Running head: Evolution of $R$ Gene Specificity in Soybean

Corresponding author: Roger W. Innes
Myers Hall Rm 150
915 E. Third St.
Bloomington,
IN 47405-7107
Tel. 812 855 2219
Fax 812 855 6082
e-mail rinnes@indiana.edu

Research Area: Genes, Development and Evolution
Evolutionary Relationship of Disease Resistance Genes in Soybean and Arabidopsis

Specific for the *Pseudomonas syringae* Effectors AvrB and AvrRpm1

Tom Ashfield, Thomas Redditt, Andrew Russell, Ryan Kessens, Natalie Rodibaugh, Lauren Galloway, Qing Kang, Ram Podicheti, and Roger W. Innes*

Department of Biology, Indiana University, Bloomington, Indiana 47405

One-sentence Summary: Polymorphisms in the LRR domains of two closely related soybean NB-LRR genes allow them to distinguish between two pathogen effectors detected by a single disease resistance gene in Arabidopsis.
This work was supported by the National Science Foundation (Plant Genome Research Program grant no. DBI–0321664 to R.W.I. and the National Institute of General Medical Sciences at the National Institutes of Health (grant nos. R01GM046451 and R01GM063761 to R.W.I.)

*Corresponding author; email rinnes@indiana.edu
ABSTRACT

In Arabidopsis, the *Pseudomonas syringae* effector proteins AvrB and AvrRpm1 are both detected by the RPM1 disease resistance (R) protein. In contrast, soybean can distinguish between these effectors, with AvrB and AvrRpm1 being detected by the Rpg1b and Rpg1r R proteins, respectively. We have been using these genes to investigate the evolution of R gene specificity and have previously identified RPM1 and Rpg1b. Here, we report the cloning of Rpg1r, which like RPM1 and Rpg1b, encodes a coiled-coil-nucleotide binding-leucine rich repeat (CC-NB-LRR) protein. As previously found for Rpg1b, we determined that Rpg1r is not orthologous with RPM1, indicating that the ability to detect both AvrB and AvrRpm1 evolved independently in soybean and Arabidopsis. The tightly-linked soybean Rpg1b and Rpg1r genes share a close evolutionary relationship, with Rpg1b containing a recombination event that combined a NB domain closely related to Rpg1r with CC and LRR domains from a more distantly related CC-NB-LRR gene. Using structural modeling, we mapped polymorphisms between Rpg1b and Rpg1r onto the predicted tertiary structure of Rpg1b, which revealed highly polymorphic surfaces within both the CC and LRR domains. Assessment of chimeras between Rpg1b and Rpg1r using a transient expression system revealed that AvrB versus AvrRpm1 specificity is determined by the C-terminal portion of the LRR domain. The *P. syringae* effector AvrRpt2, which targets RIN4 proteins in both Arabidopsis and soybean, partially blocked recognition of both AvrB and AvrRpm1 in soybean, suggesting that both Rpg1b and Rpg1r may detect these effectors via modification of a RIN4 homolog.
INTRODUCTION

Effector triggered immunity in plants involves highly specific recognition events in which plant resistance (R) proteins detect pathogen effector proteins directly, or alternatively, the modifications that they induce on host proteins (Bonardi et al., 2012). The largest group of R proteins belongs to the nucleotide binding-leucine rich repeat (NB-LRR) family (McHale et al., 2006). The NB-LRR family can be further subdivided based on N-terminal domains into the Toll-Interleukin and R protein (TIR-NB-LRR) class and non-TIR-NB-LRR class (McHale et al., 2006). The latter most often contain a coiled coil (CC) domain at the N-terminus. The contributions of the TIR, CC, and LRR domains to R protein specificity, and how new specificities evolve, remain important questions.

There are relatively few NB-LRR R proteins characterized to date that are thought to detect pathogen effectors directly; these include Pi-ta from rice, L and M variants from flax, and RRS1 and RPP1 from Arabidopsis (Jia et al., 2000; Deslandes et al., 2003; Dodds et al., 2006; Ueda et al., 2006; Catanzariti et al., 2010; Krasileva et al., 2010). In at least some of these examples, the R genes are found in clusters of NB-LRR paralogs in which multiple recognition specificities are represented (Ellis et al., 1995; Botella et al., 1998), or belong to allelic series (Ellis et al., 1995), arrangements which may promote evolution of recognition specificity via recombination between alleles and paralogs. Interestingly, sequence comparisons and domain swaps involving alleles at the L locus implicate both the LRR and TIR regions as determinants of recognition specificity (Ellis et al., 1999; Luck et al., 2000). Subsequently, domain swaps involving paralogs clustered at the barley MLA and potato Rx/Gpa loci have provided additional support for the LRR domain playing a key role in conferring recognition specificity (Ellis et al., 1999; Luck et al., 2000; Shen et al., 2003; Rairdan and Moffett, 2006).

Several R proteins are known to detect the presence of pathogen effectors indirectly, by monitoring the activity of pathogen effectors within the plant cell. For example, the Arabidopsis RPM1 and RPS2 R proteins detect modification of the effector target RIN4, while the Arabidopsis RPS5 protein detects modification of the effector target PBS1 (Mackey et al., 2002; Axtell and Staskawicz, 2003; Mackey et al., 2003; Shao et al., 2003). At least for the well-studied examples in Arabidopsis, R proteins that employ indirect recognition mechanisms are encoded by NB-LRR genes that are not members of large clusters, or allelic series, with variants
encoding distinct recognition specificities. Correlated with this genomic structure, such loci are typically relatively stable, with \textit{RPM1} and \textit{RPS5} existing as presence/absence polymorphisms that have been maintained over long evolutionary periods (Stahl et al., 1999; Tian et al., 2002). Both functional and nonfunctional alleles of \textit{RPS2} have been isolated, but only a single recognition specificity has been detected at this locus, despite sequence polymorphisms between alleles (Caicedo et al., 1999).

Most likely, specificity for this class of R proteins is determined by a combination of the ability to associate with the host protein targeted by the effector, and the ability to detect effector-induced modification of this target. Consistent with this hypothesis, it has been shown that the CC domains from at least some R proteins interact with the host proteins they are monitoring, even in the absence of pathogen effectors, in a pre-recognition complex (Mackey et al., 2002; Ade et al., 2007). Hence, evolution of recognition specificity in R proteins that employ indirect recognition mechanisms may involve evolution of both the N-terminal CC and LRR domains.

To better understand the evolution and function of R proteins that detect pathogen effectors indirectly, we have been studying two soybean R genes, with known recognition specificities, that are members of a complex NB-LRR cluster. The R genes involved, \textit{Rpg1b} and \textit{Rpg1r}, mediate detection of the \textit{Pseudomonas syringae} effector proteins AvrB and AvrRpm1, respectively (Staskawicz et al., 1984; Ashfield et al., 1995). We have previously cloned \textit{Rpg1b}, which is a CC-NB-LRR gene that maps to a cluster of R genes effective against a diverse range of pathogens (Ashfield et al., 1998; Ashfield et al., 2004). \textit{Rpg1r} is present in the same cluster and maps 0.56 cM from \textit{Rpg1b} (Ashfield et al., 1995), however, the evolutionary relationship shared by the two R genes is not known. The cluster is associated with numerous NB-LRR genes, of both the CC and TIR sub-groups, spread over more than a megabase of soybean chromosome 13 (Penuela et al., 2002; Hayes et al., 2004; Innes et al., 2008; Ashfield et al., 2012; Wen et al., 2012). The NB-LRR family in this region is evolving rapidly with duplications/deletions of paralogs, recombination, and positive selection all playing a role (Ashfield et al., 2012).

While soybean can distinguish between AvrB and AvrRpm1, both effectors are detected by a single R protein, RPM1, in Arabidopsis (Bisgrove et al., 1994; Grant et al., 1995). It is known that RPM1 recognizes the effector proteins indirectly, by detecting effector-dependent
phosphorylation of a second Arabidopsis protein, RIN4 (Mackey et al., 2002; Chung et al., 2011; Liu et al., 2011). The available evidence suggests that a related strategy is employed by soybean, at least for the Rpg1b protein, despite the AvrB-recognition specificity having evolved independently in these plant species (Ashfield et al., 2004; Selote and Kachroo, 2010; Selote et al., 2012). Soybean contains four RIN4 homologs (Chen et al., 2010), three of which interact physically with Rpg1b, with two required for full resistance conferred by this R gene (Selote and Kachroo, 2010; Selote et al., 2012). It is not known whether RIN4 homologs are required for Rpg1r function.

Here we report the map-based cloning of the soybean Rpg1r gene. Comparison of the Rpg1r protein to Rpg1b, combined with structural modeling, revealed highly polymorphic surfaces in the CC and LRR domains. Transient expression of chimeric Rpg1 proteins demonstrated that specificity for AvrB versus AvrRpm1 is determined by the C-terminal LLR region. Finally, we provide evidence that Rpg1r, like Rpg1b, detects its corresponding pathogen effector indirectly, most likely by monitoring a RIN4 homolog, indicating convergent evolution of recognition mechanisms in separate plant families.

RESULTS

Genetic and Physical Mapping of Rpg1r

We have previously scored a small recombinant inbred line (RIL) population, derived from a cross between cultivars (cvs) Flambeau (rpg1b Rpg1r) and Merit (Rpg1b rpg1r), for the segregation of both Rpg1b and Rpg1r. Based on these data we determined that Rpg1b and Rpg1r are not allelic, but are tightly linked, with Rpg1r mapping 0.56 (± 0.77) cM from Rpg1b (Ashfield et al., 1995). Rpg1b was subsequently mapped to soybean chromosome 13 (formally, molecular linkage group F), flanked by the microsatellite markers HSP176 and Sct33 (Ashfield et al., 1998), allowing the inference of a closely linked position for Rpg1r.

To further delimit the genetic interval that contains Rpg1r we scored informative recombinants from the cv. Flambeau x Merit RIL population with PCR-based markers corresponding to the BAC contigs generated during the cloning of Rpg1b (Ashfield et al., 2003; Ashfield et al., 2004). In this manner Rpg1r was delimited to an interval flanked by the markers
Origin of \( R \) Gene Specificity in Soybean

EP-52d1r and EP-221b6r (Supplemental Fig. S1). We have previously sequenced most of this interval in cv. Williams 82 (Innes et al., 2008) using a BAC-based strategy with one gap being filled from the soybean whole genome sequence (WGS) (Schmutz et al., 2010). The genetically defined interval corresponds to a physical distance of \(~ 291.5 \) Kb in the Williams 82 sequence assembled from the BAC contigs. Interestingly, this region contains four intact CC-NB-LRR (CNL) genes, and one pseudogene belonging to the NB-LRR family (Fig. 1 and Supplemental Fig. S1). In the soybean WGS, which is also derived from cv. Williams 82, the genetically defined interval extends from gene Glyma13g26000 to gene Glyma13g26300 (chromosome 13, position 29,225,278 to 29,524,087 in the Glyma1 assembly of the soybean WGS), a physical distance of \(~ 298.8 \) Kb. While the BAC-derived sequence and WGS are mostly co-linear, small indels are apparent when the sequences are aligned, thus explaining the minor difference in lengths (data not shown). Comparison of the predicted CNL gene models from the BAC-derived and WGS-derived sequences revealed sequence polymorphisms and/or indels in two of the four intact CNL genes, likely reflecting sequencing errors (Supplemental Fig. S2). When available, BAC-derived sequence and gene models were used for subsequent analyses.

