**nature biotechnology**

# A pigeonpea gene confers resistance to Asian soybean rust in soybean

Cintia G Kawashima[1], Gustavo Augusto Guimarães[2], Sônia Regina Nogueira[2], Dan MacLean[1], Doug R Cook[3], Burkhard Steuernagel[1,6], Jongmin Baek[3], Costas Bouyioukos[1,6], Bernardo do V A Melo[2], Gustavo Tristão[2], Jamile Camargos de Oliveira[2], Gilda Rauscher[4], Shipra Mittal[4], Lisa Panichelli[4], Karen Bacot[4], Ebony Johnson[4], Geeta Iyer[4], Girma Tabor[4], Brande B H Wulff[1,6], Eric Ward[5,6], Gregory J Rairdan[4], Karen E Broglie[4], Gusui Wu[4], H Peter van Esse[1,7], Jonathan D G Jones[1,7] & Sérgio H Brommonschenkel[2,7]

**Asian soybean rust (ASR), caused by the fungus *Phakopsora pachyrhizi*, is one of the most economically important crop diseases, but is only treatable with fungicides, which are becoming less effective owing to the emergence of fungicide resistance. There are no commercial soybean cultivars with durable resistance to *P. pachyrhizi*, and although soybean resistance loci have been mapped, no resistance genes have been cloned. We report the cloning of a *P. pachyrhizi* resistance gene *CcRpp1* (*Cajanus cajan* Resistance against *Phakopsora pachyrhizi 1*) from pigeonpea (*Cajanus cajan*) and show that *CcRpp1* confers full resistance to *P. pachyrhizi* in soybean. Our findings show that legume species related to soybean such as pigeonpea, cowpea, common bean and others could provide a valuable and diverse pool of resistance traits for crop improvement.**

Soybean (*Glycine max* (L.) Merr.) is one of the main sources of plant protein and vegetable oil worldwide, with only corn, rice and wheat ranking higher in terms of total cultivated area (http://faostat.fao.org/). ASR is one of the most damaging diseases of soybean, and is caused by the obligate biotrophic fungus *P. pachyrhizi* Sydow & P. Sydow[1]. Yield losses caused by ASR can be as high as 40–80%, and even a disease incidence as low as 0.05% can affect yields[2,3]. The disease is ubiquitously present in the soybean growing areas of Brazil, which is the second largest producer of soybean, contributing 30% of world soybean production. In addition, the pathogen has established itself in Florida and can spread in summer to other parts of North America; the disease, therefore, requires constant monitoring for changes in seasonal distribution[4]. Currently, no commercially grown soybean cultivars are available that are fully resistant to all *P. pachyrhizi* isolates, and thus routine fungicide application is required to control the disease. Chemical control of ASR in Brazil began in 2002/03 after introduction of the disease in 2001. By 2004 approximately 20 million hectares were sprayed with fungicides, with a mean application of three treatments per hectare[5]. Other management strategies include a host-free period to break the continuous cycle of fungal infection and delay the onset of epidemics. The total costs incurred by Brazil to control the disease, when accounting for direct and indirect losses, are estimated to be around $2 billion per year[2].

Given the cost to growers and the environment, genetic resistance against *P. pachyrhizi* is highly desirable. Eight major resistance loci have been mapped in soybean (*Rpp1* to *Rpp6*, *Rpp1b* and *Rpp?*(*Hyuuga*)), but none has yet been cloned and only *Rpp4* has been characterized in any detail[6–14]. In addition, many of these loci have been introgressed into commercial soybean cultivars separately and, as a result, ASR isolates that can overcome *Rpp1* to *Rpp6* have evolved and can be readily identified in the field[15,16]. Screening the available soybean germplasm for additional sources of resistance has not revealed genes that, individually, confer adequate resistance in an agronomic setting. A screen of the US Department of Agriculture soybean germplasm (16,595 accessions) identified no plants with immunity to ASR and only 33 plants with moderate resistance in which later arrest of the pathogen results in macroscopically observable reddish-brown lesions, so called RB (reddish-brown) type resistance[17]. Given the rapid breakdown of *Rpp1-6*–mediated resistance[15,16], there is concern that new ASR resistance genes may be specific to one isolate and therefore rapidly overcome in the field.

Due to the limited resistance available in soybean germplasm, heterologous expression of transgenes from unrelated plant species in soybean has been investigated as an alternate source of ASR resistance. For example, overexpression in soybean of several genes originally induced in the non-host plant *Arabidopsis* by ASR led to a slight increase in level of resistance in soybean[18]. However, the resistance conferred was quantitative, as might be expected from overexpressing only one out of a plethora of responses induced in *Arabidopsis* during non-host resistance[19]. Because ASR affects yields even at low levels of infection, it seems unlikely that this approach will provide commercially relevant levels of resistance in the near future[2,3].

