation was the glutamic acid-rich region in the NB-ARC domain where the DS protein harbored a much shorter length of glutamic acid repeats with the addition of several other distinct amino acid residues (Figure 2.4b). Although the effect of this variation is not known, it is likely to be non-negligible as a long string of negatively charged glutamic acid repeats could play important roles in protein function, such as interaction with metal ions and DNA mimicry (Chou and Wang 2015). Therefore, it is reasonable to assume special biological roles for this sequence in the Rpp1 proteins, and that missing a large part of it could negatively affect normal protein function.

Besides exploring the dominant susceptible allele, *Rpp1* candidate protein sequences were also compared between three resistant alleles and the William 82 susceptible allele. Based on our results, the R2 gene appears to be the best candidate among the three *Rpp1* candidates because its corresponding protein showed high similarities in three resistant genotypes and was more different from susceptible William 82. R1 showed high similarity between all the genotypes while R3 seemed to be too variable considering B and J alleles behaved similarly to isolates. Another interesting finding came from protein comparisons between PI200492 and J. R1 from J and

PI200492 had no polymorphism in the functional domains. The same observation was made for R3. The R2s had six polymorphisms within functional domains: one at the Ulp1 domain, three at the NB-ARC domain and two at the LRR domain. Even though the protein sequences of these two *Rpp1* alleles are very similar, their rust isolate specificity and phenotypic resistance responses differ drastically (Pham et al. 2009; Miles et al. 2011). Resistance in J is expressed as visible 'RB' non-sporulating legions, whereas resistance in PI200492 is expressed as an 'immune' response with microscopic non-sporulating lesions. Thus, this finding exposed polymorphisms that could be targets for relevant studies to explore the mechanism of allelic differences, for example, for investigating effector recognition specificity. The polymorphism in the LRR domain would be interesting as many studies suggested LRR domain determined the resistance specificity (Ellis et al. 1999; Zhou et al. 2006; Peter N. Dodds and Rathjen 2010).

Knowledge of the precise sequence of the negative-dominant allele allowed for more accurate evaluation of the hypothesis of transcription-level suppression. The gene expression levels and small RNA abundances were estimated. Based on the comparison of small RNA abundances, the DS genotype did not generate significantly more abundant small RNAs targeting the *Rpp1* locus of B than B itself. This reduced the probability of suppression by siRNAs produced from the DS *Rpp1* locus. As more direct evidence, the mRNA expression of B-R1 to B-R3 genes did not decrease significantly from homozygous status to heterozygous status (with DS-*Rpp1* allele). We would expect a more dramatic expression change if the putative transcriptional suppression imposed by DS did exist and could reverse the resistance phenotype. Hence, a transcriptional mechanism for the dominant suppression phenotype is not favored based on the current evidence. Intriguingly, the gene expression of DS-R was constantly significantly higher than all *Rpp1* candidate genes under different conditions. This evidence is consistent with our second hypothesis of protein-level suppression. Assuming a higher expression of protein as a result of higher mRNA levels, higher abundance of the suppressor would naturally exert a stronger dominant negative effect.

To validate the candidate DS-R gene, VIGS was performed on B x DS *Rpp1* heterozygous plants to study the effect of reducing DS-R levels in leaves. Precise genomic sequence data allowed for design of specific VIGS constructs that could differentiate DS-R from highly similar B *Rpp1* genes. Targeting different regions of DS-R resulted in different silencing effects and the construct with the strongest effect was selected. Looking at B x DS *Rpp1* heterozygous plants, leaves are normally highly susceptible due to the DS *Rpp1* allele, but in the leaves infected with the VIGS DS-R construct, a significant reduction in the number of sporulating uredinia and spores per lesion was observed, reflective of a partial restoration of resistance. The restoration was detected in both PI200492 and B *Rpp1* heterozygous leaves (with DS *Rpp1*), supporting that *DS-R* is the causal gene for dominant susceptibility. Ideally, after silencing, the phenotype would have shown complete resistance, like what was shown in Figure 2.6c. However, considering the low silencing efficiency of VIGS (gene expression decreased by about 2-fold) (Figure 2.5b) and the high gene expression level of *DS-R* (Figure 2.9), it was consistent with the hypothesis to see more subtle changes that were of a quantitative fashion. Intermediate phenotypes were also seen in a similar study after VIGS of the PI200492 *Rpp1* gene (Pedley et al. 2018). Silencing of the three PI200492 *Rpp1* candidate genes simultaneously was expected to change the phenotype from immune to TAN lesions with sporulation, but instead, RB lesions with limited sporulation was observed (Pedley et al. 2018). One improvement for VIGS efficiency could be made by using an inverted-repeat design of the insert in the VIGS construct, which was shown to reduce the expression up to three-fold (Lacomme, Hrubikova, and Hein 2003). However, the silencing effect is strongly affected by genotypes (Gedling et al. 2018) and might be difficult to achieve a large improvements of VIGS with an insensitive genotype. In that case, gene over-expression by soybean stable transformation could be a better option to validate the function of the DS-R gene, and these efforts are on-going.

The VIGS results further suggested that the mechanism of dominant susceptibility could be the DS-R protein disrupting the functional, Rpp1 resistant protein. We also hypothesized the disruption to be achieved by protein-protein interaction. Accordingly, the possibility of interaction of DS-R protein with B-R1 to B-R3 was

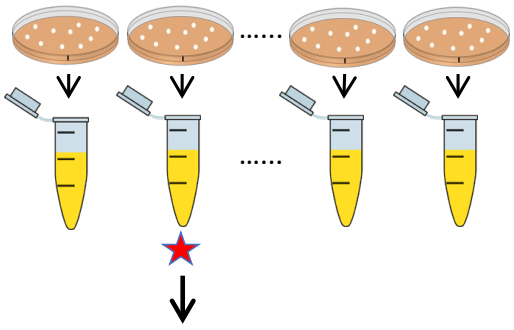
examined by the Y2H assays (Figure 2.11). The results suggested the expressed full-length DS-R protein could interact with B-R2 and B-R3 in yeast, but not B-R1. The interaction was very likely mediated by the NB-ARC domain as this domain of DS-R protein was able to interact with the same from B-R1, B-R2 and B-R3. In the past two decades, evidence showed that NBS-LRR receptors (NLRs) form oligomers after activation by pathogen effectors and the oligomerization is required for downstream immune signaling (Ade et al. 2007; Mestre and Baulcombe 2006b; Schreiber et al. 2016). Very recently, the structure of activated NLRs were captured by cryo-electron microscopy and validated the oligomerization of NLRs with presence of effectors (Wang, Hu, et al. 2019; Martin et al. 2020). The NB-ARC domain was first hypothesized to facilitate the oligomerization partly because of the analogy to the NB and oligomerization domain (NOD) in the animal NLRs. Later, several studies in plant supported that the NB-ARC domain plays a determinant role in NLR protein oligomerization (Mestre and Baulcombe 2006b; Li et al. 2020; Schreiber et al. 2016). Based on the results of *in vivo* yeast two-hybrid assays and evidence from other plant systems, it is reasonable to further hypothesize that the dominant susceptibility is mainly the result of the DS-R protein forming non-functional oligomers with the Rpp1 protein. NBS-LRR proteins have been reported to oligomerize with the presence of a ligand that could be pathogen effectors or a decoy associated with the R protein (Martin et al. 2020; Wang, Hu, et al. 2019; Wang, Wang, et al. 2019). However, in this study, the DS-R was shown to bind to B-R proteins without additional ligands in yeast. It is possible that in yeast these NBS-LRR proteins are not in their native confirmation where the binding interfaces should be tightly regulated and not exposed (Martin et al. 2020). Hence, *in planta* assays could be conducted in future to further validate the oligomerization.

There were no reports of dominant negative NLR proteins in crop systems, but in model systems like Arabidopsis, there are a few (Lolle et al. 2017; Palma et al. 2010; Dinesh-Kumar, Tham, and Baker 2000). However, in contrast to our scenario, most of the dominant negative alleles already known were created by induced mutations. Kumar et al. (2000) found some loss-of-function mutations of the tobacco N protein such as

TIR deletion or NBS point mutation could behave as dominant negative and F1 plants could not restrict systemic movement of Tobacco mosaic virus. Palma et al. (2010) generated 108 dominant negative alleles of various NLR proteins in Arabidopsis by mutagenizing the conserved p-loop motif of the NB-ARC domain. Therefore, there could be various causes of the dominant negative phenotype by the protein sequence mutation as NLR protein signaling is such a sophisticated process that involves multiple steps of structural changes (Wang, Wang, et al. 2019). As mentioned above, in spite of high homology, many polymorphisms were found between DS-R protein and Rpp1 candidates; for example, the lacking of long glutamic acid repeats in DS-R. As the Ulp1-NBS-LRR has a non-canonical N terminal domain and most of its signaling pathway remains unknown, investigating the effects of these polymorphisms could potentially reveal new knowledge about R protein signaling.

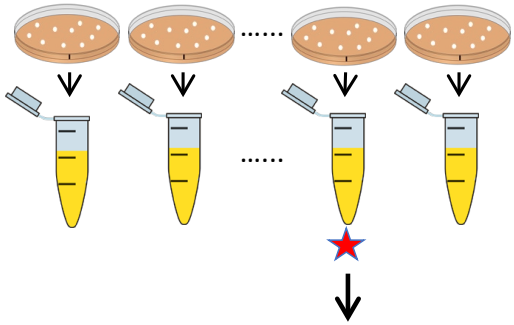
Figures and Tables

Primary pools ~ 3,000 CFU per plate, 50 plates 150 mm x 15 mm. Collect cells in LB, run PCR



Dilution plate the PCR+ culture

Secondary pools ~ 1,000 CFU per plate, 10 plates 100 mm x 15 mm. Collect cells in LB, run PCR



Dilution plate the PCR+ culture

Continue : ~ 500 CFU per plate, 10 plates 100 mm x 15 mm. Collect cells in LB, run PCR.
Continue : ~ 250 CFU per plate, 10 plates 100 mm x 15 mm. Collect cells in LB, run PCR.
Continue : ~ 100 CFU per plate, 10 plates 100 mm x 15 mm. Collect cells in LB, run PCR.

Last plates : ~ 100 CFU per plate, 3 plates.
Pick into 96 well plates, pool rows run 24 PCR reactions.
Then run PCR on individual wells of positive row(s) to identify the target fosmid

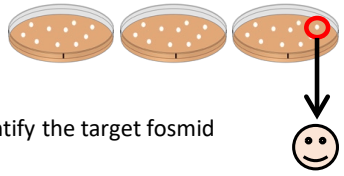


Figure 2.1. flowchart of fosmid screening.

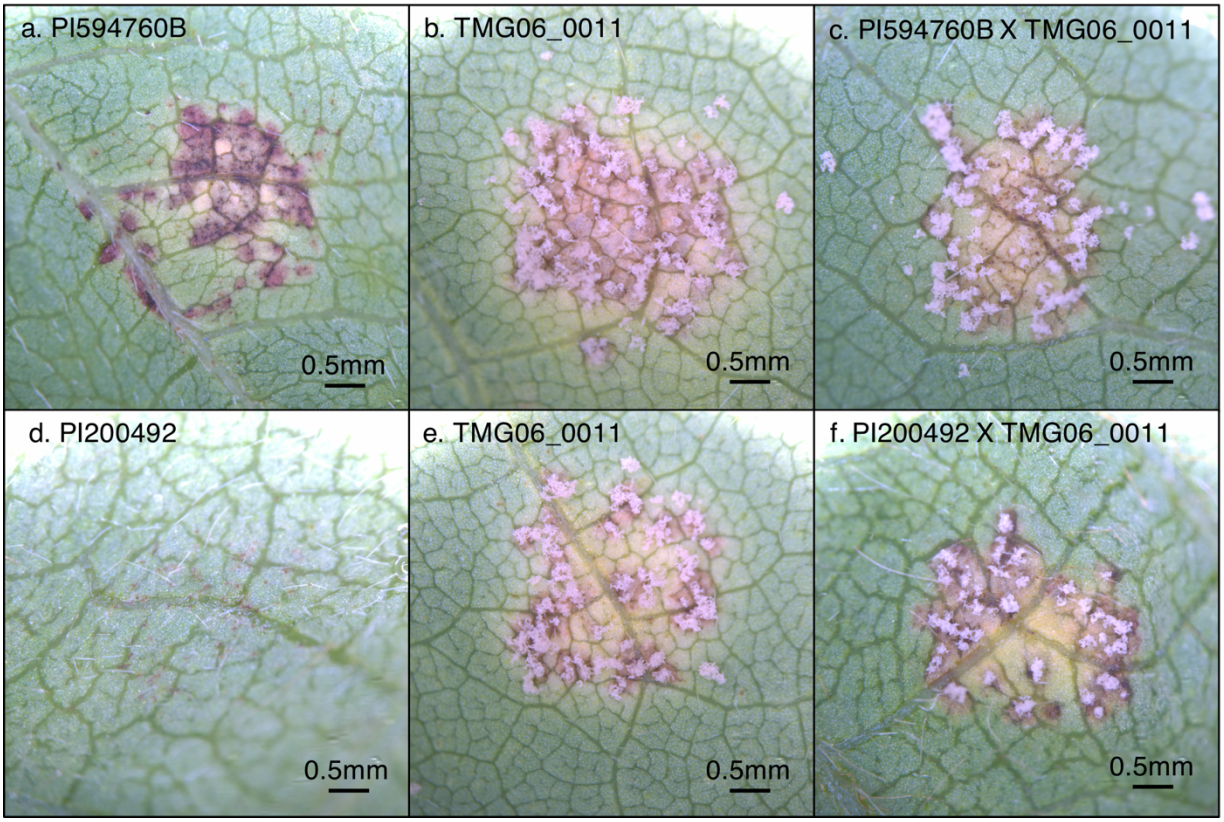


Figure 2.2. Validation of the dominant susceptible phenotypes of TMG06_0011 *Rpp1* over multiple resistant *Rpp1* alleles. (a,b) homozygous resistant *Rpp1* genotypes from PI594760B and PI200492, respectively. (b,e) homozygous dominant susceptible *rpp1* genotypes and (c,f) their corresponding heterozygous genotypes. (a-c), rust isolate "MAL2019" was used to inoculate; (d-f), rust isolate "BC09-1" was used.