Because \( Rpg1b \) belongs to the CNL family, we reasoned that \( Rpg1r \) would as well, and thus focused our attention on CNL genes in the genetic interval containing \( Rpg1r \). As cv. Williams 82 does not express \( Rpg1r \), putative alleles for these genes were isolated from the \( Rpg1r \)-expressing soybean accession PI 96983. This was accomplished using two strategies. Firstly, one gene (P92i7_8; with the P designating a CNL from PI 96983) was obtained from a previously reported contig of PI 96983 BACs that extends into the genetically defined region that contains \( Rpg1r \) (Fig. 1 and Supplemental Fig. S1) (Innes et al., 2008). Secondly, an additional three intact genes (P21f22_7, P21f22_27, P21f22_29), and one pseudogene (P21f22_26), were amplified from PI 96983 genomic DNA using primers that flank CNLs in the target region of the Williams 82 sequence (Fig. 1A and Supplemental Fig. S1). Alignment of the amino acid sequences of the three intact CNLs amplified from PI 96983 with the corresponding Williams 82 CNLs used for primer design revealed that while P21f22_7 and P21f22_27 are very similar to their presumed Williams 82 alleles, the P21f22_29 / W21f22_29 pair are more divergent, consistent with differing recognition specificities in the two soybean genotypes (Supplemental Fig. S3). Interestingly, one of the PI 96983 CNLs (P92i7_8) from the \( Rpg1r \) target region has been previously reported as a candidate to encode the \( Rsv1 \) gene, a resistance gene effective
against certain strains of soybean mosaic virus (P92i7_8 is referred to as 3gG2 in previous studies (Hayes et al., 2004; Wen et al., 2013)).

**Functional Complementation Confirms that Rpg1r Is Encoded by a CNL Type R Protein**

Given that the pathogen effectors detected by Rpg1b and Rpg1r are recognized by a single R protein in Arabidopsis, we hypothesized that these two soybean R genes may guard the same (or related) plant protein(s) and that this might be reflected in a close evolutionary relationship with one gene derived from the other. We therefore assessed the phylogenetic relationship of the Rpg1r candidate CNL genes to the previously cloned Rpg1b using a phylogenetic tree based on the nucleotide binding domain (hereafter referred to as the NB-ARC domain for NB domain found in Apaf1, R genes and CED4 (van der Biezen and Jones, 1998)) of all CNL genes in the one megabase Rpg1b region found in both Williams 82 (Rpg1b genotype) and PI 96983 (Rpg1r genotype) (Supplemental Fig. S4). Only a single cv. Williams 82 CNL gene (W21f22_29; with the W designating a CNL from cv. Williams 82) located within the genetically defined region containing Rpg1r was found to be in the strongly supported clade containing Rpg1b. Based on this observation, further analysis was focused on the PI 96983 CNL (P21f22_29) that was amplified using primers that flank this gene.

To test for Rpg1r function, we transiently co-expressed this CNL gene (P21f22_29) with AvrRpm1 in leaves of Phaseolus vulgaris (common bean) using an Agrobacterium tumefaciens delivery system. Encouragingly, co-expression of P21f22_29 with AvrRpm1 resulted in a weak hypersensitive response (HR). To unambiguously confirm the identity of Rpg1r, we then transformed a soybean cultivar (cv. Jack) that lacks Rpg1r function with a genomic clone of P21f22_29 that included the presumed promoter and transcriptional terminator regions and tested for gain in the ability to respond to AvrRpm1. Seeds from three independent transformants were grown and screened for the ability to respond with a hypersensitive response (HR) after inoculation with P. syringae expressing AvrRpm1. Progeny from two of the independent transformants segregated for the ability to mount an HR in response to AvrRpm1 (Fig. 1B). Segregation ratios obtained using PCR with T-DNA specific primers were consistent with both these transformed lines containing multiple insertions. All T2 progeny found to lack an insertion
failed to respond to AvrRpm1, consistent with recognition of the effector being dependent on the presence of the T-DNA, and indicating that CNL P21f22_29 represents a functional Rpg1r allele.

**Structure of the Rpg1r Gene**

The Rpg1r allele from PI 96983 (P21f22_29) encodes a CNL protein 1,212 amino acids in length (Fig. 2). The COILs software predicts two regions within the Coiled-Coil (CC) domain that may participate in coiled-coil structures (predictions supported with probabilities of 88% and 94%, respectively; Fig. 2A). However, additional residues between the first two heptad repeats, and the presence of hydrophilic residues at expected hydrophobic positions, argue against the second region participating in a CC and it is not marked in Figure 2. Also present is the "EDVID" motif (present as "EDLLD" in Rpg1r), a motif previously implicated in intradomain interactions (Rairdan et al., 2008) (Fig. 2A).

The NB-ARC domain contains the expected P-loop (kinase 1a), kinase 2, RNBS-D, GLPL and MHD motifs described previously for proteins of this class (Fig 2B) (McHale et al., 2006). Interestingly, the kinase 2 motif has the sequence "LLVLDN", with an asparagine (printed in bold) in place of the second acidic residue present in most CNL proteins. It is thought that this acidic residue may be important for the catalytic activity (NTP hydrolysis) of the domain (Tameling et al., 2006).

Analysis of the LRR domain primary sequence, together with the predicted secondary structure for the region, indicates that the domain contains 24 canonical repeat units at least partially matching the consensus (LxxLxxLxLxxN/C/Tx(x) LxxIPxx) for intracellular LRR domains (Fig. 2C,(Jones and Jones, 1997)). The repeats show considerable divergence from the consensus outside the core LxxLxLxxN/C/T region associated with the repeating β-strand and adjacent loops. Strikingly, a cysteine is present in 13 of the 24 repeats at the N/C/T position in the consensus.

Rpg1r Belongs to the Same Subclade of Arabidopsis CNLs as Rpg1b, but Is Not Orthologous to RPM1
To assess the evolutionary relationship of Rpg1r to the Arabidopsis RPM1 gene, we constructed a phylogenetic tree that included the NB-ARC domains from the entire set of intact Arabidopsis CNLs, together with the soybean Rpg1b and Rpg1r genes. The NB-ARC domain was used as its conserved structure facilitates alignment of distantly related CNLs, it has historically been used to define NB-LRR clades, and because we have previously observed that this domain is less prone to inter-paralog recombination events that can confound phylogenetic analyses (Ashfield et al., 2012). Significantly, as we previously observed for Rpg1b, Rpg1r and RPM1 belong to distinct clades, consistent with the AvrRpm1 recognition specificity having evolved independently in soybean and Arabidopsis (Fig. 3). The two Rpg1 genes are closely related and are found in a well-supported clade containing two Arabidopsis CNLs of unknown function, AT3G14460 and AT3G14470. It should be noted that while Rpg1r and Rpg1b, together with one additional sequence described below, are the only soybean genes included in the tree, the Rpg1 genes are members of a larger group of closely related CNLs in the soybean genome. The majority of these closely-related paralogs are located in the 1 Mb interval centered on the Rpg1b gene (shown in Figure 1) and in segmental duplications of this region (Ashfield et al., 2012).

Also included in the tree was the top BLAST hit, Glyma06g46800, from a search of the whole soybean genome proteome using the Arabidopsis RPM1 NB-ARC domain sequence as the query. RPM1 forms a well-supported clade with Glyma06g46800 indicating that, at least with respect to the NB-ARC domain, RPM1 is more closely related to this soybean gene than to any other Arabidopsis CNL. As with the Rpg1 genes, Glyma06g46800 has a number of close relatives in the soybean genome (Zhang et al., 2011), revealing that this sub-family of CNLs has either expanded in soybean, or contracted in the Arabidopsis lineage. No known AvrB or AvrRpm1-specific R genes map close to Glyma06g46800 in the soybean genome consistent with it having a recognition specificity different from RPM1. However, it is interesting to note that Glyma06g46800 and the Rpg1b/Rpg1r genes are located in homeologous segments of the soybean genome (i.e. regions duplicated by polyploidy), consistent with their lineages have been present in the same ancestral cluster (Chen et al., 2010).

**Rpg1r and Rpg1b Share Closely Related NB-ARC Domains as a Consequence of Sequence Exchange During their Recent Evolution**
The cloning of Rpg1r allowed a comparison of its amino acid sequence to those from both Rpg1b, and the Arabidopsis RPM1 protein. An alignment of the sequence of Rpg1r with the previously cloned Rpg1b revealed striking differences in the percent nucleotide identity observed in different domains (Fig. 4A and Supplemental Fig. S5). Rpg1b and Rpg1r are most similar over a stretch that includes the C-terminal portion of the CC and the complete NB-ARC domain, with the N-terminal half of the CC domain and the LRR domain being less similar (Fig. 4A). Such a pattern could be explained by different strengths of selection acting on different domains, or by sequence exchange between the ancestral lineages. To distinguish between these possibilities, we used the Recombination Detection Program (RDP) to analyze an alignment of Rpg1b, Rpg1r, and a single allele from each of the other CNL loci distributed across the 1 Mb interval in cv. Williams 82 and accession PI 96983 that contains the Rpg1 loci. For the purposes of this analysis, we define “recombination” as any exchange of DNA sequence between paralogs or alleles, including gene conversion, and thus do not attempt to address whether flanking markers have undergone reciprocal exchanges. To reduce the chance of artifacts resulting from poor alignment, the C-terminal part of the LRR domain was excluded from the analysis as not all the sequences could be aligned accurately in this region.

The RDP analysis indicated that the Rpg1b NB-ARC domain does indeed have a recombinant origin. The initial RDP screen identified three possible recombination events within the analyzed region of Rpg1b, with the most strongly supported event encompassing the C-terminal portion of the CC domain and the entire NB-ARC (Region 2 in Fig. 4A and B). This event was supported by all five recombination detection programs deployed in the RDP screen (P<0.001 threshold). For example, using the "Bootscan" method, which assesses changes in the phylogenetic relationship between the paralogs using a sliding-window approach, Rpg1b groups with CNL P92i7_8 over the N-terminal portion of the CC domain (Fig. 4B, Region 1), but with Rpg1r (P21f22_29) over the remainder of the CC and the entire NB-ARC domain (Fig. 4B, Region 2). Across the LRR domain, Rpg1b groups with P92i7_8 for only a short segment, and not at all with Rpg1r (Fig. 4B, Region 3), suggesting that additional recombination events have occurred in this region. Interestingly, RDP identified a second CNL in the cluster, W173d12_25, as potentially sharing the same recombination event as Rpg1b, implying duplication and
translocation of the ancestral \textit{Rpg1b} locus subsequent to the recombination event discussed above (Fig. 4C).

As a consequence of the event described above, the recombinant and non-recombinant segments of \textit{Rpg1b} cluster with different subsets of CNL paralogs in neighbor-joining trees (Supplemental Fig. S6). It should be noted that, at least in the trees corresponding to Regions 1 and 2, the two potential parental sequences identified by RDP (\textit{Rpg1r} and P92i7_8) belong to distinct clades that each contain several closely related paralogs (colored green and blue, respectively, in Figs. 4C and S6). It is possible that these related CNLs, or others not present in the soybean genotypes sampled, could be the actual parental sequences.