**Figure 1** *C. cajan* accessions display different reactions to *P. pachyrhizi* field isolates from Brazil. (**a**) Differential responses observed in different *C. cajan* accessions against *P. pachyrhizi* isolate PPUFV02. The reactions range from left to right: immune, immune with flecks, reddish-brown lesions without sporulation, reddish-brown lesions with sporulation, susceptible 'tan' lesions with abundant sporulation on the abaxial side of leaves. Scale bars, 500 μm. (**b**) Locations where *P. pachyrhizi* isolates used to challenge the *C. cajan* resistance response in accessions G119-99, G59-95, G146-97 were collected.
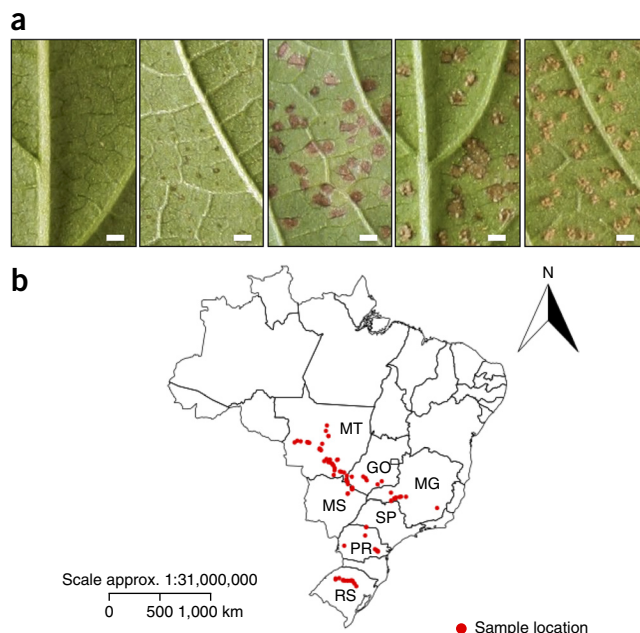


Here, we report the discovery and transfer of an intracellular immune receptor from pigeonpea into soybean. This enables soybean to recognize *P. pachyrhizi* and, as a consequence of this recognition, activate a diverse set of responses to the pathogen. By using a transgene that functions at the point of pathogen perception we can achieve a level of resistance which might provide commercial control that is superior to current management strategies.

Although many rust fungi have a limited host range, *P. pachyrhizi* is known to infect leaf tissue from diverse leguminous plants in the field (at least 31 species in 17 genera)[20]. In total, 152 leguminoseous species in 56 genera have been described as potential hosts for *P. pachyrhizi*[20–22]. Resistant and susceptible accessions are present in many of these species, and we hypothesized that such species might contain valuable sources of resistance genes that would segregate and could be cloned and transferred to soybean. Pigeonpea (*Cajanus cajan* (L.) Millsp.), a close relative of soybean, is a perennial, diploid, self-pollinating legume with a genome size of ~830 Mbp. We screened 52 accessions of *C. cajan* that were introduced into Brazil during *C. cajan* breeding programs[23]. When challenged with the Brazilian monopustule *P. pachyrhizi* isolate PPUFV02 (isolated from a mixed field sample) 52 accessions displayed a range of phenotypes, from immune (where symptoms cannot be macroscopically observed), to susceptibility levels similar to soybean's (**Fig. 1a**). Three segregating populations were established by crossing resistant genotypes G119-99, G59-95 and G146-97 to the susceptible accession G48-95. Resistance derived from G119-99, G59-95, and G146-97 segregated in a 3:1 ratio in the respective F$_2$ populations, suggesting that each accession carries a single dominant gene for resistance (**Supplementary Table 1**). We prioritized the cross G119-99 × G48-95 because it yielded a higher number of seeds compared to the other crosses, which would allow for more rapid fine mapping and therefore cloning of the causal gene in this accession. To ascertain whether resistance in *C. cajan* G119-99 could be effective against additional field isolates of *P. pachyrhizi*, we challenged this accession with 77 Brazilian field isolates collected from different geographical locations, two US isolates and a Japanese isolate (**Fig. 1b** and **Supplementary Table 2**). Remarkably, no *P. pachyrhizi* isolates were identified that could overcome the resistance in G119-99.

We next set out to identify and clone the resistance gene from the locus *CcRpp1*, from G119-99. By scoring progeny from 2,282 gametes for *P. pachyrhizi* resistance and susceptibility, we identified an interval of <154 kb delineated by the markers dCAPS52491 and SSR2152 (**Supplementary Fig. 1** and **Supplementary Table 3**) on *C. cajan* linkage group 5. Interestingly, this region is syntenic to regions within chromosome 12 and 9 in soybean that contain genes of the NB-LRR (nucleotide binding leucine rich repeat) class that are associated with disease resistance[24,25]. Despite the synteny, these regions in the soybean genome have not been shown to confer ASR resistance[6–14].

To determine the genomic organization of the *CcRpp1* locus in G119-99, we generated a G119-99 bacterial artificial chromosome (BAC) library and screened it with probes derived from the flanking markers dCAPS140555 and SSR2152. Two BAC clones (3F and 6G) were identified that span the complete interval between dCAPS52491 and SSR2152. DNA sequencing produced a single contig of 205,344 bp that contains 16 open reading frames including four predicted NB-LRR-encoding gene sequences (**Fig. 2a** and **Supplementary Dataset 1**). Southern blot analyses, using the P-loop region present in the four NB-LRR genes (NB-1, 2, 3, 4) as a probe, confirmed that the *CcRpp1* locus in G119-99 contains four NB-LRR genes (**Supplementary Fig. 2a**). Using the markers generated to fine-map the *CcRpp1* locus, we analyzed the other resistant lines (G59-95, G146-97) with *P. pachyrhizi* isolate PPUFV02 (**Supplementary Table 3**). Resistance mapped to the same region in the *C. cajan* genome, suggesting that resistance in G59-95 and G146-97 is conferred by the *CcRpp1* locus. To ascertain the level of variation at the *CcRpp1* locus we cloned NB-1, 2, 3 and 4 from all three accessions by PCR followed by Sanger sequencing (**Supplementary Table 3**). Interestingly, variation was observed both in the number of NB-LRR paralogs and their sequences (**Fig. 2b–e**). Southern blot analysis for G59-95 with the P-loop region probe suggests that the PCR analysis accurately predicts the number of paralogous genes (**Supplementary Fig. 2b**). Mutations in the LRR region of NB-LRR proteins can result in differential recognition specificity; therefore, the NB-LRR genes present at the *CcRpp1* locus in different accessions may contain nonidentical recognition specificities, as has been described in other pathosystems[26,27]. However, because we now had characterized the resistance present in G119-99 × G48-95 in the most detail (G119-99 was tested and effective against 77 isolates of *P. pachyrhizi*), we focused our additional analyses on this population.