a

b

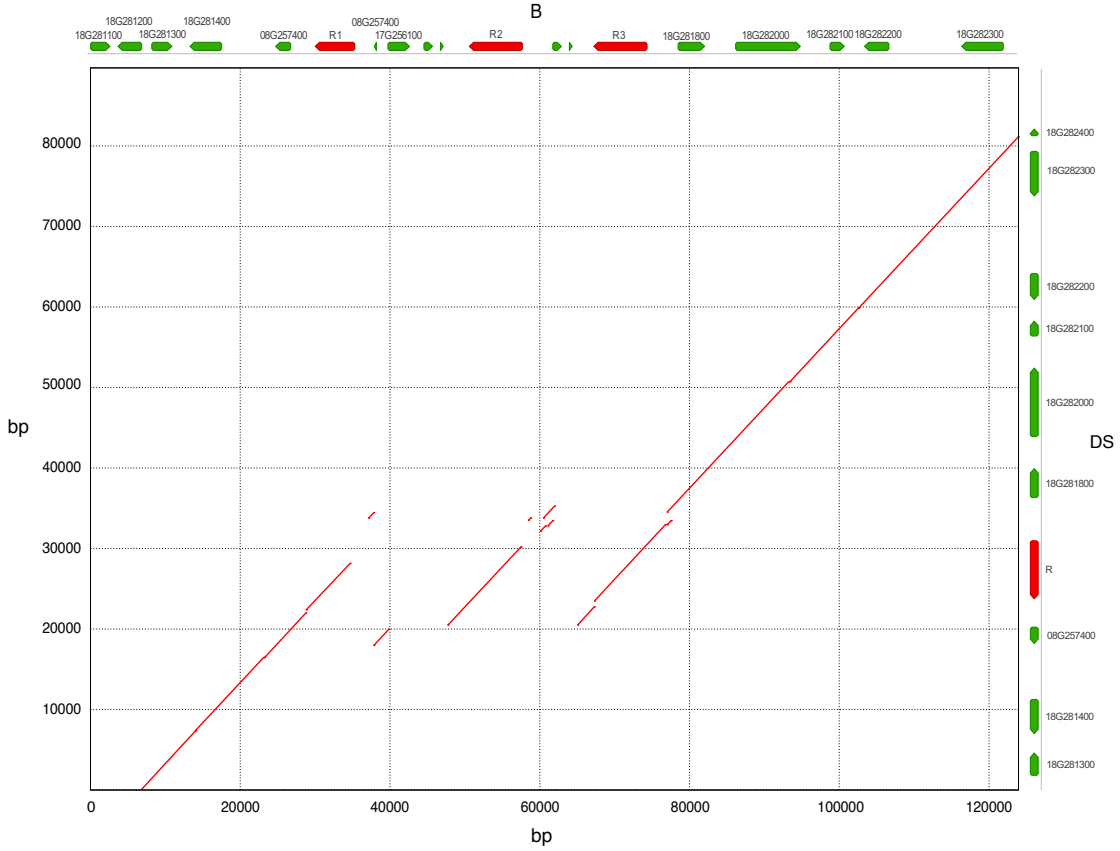
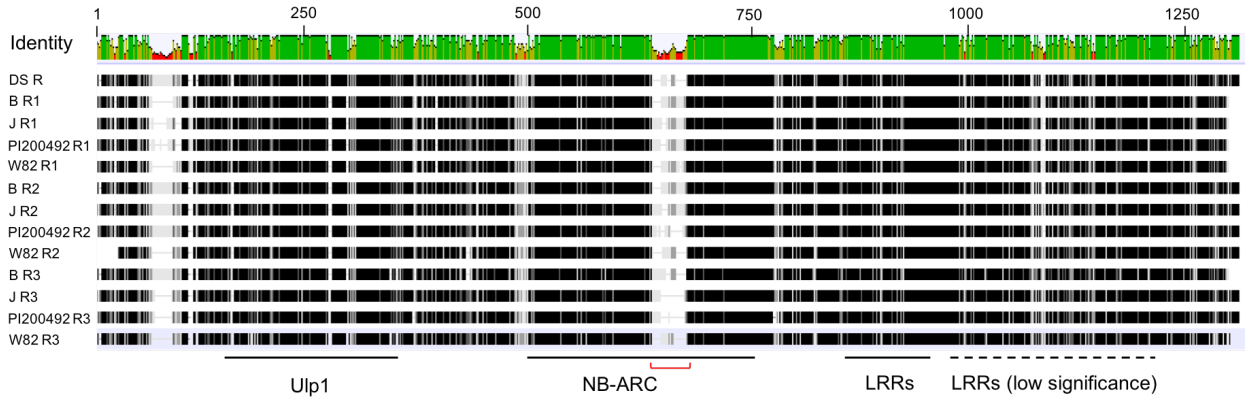
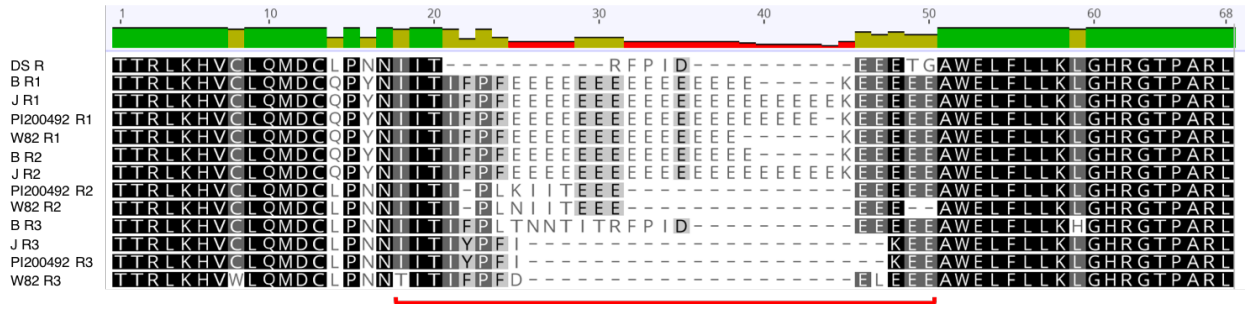


Figure 2.3. a, Assembled genomic sequence of *Rpp1* locus in TMG06_0011(DS), PI594760B (B) and PI561356 (J). **b,** Alignment of the *Rpp1* locus between B and DS.

a


1
250
500
750
1000
1250
Identity
DS R
B R1
J R1
PI200492 R1
W82 R1
B R2
J R2
PI200492 R2
W82 R2
B R3
J R3
PI200492 R3
W82 R3
Ulp1
NB-ARC
LRRs
LRRs (low significance)

b


1
10
20
30
40
50
60
68
DS R
B R1
J R1
PI200492 R1
W82 R1
B R2
J R2
PI200492 R2
W82 R2
B R3
J R3
PI200492 R3
W82 R3

c


1x10-6
J_R1
7.29x10-4
PI200492_R1
0.0859
0.0023
B_R1
0.0054
9.44x10-4
W82_R1
0.0024
J_R3
0.0213
0.0041
PI200492_R3
0.019
0.0275
DS_R
0.0033
J_R2
0.004
0.0022
B_R2
0.0359
0.0059
PI200492_R2
0.0117
0.026
W82_R2
0.0045
0.0454
B_R3
0.0019
0.0663
W82_R3
0.01

Figure 2.4. a, Protein alignments of thirteen NBS-LRR proteins located at the *Rpp1* locus of DS, B, J, PI200492 and William 82. The proteins are numbered by the positions of their corresponding genes in the genome. For instance, “B_R3” represents the third R gene of B genotype. Identity plot represents mean pairwise identity over all pairs in the column at each position. Green, 100% identity; yellow, 30%-100%; red, <30%. **b,** A 40 amino acid region in the NB-ARC domain that harbored variable numbers of glutamic acid repeats. **c,** Phylogenetic tree showing the distances between the NBS-LRR proteins. The method used for tree building is neighbor joining.

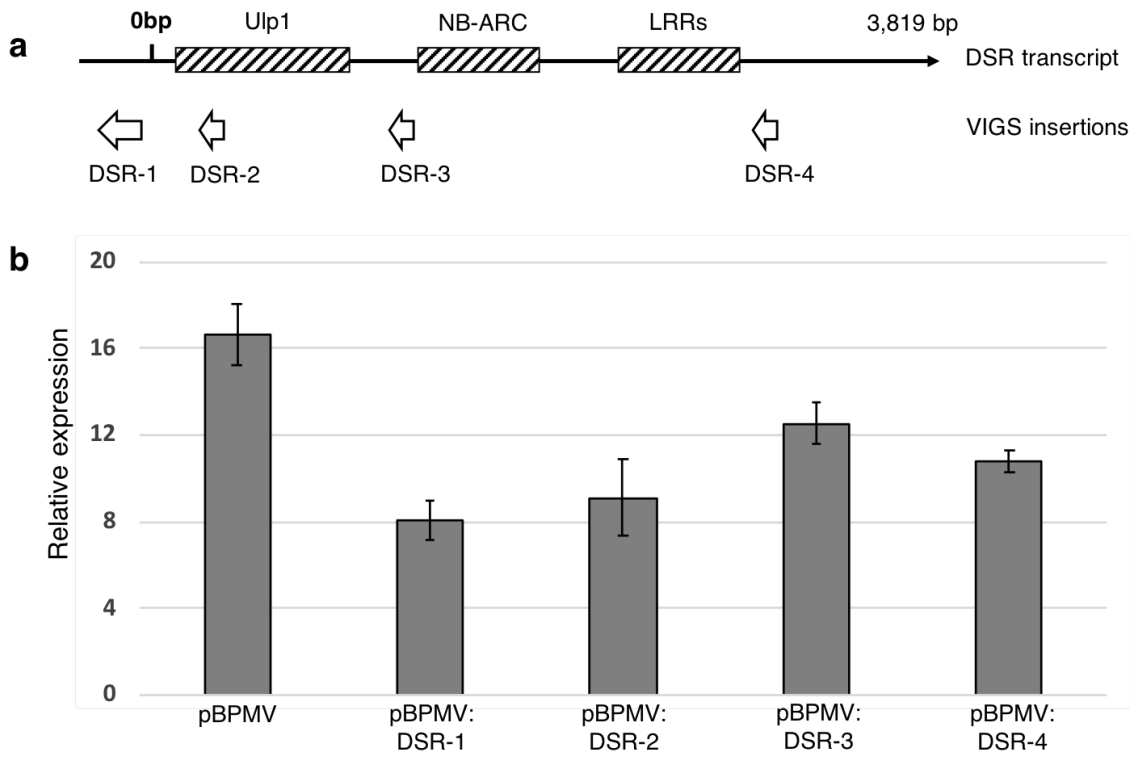


Figure 2.5. Selection of VIGS constructs. **a,** Target fragments to be inserted into the pBPMV vector for VIGS on DS-R transcript. All four insertions were in anti-sense orientation. **b,** Relative expressions of *DS-R* gene after inoculation with different VIGS constructs. Gene expressions were measured at three weeks post-inoculation, using new fully-expanded leaves on the top of plants. Relative expressions were calculated by 2^delta-Ct . Delta Ct was obtained by $Ct^{housekeeping} – Ct^{DSR}$.

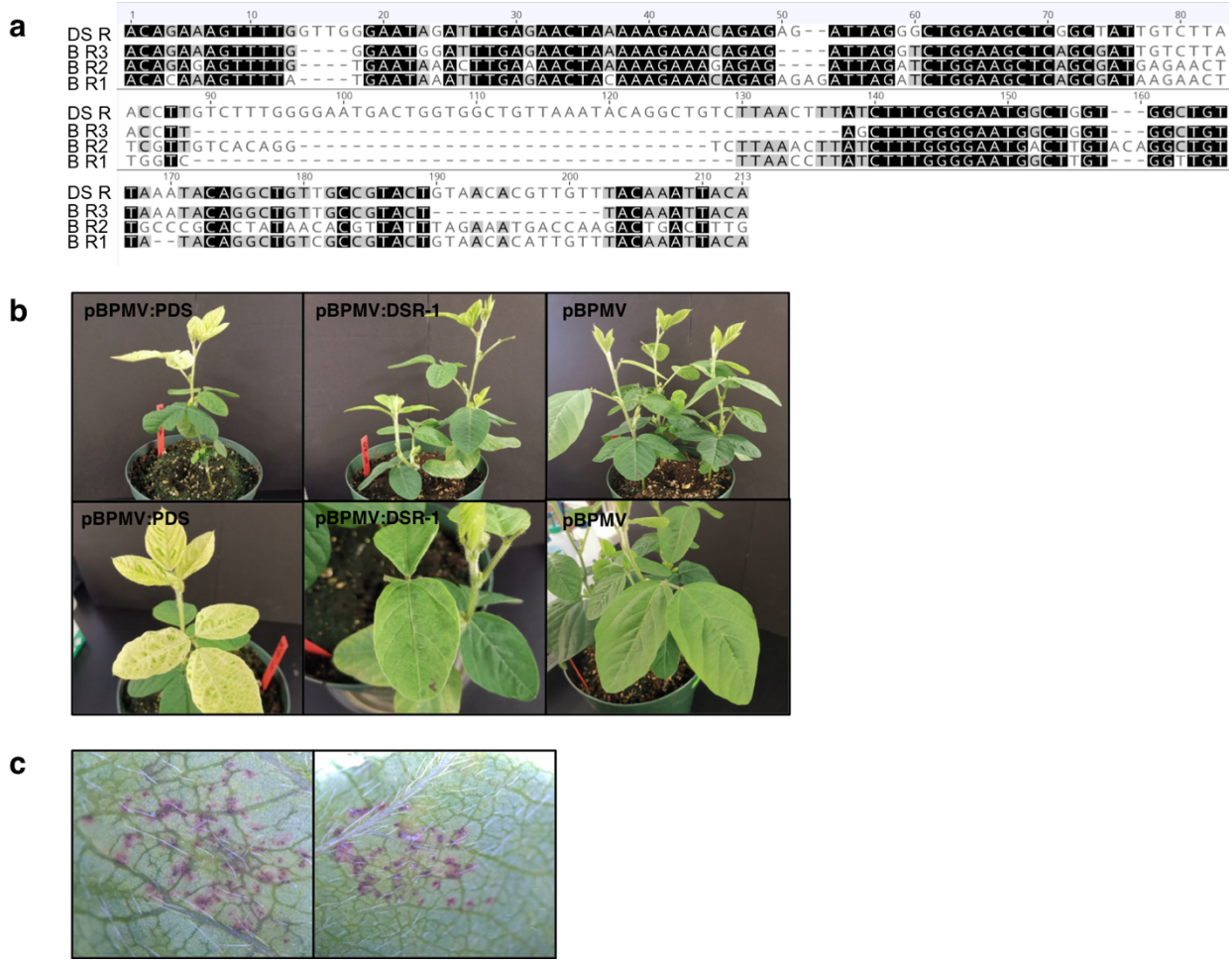


Figure 2.6. Evaluation of specificity of DSR-1 viral construct on the B X DS heterozygous genotype at the *Rpp1* locus. **a,** Sequence alignment of the VIGS fragment in DS-R and three B-R genes. **b,** Soybean plants at three weeks post inoculation of different pBPMV constructs. pBPMV construct targeting PDS (phytoene desaturase) served as a positive control of VIGS. **c,** RB lesions were displayed after rust inoculation on B plants that were inoculated with DSR-1 viral construct (i.e. if pBPMV:DSR-1 unintendedly silenced the *B-Rpp1*, the lesions would be TAN with sporulation).