To determine if different strengths of selection acting across the \textit{Rpg1} genes might also be contributing to the observed pattern of sequence similarity in the \textit{Rpg1b}/\textit{Rpg1r} alignment (Fig. 4A), dN/dS ratios were calculated from the alignment for the three regions described in Figure 4 (Supplemental Table S1). In fact, the recombinant region in \textit{Rpg1b} (Region 2 in Fig. 4) was under almost neutral selection (dN/dS=1.0853) while both the flanking regions (Regions 1 and 3 in Fig. 4) had dN/dS values less than 1 (0.4871 and 0.8494, respectively), consistent with purifying selection. It is therefore clear that the island of high sequence similarity observed in the alignment of \textit{Rpg1b} and \textit{Rpg1r} is not the result of strong purifying selection acting on this region, but is instead the consequence of recombination. In summary, during the evolution of the \textit{Rpg1b} gene, recombination between CNLs in the cluster has combined an \textit{Rpg1r}-like NB-ARC with the N-terminal section of a CC domain closely related to the other CNLs immediately adjacent to the present day \textit{Rpg1b} (Fig. 4C).

\textbf{Dating the Recombination Event in \textit{Rpg1b} Suggests A Recent Origin for this \textit{R} Gene}

To estimate the date of the recombination event spanning the NB-ARC region of \textit{Rpg1b}, we used the dS value for this section of the \textit{Rpg1b}/\textit{Rpg1r} alignment (Region 2 in Figure 4) (Supplemental Table S1) and a mutation rate estimate of \(5.2 \times 10^{-9}\) Ks per year (Pfeil et al., 2005). This calculation gives an estimated age for the recombination event of 1.46 million years ago (mya). Although it is a formal possibility that the capacity to recognize AvrB predates this recombination event, the rapid rate of duplication and deletion of NB-LRR genes in this region (Ashfield et al., 2012), supports a recent origin of AvrB specificity. Consistent with this view,
we have not observed AvrB recognition capacity in *Phaseolus vulgaris* (common bean) (Chen et al., 2010), which diverged from the *Glycine* lineage approximately 19 mya (Lavin et al., 2005).

**Comparison of the RPM1 Protein with Rpg1b and Rpg1r Reveals Only Limited Sequence Conservation Between the Arabidopsis and Soybean CNLs Specific for AvrB And AvrRpm1**

Although the ability to detect AvrB and AvrRpm1 appears to have evolved independently in the Arabidopsis and soybean lineages, it is possible that R proteins specific for these effectors will have conserved features. To investigate this possibility, Rpg1r was aligned with both the Rpg1b and RPM1 amino acid sequences. This analysis revealed only limited sequence conservation between the soybean R proteins and RPM1 outside the "EDVID" motif in the CC domain and conserved motifs of the NB-ARC domain (Supplemental Fig. S7). While the alignment across the LRR domain is highly fragmented, reflecting the presence of 24 canonical repeats in both the soybean Rpg1 proteins versus only 15 in RPM1 (Grant et al., 1995; Ashfield et al., 2004), the first four repeats align without insertions or deletions raising the possibility of a conserved function for this region. Based on functional analyses of the Arabidopsis RPS5 CNL, this is likely associated with keeping the R proteins in an ‘off’ state in the absence of elicitation (Qi et al., 2012).

**Plotting Polymorphic Positions On Predicted Tertiary Structures for the Rpg1b CC and LRR Domains Reveals Highly Polymorphic Surfaces**

As described above, a comparison of the *Rpg1r* and *Rpg1b* sequences revealed that both the CC and LRR domains are highly polymorphic, suggesting a role for one, or both, of these domains in conferring the ability to distinguish between AvrB and AvrRpm1. We hypothesized that polymorphic residues involved in recognition specificity would be located on solvent-exposed surfaces of Rpg1b. We therefore mapped the polymorphic positions in the Rpg1r/Rpg1b amino acid alignment onto models of the tertiary structures of the Rpg1b CC and LRR domains.

We have previously described modeling of the Rpg1b CC domain tertiary structure using the known MLA10 CC structure as the template (Ashfield et al., 2012). The MLA10 domain is
known to form a homo-dimer with a conserved sequence motif ("EDVID"), previously shown to interact with the NB-ARC domain, being exposed on one surface of the dimer (Maekawa et al., 2011) and we propose that Rpg1b will form a similar structure. A model for the Rpg1b CC homo-dimer, generated using the Phyre2 modeling service, is shown in Figs. 5A to D. Based on the MLA structure we predict that the α1 helices of the two Rpg1b monomers will form a coiled-coil structure. This hypothesis is supported by analysis of the Rpg1b sequence using the COILs program, which predicts that the region from Q38 to K61 in Rpg1b will participate in such a structure (P > 98%). Hydrophobic residues predicted to reside at the key "a" and "d" positions in the CC heptad repeats are shown in green in the modeled structure (Fig. 5A to D). These include residues L41, L44, and I51 which correspond, in the alignment of the Rpg1b and MLA10 sequences, to three residues in MLA10 (I32, L36, and M43, respectively) shown previously to be required for dimerization in yeast-two-hybrid assays (Maekawa et al., 2011). Intriguingly, when residue positions that differ between the Rpg1b and Rpg1r proteins were mapped onto the Rpg1b homo-dimer model, an uneven distribution was observed. While the surface that contains the conserved "EDVID" motifs (EDILD in Rpg1b) contains relatively few non-conservative substitutions (Fig. 5A and B, Movies 1 and 2; residues of the EDVID motif printed in orange, polymorphic residues in blue), a highly polymorphic surface was observed on the opposite side of the dimer (Fig. 5C and D). This highly polymorphic surface is generated by a region of the Rpg1b model that corresponds to the α1 helix in the MLA10 structure, a region previously shown to interact with WRKY domain–containing transcription factors (Shen et al., 2007).

Next we generated a model for the Rpg1b LRR domain tertiary structure using the Modeler software and the known Toll-Like Receptor 3 (TLR3) structure as the template (Fig. 5E and F). The TLR3 LRR ectodomain forms a horseshoe structure characteristic of other LRR proteins for which crystal structures are known, with a series of β strands forming a solvent exposed β sheet across the concave surface. Regular β strands are also predicted in the Rpg1b LRR sequence using the PSIPHED secondary structure prediction server and, as expected, the majority of these are positioned across the concave surface of the model (Fig. 5E). As with the CC domain, mapping of residues that differ between the two Rpg1 proteins onto the Rpg1b LRR model revealed an uneven distribution. Clustering of polymorphisms was particularly striking across the concave surface of the structure, a region known to interact with ligands for at least some LRR proteins (Fig. 5E and F, Movies 3 and 4). Interestingly, most of the polymorphic
Origin of R Gene Specificity in Soybean

positions are located within the C-terminal portion of the concave surface where they form a highly polymorphic pocket with respect to Rpg1r.

AvrB Versus AvrRpm1 Specificity is Determined By the C-terminal Portions of the LRR Domains of Rpg1b and Rpg1r

To test the predictions made by our structural modeling, we developed a transient expression system that enabled us to reconstitute the Rpg1b and Rpg1r signaling pathways in Nicotiana glutinosa (R. Kessens, T. Ashfield, S.H. Kim and R.W. Innes, submitted). We employed N. glutinosa rather than the more commonly used N. benthamiana because transient expression of AvrB and AvrRpm1 in the latter induces rapid cell death, indicating that N. benthamiana contains R genes that recognize these effectors. To establish the efficacy of N. glutinosa, we co-expressed Rpg1r-sYFP and AvrRpm1 using Agrobacterium-mediated transient expression under control of a dexamethasone-inducible promoter. Co-expression of the two genes together induced visible tissue collapse, which was not observed when either one was expressed alone (Fig. 6A). Importantly, co-expression of Rpg1r-sYFP with AvrB did not induce tissue collapse, demonstrating that Rpg1r specifically recognizes AvrRpm1. Repeating the experiment using untagged Rpg1r revealed significant R gene autoactivity, independent of the presence of AvrRpm1, indicating that the YFP tag attenuates the signaling activity of Rpg1r, which in the context of Rpg1r overexpression, was useful. Analogous experiments with Rpg1b, however, revealed that sYFP nearly eliminated Rpg1b signaling ability (data not shown).

To circumvent the dual problems of attenuation of Rpg1b signaling by sYFP, and Rpg1r autoactivity when highly overexpressed, we switched to a vector in which the Rpg1 genes were expressed without an epitope tag, and in which expression was driven by a cauliflower mosaic virus 35S promoter to avoid the rapid accumulation resulting from an inducible promoter. Using this system, co-expression of Rpg1b and AvrB induced a strong tissue collapse, but only when we also co-expressed any one of the four soybean RIN4 genes (Fig. 6B and Kessens et al., submitted). No such requirement for a GmRIN4 was noted for Rpg1r in this transient expression system (Kessens et al., submitted). Untagged Rpg1b did not induce tissue collapse when co-expressed with AvrRpm1 and GmRIN4b (Fig. 6B), thus the specificity of Rpg1r and Rpg1b are maintained in this heterologous system. Untagged Rpg1r induced some tissue collapse when
expressed by itself, or with AvrB and GmRIN4b (Fig. 6B), but this collapse was consistently weaker than that observed when co-expressed with AvrRpm1. We thus adopted a scoring system in which the responses to AvrB and AvrRpm1 were always compared within a single leaf, and assessed for which induced the stronger collapse. Figure 6B summarizes data from three independent experiments in which 32-41 leaves in total were injected for each construct. For Rpg1r, 22 of 36 leaves responded more strongly to AvrRpm1, and none responded more strongly to AvrB (14 leaves gave a similar response to both effectors, with levels of collapse similar to that observed for Rpg1r alone).

Having established the efficacy of the N. glutinosa transient system, we tested the predictions of our structural modeling by assessing the effector specificity of two sets of Rpg1b-Rpg1r chimeras. In the first set of chimeras, we swapped the N-terminal regions containing the CC domain, with the junction located just N-terminal to the P-loop. We refer to these chimeras as B-R-R (indicating CC domain from Rpg1b and NBS-LRR domains from Rpg1r) and R-B-B (CC domain from Rpg1r and NBS-LRR domains from Rpg1b). In the second set we swapped the C-terminal portion of the LRR domain starting at LRR repeat number 8 (R-R-B and B-B-R). We made the junction at this position because the majority of the polymorphisms between Rpg1b and Rpg1r were localized distal to this point. As predicted by the structural modeling, swapping the LRR domains (starting at repeat 8) was sufficient to swap recognition specificity, with R-R-B recognizing AvrB and B-B-R recognizing AvrRpm1, albeit weakly (Fig. 6B). These findings indicate that the CC domain is not required for differential recognition of AvrB and AvrRpm1. Consistent with this conclusion, swapping just the CC domains did not alter specificity (Fig. 6B).

In summary, we conclude that the ability to distinguish between AvrB and AvrRpm1 is controlled by polymorphisms located within the C-terminal 17 LRR repeats of Rpg1b and Rpg1r.

Soybean Contains Four RIN4 Homologs and All Can Be Cleaved by AvrRpt2

*RPM1*-dependent recognition of AvrB and AvrRpm1 in Arabidopsis requires effector-dependent phosphorylation of a second plant protein, RIN4 (Mackey et al., 2002; Chung et al., 2011; Liu et al., 2011; Baltrus et al., 2012). RIN4 is also targeted and cleaved at two sites (RCS1 and RCS2) by a third *P. syringae* effector, AvrRpt2, which is a cysteine protease (Axtell et al., 2003; Mackey et al., 2003). One consequence of this activity is that AvrRpt2 blocks recognition
of AvrB and AvrRpm1 in Arabidopsis and has the potential to be a useful tool for assessing the role of RIN4 homologs, or additional targets of AvrRpt2, in effector recognition in other species.