Using RNA-Seq data from the resistant accession G119-99, and aligning the reads with the *CcRpp1* locus, we found that only NB-2 had RNA-Seq reads aligned to it (**Supplementary Fig. 3**). These data suggest that NB-2 is the only gene that has a basal expression in G119-99. To test whether transfer of any of the four NB-LRR genes present in the *CcRpp1* locus into soybean would confer ASR resistance, we designed plant transformation constructs that contained the individual NB-LRR genes (NB-1, 2, 3, 4) driven by the *G. max* SUBI-1 (polyubiquitin, Glyma10G39780) promoter. These constructs were introduced into soybean by biolistic transformation. When challenged with *P. pachyrhizi* (isolates G05 and MS08), no differential phenotypes

**Figure 2** Characterization of the CcRpp1 locus in four accessions of *C. cajan*. (**a**) Genetic and physical fine-map of *CcRpp1* locus in G119-99. BAC clones identified using dCAPS52491 and SSR2152 are indicated. Sequenced BAC clones are indicated in bold letters; the position of the four CC-NB-LRR sequences (NB-1 to NB-4) on the BAC clones is also indicated. (**b**–**e**) Nonsynonymous ($K_a$) and synonymous ($K_s$) nucleotide substitutions in the CC-NB-LRR genes NB-1 (**b**), NB-2 (**c**), NB-3 (**d**) and NB-4 (**e**) present at the *CcRpp1* locus in the G59-95, G146-97 resistant and the GS48-95 susceptible accession(s) when compared to accession G119-99. The $K_a$ mutations are indicated in red; the $K_s$ mutations are indicated in blue. NB-LRR genes present in G119-99 but absent in other accessions are indicated either by absence (confirmed by PCR and Southern blot analysis), or by dashed boxes (absence confirmed by PCR only).
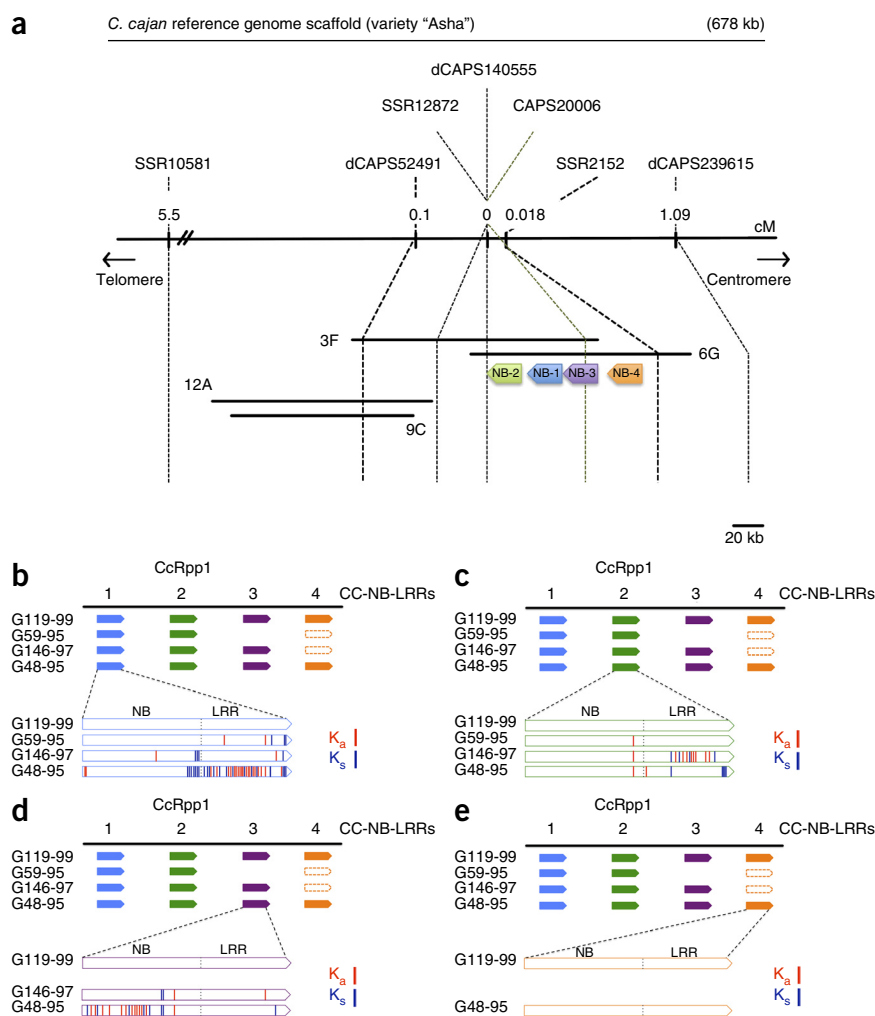
were observed for NB-1, NB-3 or NB-4, although all three resistance genes were expressed, as confirmed by qRT-PCR (data not shown). However, full resistance (no fungal sporulation or macroscopic observable lesions) was observed on the leaves of three plants representing two independent events (Soy 3194.5.1 and Soy 3194.7.1) that were confirmed by qRT-PCR to contain an expressed NB-2 gene.