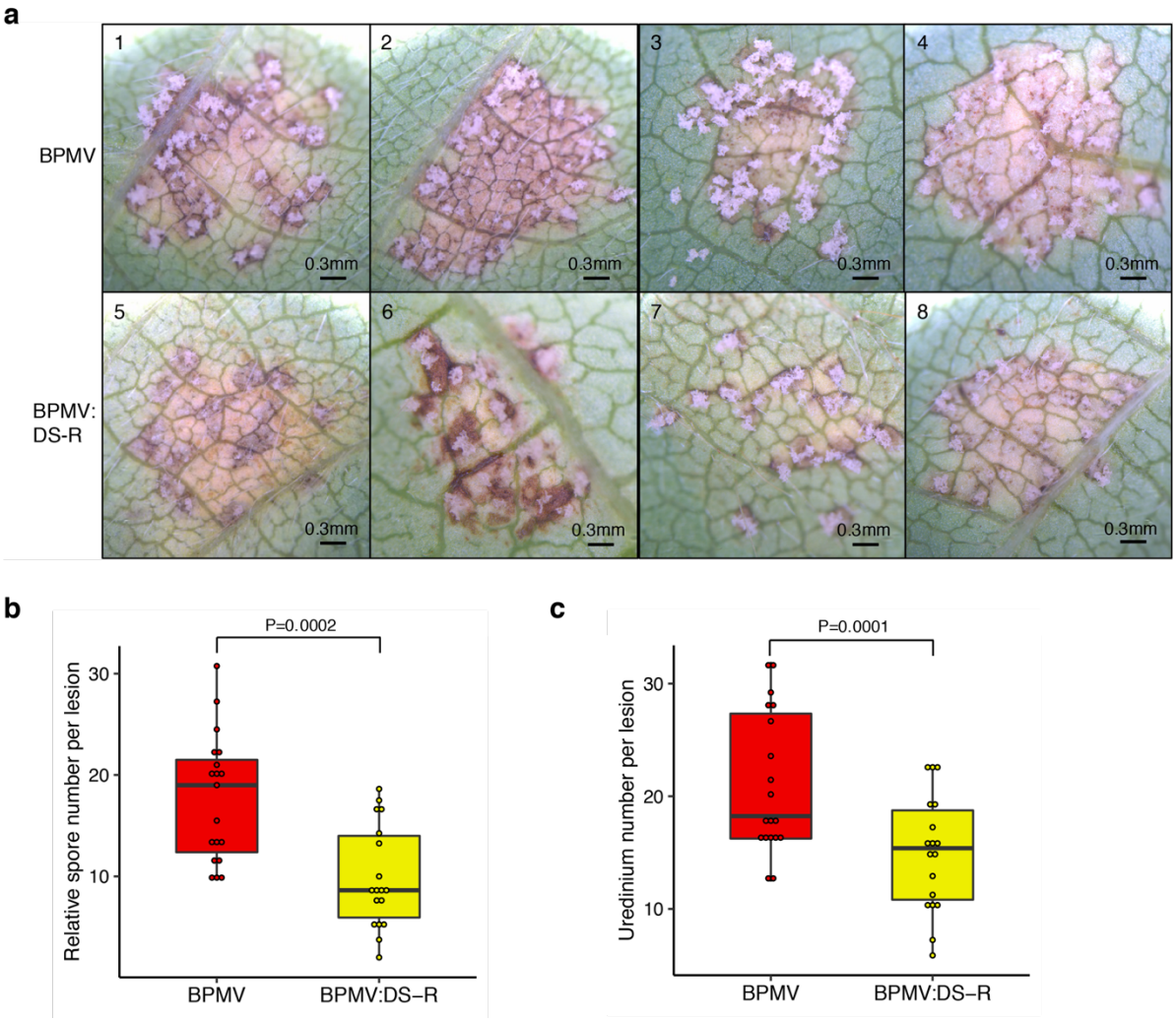


Figure 2.7. Partial restoration of resistance on B X DS plants (heterozygous *Rpp1* locus) after silencing the *DS-R* gene. **a,** Lesions generated by rust inoculation on BPMV:DSR-1 and BPMV empty vector treated leaves. **b,** Comparison of the relative spore number per lesion BPMV:DSR-1 and BPMV empty vector treated leaves. **c,** Comparison of the number of uredinia per lesion. Statistical analysis was conducted for relative spore number per lesion and number of uredinia per lesion using ANOVA and Friedman's two-way nonparametric ANOVA, respectively. Data pooled two batches of repeat experiment and included 10 biological replicates for each treatment in each batch.

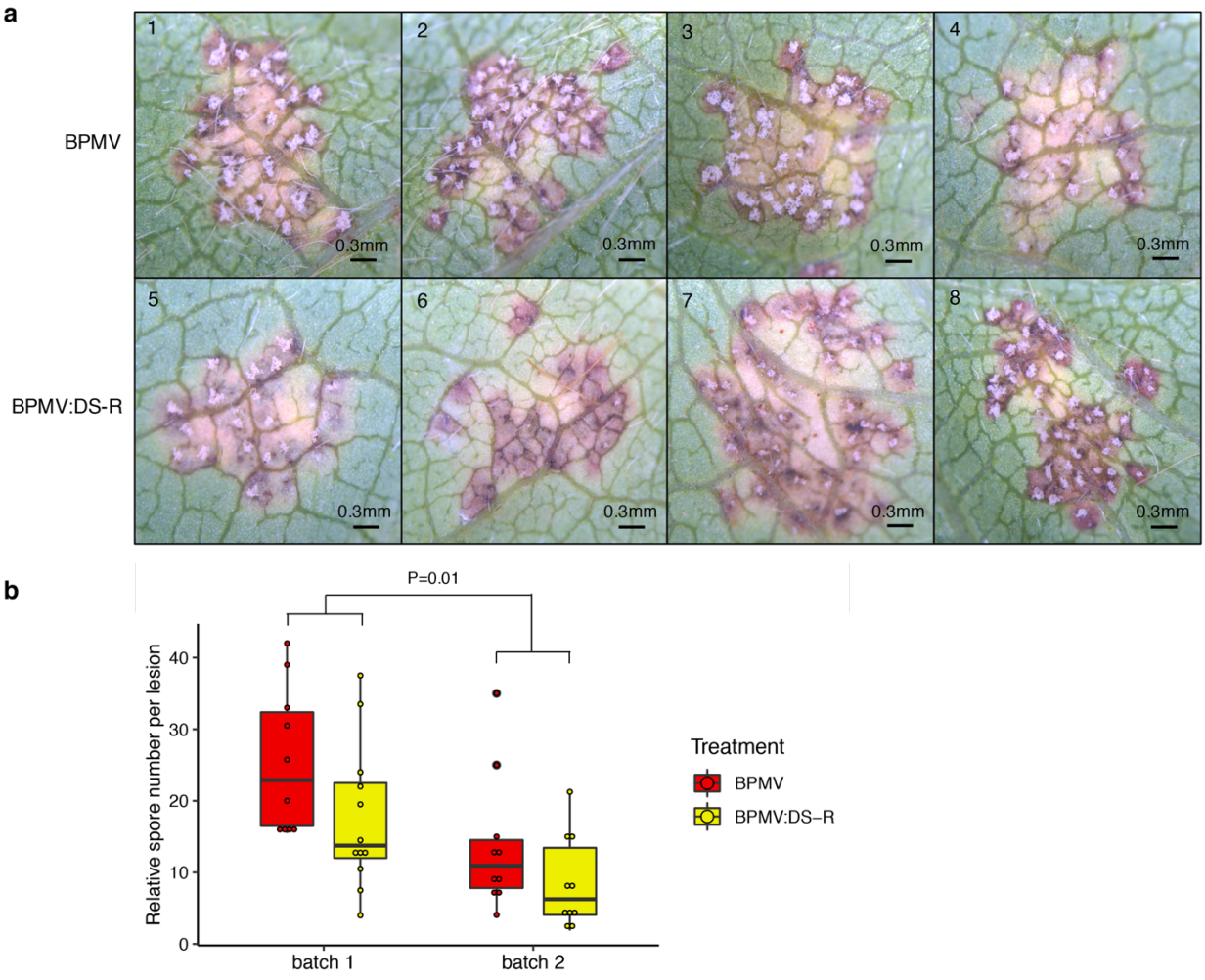


Figure 2.8. Partial restoration of resistance on PI200492 X DS plants (heterozygous *Rpp1* locus) after silencing the DS R gene. **a,** Lesions generated by rust inoculation on BPMV:DSR-1 and BPMV control treated leaves. **b,** Comparison of the relative spore number per lesion between BPMV:DSR-1 and BPMV control treated leaves. Data used two batches of experiment and included 10 biological replicates for each treatment in each batch. Friedman's two-way nonparametric ANOVA was performed and the model contained treatment and batch as factors. The p-value refers to the significance level of BPMV:DSR-1 samples compared to BPMV control samples.

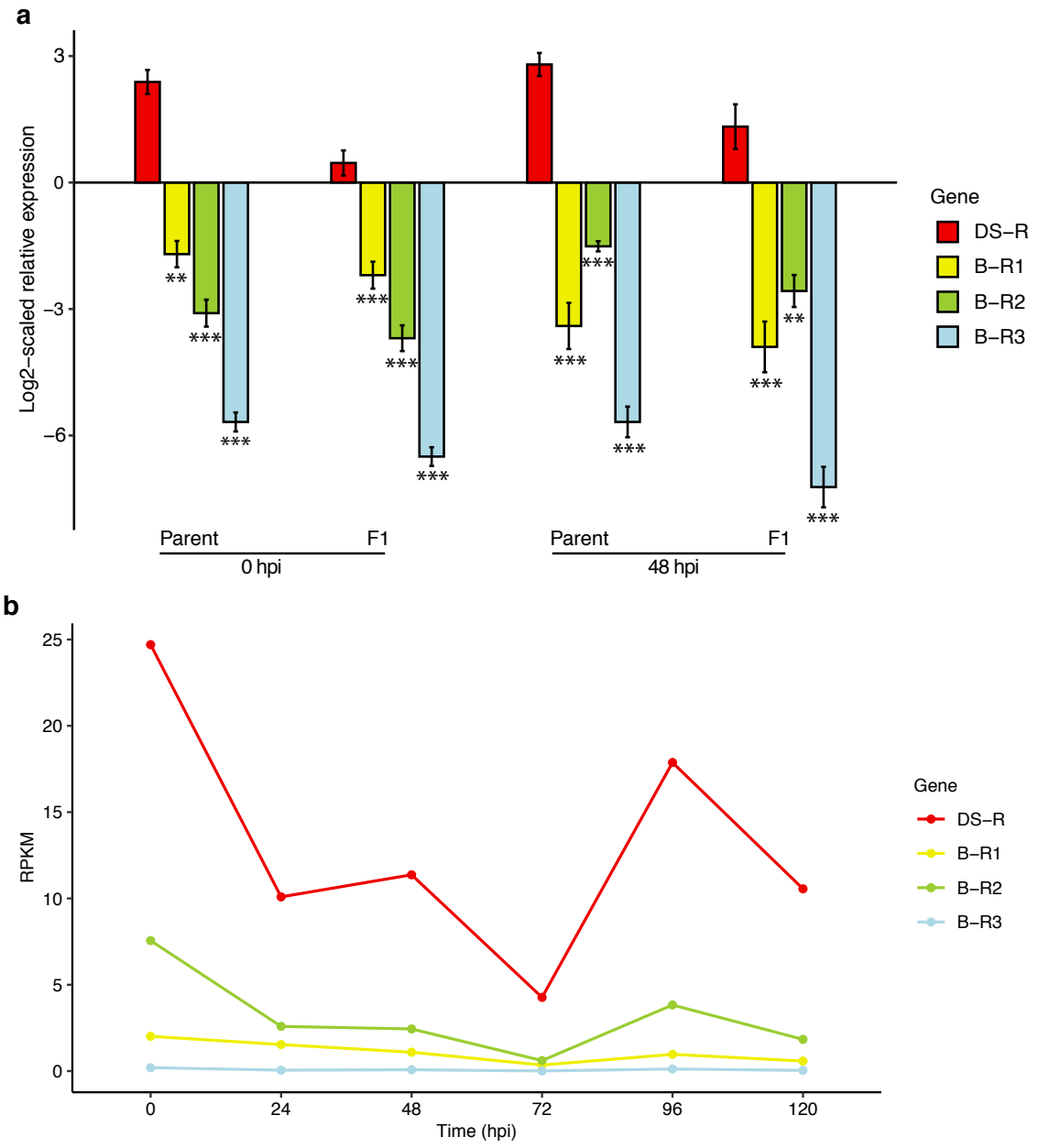
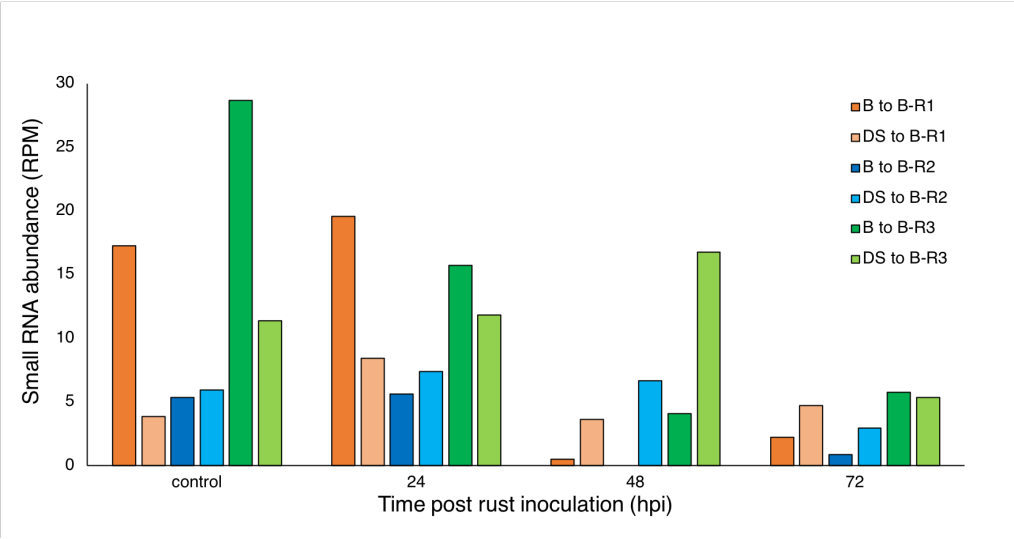


Figure 2.9. Gene expression of DS-R compared to BR1-3. **a**, Gene expression measured by qRT-PCR. Log2-scaled relative expression were generated by comparing to the reference gene *Cons15*. "Parent" refers to either DS or B genotype that harbored a homozygous *Rpp1* locus, and "F1" refers to the F1 progeny of DS and B that harbored the heterozygous *Rpp1* locus. For statistical analysis, one-way ANOVA was performed

independently for each genotype and time point combination. Asterisks indicate significant difference by pairwise comparisons with the DS-R gene. ***, P<0.0001; **, P<0.001 (multiple comparison adjusted by Tukey's method). **b**, Gene expression measured by RNA-Seq. RPKM, reads per kilobase of transcript, per million mapped reads.

a



b

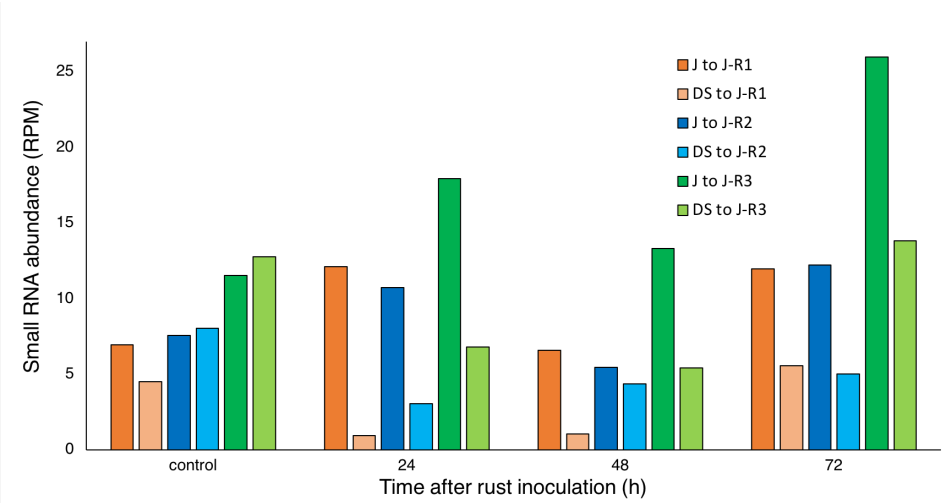


Figure 2.10. Comparison of abundances of small RNAs aligned to *Rpp1* candidate genes between DS, B and J. **a,** Small RNAs aligned to *B-R* genes. “DS to B” refers to small RNAs generated from the DS genotype aligned to the B genotype; “B to B” refers to small RNAs generated from the B genotype aligned to the B genotype itself. RPM, reads per million mapped reads. **b,** Small RNAs aligned to *J-R* genes.