Soybean expresses four RIN4 homologs (GmRIN4s) and all four contain at least the C-terminal RCS2 with no more than a single conservative substitution (GmRIN4a-d; (Chen et al., 2010) and Supplemental Fig. S8). To test whether AvrRpt2 can cleave one or more of the GmRIN4s, each was transiently co-expressed in *Nicotiana benthamiana* with either AvrRpt2, or a proteolytically inactive variant of the effector (C122A) (Axtell et al., 2003). To facilitate detection, each GmRIN4 was fused to an N-terminal 5x-Myc epitope tag and overexpressed using a dexamethasone-inducible expression vector. When expressed alone, or co-expressed with C122A, all four GmRIN4s accumulated strongly (Fig. 7A). However, co-expression with a functional AvrRpt2 transgene resulted in the complete, or near complete, disappearance of all four GmRIN4s (Fig. 7A). In some experiments, a cleavage product was detectable for GmRIN4c and/or d (Fig. 7A). Examination of the alignment of the GmRIN4s (Supplemental Fig. S8) reveals substitutions in the RCS1 cleavage site for both GmRIN4c and d that could inhibit cleavage, and thus enable accumulation of a product cleaved at only RCS2.

**AvrRpm1 Recognition in Soybean is Partially Blocked by AvrRpt2**

We have previously shown that *Rpg1b*-dependent detection of AvrB in soybean is inhibited by the presence of AvrRpt2 and have hypothesized that this was caused by cleavage of soybean RIN4 homologs (Ashfield et al., 2004). The cleavage results described above support this hypothesis. If *Rpg1r* function is also dependent on RIN4 homologs, AvrRpm1 detection in soybean should also be inhibited by AvrRpt2.

To test this prediction, mixtures of *P. syringae* strains were injected into leaves of an *Rpg1r*-expressing soybean cultivar, Flambeau. Co-injection of a strain expressing *avrRpm1* with a strain expressing a disrupted *avrRpt2* gene induced a typical HR by 24-48 hrs post-injection (hpi) (Fig. 7B). In contrast, co-injection of a strain expressing *avrRpm1* with a strain expressing a functional *avrRpt2* gene induced a weaker HR that was slower to develop, indicating that AvrRpt2 partially suppresses AvrRpm1 recognition (Fig. 7B). Co-injection of strains expressing functional AvrRpt2 and disrupted AvrRpt2 genes resulted in no tissue collapse, demonstrating that AvrRpt2 does not elicit a visible HR in cv. Flambeau.
The above observations indicate that Rpg1r function is at least partially dependent on a target(s) of the AvrRpt2 protease, supporting the hypothesis that recognition of AvrRpm1 in soybean requires at least one soybean RIN4 homolog. The lack of such a requirement in the G. glutinosa transient system is likely explained by the presence of endogenous RIN4 homologs in N. glutinosa (Kessens et al., submitted).

DISCUSSION

While the Arabidopsis RPM1 R gene can mediate detection of both the AvrB and AvrRpm1 pathogen effector proteins (Bisgrove et al., 1994), in soybean, two tightly-linked R genes can distinguish between them (Ashfield et al., 1995). The soybean Rpg1b gene can only mediate detection of AvrB, while the Rpg1r gene can only mediate detection of AvrRpm1. This was an intriguing observation because it demonstrated that, at least in soybean, the two effectors likely have distinct activities or targets and yet are detected by a single R protein in Arabidopsis. Cloning RPM1, and the two soybean R genes, has been an important goal because it has the potential to provide insights into the evolution of specificity, persistence over time, and mode of activity, of R genes that detect pathogen effectors indirectly. RPM1 and Rpg1b have been cloned previously (Grant et al., 1995; Ashfield et al., 2004). Here we describe the map-based cloning of the soybean Rpg1r gene, allowing a detailed comparison of the three functionally related R genes.

Like the previously cloned Rpg1b gene, Rpg1r belongs to the CNL family of R genes. Overall, the two genes display high overall sequence similarity, sharing 86% amino acid identity. However, as discussed below, the level of sequence identity varies dramatically between domains. Interestingly, while the Rpg1r protein contains the domains and motifs typical of CNL R proteins, its NB-ARC domain contains an unusual substitution. Specifically, the second aspartate residue of the kinase 2 (Walker B) domain (consensus hhhhDD for R proteins, where h is usually a hydrophobic residue (Takken et al., 2006)) has been replaced with an asparagine. An acidic residue is highly conserved, but not invariant, in this position in both plant R proteins and other members of the STAND class of the P-loop NTPase family (Leipe et al., 2004). This residue is proposed to play a key role in the NTPase activity of the domain in at least some classes of P-loop NTPases, priming water for nucleophilic attack of the ATP (Muneyuki et al.,
At least some R proteins have known ATPase activity and it has been proposed that this activity is required for returning the protein to an inactive state after elicitation (Tameling et al., 2002; Takken and Tameling, 2009). Consistent with this hypothesis, a D283E substitution in the tomato I-2 protein affected its ATPase activity, but not ATP binding, and leads to autoactivity (Tameling et al., 2006). The observation that Rpg1r can function with an asparagine in this location suggests either that ATPase activity is not required for its successful operation or, in this case, the domain retains its enzymatic activity despite the absence of the 2nd conserved aspartate.

Given that both AvrB and AvrRpm1 are detected by the RPM1 protein in Arabidopsis (Bisgrove et al., 1994; Grant et al., 1995), it was possible that one or both of the soybean Rpg1 genes would be an orthologue of RPM1 (i.e. an ancestral R gene with the ability of respond to AvrB and/or AvrRpm1 that has survived since before the divergence of the soybean and Arabidopsis lineages). We have previously reported that Rpg1b and RPM1 are not orthologues (Ashfield et al., 2004), and here we confirm phylogenetically that Rpg1r is not orthologous with RPM1 either. We therefore conclude that the ability to recognize both AvrB and AvrRpm1 evolved independently in Arabidopsis and soybean, consistent with the general observation that NB-LRR genes experience a high rate of turnover through species-specific duplications and deletions (Cannon et al., 2002; McHale et al., 2006).

Curiously, the soybean CNLs with NB-ARC domains phylogenetically most closely related to RPM1 are found in a region homeologous to the Rpg1 cluster; that is, the two regions were generated by duplication of an ancestral cluster as a consequence of polyploidization (Chen et al., 2010). This suggests that in the common ancestor of soybean and Arabidopsis the ancestral Rpg1 region contained CNL lineages closely related to both RPM1 and the Rpg1 genes, and these lineages were subsequently parsed between the two duplicated regions in soybean. This raises the possibility that while the functional AvrB and AvrRpm1 specific R genes evolved independently in Arabidopsis and soybean, some property of the ancestral cluster predisposed the subsequent evolution of R genes able to detect these pathogen effectors. As has been discussed previously, such a property might be the ability of at least some of the CNLs present in the cluster to monitor RIN4-like proteins for pathogen-induced modifications (Chen et al., 2010).

The hypothesis that the CNL cluster in which the Rpg1 genes reside is predisposed to evolve AvrB/AvrRpm1-specific R genes is further supported by the observation that an R gene
(Rpsar-1) with the ability to detect AvrRpm1 is also found in P. vulgaris (common bean) in a region orthologous to the Rpg1b/Rpg1r R gene cluster in soybean (Chen et al., 2010). While it is possible that the Rpg1r gene has survived for at least the ~19 million years since the divergence of the Glycine and Phaseolus lineages, the available evidence from mapping CNL sub-families relative to the AvrRpm1-specific R gene in Phaseolus argues against this (Chen et al., 2010). The alternative explanation is that the AvrRpm1-specific genes evolved independently in soybean and bean. Only the cloning of the AvrRpm1-specific gene in Phaseolus will allow the relationship between the functionally analogous genes to be unambiguously resolved. In contrast to the situation observed for AvrRpm1, no P. vulgaris cultivars with the ability to respond to AvrB have been identified (Chen et al., 2010). Interestingly, the ability to respond to AvrB and/or AvrRpm1 has been observed in Nicotiana benthamiana (Chung et al., 2011), Lactuca sativa (lettuce), and Capsicum annuum (pepper) (Wroblewski et al., 2009) suggesting that these recognition specificities have evolved independently in multiple plant lineages. Whether the AvrB/AvrRpm1 specific R genes found in the N. benthamiana and L. sativa reside in Rpg1-orthologous regions is not known.

It should be noted that while functional R genes typically do not persist long enough to be found in different plant families, the available evidence indicates that the downstream signaling components are far more conserved. For example, the MLA1 R gene from the monocot barley, has been shown to function in the dicot Arabidopsis, demonstrating that the downstream components of the MLA1 triggered pathway have been conserved since before the monocot / dicot split (Maekawa et al., 2012). This observation has practical implications because it raises the possibility of cloned R genes effective against economically important pathogens being transferred transgenically to distantly related crop species.

Many of the sites that differ between Rpg1b and Rpg1r were found to localize to the C-terminal portion of the LRR domain, in a region predicted to form a concave surface of a horseshoe-like structure (Fig. 5). This is consistent with previous observations that this surface binds ligands for at least some LRR proteins (Kobe and Deisenhofer, 1995). We thus predict that this surface on Rpg1b and Rpg1r mediates detection of AvrB- and AvrRpm1-mediated modifications of soybean RIN4 homolog(s), respectively. In support of this model, transient expression of chimeric Rpg1b-Rpg1r proteins confirmed that AvrB versus AvrRpm1 specificity is controlled by the C-terminal 17 LRR repeats (Fig. 6). The C-terminal half of the LRR domain
has recently been implicated in determining the specificity of the CNL N’ and L proteins of *N. sylvestris* and *Capsicum spp.*, which mediate detection of tobamovirus coat proteins (Sekine et al., 2012). Similarly, structural modeling of the RPP1 protein of Arabidopsis, a TNL that likely directly binds its corresponding effector ATR1, identified the C-terminal LRR region as containing hotspots of polymorphism that correlate with effector specificity (Steinbrenner et al., 2012). Like Rpg1b and Rpg1r, these polymorphisms localize to the concave surface of the LRR horseshoe.

In summary, we conclude that the AvrB and AvrRpm1-specific *R* genes in Arabidopsis and soybean evolved independently, but probably from CNL lineages originally present in the same ancestral *R* gene cluster. Despite both effectors being detected by a single *R* gene (*RPM1*) in Arabidopsis, the AvrB and AvrRpm1-specific genes in soybean contain numerous polymorphisms in both the CC and LRR domains. The work presented here indicates that the LRR domain is the primary specificity determinant, which raises the question of whether the polymorphisms in the CC domain are functionally important. We are currently investigating whether the CC domain polymorphisms impact the affinity of Rpg1b and Rpg1r for the four soybean RIN4 homologs. This information should provide insights into the apparent propensity of CNL paralogs present in the *Rpg1*, and homologous, clusters to develop into AvrB and AvrRpm1-specific *R* genes.