NB-2 conferred resistance against *P. pachyrhizi* in events Soy3194.5.1 and Soy3194.7.1, and was renamed as *CcRpp1*. The *CcRpp1* gene product has a typical NB-LRR structure with an N-terminal coiled coil (CC) domain[26]. To further characterize the resistance we advanced the Soy3194.5.1 (*CcRpp1*.5.1) and Soy3194.7.1 (*CcRpp1*.7.1) events to the $T_1$ stage (the first segregating generation of events) and compared ASR phenotypes to nontransgenic plants segregating from the same event (null plants). Most

samples (40/46) from plants lacking the transgene were scored 15 d after inoculation, whereas 71 plants hemizygous and 37 plants homozygous for the transgene were scored 29 d after inoculation. *CcRpp1* homozygous lines displayed high levels of resistance, with no visible lesions. When averaged across all of the homozygous plants, presence of the transgene correlated with >99% reduction in lesion counts per unit leaf area. Hemizygous plants displayed RB type resistance and showed 60–71% reduction in lesion count per $cm^2$. Null plants contained tan colored lesions and high sporulation, typical of a susceptible reaction to *P. pachyrhizi* (**Fig. 3a,b**, **Supplementary Datasets 2–4** and **Supplementary Table 4**). To determine whether the partial resistance in hemizygous plants is correlated with differences in *CcRpp1* expression level, we quantified *CcRpp1* expression in null
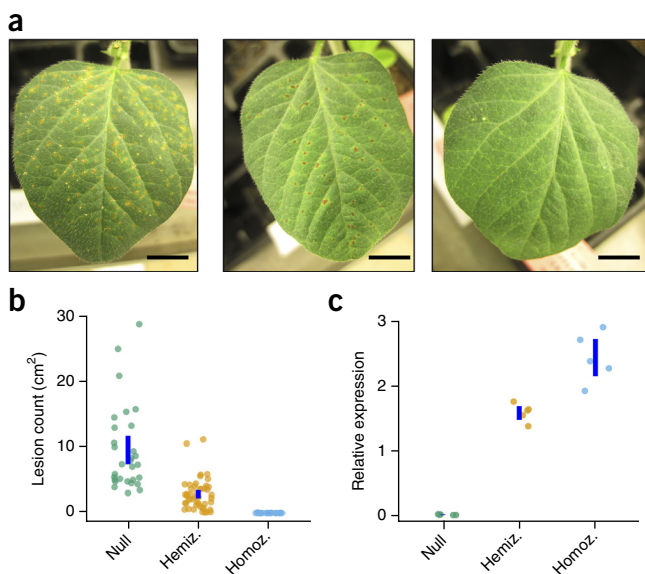


**Figure 3** Characterization of soybean plants heterologously expressing CcRpp1. (**a**) Soybean plants that express the *CcRpp1* gene driven by the *G. max* SUBI-1 (polyubiquitin, Glyma10G39780) promoter challenged with *P. pachyrhizi*. From left to right, $T_1$ null plants, $T_1$ hemizygous plants and $T_1$ plants homozygous for the transgene. Scale bars, 1 cm. (**b**) Average lesion count per $cm^2$ of leaf monitored in $T_1$ null plants, $T_1$ hemizygous plants and $T_1$ plants homozygous for the transgene. (**c**) Expression level differences between $T_1$ null plants, $T_1$ hemizygous plants and $T_1$ plants homozygous for the SUBI-1 driven *CcRpp1* gene as determined by qRT-PCR. Error bars, nonparametric (bootstrapped), limited 95% confidence interval of the mean (see **Supplementary Datasets 2–7** for details on the statistical analyses and figure generation).

plants and plants hemizygous and homozygous for the transgene by qRT-PCR (**Fig. 3c**, **Supplementary Datasets 5–7** and **Supplementary Table 5**). Increased transgene expression was observed in the homozygous plants compared with the hemizygotes, suggesting that expression level influences efficacy of *CcRpp1*.

To test whether *P. pachyrhizi* disease resistance is caused by auto-activity of CcRpp1, we tested plants homozygous for the transgene against *Fusarium virguliforme,* a filamentous fungal plant pathogen that causes sudden death syndrome (SDS) in soybean. We did not observe increased resistance against this pathogen when we compared plants homozygous for the transgene with null segregants, indicating that the response observed against *P. pachyrhizi* is specific (**Supplementary Fig. 4** and **Supplementary Datasets 8–10**). In addition, preliminary agronomic data indicate there is no adverse affect of *CcRpp1* expression on plant development, a phenotype often associated with auto activity of plant immune receptors[28] (**Supplementary Fig. 5a,b**). Finally, we did not observe segregation distortion in germination rate when multiplying hemizygous plants originating from the *CcRpp1*.7.1 event (**Supplementary Fig. 5c**, **Supplementary Datasets 11–16** and **Supplementary Table 6**).

In conclusion, we have identified and cloned a gene from *C. cajan* that confers resistance to *P. pachyrhizi* when expressed in soybean. Previous work did not identify novel soybean germplasm that displayed immunity to ASR and identified only 33 accessions with moderate RB type resistance, and thus revealed that the number of ASR resistance genes in soybean germplasm is limited[16]. Resistance genes that provide immunity to ASR are a valuable resource.

Thus, the significance of this work is the demonstration that it is possible to effectively transfer a dominant resistance gene from a related legume into soybean. The Fabaceae (Leguminosae) is a large and diverse plant family, with around 700 genera and 20,000 species[29]. Our results suggest that this tremendous natural resource can be used to identify additional resistance genes against ASR that are absent from the soybean gene pool. These legume resistance genes could be used to develop durable and environmentally sustainable ASR control strategies.