a

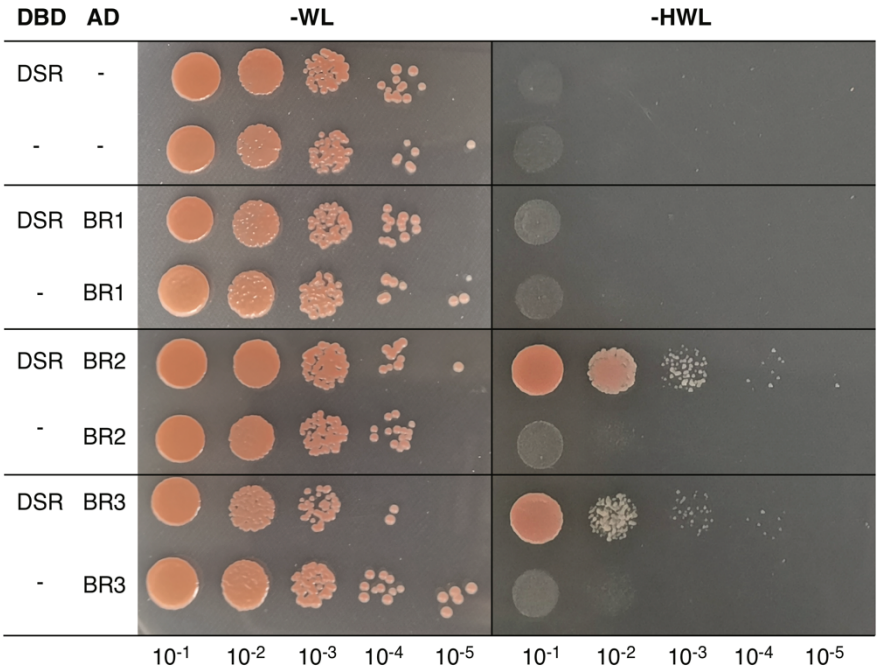


Figure 2.11. Evaluation of interaction between the DS-R protein and the B-R proteins using the yeast-two-hybrid system. **a,** Interaction between the full-length DS-R protein and full-length B-R proteins. DBD, DNA binding domain; AD, activation domain. The column on the left indicates which protein is fused to DBD and which is fused to AD. "-" indicates corresponding empty vectors. "-WL" indicated SD medium lacking Leu and Trp, "-HWL" lakcing Leu, Trp and His. Numbers in the lower row indicates 10-fold serial dilutions of different yeast co-transformants. Photos were taken after 3 days of growth. **b,** Interaction between individual domains of the DS-R protein and B-R proteins. The left column indicates which protein the domain come from; the row of texts above photos indicates which domain is transformed.

b

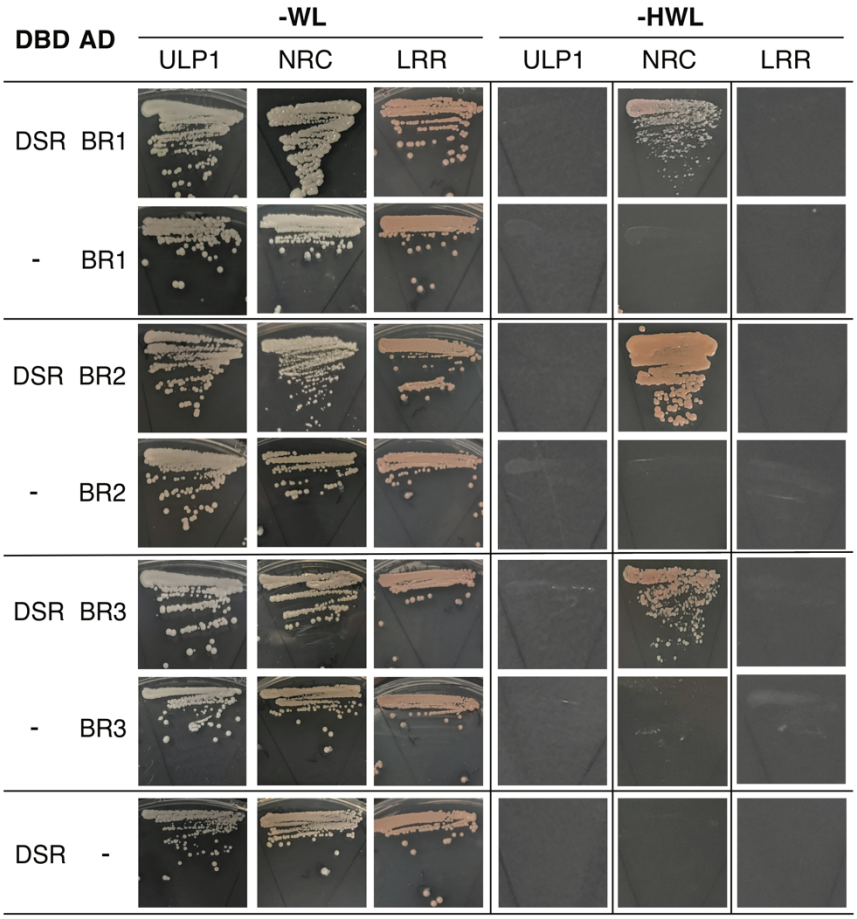


Figure 2.11. (continued)

Table 2.1. Primers for fosmid library screening.

Primer name	Primer sequence	Genotype	Position on W82 genome
DS-1-F	AAGCACTAACAACTTTGTGAAATGA	DS	no 100% hit on W82
DS-1-R	GGAGGTACATTCTAGTTGCCATTC	DS	
16-1700-F	TGTGTATGCGCAATTTCCCG	J	between Glyma.18G281600 and Glyma.18G281700
16-1700-R	GTTGCCCGCACTATAACACG	J	
B-5-F	AACTCAAGACAGTTAGTCAACCACA	B	between Glyma.18G281600 and Glyma.18G281700
B-5-R	GAGTTTGGTAGTGACTCCTTGCT	B	
B-8-F	TTACTTTTGGGCCGACCTCC	B	On Glyma. 18G281500
B-8-R	CCCAAAGATAAGGTTAAGACCAAGT	B	
13-1400-F	CACCAGTTCCTTGTTACGCACTA	J, B, DS	between Glyma.18G281600 and Glyma.18G281700
13-1400-R	GCTTTAGGCTTGGCCCTCAAAC	J, B, DS	
1800-1-F	ATGGCAAGCTGGGTCTTAGG	J, B, DS	On Glyma. 18G281800
1800-1-R	GGTTGCTAACCGCTTCCTCT	J, B, DS	

Table 2.2. Metrics of fosmid clone sequencing and assembly. Assembled length does not include the vector.

Library name	Oxford Nanopore				Illumina Miseq	Assembled length (kb)
	Total number of reads	Mean read length (nt)	20kb<read<50kb	30kb<read<50kb	Total number of reads	
B13	150,477	8,018	6,776	4,239	40,554	34,215
B15	95,328	8,294	4,920	3,381	51,773	37,408
B23	155,463	7,143	3,065	1,076	40,994	37,414
B29	74,458	8,622	3,913	2,984	45,839	41,708
B12	-	-	-	-	46,800	36,334
DS4	93,038	7,170	2,693	1,153	44,958	40,574
DS17	110,228	7,326	3,304	1,688	42,778	34,682
DS40	73,323	6,927	1,439	449	42,398	40,171
DS32	-	-	-	-	59,705	18,972
J2	77,399	8,570	3,958	2,824	27,037	42,043
J7	121,108	7,297	3,608	1,395	58,035	40,391
J14	152,436	6,136	1,469	239	41,918	37,960
XJ27	136,615	6,561	1,503	238	43,851	33,663

Table 2.3. summary of nucleotide polymorphisms between R protein candidates in PI200492 and J genotype. The amino acid position is based on the JR2 protein.

Amino acid position (aa)	Protein comparison			Domain
	J R1 vs PI200492 R1	J R2 vs PI200492 R2	J R3 vs PI200492 R3	
260	-	H/Q	-	Ulp1
355	-	-	E/-	-
469-478	-	KFIYLMIV / NFIYLLIE	-	-
519	-	I F	-	NB-ARC
622-662	-	QPYNIITIFPFEEEEEEEEEEEEEEEEEEEKEEEEAWE / LPNNIITIPLKIITEEEEEEEEAWE		NB-ARC
652	E/-	-	-	NB-ARC
678	-	P/S	-	NB-ARC
689-697	-	-	HISKGQWV/-ADEGKWA	-
835	-	K/N	-	-
1132	-	P/S	-	LRR (low significance)
1147	-	F/N	-	LRR (low significance)
1229	-	-	N/K	-

CHAPTER III GENOME WIDE ASSOCIATION STUDY OF SOYBEAN RESISTANCE TO ASIAN SOYBEAN RUST

Introduction

Asian soybean rust (ASR) is one of the most devastating soybean diseases worldwide. It is caused by the pathogen *Phakopsora pachyrhizi*. This fungus can produce a high amount of airborne spores that infect millions of acres of soybeans and lead to tremendous yield loss (Langenbach et al. 2016). It is currently the main threat to soybean production in many south American countries, such as Brazil, Argentina and Paraguay (B. and H. Villavicencio 2011). In Brazil, it has resulted in more than 10 billion dollars loss since its first outbreak in 2001, and cost 1-2 billion dollars per year from 2002 to 2014 for disease control (Langenbach et al. 2016; Godoy et al. 2016).

Deploying soybean genetic resistance is a common strategy to control ASR, in addition to chemical control. Its advantages as a complementary strategy are obvious, including reducing the expense of fungicides, reducing adverse impacts of chemicals on the environment and mitigating the selection pressure of fungicides on rust populations (Godoy et al. 2016). The majority of genetic resistance identified in soybean so far was conferred by single genes. To date, a total of seven loci distributed on five chromosomes (*Rpp1*-*Rpp7*) were mapped in soybean. For many loci, multiple alleles with differential pathogen profiles were identified, such as *Rpp1-b*, *Rpp-b-like* and *Rpp1-?*. Although they could provide complete resistance against corresponding avirulent rust strains, due to the highly variable rust populations, single-gene resistance can lose its effectiveness over years. *Rpp1* and *Rpp1-b* have been overcome by Brazilian field isolates (Kim et al. 2012; Ribeiro et al. 2007). The resistance conferred by *Rpp1* and *Rpp2* genes became ineffective in many areas in Asia between 2000 and 2010, with these two genes being incorporated in the breeding programs (Echeveste da Rosa 2015). The resistance of the *Rpp2* gene carried by PI230970 was found to be overcome in South America from 2000 to 2010 (Kato 2017). Thus, new resistance loci or new resistance alleles of known loci are always needed to expand the sources of

genetic resistance. Also, strategies such as R gene pyramiding to make resistance genes more durable are desired (Yamanaka et al. 2015).

Quantitative resistance is effective against a broad spectrum of rust isolates (Niks, Qi, and Marcel 2015). It is more durable than single gene mediated resistance because it will not lose effectiveness due to the loss of a single effector in rapidly evolving pathogen populations. Since the exhibition of resistance is often not complete, quantitative resistance is also called partial resistance. It is usually featured by slowing down the progress of epidemic disease development, such as lowering infection frequency through a longer latent period, smaller or fewer lesions, and less spore production. Soybean accessions with rate-reducing features have been identified for ASR, but the quantitative trait loci (QTLs) underlying the partial resistance have yet to be located (Mcebisi and Phinehas 2011). Although quantitative resistance genes may not provide immune-like resistance as R genes might, its higher durability and broad resistance spectrum is advantageous, and could be used as a supplementary source of genetic resistance. However, so far, little research has been reported to uncover the genetic components of quantitative resistance to rust in soybean genome and the mechanisms of this form or resistance are still largely unknown.

In the early 2000's, over 16,000 soybean accessions in the USDA Germplasm Collection located at the Urbana, Illinois were evaluated for resistance using a mixture of four foreign *P. pachyrhizi* isolates in USDA Biosafety Level 3 containment greenhouses at Fort Detrick, Maryland (Miles, Frederick, and Hartman 2006). Based on the initial screening, 3,215 accessions were selected for the second round of replicated evaluations, during which reaction types were recorded and leaf disease severity were measured in a scale of 1-5 (Figure 3.1). One of the main findings from the screening was that the leaf disease severity was not highly correlated with the lesion types. In other words, leaves exhibiting RB lesions could have as high lesion numbers as the leaves exhibiting TAN lesions. If leaf disease severity was considered as an index of quantitative resistance, possibly it would be controlled by different genetic mechanisms largely independent from R gene-controlled resistance. What is more, since multiple rust isolates from diverse sources were used, the leaf disease severity data was

expected to evaluate the average quantitative resistance to a broad spectrum of isolates. Thus, we hypothesized that this data could be utilized to study the genetic mechanism of soybean quantitative resistance to ASR. On the other hand, the reaction type phenotypes could also be used for discovering new resistance loci. Since multiple isolates were involved, it would be possible to identify more than one resistance loci in one dataset.

Genome-wide association analysis is widely used to study quantitative resistance in various crops to a diversity of pathogens (Arruda et al. 2016; Wang et al. 2012; Wei et al. 2017). With its power to dissect the genetic architecture of soybean quantitative resistance to ASR, it would be a promising tool to identify minor-effect loci contributing to quantitative resistance. Mapping of R genes for ASR usually take strategies like biparental linkage mapping or bulked segregate analysis that require building a mapping population for years (Silva et al. 2008; Hyten et al. 2009). GWAS enables mapping of resistance loci using a diversity panel without time-consuming population construction. The objectives of this chapter include: (1) to conduct a genome wide association study of the 2006 published data to identify QTLs for soybean quantitative resistance, and (2) to determine if the 2006 published data reveals any overlooked novel resistance loci to soybean rust isolates.

Materials and methods

Association analysis for leaf disease severity data

- Data filtering

Accessions with fewer than three replicates, or with a standard deviation greater than 0.58 (95 percentile), were removed from the dataset. Accessions that had not been genotyped by the SoySNP50K SNP chip were also removed. SNPs and accessions were both filtered for missing data at a maximum of 10% missing. Remaining missing data were imputed using Beagle v1.8. SNP sites and accessions with high heterozygosity were removed, with a cutoff at 10%. Alleles with low frequencies were filtered out by setting the minor allele frequency cutoff at 0.05. LD pruning was

performed on SNP data with a criterion of $r^2 < 0.8$. The filtering steps kept a total of 2,155 accessions and 15,643 SNP sites.

- Association mapping

Only the leaf severity data of accessions with TAN lesions were analyzed. The association analyses were performed with R package "GAPIT", using three different models: the Mixed Linear Model (MLM), the Multi-Loci Mixed Linear Model (MLMM), and the FarmCPU Model. To account for population structure and familial relatedness, the kinship matrix and 6 principle components were added for the "TAN" dataset, based on the scree plot (data not shown). For the FarmCPU model, significant SNPs were selected with an FDR-corrected p-value cutoff of 0.1. Only significant SNPs were retained if they also had a p-value less than 0.0001 with the MLM model.

Association analysis for reaction type data

For data filtering, accessions and SNPs were filtered with the same criteria as the leaf disease severity data, except that no SNP LD pruning was performed. A total of 2,196 accessions and 35,339 SNPs were kept in the end.

Since the type of disease reaction phenotypes were categorical and did not meet the requirements for general linear models, a logistic regression model was used to do the association analysis with the software Plink v1.9. Phenotypic data was transformed into numerical binary data, with the "RB" and "Mixed" combined as "1" and "TAN" as "0". To reduce false positives due to population structure, six principle components were added to the model as covariates. Significant SNPs were selected using a corrected p-value at 0.01 using the Bonferroni method. Linkage disequilibrium (LD) was measured using squared allele-frequency correlation (r^2) in Haploview v4.2 and the LD blocks were defined using the 95% confidence interval method (Gabriel et al. 2002).