**MATERIALS AND METHODS**

**Plant Material and Growth Conditions**

Seed for soybean (*Glycine max* (L.) Merr.) accession PI 96983 (*Rpg1r rpg1b*) was obtained from M. A. Saghai-Maroof (Department of Crop and Soil Environmental Sciences, Virginia Tech). Seed for cultivar Williams 82 (*rgl1r Rpg1b*) was ordered from the USDA Soybean Germplasm Collection via the National Plant Germplasm System web portal (www.ars-grin.gov/npgs). The soybean cv. Flambeau (*Rpg1r rpg1b*) x cv. Merit (*rgl1r Rpg1b*) RIL population has been described previously (Ashfield et al., 1995). Soybean plants were grown in clay pots containing a compost:Metro-Mix 360 (Sun Gro Horticulture):vermiculite:perlite (4:2:1:1) mix supplemented with osmocote slow-release fertilizer. Pathogenicity tests were
conducted 2 - 3 weeks after planting with the growth chamber set for 23°C and a 16 hr photoperiod.

*Nicotiana glutinosa* accession 241768 was obtained from the USDA National Plant Germplasm System *Nicotiana* Collection at North Carolina State University in Raleigh, NC. Both it and *N. benthamiana* were grown in a 3:1 mixture of Metro-Mix 360 potting soil (Sun Gro Horticulture) and expanded perlite in a growth room at 22 °C. *N. glutinosa* plants were germinated under long-day conditions (16h light/8h dark) for 12 days to facilitate seedling establishment then transferred to short-day conditions (9h light/15h dark cycle; 150 µEm⁻²m⁻²) for 2-3 more weeks before transformation. The initial growth period under long-day conditions was omitted for *N. benthamiana*.

**Genetic and Physical Mapping**

The PCR-based genetic markers used for the genetic fine-mapping of *Rpg1r* in the Flambeau x Merit RIL population, as shown in Supplemental Figure S1, are described in Supplemental Table S3. The associated primers are listed in Supplemental Table S2. The PCR was conducted as described by Ashfield *et al.* 1998. The scoring of the RIL population for response to *P. syringae* expressing *AvrRpm1*, and the identification of lines containing informative recombination events in the region containing the *Rpg1* genes, has been described previously (Ashfield *et al.*, 1995; Ashfield *et al.*, 2003).

The generation of the BAC-based sequence centered on the *Rpg1b* locus for both the Williams 82 and PI 96983 soybean genotypes, and the identification of CNL genes within these sequences, was described by Innes *et al.* (2008). The whole genome sequence (WGS) (Glyma1 assembly) for soybean cultivar Williams 82 (Schmutz *et al.*, 2010), and the associated gene set (v1.1 gene set), was accessed from the Phytozome web-portal (www.phytozome.net).

**Amplifying *Rpg1r* Candidate CNLs from PI 96983 Genomic DNA**

The primers and PCR annealing temperatures used to amplify CNLs from soybean accession PI 96983 (*Rpg1r*) genomic DNA, using locus-specific primers corresponding to non-
coding sequences that flank CNLs in cultivar Williams 82 (rpg1r), are given in Supplemental Table S4. Primer sequences are listed in Supplemental Table S2.

PI 96983 genomic DNA was isolated as described by Ashfield et al. (2003) or using the QIAGEN DNaseasy plant mini kit; in both cases small (< 1 cm in length), only partially expanded, trifoliate leaves were used as the source tissue. PCR was accomplished using the TripleMaster polymerase (Eppendorf) according to the manufacture's instructions. PCR products were isolated from agarose gel slices using the QIAquick gel extraction kit (QIAGEN). The purified DNA fragments were subsequently sequenced using the ABI PRISM BigDye v3.0 Terminator cycle sequencing kit (PE Applied Biosystems) according to the manufactures instructions except that reaction volumes were reduced to 10 µl consisting of 1 µl BigDye terminator ready reaction mix, 3 µl MgCl₂ (5 mM), 2 µl primer (2 µM), 2 µl template DNA, and 2 µl water.

**Complementation of Rpg1r Function in Soybean Stable Transformants**

The ~ 5 Kb genomic region containing the Rpg1r (CNL P21f22_29) coding region and flanking sequences was PCR amplified from soybean cultivar PI 96983 (Rpg1r) genomic DNA using locus-specific primers as described above (details in Supplemental Table S4). The amplified fragment was initially cloned into the pGEM-T Easy vector (Promega), then excised as a NotI fragment, and subsequently ligated into NotI digested pGREENII 229 (www.pgreen.ac.uk) (Hellens et al., 2000) to generate the plasmid pTAI-30.9. Soybean cv. Jack (Rpg1b rpg1r) was transformed with pTAI-30.9 by the Plant Transformation Core Facility, University of Missouri.

The presence of the pTAI-30.9 T-DNA in the primary transformants and T2 progeny was monitored using PCR and the T-DNA-specific primers TA129 and TA130 (Supplemental Table S2).

**Construction of Rpg1b / Rpg1r Chimeras**

Overlap PCR was used to generate Rpg1r / Rpg1b chimeric genes in which either the CC domains, or C-terminal portions of the LRRs, were swapped. The splice sites are indicated in Supplemental Figure S5. The Rpg1r and Rpg1b CC domains were amplified using primer pairs
attB1-Rpg1r-F + CC/NB-R and attB1-Rpg1b-F + CC/NB-R, respectively, while the Rpg1r and Rpg1b NB-ARC-LRR regions were amplified using primer pairs CC/NB-F + attB2-Rpg1r-R and CC/NB-F + attB2-Rpg1b-R, respectively (Supplemental Table S2). The PCR products were gel purified, appropriate combinations mixed, and the chimeric Rpg1 coding sequences amplified using flanking primers.

The Rpg1r/Rpg1b chimeric genes in which the C-terminal portion of the LRR domains were swapped were generated in a similar manner. In this instance, the Rpg1r and Rpg1b CC-NB-ARC-N-terminal LRR regions were amplified using primer pairs attB1-Rpg1r-F + LRR-R and attB1-Rpg1b-F + LRR-R, respectively. The Rpg1r and Rpg1b C-terminal LRR regions were amplified using primer pairs LRR-F + attB2-Rpg1r-R and LRR-F + attB2-Rpg1b-R, respectively.

The amplified chimeric sequences were subsequently cloned in the pEG100 destination vector (Earley et al., 2006) using Gateway™ (Invitrogen) technology to permit in planta expression under control of the 35S promoter.

**Transient Expression Assays in N. glutinosa**

For transient expression in planta, Rpg1r-sYFP, Rpg1b-sYFP, untagged AvrB, and untagged AvrRpm1 were cloned into a Gateway® compatible derivative of pTA7002, which places the transgene under control of a dexamethasone inducible promoter (Aoyama and Chua, 1997). Untagged Rpg1b, Rpg1r, and the untagged chimeric Rpg1 constructs were cloned in the pEG100 vector (Earley et al., 2006) to allow expression by the 35S promoter. The 35S - GmRIN4B construct has been described previously (Selote, 2013).

The R gene constructs, GmRIN4b, and the AvrB/AvrRpm1 effectors were transiently expressed in N. glutinosa by Agrobacterium-mediated transformation. A. tumefaciens strain GV3101(pMP90) was used for all constructs except GmRIN4b which in which case LBA4404 was used. Agrobacterium strains were grown overnight (at least 16h) at 30°C in LB media with 50 μg/mL of kanamycin and 50 μg/mL of gentamycin for the GV3101 strains, and 100 μg/mL spectinomycin and 50 μg/mL streptomycin for LBA4404 harboring the GmRIN4b construct. After overnight culture, bacterial cells were pelleted by centrifugation, washed with 10 mM
MgCl$_2$, and resuspended in 10 mM MgCl$_2$ containing 100 µM acetosyringone and incubated at room temperature for at least 2 hrs. When assessing the wild-type Rpg1 genes with C-terminal sYFP tags, inoculum was prepared by mixing the indicated strains so that each transgene-containing strain had an individual OD$_{600}$= 0.1. When assessing the untagged Rpg1 genes or chimeras, the density of the strains harboring the R genes and GmRIN4b were reduced to an OD$_{600}$= 0.05. The inoculum was injected into 3.5 to 4.5-week old N. glutinosa plants with a needleless syringe, nicking the leaf surface with a razor blade prior to injection to facilitate fluid entry when necessary. Older plants were avoided as they transformed less reliably. Transgene expression was induced ~40 hours post injection by spraying the plants with 50 µM dexamethasone and 0.02% L77 surfactant (Momentive Performance Materials, Inc. Waterford, NY). The leaves were scored for tissue collapse ~24 - 48 hours later, and leaves removed for photography shortly after. Responses were scored as follows: 0, no visible response; 1, white speckling typically associated with a non-specific response to the Agrobacterium; 2, chlorosis or tissue discoloration; 3, partial collapse of the injected panel; 4, complete collapse of the injected panel.

HR Tests and Testing for Suppression of Rpg1r Function by AvrRpt2

The response of soybean plants to P. syringae expressing AvrRpm1 was assessed using hypersensitive response (HR) hand-inoculation tests as described by Ashfield et al. (1995). In brief, P. syringae pv. glycinea race 4 (PsgR4) (Ashfield et al., 1995) was injected into the apoplast of soybean leaves at an OD$_{600}$ of either 0.1 or 0.2 using a 1 ml disposable syringe. Rpg1r-dependent recognition of AvrRpm1 was associated with brown discoloration of the leaf veins, and typically tissue collapse, within the injected panels. Injected leaves were scored 24 - 48 hrs after injection.

Suppression of AvrRpm1 recognition by AvrRpt2 was demonstrated using mixed inoculations in which the component PsgR4 strains were co-infiltrated at equal densities with each strain at an OD$_{600}$ of 0.1. The PsgR4 (AvrRpm1), PsgR4 (AvrRpt2), and PsgR4 (AvrRpt2::Ω) strains have been described previously (Whalen et al., 1991; Ashfield et al., 1995). The non-functional AvrRpt2::Ω allele was generated by insertion of the omega fragment into the
AvrRpt2 clone pLH12 to generate plasmid pLH12Ω (Whalen et al., 1991). HR responses were scored 48 hrs after injection.

**Sequence Analyses**

Sequence alignments were generated either using pairwise BLAST (Altschul et al., 1990) or ClustalX (Larkin et al., 2007). Sequence conservation in Clustal alignments was highlighted using the BoxShade 3.21 server (http://www.ch.embnet.org/software/BOX_form.html).

Variation in conservation of nucleotide identity between \textit{Rpg1b} and \textit{Rpg1r} across their coding sequences was assessed using the VISTA server (http://genome.lbl.gov/vista/index.shtml) (Frazer et al., 2004) with the mVISTA and LAGAN (Brudno et al., 2003) settings.

The COILs server (http://www.ch.embnet.org/software/COILS_form.html) (Lupas et al., 1991) was used to screen the Rpg1b and Rpg1r CC domains for regions likely to participate in coiled-coil structures. The MTIDK matrix was used together with a 2.5-fold weighting for positions "a" and "d" in the heptad repeats.