Finally, although we have not been able to identify *P. pachyrhizi* isolates that can overcome CcRpp1, *P. pachyrhizi* has demonstrated that it can rapidly overcome resistance genes that are deployed individually. With 30 million hectares of soybean under cultivation in Brazil, it would be prudent to only deploy CcRpp1 in soybean together with additional resistance genes that have different specificity or different mechanism of ASR resistance, to increase the durability of these resources[30].

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** European Nucleotide Archive (ENA): LN879382, UCDavis BAC Cc13A15 (referred to in text as Cc13A15); LN879378, NB-1; LN879379: NB-2; LN879380: NB-3; LN879381: NB-4; and study accession PRJEB9933: raw sequencing data.

*Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).*

## AUTHOR CONTRIBUTIONS

C.G.K., G.A.G., S.R.N., B.d.V.A.M., G.T., J.C.d.O., G.R., S.M., L.P., K.B., E.J., G.I., G.T. and S.H.B. performed research. D.R.C., J.B., C.B., B.S. and D.M. contributed bioinformatic tools. C.G.K., S.H.B., B.S., G.R., G.J.R., K.E.B., D.R.C., B.B.H.W., J.D.G.J., H.P.v.E. and E.W. analyzed the data. E.W., J.D.G.J., S.H.B., K.E.B., B.B.H.W., B.S., D.R.C., G.J.R., G.R., C.B. and G.W. edited the manuscript. C.G.K. and H.P.v.E. wrote the paper. K.E.B., G.W., E.W., B.B.H.W., H.P.v.E., J.D.G.J. and S.H.B. directed aspects of the project.

1. Goellner, K. *et al.* Phakopsora pachyrhizi, the causal agent of Asian soybean rust. *Mol. Plant Pathol.* **11**, 169–177 (2010).
2. Yorinori, J.T. *et al.* Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. *Plant Dis.* **89**, 675–677 (2005).
3. Scherm, H., Christianoa, R.S.C., Eskerb, P.D., Del Pontec, E.M. & Godoy, C.V. Quantitative review of fungicide efficacy trials for managing soybean rust in Brazil. *Crop Prot.* **28**, 774–782 (2009).
4. Kelly, H.Y. *et al.* From select agent to an established pathogen: the response to Phakopsora pachyrhizi (soybean rust) in North America. *Phytopathology* **105**, 905–916 (2015).
5. Reis, E.M., Deuner, E. & Zanatta, M. *In vivo* sensitivity of *Phakopsora pachyrhizi* to DMI and QoI fungicides. *Summa Phytopathol.* **41**, 21–24 (2015).
6. Hyten, D.L. *et al.* Map location of the *Rpp1* locus that confers resistance to soybean rust in soybean. *Crop Sci.* **47**, 837–838 (2007).
7. Yu, N. *et al.* Fine mapping of the Asian soybean rust resistance gene *Rpp2* from soybean PI 230970. *Theor. Appl. Genet.* **128**, 387–396 (2015).
8. Hyten, D.L. *et al.* 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 (2009).
9. Silva, D.C.G. *et al.* Molecular mapping of two loci that confer resistance to Asian rust in soybean. *Theor. Appl. Genet.* **117**, 57–63 (2008).
10. Li, S., Smith, J.R., Ray, J.D. & Frederick, R.D. Identification of a new soybean rust resistance gene in PI 567102B. *Theor. Appl. Genet.* **125**, 133–142 (2012).
11. Garcia, A. *et al.* Molecular mapping of soybean rust (*Phakopsora pachyrhizi*) resistance genes: discovery of a novel locus and alleles. *Theor. Appl. Genet.* **117**, 545–553 (2008).
12. Kim, K.S. *et al.* 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 (2012).
13. Monteros, M.J., Missaoui, A.M., Phillips, D.V., Walker, D.R. & Boerma, H.R. Mapping and confirmation of the 'Hyuuga' red-brown lesion resistance gene for Asian soybean rust. *Crop Sci.* **47**, 829–834 (2007).
14. Meyer, J.D. *et al.* Identification and analyses of candidate genes for *rpp4*-mediated resistance to Asian soybean rust in soybean. *Plant Physiol.* **150**, 295–307 (2009).
15. Akamatsu, H. *et al.* Pathogenic diversity of soybean rust in Argentina, Brazil, and Paraguay. *J. Gen. Plant Pathol.* **79**, 28–40 (2013).
16. Paul, C., Hartman, G.L., Marois, J.J., Wright, D.L. & Walker, D.R. First report of *Phakopsora pachyrhizi* adapting to Soybean genotypes with *Rpp1* or *Rpp6* rust resistance genes in field plots in the United States. *Plant Dis.* **97**, 1379 (2013).
17. Miles, M.R., Frederick, R.D. & Hartman, G.L. Evaluation of soybean germplasm for resistance to *Phakopsora pachyrhizi*. *Plant Health Prog.* http://dx.doi:10.1094/PHP-2006-0104-01-RS. (2006).
18. Langenbach, C. *et al.* Interspecies gene transfer provides soybean resistance to a fungal pathogen. *Plant Biotechnol. J.* http://dx.doi.org/10.1111/pbi.12418 (2015).
19. Ellis, J. Insights into nonhost disease resistance: can they assist disease control in agriculture? *Plant Cell* **18**, 523–528 (2006).
20. Slaminko, T.L., Miles, M.R., Frederick, R.D., Bonde, M.R. & Hartman, G.L. New legume hosts of *Phakopsora pachyrhizi* based on greenhouse evaluations. *Plant Dis.* **92**, 767–771 (2008).