Results

Association analysis for leaf disease severity data

The filtering steps retained a total of 2,155 accessions and 15,643 SNP sites. Within the 2,155 accessions, 1,885 presented a "TAN" reaction and the distribution of leaf severity exhibited a normal distribution with a scale from 1-5 (Figure 3.2a). With six PCs as covariates in the model, association mapping using MLM did not generate any significant loci with a cutoff of FDR-corrected p value at 0.1 (Figure 3.2c). MLMM had an output similar to the MLM. Association analysis using the FarmCPU model identified 20 significant loci with a cutoff of FDR-corrected p value at 0.1 (Figure 3.2c). Although association analysis using the MLM did not identify significant SNP, six significant SNPs by the FarmCPU model overlapped with some SNPs with lowest p-values by MLM: two on chromosome 2, one on chromosome 9, one on chromosome 13, one on 14 and one on 17 (Figure 3.2c). These overlapped loci might have a greater possibility of being true positives rather than artifacts. Looking at the effect size of these SNPs (Figure 3.2d), all of the significant loci were between -0.1 -0.1, which are very minor effects. The significant loci all together explained about 5% of the phenotypic variation. This result suggests the complexity of ASR quantitative resistance might be controlled by an array of small-effect genes, and that attempts to breed for effective quantitative resistance to ASR might be futile.

Association analysis for reaction type data

Logistic regression was performed for binary phenotypes consisting of 330 resistance accessions (63 "RB" and 267 "Mixed") and 1,886 susceptible accessions ("TAN"). A total of 57 SNPs reached the significance threshold of Bonferroni-corrected p-value at 0.01. As seen by the Manhattan plot (Figure 3.3a), "ss715632290", the most significant SNP and an array of following SNPs came from a 1.5 Mb region on chromosome 18, from 54,936,151-56,397,742bp. To remove the redundancy of significant SNPs or false positives, the most significant SNP was placed in the model as a covariate. After doing that, "ss715594493" and "ss715610517", the significant SNPs

on chromosome 6 and chromosome 11 became more significant, but in the meantime, some SNPs were removed; for example, the signal on chromosome 1 (Figure 3.3a). They were significant possibly due to correlation with the SNP on chromosome 18. When both the significant SNPs on chromosome 18 and chromosome 6 were fixed as covariates, the signal on chromosome 11 was still significant, as well as two loci on chromosome 18. Thus, a total of five significant loci were obtained and selected for further investigation. The "ss715632290" and "ss715632304" were at the *Rpp1* locus and the "ss715594493" was at the *Rpp3* locus (Figure 3.3b). The other two loci were not close to any known *Rpp* loci.

A potentially new *Rpp* locus close to *Rpp1*

Two significant loci from chromosome 18, represented by peak SNPs "ss715632290" and "ss715632304", were located near the *Rpp1* locus (Figure 3.4a). The SNP "ss715632290" was located within the 1.9Mb *Rpp1-?* locus from PI587886 and the 1.1Mb *Rpp1-b* locus from PI594538A, but it was outside of the 150 kb *Rpp1* locus from PI200492 and the 100 kb *Rpp1-b* like locus from PI561356. The "ss715632304" was located within the fine-mapped region of *Rpp1* from PI200492. When pairwise LD was calculated, these two SNPs were not in LD (r^2=0.02) with each other, which supports they were two independent loci. The SNP "ss715632290" was in a 27 kb LD block that was comprised of five SNPs and five different haplotypes (Figure 3.4b&c). The frequency of these haplotypes varied from 0.021 to 0.441, and among them, haplotype "AGGCT" with a frequency of 0.089 was significantly enriched in ASR resistant accessions. Based on the r-squared analysis, "ss715632290" showed LD (r^2>0.4) with several SNPs on its centromere proximal side between 56.04 Mb and 56.09 Mb, but not any SNP on its centromere distal side, where the *Rpp1* locus is more finely mapped. This finding raised the possibility that ss715632290 might be associated with a novel *Rpp* resistance locus. The SNP "ss715632304" did not lie in any LD blocks, nor exhibit LD ($r^2 > 0.4$) with any of the SNPs nearby (Figure 3.4b).

Accessions known to harbor various *Rpp1* alleles were examined for their genotypes at these two SNPs (Figure 3.4d). As seen in Figure 3.4d, none of the

accessions had the resistant allele at the SNP site “ss715632290”; neither did Williams 82, a susceptible genotype. Different haplotypes at the LD block containing “ss715632290” could be observed for accessions with the known resistant *Rpp1* gene but none of them were the haplotype mostly enriched in the resistant accessions. Therefore, SNP ss715632290 was not associated with any resistant *Rpp1* alleles identified previously. For the SNP site “ss715632304”, several accessions had the resistant allele that was different from the susceptible allele in William 82, such as PI200492, PI587886 and PI594767A. Thus, SNP ss715632304 could be associated with these resistant *Rpp1* alleles.

On the Williams 82 reference genome, there were four genes within the 27 kb LD block where the “ss715632290” located: Glyma.18g280200 (Leucine-rich repeat receptor-like protein kinase), Glyma.18g280300 (Ubiquitin-like protease protein 2, containing NB-ARC and LRR domain), Glyma.18g280400 (Leucine-rich repeat-containing protein) and Glyma.18g280500 (E3 ubiquitin ligase). These three genes were the best candidates for the signal “ss715632290”. Expanding the searching area to cover all the associated significant SNPs at that locus (54,936,151-56,397,742 bp), there were eleven NBS-LRR or leucine-rich repeat receptor-like kinase (LRR-RLK) genes (Appendix B). Besides the four mentioned, the closest other gene was LRR-RLK (Glyma.18g278300), lying 100 kb away from this SNP.

A potentially new resistance locus on chromosome 11

The SNP “ss715610517” on chromosome 11 was not within any known *Rpp* loci. However, it is the third most significant SNP identified in this study and with the top two both being true signals, its likelihood of being truly associated with a resistance gene is high. From the LD plot, the SNP “ss715610517” laid in a LD block with “ss715610518” that is only 1kb away from it (Figure 3.5a). Besides that, “ss715610517” was not in LD with any SNP. Three haplotypes detected for this two-SNP block: “GC”, “AT”, “GT”, with frequencies being 0.152, 0.384 and 0.465. Among them, haplotype “GC” was obviously enriched in resistant accessions and might be associated with a resistance gene (Figure 3.5b).

The candidate genes were searched 100 kb away from left to right of the significant SNP, due to relatively fast LD decay in this region (Figure 3.5a). There were a total of 32 genes in this window. Gene Glyma.11g254000 was annotated as “Leucine-rich repeat receptor-like protein kinase” and considered as a good candidate for an R gene. It was about 70 kb away from the SNP.

Discussion

- Comparison of different models for doing GWAS

This study conducted several genome-wide association mappings using the screening data with more than two thousand soybean accessions for resistance against ASR. Two types of resistance were looked at using corresponding phenotypes: leaf disease severity phenotype for quantitative resistance and lesion reaction type phenotype for qualitative resistance. For quantitative resistance, three models were adopted to conduct the mapping: MLM, MLMM and FarmCPU. Neither MLM nor MLMM identified significant SNPs with the cutoff of FDR-corrected p-value at 0.05 and 0.1, while FarmCPU model had five loci with p-value less than 0.05 and nineteen loci less than 0.1. The SNPs identified by the FarmCPU model were all minor-effect loci that account around or less than 1% of the phenotypic variation. By this result, FarmCPU seems to be a better model than MLM at identifying minor-effect loci that affect quantitative trait with a highly complex genetic architecture. Quantitative resistance to ASR could be this kind of trait and probably requires more sophisticated models than MLM to do association mapping.

MLM with population structure and familial relatedness accounted for has been applied widely in research on human, animals and plants for doing GWAS, due to its ability to reduce spurious associations and unmask true positive signals (Yu et al. 2006). However, it was found that for complex traits associated with population structure, true positive signals could be weakened when population structure and familial relatedness were accounted for in the model (Yang et al. 2014). The decreased power could be due to double-fitting of the associated marker, both as tested marker

and a part of the genetic relationship matrix (Yang et al. 2014). To mitigate this problem, many improvements based on MLM have been made, such as the compressed MLM (Lipka et al. 2012), SUPER (Wang et al. 2014) and FarmCPU (Liu et al. 2016). Our study found six significant SNPs (FDR-corrected $p < 0.1$) using FarmCPU that overlapped with SNPs with lowest p-values generated by MLM. With MLM, they could not pass the significance threshold but with FarmCPU they did. At the same time, FarmCPU also generated many significant loci that did not show signs of association with MLM and it is not known whether they were false positives or not. Thus, considering the pros and cons of different models, it is always beneficial to carefully chose the model for conducting GWAS based on species and trait being studied, or to conduct GWAS with many models and compare.

This study utilized a logistic regression model to do association mapping with the qualitative reaction type phenotype, with the aim to identify resistance loci that could confer qualitative resistance. The lesion type phenotype was turned into binary values, with “RB” and “Mixed” in the same group and “TAN” in the other. As a result, five significant loci were detected on three different chromosomes: chromosome 6, chromosome 11 and chromosome 18. Three of them were located at known *Rpp* loci, two at *Rpp1* (potentially different alleles) and one at *Rpp3*. Chang et al. (2016) conducted GWAS on the same dataset using MLM, with “RB”, “Mixed” and “TAN” as three-category phenotypes. The most significantly associated SNP in this study (ss715632290) was also identified in their study but not the SNP (ss715594493) on *Rpp3* and another SNP in the *Rpp1* locus that might refer to another *Rpp1* allele. Increased statistical power in this study might justify the different way of categorizing phenotypes and the logistic regression model. To control false positives due to population structure, six principle components were used as fixed covariates in the model. Furthermore, to reduce background signals from correlated loci with the large-effect loci, the first and second most significant SNPs were added to the model as covariates. These two strategies reduced false positives, but the Q-Q plot was still quite inflated (Figure 3.3 c&d). Analogous to the unified MLM that accounted for population structure and familial relatedness, generalized linear mixed model association test

(GMMAT) that adopted a logistic mixed model could solve the problem of spurious association (Chen et al. 2016). This model was compared to the unified MLM for binary traits with a maize diversity panel, and found it presented superior control of spurious association largely because the binary traits could violate the assumption of the constant error variance (Shenstone et al. 2018). Thus, our study could potentially be improved by using this logistic mixed model.

- Quantitative resistance to ASR

In this study, the identified significant SNPs associated with quantitative resistance accounted for a fairly low percentage of phenotypic variation, mostly less than 1%. Each SNP presented a predicted effect ranging from 0.04 to 0.1, which was fairly small in the context of the 1-5 scale. That matched the nature of quantitative resistance that was constituted by an array of minor-effect genes. However, this brings great difficulties for breeding for this trait because each locus is so minor and makes marker assisted selection futile. What makes breeding for quantitative resistance even less effective is that our model in total only explained 13% of the phenotypic variation, which means in our study, the variation was predominantly affected by other non-genetic effects, such as inoculation handlings, environmental factors and rating. Since the phenotyping procedure was conducted in the greenhouse, the environmental factors should be controlled relatively well compared to a field study, although variations in micro-environments are hard to avoid. Another major source of variation I propose might come from the way of inoculating. Since the lesion density on the leaf should also be influenced by how much spore could land on the leaf surface and initiate infection, it is very important to make sure this parameter is constant across the whole inoculation procedure. However, this might be hard to do with spraying spore suspensions onto plants because different soybean plants have different leaf morphologies. Even though the experimental protocol specifies “spray until several runoffs occurred” (Miles, Frederick, and Hartman 2006), the spore suspensions on the leaf surface could drop during movement of the plant into the moisture dew chamber. It could be more consistent to do detached leaf inoculation in a plastic box. The rating procedure could

also bring in variation. Using a “1-5” scale to measure the leaf disease severity has the advantage of being easy and fast but it is hard to be accurate, especially for intermediate levels like 2-4 and when doing thousands of leaves. Thus, it might lead to higher accuracy to change the “1-5” eyeballing rating into precise quantitative measurements, such as counting the lesion numbers in a certain leaf area. Last but the most important, the low genetic heritability on this data could also mean the leaf disease severity phenotype alone might not be a good index to assess genetic quantitative resistance. It would be worth trying another disease index more reflective of the pathogen reproduction, such as sporulation level. Walker et al. (2014) used a rust index score calculated from both rust leaf severity and sporulation level to evaluate the resistance of soybean germplasm to ASR in southern United States. Yamanaka et al. (2010) suggested using four criteria for selecting resistance materials: frequency of lesions having uredinia, number of uredinia per lesion, frequency of open uredinia and sporulation level, as sporulation was considered as more important than lesion color in evaluating resistance. Sixty-three genotypes were assessed using his criteria and variations were detected for number of uredinia per lesion, frequency of open uredinia and sporulation level. One drawback of these indexes is that they could be confounded by R gene-controlled resistance.

- Possible novel *Rpp* loci

This study identified one ASR resistance locus at *Rpp3*, a novel locus on chromosome 11 and another locus on chromosome 18 that was not close to any of *Rpp1*, *Rpp4* or *Rpp6*. It also identified two independent resistance loci on chromosome 18 near or at the *Rpp1* locus. Near *Rpp1*, the most significant SNP was close (~50-100 kb), but not located within the finely mapped *Rpp1* locus from PI200492 and PI561356 (Hyten et al. 2007; Kim et al. 2012). The LD block of this locus did not overlap with the fine mapped *Rpp1* locus, either. In addition, none of the soybean accessions with a known resistant *Rpp1* allele had the resistant allele of this SNP. A combination of these findings suggested that this SNP might represent a newly discovered ASR R gene locus close to the previously mapped *Rpp1* locus. However, it is also possible that the

significant signal should have located at the *Rpp1* locus, but instead was mis-mapped outside *Rpp1* due to insufficiency in SNP resolution or possible spurious artifacts of GWAS. If so, this SNP would then most likely be associated with a novel allele of *Rpp1*, rather than a novel locus.

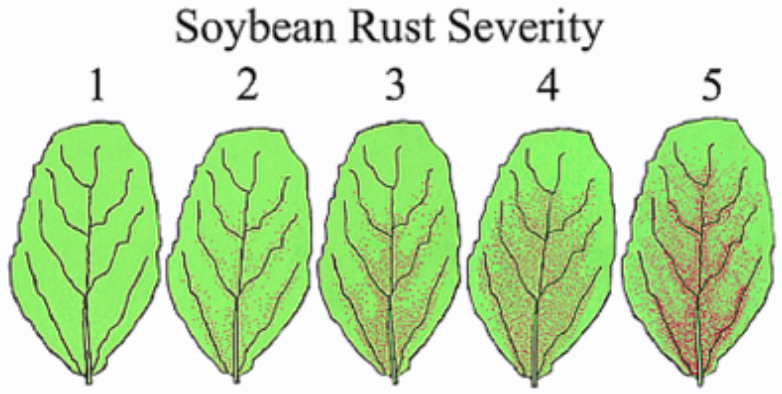
Pedley et al. (2018) cloned and sequenced more than 300 kb around the *Rpp1* locus from PI200492, and found eight candidate genes belonging to the NBS-LRR gene family (Figure 3.6). Among them, R1, R3, R4 and R5 encoded NBS-LRR proteins with a unique N terminal, Ubiquitin-like protease 1 domain (ULP1). Based on protein alignments, these are the only four genes in the soybean Williams 82 reference genome encoding ULP1-NBS-LRR, and they shared around 88% similarity. R3-R5 were considered as the best candidates for the original *Rpp1* gene from PI200492 since the resistance was partially lost after silencing these three genes simultaneously using VIGS (Pedley et al. 2018). In our study, the putative newly identified locus associated with SNP ss715632290 was closest to R1 and R2, and 130 kb away from R3-R5, raising the possibility that either R1 or R2 could be the causal R gene for this new locus.