**Phylogenetic Analyses**

The phylogenetic tree presented in Fig. 3 is based on an amino acid alignment of the NB-ARC domains (VGMGG to MHDLL in \textit{Rpg1b}) from the complete set of intact Arabidopsis CNLs together with three soybean genes. The soybean CNLs are the Williams 82 allele of \textit{Rpg1b} (GenBank code AY452685), the PI 96983 allele of \textit{Rpg1r} (P21f22_29), and the top hit in a BLASTp search (using the BLAST tool available at www.phytozome.org) of the soybean whole genome proteome (Schmutz et al., 2010) using the NB-ARC domain sequence from the Arabidopsis \textit{RPM1} gene (AT3G07040). The Arabidopsis TIR-NB-LRR (TIR) gene AT3G44400 was also included in the analysis as an outgroup. The Arabidopsis CNL sequences are available from the Arabidopsis Information Resource website (www.Arabidopsis.org). The soybean sequences were added to a previously described alignment of Arabidopsis CNL genes (Meyers et al., 2003) using Clustal (Larkin et al., 2007) and the BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) interface. The Bayesian tree was then generated using the MrBayes 3.1.2 software (Ronquist and Huelsenbeck, 2003) and the
following parameters: 2 parallel runs with 4 chains in each; model jumping between fixed-rate amino acid models; 1 million generations; sampling every 100 generations; first 25% of trees discarded before the consensus tree calculated. The runs were considered to have converged when the average standard deviation of split frequencies between the two runs had reduced to < 0.01.

The Neighbor-joining tree shown in Supplemental Figure S4 is based on an alignment of the NB-ARC domains of the soybean CNL genes from the 1 Mb interval centered on the $R_{pg1b}$ gene in cultivar Williams 82, and the corresponding region in accession PI 96983 (Ashfield et al., 2012). The tree was generated using ClustalX 2.1 (Larkin et al., 2007) and clades assessed using 1000 bootstrap repeats. Trees were visualized using TreeView X (Page, 1996).

**Detection of Recombination**

An alignment was generated using ClustalX for $R_{pg1b}$ (W52d1_8), $R_{pg1r}$ (P21f22_29), and the remaining intact CNLs from the 1 Mb interval in cv. Williams 82 shown in figure 1. Also included were two additional genes from the corresponding region in accession PI 96983 (P98f7_6 and P92i7_8) that lack alleles in the Williams 82 sequence. The alignment was then trimmed to remove the 3’ end of the LRR domain (leaving nucleotides 1 - 2664 for $R_{pg1b}$) as this region was hard to align accurately. The alignment was then screened for evidence of recombination using the Recombination Detection Program (RDP) v.4.9 (Martin et al., 2010) using the parameters described by Ashfield and associates (Ashfield et al., 2012) and the following methods: RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), Bootscan (Martin et al., 2005), Chimaera (Posada and Crandall, 2001), and MAXCHI (Smith, 1992). Only events supported by two or more methods at a Bonferroni-corrected p-value of 0.001 or better were considered. After manual assessment of the RDP output associated with the NB-ARC-encompassing recombination event detected in $R_{pg1b}$, the position of the N-terminal breakpoint was refined and the alignment then rescanned.

**Calculating Ka and Ks Values and Estimating the Age of Recombinant Segments**
Ka, Ks, and Ka/Ks values were calculated for three segments of the alignment of non-redundant CNL genes used in the RDP recombination analysis (segments shown in Figure 4A and B). Region 1, the interval between the start of the coding regions and the recombinant region detected in Rpg1b (W52d1_8); Region 2, the part of alignment corresponding to the recombinant region in Rpg1b (ending with the first of the two possible breakpoints; see Supplemental Figure S9); Region 3, the region from the end of recombinant segment in Rpg1b to the end of the trimmed alignment (starting with the second of the two possible breakpoints; see Supplemental Figure S9). Ka, Ks, and Ka/Ks values were calculated using the Yang & Nielsen (2000) method (Yang and Nielsen, 2000) as implemented by the PAML software package (Yang, 2007).

The age of the recombinant section identified in Rpg1b was estimated using a substitution rate of $5.2 \times 10^{-9}$ Ks per year (Pfeil et al., 2005), and the formula $T = K/2r$, where $T$ = time, $K$ = the Ks value for the Rpg1b (W52d1_8) / Rpg1r (P21f22_29) pairwise comparison for this region, and $r$ = substitution rate.

**Tertiary Structure Prediction and Plotting Polymorphic Positions**

The Rpg1b CC domain monomer structure was modeled using the Phyre2 server with the normal mode modeling setting (http://www.sbg.bio.ic.ac.uk/phyre2/) (Kelley and Sternberg, 2009). Phyre2 modeled 120 residues in Rpg1b (A13 to K132) using the MLA10 CC structure (PDB code 3qfl) (Maekawa et al., 2011) as the template. Before assembly of the Rpg1b homodimer, a 14 residue stretch (Q99 to T112; associated with an insertion in Rpg1b and a gap in the MLA10 structure; shown in blue in Supplemental Figure S10A) that could not be threaded onto known MLA10 structure, was deleted. The Rpg1b dimer was assembled and visualized using the UCSF Chimera package (v. 1.7) (http://www.cgl.ucsf.edu/chimera/) (Pettersen et al., 2004) and the "biological unit" information present in the MLA10 PDB structure file.

The Rpg1b LRR domain structure was modeled using the Modeller server, (Eswar et al., 2006) accessed remotely using the UCSF Chimera package (v 1.7) (Pettersen et al., 2004). Default settings were used. The known structure of the Human Toll-like Receptor 3 ectodomain (TLR3; PDB code 1ziw) (Choe et al., 2005) was identified as a suitable template using the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/) (Kelley and Sternberg, 2009) and the structure file obtained from the RCSB Protein Data Bank (www.pdb.org) (Berman et al., 2000).
The Rpg1b LRR and TLR3 amino acid sequences were first aligned using ClustalX (Larkin et al., 2007) as implemented by the BioEdit sequence alignment editor (www.mbio.ncsu.edu/BioEdit/bioedit.html). The alignment was then refined to improve the alignment of the β strands (secondary structure known in TLR3, predicted in Rpg1b), and associated xxLxLxx sequence motifs (where L can be substituted by another hydrophobic residue), characteristic of the concave surfaces typical of LRR domains (see Supplemental Fig. S11 for alignment). TLR3 secondary structure information was sourced from the 1ziw PDB file. Rpg1b secondary structure was predicted using the PSIPRED protein structure prediction server (www.bioinf.cs.ucl.ac.uk/psipred/) (Jones, 1999; Bryson et al., 2005). Secondary structure visualization and alignment editing was accomplished using the SBAL program (www.structuralchemistry.org/pcsb/) (Wang et al., 2012). After modeling, two insertions present in Rpg1b (886Q-898C and 1059P-1075I; shown in blue in Supplemental Figure S10C), likely representing segments of additional repeat units absent in the TLR3 template structure, were deleted from the model to generate the structure shown in Supplemental Figure S10D.

Sequence conservation between the Rpg1 proteins over the modeled regions was determined by aligning the Rpg1b and Rpg1r amino acid sequences using ClustalX (Larkin et al., 2007). Subsequently, a custom Perl script was used to substitute the b factor values in the Rpg1b CC and LRR model PDB files with values reflecting the substitution pattern. For this analysis, substitutions between the following amino acid pairs were considered to be conservative: FW, IL, IV, IM, LV, LM, VM, RK, RH, KH, GA, TS, DE, NQ. The models were viewed, and images and movies generated, using the UCSF Chimera package (v 1.7) (Pettersen et al., 2004).

Cleavage of Soybean RIN4 Homologs by AvrRpt2

N-terminally Myc-tagged GmRIN4 expression constructs were generated using multisite Gateway™ (Invitrogen) technology using protocols provided by the manufacturer. Briefly, the soybean GmRIN4a (Glyma03g19920), GmRIN4b (Glyma16g12160), GmRIN4c (Glyma18g36000), and GmRIN4d (Glyma08g46400) cDNA sequences (see Supplemental Figure S8 for an alignment) were amplified from EST clones for GmRIN4a-c, and cDNA for GmRIN4d, derived from soybean cv. Williams 82 using primers that included Gateway™ attB sites. The gene-specific primers (minus the attB tails) are provided in Supplemental Table S2. BP
cloning was used to introduce these PCR products into a modified Bluescript vector (Agilent Technologies) containing the appropriate attP sites (Qi et al., 2012). Subsequently, multi-site LR cloning was used to combine the GmRIN4 sequences with 5x Myc N-terminal tags in a Gateway™ compatible variant of the pTA7002 expression vector (Aoyama and Chua, 1997). All constructs were sequence verified. Cloning in pTA7002 allowed dexamethasone-inducible expression of the GmRIN4s in transformed plant tissue. AvrRpt2, and the proteolytically inactive variant C122A, were expressed under the control of the 35S promoter and the constructs used (pMD1-AvrRpt2-HA and pMD1-AvrRpt2-C122A-HA) have been described previously (Axtell et al., 2003).

Agrobacterium tumefaciens GV3101 strains carrying the various dexamethasone-inducible GmRIN4 and AvrRpt2 constructs were grown, and inoculum prepared for transient transformation of N. benthamiana leaves, based on protocols described by Wroblewski and associates (Wroblewski et al., 2005). All strains were resuspended in water at an OD600 = 0.8. Suspensions for mixed inoculations were combined in a 1:1 ratio prior to infiltrating the leaves of 4-week old N. benthamiana. For treatments where each GmRIN4 was expressed alone, Agrobacterium suspensions were mixed in a 1:1 ratio with water prior to infiltration. Plants were sprayed with 50 µM dexamethasone (Sigma) 36 h following injection to induce expression. Samples were collected 3.5 h after dexamethasone treatment and flash-frozen in liquid nitrogen. Protein extraction, SDS-PAGE, and immunoblotting were conducted as described by Qi and associates (Qi et al., 2012).

Sequence data from this article can be found in the GenBank data library under accessions: AY452685 (Rpg1b from cv Williams 82); KF958751 (Rpg1r from PI96983); XXXX (Rpg1r from Williams 82); KF958753 (P21F22_7); YYYY(P21f22_7); KF958752 (P21f22_27); ZZZZ (W21f22_27); X87851 (Arabidopsis RPM1); GU132851 (GmRIN4a); GU132855 (GmRIN4B); GU132852 (GmRIN4C); GU132853 (GmRIN4D).

Supplemental Data

The following materials are available in the online version of this article.
Supplemental Figure S1. Genetic and physical mapping of Rpg1r region.

Supplemental Figure S2. Comparison of corresponding CNL gene sequences from BAC-derived, and soybean Whole Genome Sequence (WGS)-derived sequence.

Supplemental Figure S3. Comparison of cv. Williams 82 and PI 96983 CNL alleles.

Supplemental Figure S4. Phylogenetic relationships of the CNL genes in cultivars Williams 82 (rpg1r Rpg1b) and PI 96983 (Rpg1r rpg1b) from the ~1 Mb region containing the Rpg1 genes.

Supplemental Figure S5. Alignment of Rpg1r and Rpg1b amino acid sequences.

Supplemental Figure S6. The Rpg1b CC and NB-ARC domains group with different paralogs from the Rpg1b region in phylogenetic analyses.

Supplemental Figure S7. Comparison of the Arabidopsis RPM1 and soybean Rpg1r and Rpg1b amino acid sequences.

Supplemental Figure S8. Alignment of the Arabidopsis RIN4 (AtRIN4) amino acid sequence with the four soybean homologs (GmRIN4a, b, c and d).