21. Bonde, M.R. *et al.* Comparative susceptibilities of legume species to infection by *Phakopsora pachyrhizi. Plant Dis.* **92**, 30–36 (2008).
22. Ono, Y., Buritica, U.P. & Hennen, J.F. Delimitation of *Phakopsora, Physopella* and *Cerotelium* and their species on Leguminosae. *Mycol. Res.* **96**, 825–850 (1992).
23. Provazi, M., Camargo, L.H.G., Santos, P.M. & Godoy, R. Descrição botânica de linhagens puras selecionadas de guandu. *Rev. Bras. Zootec.* **36**, 328–334 (2007).
24. Jones, J.D.G. & Dangl, J.L. The plant immune system. *Nature* **444**, 323–329 (2006).
25. Dodds, P.N. & Rathjen, J.P. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* **11**, 539–548 (2010).
26. Allen, R.L. *et al.* Natural variation reveals key amino acids in a downy mildew effector that alters recognition specificity by an *Arabidopsis* resistance gene. *Mol. Plant Pathol.* **9**, 511–523 (2008).
27. Ravensdale, M., Nemri, A., Thrall, P.H., Ellis, J.G. & Dodds, P.N. Co-evolutionary interactions between host resistance and pathogen effector genes in flax rust disease. *Mol. Plant Pathol.* **12**, 93–102 (2011).
28. Gurr, S.J. & Rushton, P.J. Engineering plants with increased disease resistance: what are we going to express? *Trends Biotechnol.* **23**, 275–282 (2005).
29. Doyle, J.J. & Luckow, M.A. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiol.* **131**, 900–910 (2003).
30. Jones, J.D.G. *et al.* Elevating crop disease resistance with cloned genes. *Phil. Trans. R. Soc. B* **369**, 20130087 (2014).

# ONLINE METHODS

**Fine mapping of the *CcRpp1* locus.** We scored $F_2$ plants from the cross G119-99 × G48-95 segregating for resistance and susceptibility to *P. pachyrhizi* race PPUFV02, and extracted mRNA from 50 susceptible and 50 resistant *C. cajan* plants. Resistant and susceptible samples were bulked (pooled). The SMART kit and SMARTIV oligonucleotide (Clontech) were used for the first-strand cDNA synthesis. Duplex-specific Kamchatka crab nuclease was used after melting and re-association to normalize gene copy number in the cDNA generated from both resistant and susceptible bulks (TRIMMER cDNA Normalization Kit, Evrogen). Normalized cDNAs were rendered double stranded, digested with SfiI (New England BioLabs) and size-fractionated by CHROMA spin-1000 columns (Clontech) to select for transcript fragments above 1 kb. Resistant and susceptible bulk cDNAs were subsequently sheared using a Covaris S20 to obtain average fragment sizes of 200 bp. These sheared cDNAs were used to generate Illumina libraries using the sequencing generation kit (Cat no. FC-103-1002) and the sequencing kit (v3) (Cat no. FC-104-1003), and sequenced with the Illumina GAII sequencer. We generated data by sequencing on two lanes, one for each $F_2$ bulk: (i) the resistant bulk containing 3.07 Gb in 20 million paired reads and (ii) the susceptible bulk containing 2.96 Gb in 19.5 million paired reads. The raw sequencing data are deposited at the European Nucleotide Archive (ENA), study accession PRJEB9933. We pooled raw data from the susceptible and resistant bulks and generated a filter to select only NB-LRR-associated reads before assembly. We performed a reciprocal BLAST search between the *Arabidopsis*[31] and Medicago[32] NB-LRR-associated genes (205 and 369 sequences, respectively) against the 75,778 soybean predicted genes (JGI; http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax) as a database with TBLASTN[33]. This database was queried with the soybean predicted genes using BLASTX[33]. The resulting NB-LRR-related reads were then *de novo* assembled with ABySS[34] to obtain a common reference comprising 49 NB-LRR genes. Based on mapping the resistant and the susceptible bulk reads to this common reference respectively, using MAQ[35] (version 0.7.1), we mined our data for informative single-nucleotide polymorphisms (SNPs). Using a custom-made Perl script (find_proportional_snps.pl; https://github.com/danmaclean/2bl), we identified several SNPs with a minimum coverage of ten nucleotides that had an allele frequency profile consistent with that expected at the *CcRpp1* locus in the $F_2$ population (a ~66% (0.55 to 0.75 nucleotide proportion) and ~0.33% (0.25 to 0.45 nucleotide proportion) reference allele frequency in the bulk resistant and susceptible data, respectively). These polymorphisms were used to generate markers for fine mapping. All the dCAPS markers in this study were generated using the program dCAPS Finder 2.0 (ref. 36; http://helix.wustl.edu/dcaps/dcaps.html). Marker dCAPS140555 was completely linked to the *CcRpp1* locus after analyzing 2,282 gametes (1,141 $F_2$ susceptible plants).