To validate the potentially novel R genes, the best strategy would be to select soybean accessions with the resistance allele, inoculate with different rust isolates and compare to known resistance materials. Pham et al. (2009) selected 20 soybean entries and tested their reaction to 10 rust isolates, including the 4 isolates used in Miles et al, (2006). Among the 20 soybean entries, PI437323 that harbors the potentially novel *Rpp1* allele showed RB lesions with moderate sporulation to isolates from Paraguay (PG 01-2) and Louisiana (LA 04-1). Its reaction profile was found to be different from PI200492 (*Rpp1*) and PI587886 (*Rpp1-?*). Miles et al. (2011) selected 34 soybean accessions and tested them with the same four isolates in Miles et al (2006). PI467323A showed mixed lesions when inoculated with isolates from Brazil (BZ01-1) and Paraguay (PG01-2). PI427241 showed TAN lesions but significantly lower number of sporulating uredinia compared to most accessions such as PI200492 (*Rpp1*). Combined with evidence stated, the possibility of a novel R gene or *Rpp1* allele is high but the effectiveness of the R gene might be weak. To better map R genes using GWAS, higher resolution of markers might be needed in that region as the Soy50K SNP

chip might not capture enough markers associated with low-frequency resistance genes.

Figures and Tables

a

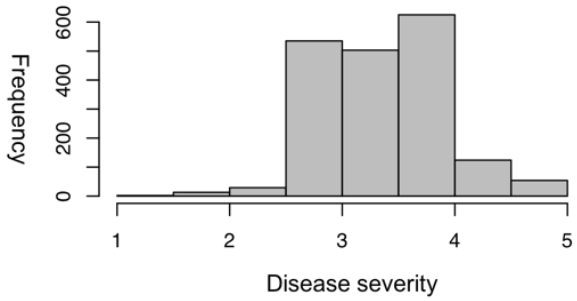


b

Line	Replicate#	Mean rating	Standard deviation	Reaction Type
PI240667A	3	2.0	0.25	RB
PI518759	2	2.5	0.68	TAN
PI567565	3	1.3	0.33	Mixed

Figure 3.1. a, Soybean leaf rust severity scales (from: Miles, Frederick, and Hartman 2006). "1-5" was used for visual assessment. 1 = no visible lesions, 2 = few scattered lesions, 3 = moderate number of lesions on at least part of the leaf, 4 = abundant number of lesions on at least part of the leaf, and 5 = prolific lesion development over most of the leaf. **b,** the raw data contained both disease severity rating and reaction type. RB, red-brown lesions; TAN, tan lesions; Mixed, a mixture of red-brown and tan lesions.

a



b

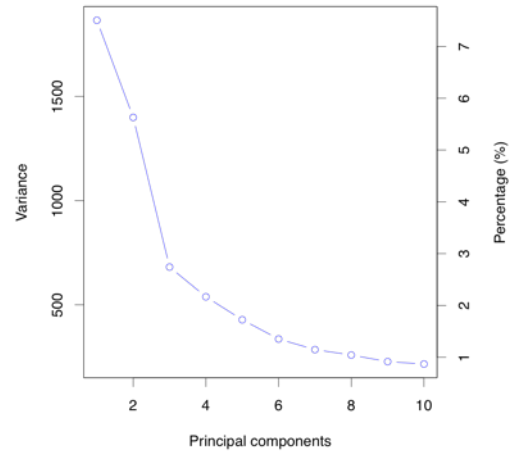
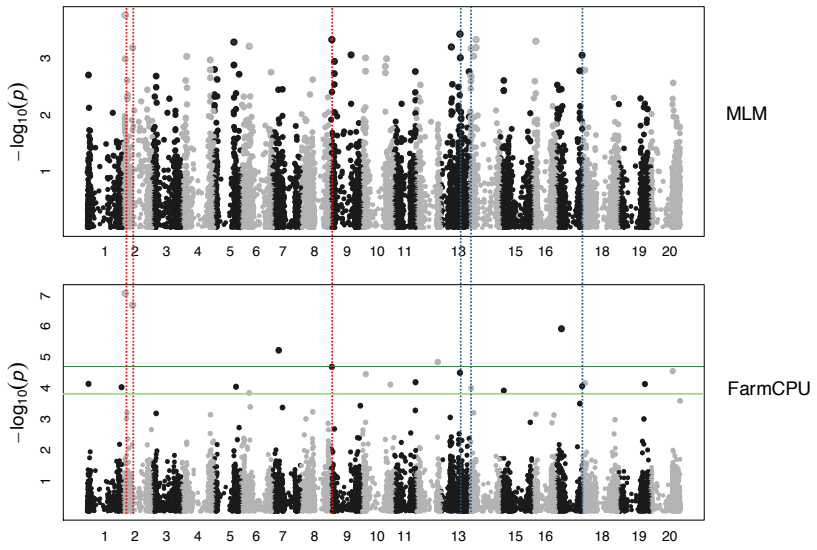


Figure 3.2. Association mapping results for leaf rust severity data. **a,** phenotype distribution. **b,** Eigen value plot of principle component analysis. **c,** Manhattan plots generated by the MLM model and the FarmCPU model. Green horizontal line corresponds to FDR-corrected p values of 0.05 and 0.1. Vertical dashed lines refer to significant loci by FarmCPU that were also the SNPs with lowest p values by MLM ($P<0.0001$). **d,** Six significant loci that overlapped between the two methods, corresponding to vertical dashed lines in panel c.

c

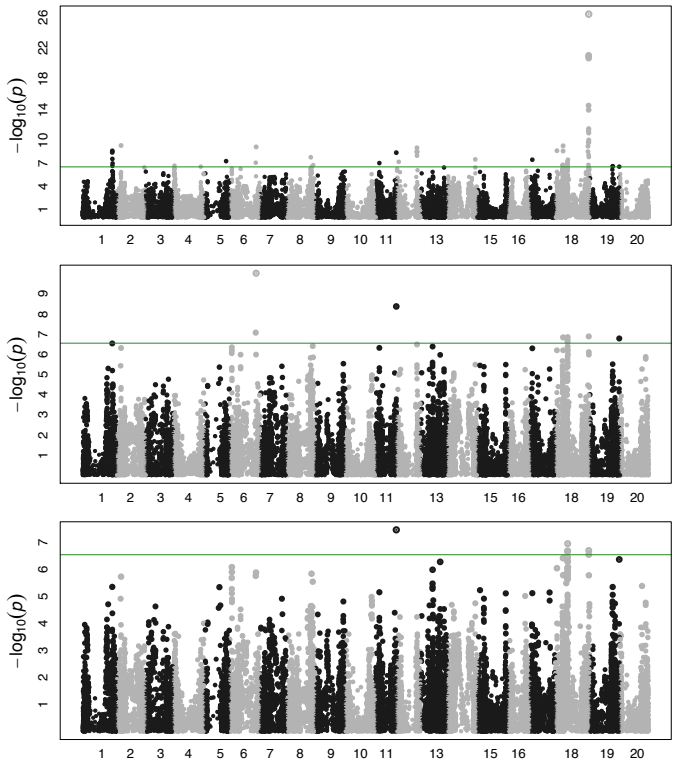


d

SNP	Chromosome	Position	P.value	FDR-corrected	effect	maf	Phenotypic variation explained
ss715582566	2	4385134	8.45E-08	0.0013	0.097	0.15	0.63%
ss715581367	2	16140341	2.05E-07	0.0016	0.077	0.16	0.54%
ss715603233	9	1943831	2.04E-05	0.0532	0.072	0.12	0.69%
ss715614921	13	30157970	3.18E-05	0.0621	-0.052	0.27	1.05%
ss715618050	14	2278087	1.00E-04	0.0992	-0.043	0.29	1.05%
ss715627808	17	40935537	8.68E-05	0.0856	-0.063	0.26	0.72%

Figure 3.2. (cont.)

a

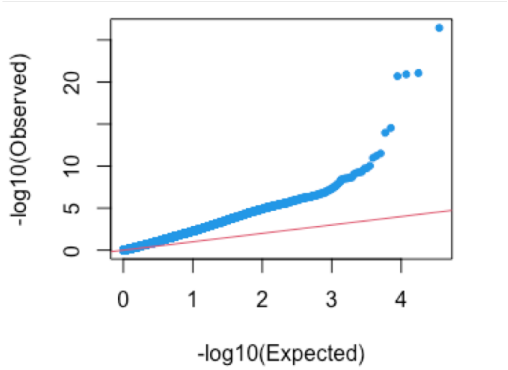


b

SNP	Chromosome	Position	P value	Maf	At or close to known Rpp locus?
ss715632290	18	56100116	3.74E-27	0.091	Rpp1
ss715594493	6	44867962	7.10E-10	0.074	Rpp3
ss715610517	11	34384415	4.00E-09	0.15	-
ss715629874	18	20835411	6.00E-08	0.258	-
ss715632304	18	56207755	9.18E-11	0.477	Rpp1

Figure 3.3. Association mapping using qualitative reaction-type data. **a,** Logistic regression model was used with different covariates. The green line in each plot corresponds to a Bonferroni corrected p-value of 0.01. **b,** Significant loci identified by logistic regression. The SNPs included the most and the second most significant SNPs and three other loci that passed the cutoff of corrected p-value < 0.01 after fitting the two most significant SNPs as covariates. Maf, minor allele frequency. **c,** Q-Q plot for logistic regression model with no SNP as covariates. **d,** Q-Q plot for logistic regression model with the first and second most significant SNP as covariates ("ss715632290" on chromosome 18 and "ss715594493" on chromosome 6)

c



d

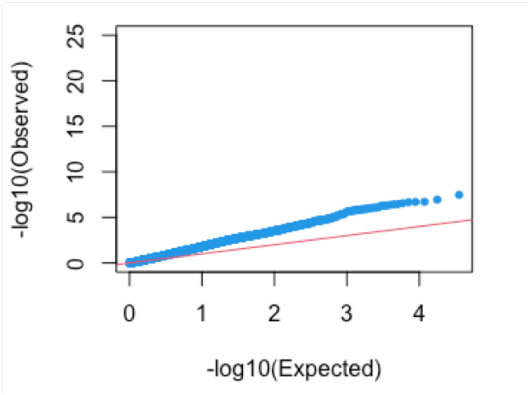
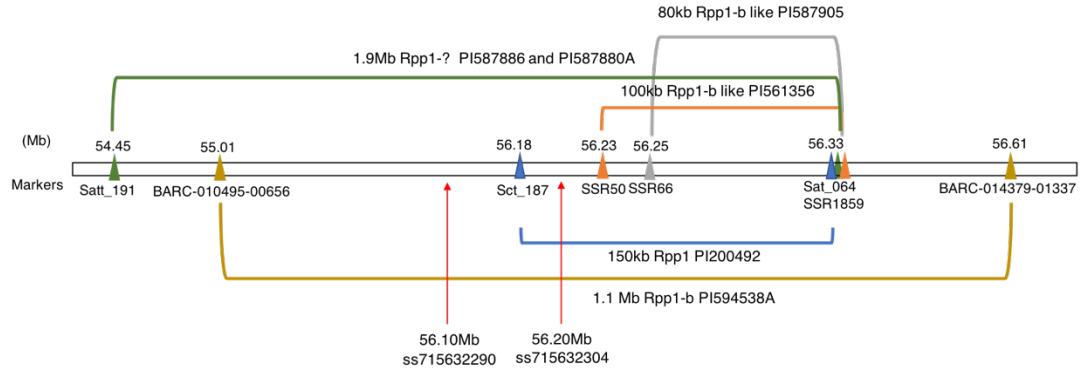


Figure 3.3. (cont)

a



b

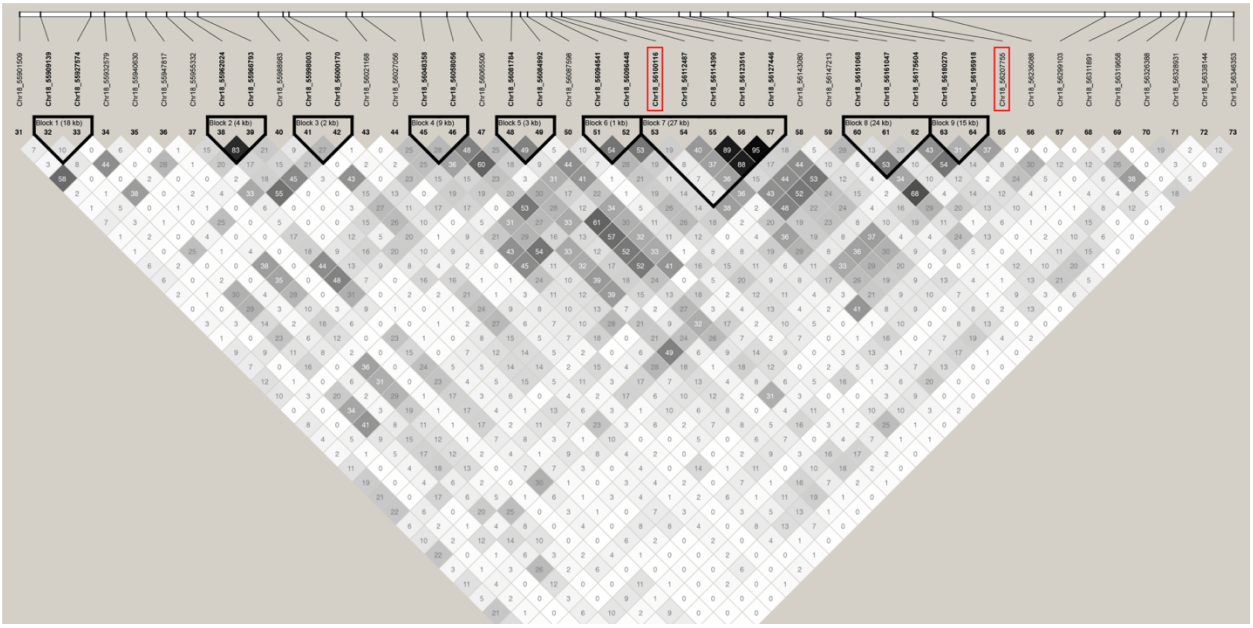
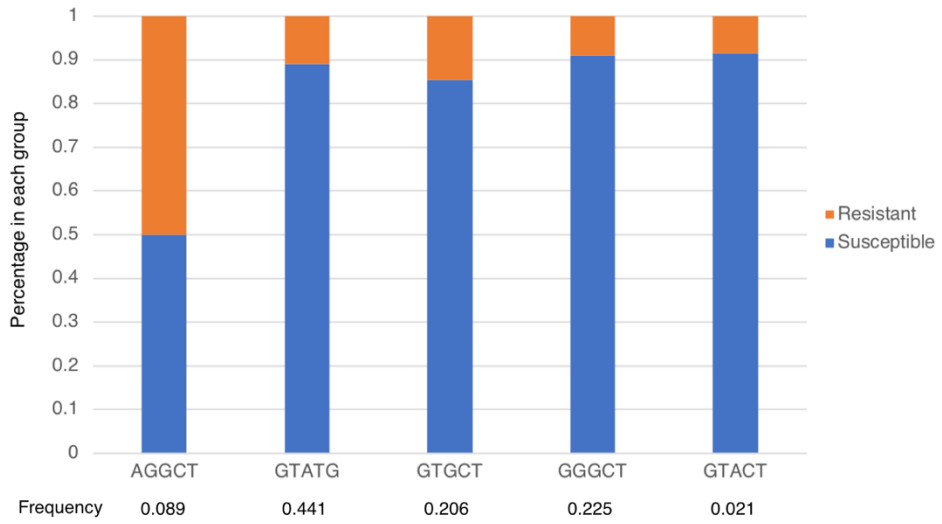


Figure 3.4. LD analysis for two SNPs on chromosome 18. **a,** physical positions of *Rpp1* alleles mapped using different soybean accessions and the physical positions of the two significant loci identified in this study. **b,** pairwise LDs of SNPs between 55.90Mb and 56.34Mb on chromosome 18. The value of LD was measured by r squared and presented by shades of grey. SNPs in the red rectangle refer to the two significant SNPs. **c,** distribution of each haplotype in susceptible and resistant accessions. **d,** Alleles at the two significant SNPs of different soybean accession that contained the Rpp1 resistance gene, compared to Williams 82. Yellow color indicates alleles that were significantly enriched in susceptible accessions and red color indicates alleles that were significantly enriched in resistant accessions.