Supplemental Figure S9. Identifying the location of the breakpoints that flank the recombination event in Rpg1b that encompasses the NB-ARC domain.

Supplemental Figure S10. Comparison of the Rpg1b CC and LRR domain modeled structures with the templates on which the models are based.

Supplemental Figure S11. Secondary structure guided alignment of the Rpg1b and TLR3 LRR domains using the SBAL aligner and manual editing.
Supplemental Table S1. dN, dS, and omega (dN/dS) values calculated from an alignment between Rpg1b and Rpg1r.

Supplemental Table S2. Primer sequences used in this study.

Supplemental Table S3. Details of markers used for genetic mapping.

Supplemental Table S4. Details of amplification of CNLs from soybean accession PI 96983 (Rpg1r) using primers that flank CNLs from cv. Williams82 (rpg1r).

Supplemental Movie S1. Model of the proposed Rpg1b CC domain dimer showing the location of polymorphic residues with respect to Rpg1r with the structure represented as a ribbon.

Supplemental Movie S2. Model of the proposed Rpg1b CC domain dimer showing the location of polymorphic residues with respect to Rpg1r with the structure surface shown.

Supplemental Movie S3. Model of the Rpg1b LRR domain showing the location of polymorphic residues with respect to Rpg1r with the structure represented as a ribbon.

Supplemental Movie S4. Model of the Rpg1b LRR domain showing the location of polymorphic residues with respect to Rpg1r with the structure surface shown.

ACKNOWLEDGEMENTS
We thank Michelle Metcalf and Brian Rutter for technical assistance, Blake Meyers for an alignment of the Arabidopsis CNL genes, and the USDA Soybean Germplasm Collection for soybean seed. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and
Informatics at the University of California, San Francisco (supported by NIGMS 9P41GM103311).

LITERATURE CITED


Ashfield T, Keen NT, Buzzell RI, Innes RW (1995) Soybean resistance genes specific for different Pseudomonas syringae avirulence genes are allelic, or closely linked, at the Rpg1 locus. Genetics 141: 1597-1604


Catanzariti AM, Dodds PN, Ve T, Kobe B, Ellis JG, Staskawicz BJ (2010) The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. Mol Plant-Microbe Interact 23: 49-57

Chen NW, Sevignac M, Thareau V, Magdeletan G, David P, Ashfield T, Innes RW, Geffroy V (2010) Specific resistances against Pseudomonas syringae effectors AvrB and AvrRpm1 have evolved differently in common bean (Phaseolus vulgaris), soybean (Glycine max), and Arabidopsis thaliana. New Phytol 187: 941-956


Origin of R Gene Specificity in Soybean


Origin of R Gene Specificity in Soybean


Liu J, Elmore JM, Lin ZJ, Coaker G (2011) A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. Cell Host Microbe 9: 137-146


Proc Natl Acad Sci USA 81: 6024-6028


FIGURE LEGENDS

**Figure 1.** Cloning of the soybean *Rpg1r* gene. A, Distribution of NB-LRR genes across an ~ 1 Mb interval encompassing the *Rpg1r* locus in Williams 82 and PI 96983. The vertical black lines represent genomic sequence from soybean cultivars Williams 82 (*rpg1r Rpg1b*) and PI 96983 (*Rpg1r rpg1b*) as described by Innes and associates (2008). The region corresponds to the interval from Glyma13g25440 to Glyma13g26530 in the soybean whole genome sequence (Schmutz et al., 2010). Breaks in the lines indicate lack of available sequence. The red and green polygons correspond to CC-NB-LRR (CNL) and TIR-NB-LRR (TNL) family members, respectively. Black polygons represent CNL pseudogenes or paralogues that extend beyond the available sequence. Gene names printed in black correspond to CNL genes identified in the available genomic sequence, while those printed in blue correspond to CNL genes amplified from PI 96983 total genomic DNA. The blue dashed rectangle indicates the genetically defined region containing *Rpg1r* (see Supplemental Fig. S1). The blue lines linking CNL genes in the Williams 82 and PI 96983 sequences indicate probable alleles. * W52d1_5 has the atypical structure CNL:CNL. ** gene 21f22_26 is a pseudogene in both Williams 82 and PI 96983. B, Functional complementation of *Rpg1r* recognition specificity in soybean stable transformants. Primary leaves from accession PI 96983 (*Rpg1r*), untransformed cv. Jack (*rpg1r*), and cv. Jack transformants expressing an *Rpg1r* transgene, were injected with *P. syringae* pv. *glycinea* race 4 expressing *avrRpm1* or a non-functional *avrB* allele disrupted with an omega sequence (*avrB::Ω*). The responses in two independent transformants (ND-23-1 and ND-23-3) are shown. Photographs were taken ~ 24 hrs after injection.

**Figure 2.** Structural domains of the Rpg1r protein. A, The CC domain. The predicted coiled-coil is printed in orange, with the hydrophobic residues at the "a" and "d" positions in the heptad repeats underlined. The conserved "EDVID" motif (Rairdan et al., 2008) is printed in blue. B, The nucleotide binding domain. Shown is the region corresponding to the known structure of the human Apaf-1 NB-ARC1-ARC2 region. Previously described conserved motifs (van der Biezen and Jones, 1998; Meyers et al., 1999) are printed in blue in the order: P-loop (kinase 1a), kinase 2, GLPL, RNBS-D, and MHD. An unusual substitution with a non-acidic residue within the kinase 2 motif is printed in red. C, LRR domain. Regions predicted to form α helices and β
strands are highlighted in green and blue, respectively. Residues matching the consensus for intracellular LRRs are printed in red. D, Consensus for intracellular LRRs (Jones and Jones, 1997), where x can be any residue and L can be substituted with I, V or F.

**Figure 3.** The soybean Rpg1r and Rpg1b genes are closely related, but neither is orthologous to the RPM1 gene from Arabidopsis. A Bayesian phylogenetic tree based on an amino acid alignment of the NB-ARC domains from the soybean Rpg1r and Rpg1b genes, the complete set of Arabidopsis CNLs, and the soybean CNL (Glyma06g46800) most closely related to RPM1 (based on a BLAST search of the soybean proteome). Glyma13g25950.2 represents the Rpg1b locus in the soybean whole genome sequence (Schmutz et al., 2010) but contains a small internal deletion relative to the known BAC-derived sequence. Sequence AT3G44400 belongs to the TIR-NB-LRR sub-class of NB-LRR genes and was used as an outgroup. Branches supported with posterior probabilities greater than 90% are printed in bold.

**Figure 4.** As a consequence of recombination, Rpg1r and Rpg1b share closely related NB-ARC domains but more divergent CC and LRR regions. A, Comparison of the nucleotide sequences of Rpg1r and Rpg1b. The Rpg1 sequences were compared using the VISTA program and a 100nt sliding window. The y-axis shows the percent nucleotide identity between the two genes and the x-axis shows the position in the Rpg1b ORF. The interval between the two dashed green lines corresponds to the region subsequently analyzed for evidence of recombination (as described in part B). A region found to have a strong signal for recombination is indicated with a blue dashed box (corresponds to the pink shaded region in part B). B, Evidence for a recombinant origin for the Rpg1b NB-ARC domain. Shown is output from the Bootscan program. The recombinant region in Rpg1b (region 2) completely encompasses the NB-ARC domain and is indicated by pink shading. C, Location of recombinant Rpg1b, and the predicted parental sequences, within the 1 Mb interval that contains the Rpg1 genes. The locations of CNL paralogs are indicated with colored polygons arranged on the horizontal line. The image shown represents a hypothetical ancestral chromosome that includes all the unique CNL loci represented in both Williams 82 and PI 986983. Rpg1b is shown in red and a second paralog (W173d12_25) predicted to carry the same recombination event is shown in magenta. The two parental sequences, P92i7_8 and
P21f22_29 (Rpg1r), predicted to have recombined to generate the ancestral Rpg1b locus are printed in blue and green, respectively.

**Figure 5.** Modeling of the Rpg1b protein structure, and comparison to Rpg1r, reveals highly polymorphic surfaces on both the LRR and CC domains. A to D, Model of the proposed Rpg1b CC domain homo-dimer showing the location of polymorphic residues with respect to Rpg1r. The residues conserved between the Rpg1 proteins are shown in red and those that differ shown in blue (conservative and non-conservative substitutions are shown in light and dark blue, respectively). Residues that reside at the "a" and "d" positions in predicted coiled-coil heptad repeats are shown in green. Residues that are part of the conserved EDVID sequence motifs are shown in orange. A and B, Surface containing the conserved EDVID sequence motifs. C and D, Surface on the opposite side of the structure as the EDVID motifs (i.e. Model rotated through 180 degrees). E and F, Model of the Rpg1b LRR domain showing the location of polymorphic residues with respect to Rpg1r. Residue coloring as for the CC domain.

**Figure 6.** The LRR domains of Rpg1b and Rpg1r determine AvrB versus AvrRpm1 specificity. A, Rpg1r specificity is retained when transiently expressed in *Nicotiana glutinosa*. Agroinfiltrations were used to transiently express combinations of Rpg1r-YFP, AvrRpm1, AvrB or empty vector (EV). All transgenes were under the control of a dexamethasone inducible promoter. Photographs were taken ~ 24 hrs after gene induction. B, Expression of Rpg1b-Rpg1r chimeras reveals that the LRR domain determines effector specificity. The indicated chimeras were co-expressed with GmRIN4b and AvrB or AvrRpm1. The percentages given below each leaf indicate the percentage of leaves showing a stronger response to AvrB than AvrRpm1 (left value) and the percentage of leaves showing a stronger response to AvrRpm1 than AvrB (right value). The actual number of leaves used to calculate the percentages are shown in parentheses (number of leaves showing stronger response / total number of leaves injected). The values for leaves in which the responses to both effectors were similar are not shown. The numbers in bold indicate average HR scores. Photographs were taken ~ 48 hrs after dexamethasone application.
Figure 7. AvrRpt2 cleaves all four soybean RIN4 homologs and suppresses Rpg1r-mediated HR. A, Cleavage of GmRIN4s by AvrRpt2. Soybean 5xMyc-tagged RIN4 homologs (GmRIN4a, b, c and d) were transiently co-expressed with or without HA-tagged AvrRpt2 (wt), or a protease inactive variant (C122A), in N. benthamiana, total protein extracted, and immunoblotted with the indicated antibodies. B, Rpg1r-mediated recognition of AvrRpm1 in soybean is partially blocked by AvrRpt2. P. syringae pv. glycinea race 4 (PsgR4) expressing AvrRpm1 was co-infiltrated into soybean leaves (cv. Flambeau) expressing Rpg1r together with PsgR4 strains expressing either a functional (AvrRpt2), or non functional (AvrRpt2::Ω), avrRpt2 allele. Infiltration of a mixture of the PsgR4 (avrRpt2) and PsgR4 (avrRpt2::Ω) strains was used to confirm that AvrRpt2 failed to elicit a visible HR in cultivar Flambeau. Photographs were taken ~48 hrs after infiltration.
**Figure 1.** Cloning of the soybean Rpgfr gene. A, Distribution of NB-LRR genes across an ~1 Mb interval encompassing the Rpgfr locus in Williams 82 and PI 96983. The vertical black lines represent genomic sequence from soybean cultivars Williams 82 (pgfr1 Rpgfr1b) and PI 96983 (Rpgfr1 Rpgfr1b) as described by Innes and associates (2008). The region corresponds to the interval from Glyma13g28440 to Glyma13g26530 in the soybean whole genome sequence (Schmutz et al., 2010). Breaks in the lines indicate lack of available sequence. The red and green polygons correspond to CC-NB-LRR (CNL) and TIR-NB-LRR (TNL) family members, respectively. Black polygons represent CNL pseudogenes or paralogues that extend beyond the available sequence. Gene names printed in black correspond to CNL genes identified in the available genomic sequence, while those printed in blue correspond to CNL genes amplified from PI 96983 total genomic DNA. The blue dashed rectangle indicates the genetically defined region containing Rpgfr (see Supplemental Fig. S1). The blue lines linking CNL genes in the Williams 82 and PI 96983 sequences indicate probable alleles. *W25d1_5* has the atypical structure CNL-CN. ** gene 2122_26 is a pseudogene in both Williams 82 and PI 96983. B, Functional complementation of Rpgfr recognition specificity in soybean stable transformants. Primary leaves from accession PI 96983 (Rpgfr1), untransformed cv. Jack (pgfr1), and cv. Jack transformants expressing an Rpgfr transgene were infected with *P. syringae pv. glycinea* race 4 expressing AvrRpm1 or a non-functional avrB allele disrupted with an omega sequence (avrB::O). The responses in two independent transformants (ND-23-1 and ND-23-2) are shown. Photographs were taken ~24 h after infection.
A  CC Domain