Seven marker sequences genetically linked to *CcRpp1* were aligned to the *G. max* genome (Phytozome v5.0, http://phytozome.jgi.doe.gov/) using BLASTN[33] (default parameters). The seven linked markers identified in *C. cajan* displayed strong synteny with both chromosomes 9 and 12, with chromosome 12 being the most syntenic. *G. max* is an ancient polyploid and chromosomes 9 and 12 are homoeologous[37]. To identify additional polymorphisms and indels for marker development, a 2-Mb region starting from the telomere region on chromosome 12 that contains the identified homologous region was used to align *C. cajan* variety Asha scaffolds (Version 1; http://www.ncbi.nlm.nih.gov/nuccore/450342093?report=genbank) from the publicly available *C. cajan* assembly. Based on this sequence we designed additional primers surrounding polymorphic sequences between the resistant G119-99 and the susceptible G48-95 parents that resulted in the identification of five additional markers for fine mapping of *CcRpp1* (dCAPS3978, SSR10581, dCAPS52491, SSR2152 and dCAPS239615). The majority of these marker sequences were located within a 1.5-Mb region on chromosome 12 of *G. max*. Notably, the dCAPS140555 marker (absolutely linked to *CcRpp1* gene in *C. cajan*) is 6 kb from a predicted CC-NB-LRR gene in *G. max* (*Glyma12g01420*) and situated in a ~150-kb interval delineated by the dCAPS52491 and SSR2152 marker sequences.

In addition, two scaffolds (134711 and 132275) from the *C. cajan* reference genome sequence variety "Asha" were identified that are syntenic to the region comprising the first 2 Mb of soybean chromosome 12. We were able to link the two scaffolds by PCR, which resulted in a 300-bp product. We aligned the scaffolds as a new reference sequence and showed that the region delineated by dCAPS52491 and SSR2152 consisted of a ~154-kb interval which contained four predicted NB-LRR genes (**Supplementary Fig. 1a** and **Supplementary Table 3**). We used an NB-LRR-guided BAC library from the reference *C. cajan* variety "Asha ICPL 87119" to confirm if NB-LRR genes linked to marker dCAPS140555 are also present in *C. cajan*[38,39]. We identified a BAC clone of 78.7 kb that contained dCAPS140555 (Cc13A15, ENA:LN879382). DNA was prepared from this clone using the Qiagen Large Construct Kit and sheared to obtain average fragment sizes of 5 kb and sequenced using Sanger sequencing. The reads were subsequently assembled using Phred, Phrap and Conseddata analyses using standard parameters[40–43]. Sequence annotation of this BAC revealed the presence of four highly homologous NB-LRR gene paralogs: three full-length genes and one truncated gene. We therefore initiated high-resolution mapping of the 2,282 gametes and identified on gain- and four loss-of-function recombinants delineated by dCAPS52491 and one resistant recombinant delineated by SSR2152.

**BAC sequencing and annotation.** A BAC library of *C. cajan* accession G119-99 was ordered from BioS&T (Montreal, Canada) and screened with PCR amplicons comprising sequences surrounding markers dCAPS140555 and SSR2152. Two positive BAC clones (3F and 6G) were identified that together span the entire interval between dCAPS52491 and SSR2152 (**Fig. 2a** and **Supplementary Fig. 2**; raw sequencing data is deposited at ENA, study accession PRJEB9933). Several colonies containing BAC 3F (145 kb) and BAC 6G (130 kb) were tested for integrity by DNA fingerprinting using the restriction enzyme HindIII (New England BioLabs). In addition, the BAC sizes of these clones were verified by pulsed-field gel electrophoresis of NotI-digested DNA. Single-colony clones from 3F and 6G were sequenced with PacBio and Illumina MiSeq (The Genome Analysis Centre, Norwich, UK). The two BAC sequences were assembled into one large contig of 205,344 bp using the HGAP pipeline[44] with the PacBio data. Subsequently, we validated assembled sequences by mapping MiSeq data to the assemblies using BWA[45] (version 0.7.10; BWA-MEM, default parameters) and manually inspecting each base for consistency between assembly and mapped Illumina reads. Tablet[46] (version 1.14.04.10) was used for mapping visualization. Four complete NB-LRR candidate gene sequences were identified in this contig; BAC 3F carries three of the NB-LRR gene sequences, NB-1, -2 and -3 (Seq ID nos: 3, 5 and 7, respectively), and BAC 6G carries four NB-LRR gene sequences, NB-LRR-1–4, (Submitted at the ENA: NB-1; LN879378; NB-2, LN879379, NB-3, LN879380; NB-4, LN879381).

**Visualization of the RNA-Seq reads aligned to the *CcRpp1* locus.** Expression analysis of the four NB-LRR paralogs was investigated by aligning the non-challenged resistant parent G119-99 RNA-Seq reads against the *CcRpp1* locus sequence of 205,344 bp in Geneious[47] (version 8.1.2; http://www.geneious.com/) using the Align/assemble -> Map to reference default parameters (**Supplementary Fig. 3**). The transcriptome raw data are deposited at ENA, study accession PRJEB9933.

**Pathogen assays.** Pathogen assays were performed as described previously[48]. Briefly, spores were heat-shocked at 40 °C for 5 min and subsequently suspended in an aqueous solution of 0.01% Tween 20, and mixed thoroughly; the spore concentration was then adjusted to $2 \times 10^4$ sp/ml with a hemocytometer. Plants were spray-inoculated with the urediniospore suspension, incubated at 100% relative humidity in the dark for 24–36 h and then transferred to the greenhouse (*C. cajan*) or a growth chamber (soybean) set at 22 °C, 70% relative humidity, 16 h photoperiod. Detached $T_0$ leaves were spray-inoculated with $1 \times 10^5$ spores/ml of *P. pachyrhizi* and placed in a large Petri dish atop moist filter paper to maintain humidity. Inoculated material from control source plants and $T_0$ transgenic plants was incubated in a reach-in growth chamber at 22 °C and scored for disease symptoms 12 d after inoculation.

**Evaluation of resistance in *C. cajan*.** Plants were inoculated as previously described and rated at 14 and 21 d after inoculation using the following disease rating: 0 = no lesions observed (immune), 1 = immune with flecks, 2 = reddish-brown (RB) lesions without sporulation, 3 = RB lesions with

sporulation and 4 = tan lesions with abundant sporulation. A dissecting microscope at 50× magnification was used to observe sporulation of uredinia in both the RB and tan lesions.