c



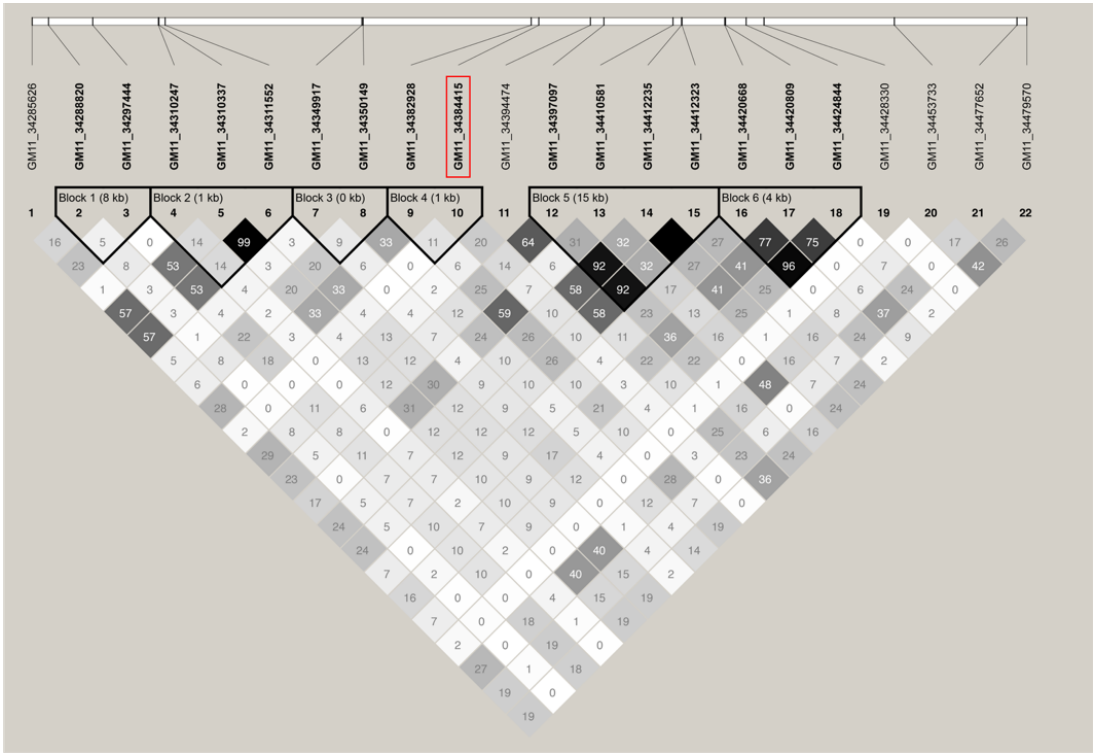
d

Accessions	Rpp1 allele	GM18_56100116 (ss715632290)	Haplotype*	GM18_56207755 (ss715632304)	Phenotype	Source
PI200492	Rpp1	G	GTATG	C	Mixed	Hyten et al. 2007
PI594538A	Rpp1-b	G	GTATG	T	Mixed	Chakraborty et al. 2009
PI561356	Rpp1-b like	G	GGGCT	T	Mixed	Kim et al. 2012a
PI587886	Rpp1-?	G	GTGCT	C	RB	Ray et al. 2009
PI587880A	Rpp1-?	G	GTACT	C	RB	Ray et al. 2009
PI594767A	Rpp1-b like	G	GGGCT	C	RB	Hossain et al. 2015
PI587905	Rpp1-b like	G	GGGCT	C	RB	Hossain et al. 2015
PI437323	unknown	A	AGGCT	C	RB	
PI467323A	unknown	A	AGGCT	T	Mixed	
Williams 82	**no Rpp1**	G	GTATG	T	TAN	

*Haplotype refers to haplotypes of the LD block comprised by “ss715632290” to “ss715632295”

Figure 3.4. (cont.)

a



b

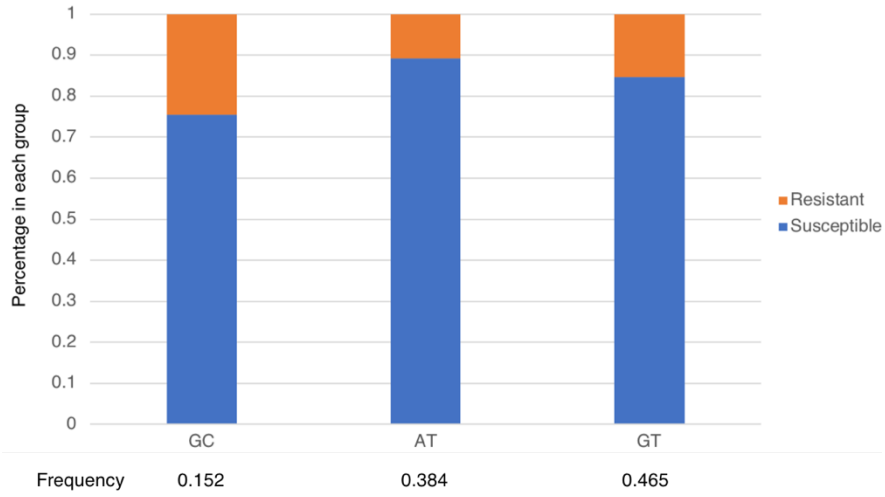


Figure 3.5. LD analysis for SNP "ss715610517" on chromosome 11. **a,** pairwise LDs of SNPs between 34.28Mb and 34.47Mb on chromosome 11. The value of LD was measured by r squared and represented by shades of grey. SNP in the red rectangle

refers to the significant SNP. **b,** Percentage of each LD block haplotype in susceptible and resistant accessions. Frequency represents the frequency of the corresponding haplotype in the population.

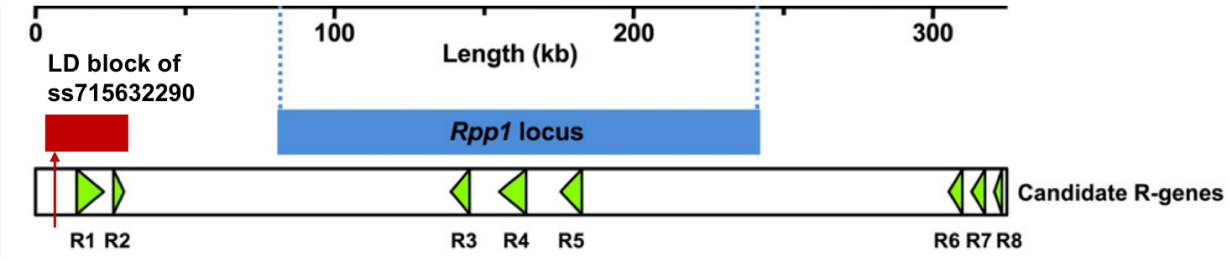


Figure 3.6. Positions of candidate genes of the *Rpp1* locus and the potentially novel locus (the red block) identified close to *Rpp1*. The figure is adapted from Pedley et al. 2018.

References

Ade, J., DeYoung, B. J., Golstein, C., and Innes, R. W. 2007. Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. Proc. Natl. Acad. Sci. U. S. A. 104:2531–2536.

Ameline-Torregrosa, C., Wang, B.-B., O'Bleness, M. S., Deshpande, S., Zhu, H., Roe, B., et al. 2007. Identification and Characterization of Nucleotide-Binding Site-Leucine-Rich Repeat Genes in the Model Plant Medicago truncatula. Plant Physiol. 146:5–21.

Arruda, M. P., Brown, P., Brown-Guedira, G., Krill, A. M., Thurber, C., Merrill, K. R., et al. 2016. Genome-Wide Association Mapping of Fusarium Head Blight Resistance in Wheat using Genotyping-by-Sequencing. Plant Genome. 9.

Ashikawa, I., Hayashi, N., Yamane, H., Kanamori, H., Wu, J., Matsumoto, T., et al. 2008. Two adjacent nucleotide-binding site-leucine-rich repeat class genes are required to confer Pikm-specific rice blast resistance. Genetics. 180:2267–2276.

Axtell, M. J. 2013. ShortStack: comprehensive annotation and quantification of small RNA genes. RNA. 19:740–51.

B., G., and H. Villavicencio, A. L. C. 2011. The Asian Soybean Rust in South America. Soybean Physiol. Biochem.

Bent, A. F., and Mackey, D. 2007. Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. Annu. Rev. Phytopathol. 45:399–436.

Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., et al. 2011. Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. Cell Host Microbe. 9:200–211.

Bettgenhaeuser, J., Gilbert, B., Ayliffe, M., and Moscou, M. J. 2014. Nonhost resistance to rust pathogens - a continuation of continua. Front. Plant Sci. 5:1–15.

Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., et al. 1997. The barley Mlo gene: A novel control element of plant pathogen resistance. Cell. 88:695–705.

Calla, B., Blahut-Beatty, L., Koziol, L., Simmonds, D. H., and Clough, S. J. 2014. Transcriptome analyses suggest a disturbance of iron homeostasis in soybean leaves

during white mould disease establishment. Mol. Plant Pathol. 15:576–88.

Casey, L. W., Lavrencic, P., Bentham, A. R., Cesari, S., Ericsson, D. J., Croll, T., et al. 2016. The CC domain structure from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. Proc. Natl. Acad. Sci. U. S. A. 113:12856–12861.

Centre, J. I., Lane, C., Kingdom, U., and Hall, K. 1998. Different requirements for EDS 1 and NDR 1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis. 95:10306–10311.

Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P. N. 2014. A novel conserved mechanism for plant NLR protein pairs: the “integrated decoy” hypothesis. Front. Plant Sci. 5:1–10.

Césari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., et al. 2014. The NB - LRR proteins RGA 4 and RGA 5 interact functionally and physically to confer disease resistance . EMBO J. 33:1941–1959.

Chakraborty, N., Curley, J., Frederick, R. D., Hyten, D. L., Nelson, R. L., Hartman, G. L., et al. 2009. Mapping and confirmation of a new allele at rpp1 from soybean pi 594538a conferring RB lesion-type resistance to soybean rust. Crop Sci. 49:783–790.

Chang, H.-X., Lipka, A. E., Domier, L. L., and Hartman, G. L. 2016. Characterization of Disease Resistance Loci in the USDA Soybean Germplasm Collection Using Genome-Wide Association Studies. Phytopathology. 106:1139–1151.

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., et al. 2016. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. Am. J. Hum. Genet. 98:653–666.

Childs, S. P., King, Z. R., Walker, D. R., Harris, D. K., Pedley, K. F., Buck, J. W., et al. 2018. Discovery of a seventh Rpp soybean rust resistance locus in soybean accession PI 605823. Theor. Appl. Genet. 131:27–41.

Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. Cell. 124:803–14.

Chou, C. C., and Wang, A. H. J. 2015. Structural D/E-rich repeats play multiple roles especially in gene regulation through DNA/RNA mimicry. Mol. Biosyst. 11:2144–2151.

Deslandes, L., Olivier, J., Peeters, N., Feng, D. X., Khounlotham, M., Boucher, C., et al. 2005. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. :600–603.

DeYoung, B. J., and Innes, R. W. 2006. Plant NBS-LRR proteins in pathogen sensing and host defense. Nat. Immunol. 7:1243–1249.

Dinesh-Kumar, S. P., Tham, W. H., and Baker, B. J. 2000. Structure-function analysis of the tobacco mosaic virus resistance gene N. Proc. Natl. Acad. Sci. U. S. A. 97:14789–14794.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21.

Dodds, Peter N, and Rathjen, J. P. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. Nat. Rev. Genet. 11:539–48.

Dodds, Peter N., and Rathjen, J. P. 2010. Plant immunity: Towards an integrated view of plantĝ€" pathogen interactions. Nat. Rev. Genet. 11:539–548.

Dow, M., Newman, M.-A., and Roepenack, E. von. 2000. The Induction and Modulation of Plant Defense Responses by Bacterial Lipopolysaccharides. 38.

Duprat, A., Caranta, C., Revers, F., Menand, B., Browning, K. S., and Robaglia, C. 2002. The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. Plant J. 32:927–934.

Echeveste da Rosa, C. R. 2015. Asian Soybean Rust Resistance: An Overview. J. Plant Pathol. Microbiol. 06.

Eitas, T. K., and Dangl, J. L. 2010. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Curr. Opin. Plant Biol. 13:472–7.

Ellis, J. G., Lawrence, G. J., Luck, J. E., and Dodds, P. N. 1999. Identification of Regions in Alleles of the Flax Rust Resistance Gene L That Determine Differences in Gene-for-Gene Specificity. Plant Cell. 11:495.

Fei, Q., Zhang, Y., Xia, R., and Meyers, B. C. 2016. Small RNAs Add Zing to the Zig-Zag-Zig Model of Plant Defenses. Mol. Plant. Microbe. Interact. 29:165–9.

Flor, H. 1971. Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. :275–296.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. 2002. The structure of haplotype blocks in the human genome. Science. 296:2225–9.

Garcia, A., Calvo, É. S., Kiihl, R. A. D. S., and Souto, E. R. De. 2011. Evidence of a Susceptible Allele Inverting the Dominance of Rust Resistance in Soybean. Crop Sci. 51:32.

Garcia, A., Calvo, E. S., de Souza Kiihl, R. A., Harada, A., Hiromoto, D. M., and Vieira, L. G. E. 2008. Molecular mapping of soybean rust (Phakopsora pachyrhizi) resistance genes: discovery of a novel locus and alleles. Theor. Appl. Genet. 117:545–53.

Gedling, C. R., Ali, E. M., Gunadi, A., Finer, J. J., Xie, K., Liu, Y., et al. 2018. Improved apple latent spherical virus-induced gene silencing in multiple soybean genotypes through direct inoculation of agro-infiltrated Nicotiana benthamiana extract. Plant Methods. 14:1–10.

Godoy, C. V., Dinali, C., Seixas, S., and Soares, R. M. 2016. Asian soybean rust in Brazil : past , present , and future. :407–421.

Green, M. R., and Sambrook, J. 2017. Isolation of high-molecular-weight DNA using organic solvents. Cold Spring Harb. Protoc. 2017:356–359.

Hammond-Kosack, K. E., and Jones, J. D. G. 1997. Plant disease resistance genes. Annu. Rev. Plant Biol. 48:575–607.

Hanae, K., Yoo, N., Naoko, I.-M., Chiharu, A.-T., Naoshi, D., Koji, T., et al. 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. Proc. Natl. Acad. Sci. 103:11086–11091.

Hartman, G., Hill, C. B., Twizeyimana, M., Miles, M. R., and Bandyopadhyay, R. 2012. Interaction of soybean and Phakopsora pachyrhizi, the cause of soybean rust. CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour. 6.

Hartman, G. L., Miles, M. R., and Frederick, R. D. 2007. Breeding for Resistance to Soybean Rust. Plant Dis. 89:664–666.

Hartman, G. L., Wang, T. C., and T., H. 1992. Sources of Resistance to Soybean Rust in Perennial Glycine Species. Plant Dis. 76:396–399.

van der Hoorn, R. A. L., and Kamoun, S. 2008. From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. Plant Cell. 20:2009–2017.

Hossain, M. M., Akamatsu, H., Morishita, M., Mori, T., Yamaoka, Y., Suenaga, K., et al. 2015. Molecular mapping of Asian soybean rust resistance in soybean landraces PI 594767A, PI 587905 and PI 416764. Plant Pathol. 64:147–156.

Hyten, D. L., Hartman, G. L., Nelson, R. L., Frederick, R. D., Concibido, V. C., Narvel, J. M., et al. 2007. Map location of the Rpp1 locus that confers resistance to soybean rust in soybean. Crop Sci. 47:837–840.

Hyten, D. L., Tucker, M. L., Cregan, P. B., Smith, J. R., Frederick, R. D., and Song, Q. 2009. Bulked segregant analysis using the goldengate assay to locate the Rpp3 locus that confers resistance to soybean rust in soybean. Crop Sci. 49:265–271.