MAVELIAGALLSSFLVQAFEKLAASPQVLFHFFHGKLDETLIRKLKIKLLSILDALADDQFERQFADEPRVRNWLPHEVIDMVFDAE
LDEMQYEFSDKWELEAESESQTCTGCTKVPNFFKSSPASSFNREIKSRMEKILSDLEFLSSQXDDGLKLNASGQVGSELL

B  NB-ARC Domain

GSAVPQIQSQSTSSVVESDIYGRDKDMMIFDLTSDNPNQPSPSILSVGMGGMKTTLAQHVFDNPDRIQUEARFVDKAWVCVSDDFD
FAVRTRTVRIT EA1KTSDKDRDLEMVHGRKarenLKTLTGRFLLVLDNVWNENRLKWEAFLHVLFQAGQGSI IATRSEVASTMR
SREHLEQLQEDHCDKLWAFKAFQFDQNDIQPNPDCEQKIKTeIJKERGMLPLAKMTGSLHDKSSITEVKSLQSEIWFSASTERD
IVPALASLYHLPSSHLKRCFAYCALFPPDKYVDFKECIIQLNMMAEKLFCIQCQSQQGKSPEEVGQYFQNDLLSRCFFQOSSNTKRTHF
VMDHLNLNDLRFICG

C  LRR Domain

DICFRDLGQDTQGTPKAPRHFVSAEIEHVRYFDGFGLTDACKLKLRSYMTSNNKIDFYDSPSDDWDNNMTHEL

| ISKFKFL | RVLSLH | CSN | LSQVPDS |
| VGNLKYL | SSLLSN | TE  | TVKLPES |
| CSLYNL | QIKLNY | CKQ | LDELPSPN |
| LHKTLTL | HRLRIID | TG  | VRKVAH |
| LGKLKQL | QVSM | SPPVKV | SRE | FSQQ |
| LGELONLH | GSLSQN | LQN | VESPISALAVD |
| LKYNKTLH | LEVELEW | D  | SDWNPDSTKREDIEIVN |
| LQPPKHL | EKLDRWN | YG  | GKOFPWLL |
| NNSLLNV | VSLLENS | CQSS | CQRLPP |
| LGLLPFL | KELLSIQ | LAG | TVSNADEFGSS |
| SCSPTSL | ESIMERHS | M  | KEWEWECKGV |
| TGAFPLL | QRLSEY | CQKL | KGHLPQE |
| LCHL | NDLKyGC | C  | EQLVPSS |
| ALSAPDI | HQLSLGD | C  | GKLQ |
| TAHPTTL | KELTITG | HN  | VAALEQIG |
| RYSJCSN | NNIPMHS | CYDF | VR VINGCGLSLTTIP |
| LDMPKRL | RGLDGC | CPN | LQRISS |
| GQAHNL | QSLYIK | CPQ | LESLPEGM |
| HVLLLD | HYLSIRD | CPK | VEMPFE |
| GGPLSPL | KGMRILHG | SYKLN | MSS KGA |
| LGNSHL | ISLICYG | VD  | VECLOPD |
| GVLPHSL | VSLWESD | CED | LKRLDYKG |
| LCHLSL | KNLFVYD | CPR | LQCLPE |
| EGPLPSM | LSLHTYD | CF  | LIQRCEPEGED |
| WPKIAHI | DXYKEENG | EVDYK |

D  LRR Consensus for intracellular LRRs

LxxLxxL  xxLxxLxx  Cx(x)  LxxIPxx

T  N

Figure 2. Structural domains of the Rpg1r protein. A, The CC domain. The predicted coiled-coil is printed in orange, with the hydrophobic residues at the "a" and "d" positions in the heptad repeats underlined. The conserved "EDVID" motif (Rainard et al., 2008) is printed in blue. B, The NB-ARC domain. Shown is the region corresponding to the known structure of the human Apaf-1 NB-ARC1-ARC2 region. Previously described conserved motifs (van der Biezen and Jones, 1998; Meyers et al., 1999) are printed in blue in the order: P-loop (kinase 1a), kinase 2, GLPL, RNBS-D, and MHD. An unusual substitution with a non-acidic residue within the kinase 2 motif is printed in red. C, LRR domain. Regions predicted to form α helices and β strands are highlighted in green and blue, respectively. Regions matching the consensus for intracellular LRRs are printed in red. D, Consensus for intracellular LRRs (Jones and Jones, 1997), where x can be any residue and L can be substituted with I, V or F.
Figure 3. The soybean Rpg1r and Rpg1b genes are closely related, but neither is orthologous to the RPM1 gene from Arabidopsis. A Bayesian phylogenetic tree based on an amino acid alignment of the NB-ARC domains from the Rpg1r and Rpg1b proteins, the complete set of Arabidopsis CNLs, and the soybean CNL (Glyma06g46800) most closely related to RPM1 (based on a BLAST search of the soybean proteome). Glyma13g25950.2 represents the Rpg1b locus in the soybean whole genome sequence (Schmutz et al., 2010) but contains a small internal deletion relative to the known BAC-derived sequence. Sequence AT3G44400 belongs to the TIR-NB-LRR sub-class of NB-LRR genes and was used as an outgroup. Branches supported with posterior probabilities greater than 90% are printed in bold.
Figure 4. As a consequence of recombination, Rpg1r and Rpg1b share closely related NB-ARC domains but more divergent CC and LRR regions. A, Comparison of the nucleotide sequences of Rpg1r and Rpg1b. The Rpg1 sequences were compared using the VISTA program and a 100nt sliding window. The y-axis shows the percent nucleotide identity between the two genes and the x-axis shows the position in the Rpg1b ORF. The interval between the two dashed green lines corresponds to the region subsequently analyzed for evidence of recombination (as described in part B). A region found to have a strong signal for recombination is indicated with a blue dashed box (corresponds to the pink shaded region in part B). B, Evidence for a recombinant origin for the Rpg1b NB-ARC domain. Shown is the output from the Bootscan program. The recombinant region in Rpg1b (region 2) completely encompasses the NB-ARC domain and is indicated by pink shading. C, Location of recombinant Rpg1b, and the predicted parental sequences, within the 1Mb interval that contains the Rpg1 genes. The locations of CNL paralogues are indicated with colored polygons arranged on the horizontal line. The image shown represents a hypothetical ancestral chromosome that includes all the unique CNL loci represented in both Williams82 and PI 986983. One allele from either Williams82 or from PI 96983 is included for each CNL locus. Rpg1b is shown in red and a second parologue (W173d12_25) predicted to carry the same recombination event is shown in magenta. The two parental sequences, P9217_8 and P21122_29 (Rpg1r), predicted to have recombined to generate the ancestral Rpg1b loci are printed in blue and green, respectively. The polygons representing potential alternative recombinant parental sequences to P9217_8 and P21122_29 are printed in blue and green, respectively.
Figure 5. Modeling of the Rpg1b protein structure, and comparison to Rpg1r, reveals highly polymorphic surfaces on both the LRR and CC domains. A to D, Model of the proposed Rpg1b CC domain homo-dimer showing the location of polymorphic residues with respect to Rpg1r. The residues conserved between the Rpg1 proteins are shown in red and those that differ shown in blue (conservative and non-conservative substitutions are shown in light and dark blue, respectively). Residues that reside at the "a" and "d" positions in predicted coiled-coil heptad repeats are shown in green. Residues that are part of the conserved EDVID sequence motifs are shown in orange. A and B, Surface containing the conserved EDVID sequence motifs. C and D, Surface on the opposite side of the structure as the EDVID motifs (i.e. Model rotated 180° about the Y axis). E, Model of the Rpg1b LRR domain showing the location of polymorphic residues with respect to Rpg1r. E, Model represented as a ribbon. F, Model surface shown. Residue coloring as for the CC domain.
Figure 6. The LRR domains of Rpg1b and Rpg1r determine AvrB versus AvrRpm1 specificity. A, Rpg1r specificity is retained when transiently expressed in *Nicotiana glutinosa*. Agroinfiltrations were used to transiently express combinations of Rpg1r-YFP, AvrRpm1, AvrB or empty vector (EV). All transgenes were under the control of a dexamethasone inducible promoter. Photographs were taken ~24 hrs after gene induction. B, Expression of Rpg1b-Rpg1r chimeras reveals that the LRR domain determines effector specificity. The indicated chimeras were co-expressed with GmRIN4b and AvrB or AvrRpm1. The percentages given below each leaf indicate the percentage of leaves showing a stronger response to AvrB than AvrRpm1 (left value) and the percentage of leaves showing a stronger response to AvrRpm1 than AvrB (right value). The actual number of leaves used to calculate the percentages are shown in parentheses (number of leaves showing stronger response / total number of leaves injected). The values for leaves in which the responses to both effectors were similar are not shown. The numbers in bold indicate average HR scores. Photographs were taken ~48 hrs after dexamethasone application.
Figure 7. AvrRpt2 cleaves all four soybean RIN4 homologs and suppresses Rpg1r-mediated HR.

A. Cleavage of GmRIN4s by AvrRpt2. Soybean 5xMyc-tagged RIN4 homologs (GmRIN4a, b, c and d) were transiently co-expressed with or without HA-tagged AvrRpt2 (wt), or a protease inactive variant (C122A), in N. benthamiana, total protein extracted, and immunoblotted with the indicated antibodies.

B. Rpg1r-mediated recognition of AvrRpm1 in soybean is partially blocked by AvrRpt2. P. syringae pv. glycinea race 4 (PsgR4) expressing AvrRpm1 was co-infiltrated into soybean leaves (cv. Flambeau) expressing Rpg1r together with PsgR4 strains expressing either a functional (AvrRpt2), or non functional (AvrRpt2::Ω), avrRpt2 allele. Infiltration of a mixture of the PsgR4 (avrRpt2) and PsgR4 (avrRpt2::Ω) strains was used to confirm that AvrRpt2 failed to elicit a visible HR in cultivar Flambeau. Photographs were taken 5 days after infiltration.