**T₁ transgenic testing.** Seeds from the first segregating generation (T₁) of events Soy3194.5.1 and Soy3194.7.1 were planted and maintained under growth chamber conditions for 17 d until the Vc growth stage (unrolled unifoliolate leaves). The plantlets were sampled for qPCR to determine the transgene copy number and inoculated with a suspension of *P. pachyrhizi* spores. The inoculation was performed as described above with urediniospores collected from a susceptible variety and stored at −80 °C. New growth was excised regularly to keep the unifoliate leaves from senescing for the duration of the experiment.

To assess the effect of *CcRpp1*, the plants were scored qualitatively as immune (I; no lesions), partial resistance (PR; red-brown (RB), low sporulating lesions) and susceptible (S; tan, highly sporulating lesions) and quantitatively, by excising and scanning leaves followed by determination of lesion counts using ImagePro7 software (**Supplementary Table 4**). Most null samples were scored 15 d after inoculation, whereas the heterozygous and homozygous plants were scored 29 d after inoculation. To determine the effect of the gene, we compared the transgenic plants to the null plants from the same event; a separate analysis was performed for each zygosity level.

**Generation of constructs for *CcRpp1* for heterologous expression.** A 2,775-bp SfiI fragment containing the *CcRpp1* coding region was ligated at the 5′ end to 1,948 bp of soybean polyubiquitin (SUBI-1) regulatory sequence that is found immediately 5′ to the SUBI-1 start codon. The 3′ end of *CcRpp1* was ligated to 890 bp of *Arabidopsis* polyubiquitin 10 terminator sequence that is found immediately distal to the At UBQ10 stop codon[49]. This vector also contains FRT1/FRT87 sequences flanking the expression cassette to mediate site-specific integration. The FRT1 site in the transformation vector is 5′ to a sulfonylurea-resistant soybean acetolactate synthase (ALS) coding sequence, whereas a FRT1 site previously introduced to the genome of the target soy line is 3′ to a strong constitutive promoter, such that a recombination event between the FRT1 sites will link the promoter and the ALS coding sequence, generating expression of ALS, which allows transgenic recombination events to be selected for with sulfonylurea herbicides[50].

**Biolistic transformation of soybean.** Transgenic soybean (*G. max*) lines were generated by delivery of DNA into embryogenic callus cultures using particle gun bombardment (US patent no. 4,945,050) with a BIORAD Biolistic PDS1000/He instrument. Site-specific integration of the *GmUbi-CcRpp1* construct into the soybean genome was achieved by recombinase-mediated cassette exchange (RMCE)[50]. Transgenic events were identified after selection on 100-p.p.b. chlorsulfuron and somatic embryos regenerated to produce T₀ seedlings that were advanced for T₁ seed production. Confirmation and characterization of RMCE events was performed using qPCR[50].

**Transgene expression analysis.** Total RNA was prepared from null, heterozygous and homozygous plants (4–5 plants each) of the Soy3194.5.1 and Soy3194.7.1 events using an RNeasy Plant Mini Kit (Qiagen). Isolated total RNA (1 µg) was DNase-treated and cDNA synthesized using the QuantiTect Reverse Transcription Kit (Qiagen). Absolute quantification of *CcRpp1* transcript levels was determined by TaqMan qPCR (TaqMan Gene Expression Master Mix; Applied Biosystems) on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems) using the primers and probe indicated in **Supplementary Table 3**. Three technical replicates of each sample were run. QuantStudio 6 and 7 Flex Software was employed for analysis. *CcRpp1* transcript levels in transgenic plants are expressed relative to those of a *G. max* zinc-metallopeptidase gene (XM_003521127) that was used as an internal control for expression studies.

31. Meyers, B.C., Kozik, A., Griego, A., Kuang, H. & Michelmore, R.W. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809–834 (2003).
32. Ameline-Torregrosa, C. *et al.* Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* **146**, 5–21 (2008).
33. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
34. Simpson, J.T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
35. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
36. Neff, M.M., Turk, E. & Kalishman, M. Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet.* **18**, 613–615 (2002).
37. Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183 (2010).
38. Varshney, R.K. *et al.* Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* **30**, 83–89 (2012).
39. Varshney, R.K. *et al.* Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Mol. Breed.* **26**, 393–408 (2010).
40. Gordon, D. Viewing and editing assembled sequences using consed. *Curr. Protoc. Bioinformatics* **2**, 11.2 (2004).
41. Gordon, D., Desmarais, C. & Green, P. Automated finishing with autofinish. *Genome Res.* **11**, 614–625 (2001).
42. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
43. Koren, S. *et al.* Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012).
44. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
46. Milne, I. *et al.* Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14**, 193–202 (2013).
47. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
48. Langenbach, C., Campe, R., Schaffrath, U., Goellner, K. & Conrath, U. UDP-glucosyltransferase UGT84A2/BRT1 is required for *Arabidopsis* nonhost resistance to the Asian soybean rust pathogen *Phakopsora pachyrhizi*. *New Phytol.* **198**, 536–545 (2013).
49. Xia, B.S. *et al.* Nucleotide sequence of a soybean (*Glycine max* L. Merr.) ubiquitin gene. *Plant Physiol.* **104**, 805–806 (1994).
50. Li, Z. *et al.* Site-specific integration of transgenes in soybean via recombinase-mediated DNA cassette exchange. *Plant Physiol.* **151**, 1087–1095 (2009).