Jones, J. D. G., and Dangl, J. L. 2006. The plant immune system. Nature. 444:323–9.

Kang, Y. J., Kim, K. H., Shim, S., Yoon, M. Y., Sun, S., Kim, M. Y., et al. 2012. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol. 12.

El Kasmi, F., and Nishimura, M. T. 2016. Structural insights into plant NLR immune receptor function. Proc. Natl. Acad. Sci. 113:12619–12621.

Kato, M. 2017. Effectiveness of resistance genes to the soybean rust pathogen Phakopsora pachyrhizi. Japan Agric. Res. Q. 51:199–207.

Kawashima, C. G., Guimarães, G. A., Nogueira, S. R., MacLean, D., Cook, D. R., Steuernagel, B., et al. 2016. A pigeonpea gene confers resistance to Asian soybean rust in soybean. Nat. Biotechnol. 34:661–665.

Kim, K. S., Unfried, J. R., Hyten, D. L., Frederick, R. D., Hartman, G. L., Nelson, R. L., et al. 2012. Molecular mapping of soybean rust resistance in soybean accession PI 561356 and SNP haplotype analysis of the Rpp1 region in diverse germplasm. Theor. Appl. Genet. 125:1339–1352.

Kim, S. H., Qi, D., Ashfield, T., Helm, M., and Innes, R. W. 2016. Using decoys to expand the recognition specificity of a plant disease resistance protein. Science (80-.). 351:684–687.

Lacomme, C., Hrubikova, K., and Hein, I. 2003. Enhancement of virus-induced gene silencing through viral-based production of inverted-repeats. Plant J. 34:543–553.

Lahaye, T. 2002. The Arabidopsis RRS1-R disease resistance gene – uncovering the

plant's nucleus as the new battlefield of plant defense? 7:425–427.

Langenbach, C., Campe, R., Beyer, S. F., Mueller, A. N., and Conrath, U. 2016. Fighting Asian Soybean Rust. Front. Plant Sci. 7:797.

Leister, D. 2004. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet. 20:113–116.

Li, L., Habring, A., Wang, K., and Weigel, D. 2020. Atypical Resistance Protein RPW8/HR Triggers Oligomerization of the NLR Immune Receptor RPP7 and Autoimmunity. Cell Host Microbe. 27:405-417.e6.

Li, S., Smith, J. R., Ray, J. D., and Frederick, R. D. 2012. Identification of a new soybean rust resistance gene in PI 567102B. Theor. Appl. Genet. 125:133–142.

Libault, M., Thibivilliers, S., Bilgin, D. D., Radwan, O., Benitez, M., Clough, S. J., et al. 2008. Identification of four soybean reference genes for gene expression normalization. Plant Genome J. 1:44.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics. 28:2397–2399.

Lipka, V., Dittgen, J., Bednarek, P., Bhat, R., Wiermer, M., Stein, M., et al. 2005. Pre- and postinvasion defenses both contribute to nonhost resistance in Arabidopsis. Science. 310:1180–3.

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome- Wide Association Studies. PLoS Genet. 12:e1005767.

Liu, Y., Du, H., Liu, Y., Du, H., Li, P., Shen, Y., et al. 2020. Resource Pan-Genome of Wild and Cultivated Soybeans Resource Pan-Genome of Wild and Cultivated Soybeans. Cell. :1–15.

Loehrer, M., Langenbach, C., Goellner, K., Conrath, U., and Schaffrath, U. 2008. Characterization of Nonhost Resistance of Arabidopsis to the Asian Soybean Rust . Mol. Plant-Microbe Interact. 21:1421–1430.

Lolle, S., Greeff, C., Petersen, K., Roux, M., Jensen, M. K., Bressendorff, S., et al. 2017. Matching NLR Immune Receptors to Autoimmunity in camta3 Mutants Using Antimorphic NLR Alleles. Cell Host Microbe. 21:518-529.e4.

Martin, R., Qi, T., Zhang, H., Liu, F., King, M., Toth, C., et al. 2020. Structure of the activated Roq1 resistosome directly recognizing the pathogen effector XopQ. Science (80-.). :2020.08.13.246413.

Mcebisi, M., and Phinehas, T. 2011. *State of knowledge on breeding for durable resistance to soybean rust disease in the developing world*.

McHale, L. K., Haun, W. J., Xu, W. W., Bhaskar, P. B., Anderson, J. E., Hyten, D. L., et al. 2012. Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol. 159:1295–1308.

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

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

Mestre, P., and Baulcombe, D. C. 2006a. Elicitor-mediated oligomerization of the tobacco N disease resistance protein. Plant Cell. 18:491–501.

Mestre, P., and Baulcombe, D. C. 2006b. Elicitor-mediated oligomerization of the tobacco N disease resistance protein. Plant Cell. 18:491–501.

Meyers, B. C., Kaushik, S., and Nandety, R. S. 2005. Evolving disease resistance genes. Curr. Opin. Plant Biol. 8:129–134.

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. 2003. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell. 15:809–34.

Miles, M. R., Bonde, M. R., Nester, S. E., Berner, D. K., Frederick, R. D., and Hartman, G. L. 2011. Characterizing resistance to Phakopsora pachyrhizi in soybean. Plant Dis. 95:577–581.

Miles, M. R., Frederick, R. D., and Hartman, G. L. 2006. Evaluation of Soybean Germplasm for Resistance to *Phakopsora pachyrhizi*. Plant Heal. Prog. 1:1–6.

Monteros, M. J., Missaoui, A. M., Phillips, D. V., Walker, D. R., and Boerma, H. R. 2007. Mapping and confirmation of the “Hyuuga” red-brown lesion resistance gene for Asian soybean rust. Crop Sci. 47:829–836.

van de Mortel, M., Recknor, J. C., Graham, M. a, Nettleton, D., Dittman, J. D., Nelson, R. T., et al. 2007. Distinct biphasic mRNA changes in response to Asian soybean rust

infection. Mol. Plant. Microbe. Interact. 20:887–899.

Niks, R. E., Qi, X., and Marcel, T. C. 2015. Quantitative Resistance to Biotrophic Filamentous Plant Pathogens: Concepts, Misconceptions, and Mechanisms. Annu. Rev. Phytopathol. 53:445–470.

Nishimura, M. T., Anderson, R. G., Cherkis, K. A., Law, T. F., Liu, Q. L., Machius, M., et al. 2017. TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in Arabidopsis . Proc. Natl. Acad. Sci. 114:E2053–E2062.

Oakeley, E. J., Jones, J. D. G., Zipfel, C., Felix, G., Robatzek, S., Navarro, L., et al. 2004. Bacterial disease resistance in Arabidopsis through flagellin perception. Nature. 428:764–767.

Paiano, A., Margiotta, A., De Luca, M., and Bucci, C. 2019. Yeast Two-Hybrid Assay to Identify Interacting Proteins. Curr. Protoc. Protein Sci. 95:1–33.

Palma, K., Thorgrimsen, S., Malinovsky, F. G., Fiil, B. K., Nielsen, H. B., Brodersen, P., et al. 2010. Autoimmunity in arabidopsis acd11 is mediated by epigenetic regulation of an immune receptor. PLoS Pathog. 6.

Pandey, A. K., Yang, C., Zhang, C., Graham, M. a, Horstman, H. D., Lee, Y., et al. 2011. Functional analysis of the Asian soybean rust resistance pathway mediated by Rpp2. Mol. Plant. Microbe. Interact. 24:194–206.

Paul, C., Harris, D. K., Li, Z., Bollich, P. A., and Walker, D. R. 2021. Reactions of 52 soybean germplasm accessions with Rpp3 alleles to a panel of 13 Phakopsora pachyrhizi (soybean rust) isolates from the southern United States. J. Gen. Plant Pathol. 87:55–70.

Paul, C., and Hartman, G. L. 2009. Sources of soybean rust resistance challenged with single-spored isolates of phakopsora pachyrhizi. Crop Sci. 49:1781–1785.

Pedley, K. F., Pandey, A. K., Ruck, A., Lincoln, L. M., Whitham, S. A., and Graham, M. A. 2018. Rpp1 Encodes a ULP1-NBS-LRR Protein That Controls Immunity to Phakopsora pachyrhizi in Soybean . Mol. Plant-Microbe Interact. 32:120–133.

Pham, T. a, Miles, M. R., Frederick, R. D., Hill, C. B., and Hartman, G. L. 2009. Differential Responses of Resistant Soybean Entries to Isolates of Phakopsora pachyrhizi. Plant Dis. 93:224–228.

Piffanelli, P., Zhou, F., Casais, C., Orme, J., Jarosch, B., Schaffrath, U., et al. 2002. The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. Plant Physiol. 129:1076–1085.

Porebski, S., Bailey, L. G., and Baum, B. R. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Report. 15:8–15.

Rao, X., Huang, X., Zhou, Z., and Lin, X. 2013. An improvement of the 2ˆ(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostat. Bioinforma. Biomath. 3:71–85.

Ray, J. D., Morel, W., Smith, J. R., Frederick, R. D., and Miles, M. R. 2009a. Genetics and mapping of adult plant rust resistance in soybean PI 587886 and PI 587880A. Theor. Appl. Genet. 119:271–280.

Ray, J. D., Morel, W., Smith, J. R., Frederick, R. D., and Miles, M. R. 2009b. Genetics and mapping of adult plant rust resistance in soybean PI 587886 and PI 587880A. Theor. Appl. Genet. 119:271–280.

Ribeiro, A. S., Moreira, J. U. V., Pierozzi, P. H. B., Rachid, B. F., De Toledo, J. F. F., Arias, C. A. A., et al. 2007. Genetic control of Asian rust in soybean. Euphytica. 157:15–25.

Ruffel, S., Dussault, M. H., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C., et al. 2002. A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). Plant J. 32:1067–1075.

Schneider, K. T., van de Mortel, M., Bancroft, T. J., Braun, E., Nettleton, D., Nelson, R. T., et al. 2011. Biphasic Gene Expression Changes Elicited by Phakopsora pachyrhizi in Soybean Correlate with Fungal Penetration and Haustoria Formation. Plant Physiol. 157:355–371.

Schreiber, K. J., Bentham, A., Williams, S. J., Kobe, B., and Staskawicz, B. J. 2016. Multiple Domain Associations within the Arabidopsis Immune Receptor RPP1 Regulate the Activation of Programmed Cell Death. PLoS Pathog. 12:1–26.

Shao, Z.-Q., Xue, J.-Y., Wu, P., Zhang, Y.-M., Wu, Y., Hang, Y.-Y., et al. 2016. Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. Plant Physiol.

170:2095–2109.

Shenstone, E., Cooper, J., Rice, B., Bohn, M., Jamann, T. M., and Lipka, A. E. 2018. An assessment of the performance of the logistic mixed model for analyzing binary traits in maize and sorghum diversity panels. PLoS One. 13:1–20.

Silva, D. C. G., Yamanaka, N., Brogin, R. L., Arias, C. A. A., Nepomuceno, A. L., Di Mauro, A. O., et al. 2008. Molecular mapping of two loci that confer resistance to Asian rust in soybean. Theor. Appl. Genet. 117:57–63.

Singh, R. J., and Nelson, R. L. 2015. Intersubgeneric hybridization between Glycine max and G. tomentella: production of F 1 , amphidiploid, BC 1 , BC 2 , BC 3 , and fertile soybean plants. Theor. Appl. Genet. 128:1117–1136.

Swiderski, M. R., and Innes, R. W. 2001. The Arabidopsis PBS1 resistance gene encodes a member of a novel protein kinase subfamily. Plant J. 26:101–112.

Takken, F. L., Albrecht, M., and Tameling, W. IL. 2006. Resistance proteins: molecular switches of plant defence. Curr. Opin. Plant Biol. 9:383–390.

Takken, F. L. W., and Goverse, A. 2012. How to build a pathogen detector: Structural basis of NB-LRR function. Curr. Opin. Plant Biol. 15:375–384.

Tao, Y., Yuan, F., Leister, R. T., Ausubel, F. M., and Katagiri, F. 2000. Mutational Analysis of the Arabidopsis Nucleotide Binding Site–Leucine-Rich Repeat Resistance Gene. Plant Cell. 12:2541 LP – 2554.

Walker, D. R., Boerma, H. R., Phillips, D. V, Schneider, R. W., Buckley, J. B., Shipe, E. R., et al. 2011. Evaluation of USDA Soybean Germplasm Accessions for Resistance to Soybean Rust in the Southern United States. Crop Sci. 51:678–693.

Walker, D. R., Harris, D. K., King, Z. R., Li, Z., Boerma, H. R., Buckley, J. B., et al. 2014. Evaluation of soybean germplasm accessions for resistance to Phakopsora pachyrhizi populations in the southeastern United States, 2009-2012. Crop Sci. 54:1673–1689.

Wang, Jizong, Hu, M., Wang, Jia, Qi, J., Han, Z., Wang, G., et al. 2019. Reconstitution and structure of a plant NLR resistosome conferring immunity. Science (80-.). 364.

Wang, Jizong, Wang, Jia, Hu, M., Wu, S., Qi, J., Wang, G., et al. 2019. Ligand-triggered allosteric ADP release primes a plant NLR complex. Science (80-.). 364.

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y., et al. 2012. Genome-wide

association study (GWAS) of resistance to head smut in maize. Plant Sci. 196:125–131.

Wang, Q., Tian, F., Pan, Y., Buckler, E. S., and Zhang, Z. 2014. A SUPER powerful method for genome wide association study. PLoS One. 9.

Wei, W., Mesquita, A. C. O., Figueiró, A. de A., Wu, X., Manjunatha, S., Wickland, D. P., et al. 2017. Genome-wide association mapping of resistance to a Brazilian isolate of Sclerotinia sclerotiorum in soybean genotypes mostly from Brazil. BMC Genomics. 18:1–16.

Yamanaka, N., Morishita, M., Mori, T., Lemos, N. G., Hossain, M. M., Akamatsu, H., et al. 2015. Multiple Rpp-gene pyramiding confers resistance to Asian soybean rust isolates that are virulent on each of the pyramided genes. Trop. Plant Pathol. 40:283–290.

Yamanaka, N., Yamaoka, Y., Kato, M., Lemos, N. G., Passianotto, A. L. de L., Santos, J. V. M. dos, et al. 2010. Development of classification criteria for resistance to soybean rust and differences in virulence among Japanese and Brazilian rust populations. Trop. Plant Pathol. 35:153–162.

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. 2014. Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. 46:100–6.

Yorinori, J. T., Soja, E., Paiva, W. M., Miranda, C., Frederick, R. D., Ars, U., et al. 2005. Symposium Epidemics of Soybean Rust (Phakopsora pachyrhizi) in Brazil and Paraguay from 2001 to 2003. :2003–2005.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–8.

Yue, J. X., Meyers, B. C., Chen, J. Q., Tian, D., and Yang, S. 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytol. 193:1049–1063.

Zhang, C., Bradshaw, J. D., Whitham, S. A., and Hill, J. H. 2010. The Development of an Efficient Multipurpose Bean Pod Mottle Virus Viral Vector Set for Foreign Gene Expression and RNA Silencing. Plant Physiol. 153:52–65.

Zhang, C., Whitham, S. A., and Hill, J. H. 2013. Virus-induced gene silencing in soybean

and common bean. Methods Mol. Biol. 975:149–156.

Zhou, B., Qu, S., Liu, G., Dolan, M., Sakai, H., Lu, G., et al. 2006. The eight amino-acid differences within three leucine-rich repeats between Pi2 and Piz-t resistance proteins determine the resistance specificity to Magnaporthe grisea. Mol. Plant. Microbe. Interact. 19:1216–28.

APPENDICES

Appendix A. Primers used for Chapter II.

Appendix B. Candidate R genes for the locus associated with SNP “ss715632290” on chromosome 18.