

# Genome-Wide Association Study of Brown Stem Rot Resistance in Soybean across Multiple Populations

Keith Rincker, Alexander E. Lipka, and Brian W. Diers\*

## Abstract

Genetic resistance to brown stem rot (BSR) of soybean [*Glycine max* (L.) Merr.], caused by *Cadophora gregata* (Allington & D.W. Chamb.) T.C. Harr. & McNew, has been identified and mapped with biparental populations. Although nearly 400 accessions have been identified with BSR resistance, this trait has been mapped in only 12 sources, and just two, PI84946-2 and PI88788, have been used to develop BSR resistant cultivars. Thus, there is a serious need to improve our knowledge of the genetic basis of BSR resistance in soybean so that resistance genes in cultivars can be diversified and markers close to resistance genes can be identified and used in marker-assisted selection (MAS). To this end, we conducted a genome-wide association study (GWAS) to identify novel genomic loci associated with BSR resistance and to gain further insight into a previously reported chromosome 16 region containing BSR resistance (*Rbs*) genes. A total of 52,041 single-nucleotide polymorphisms (SNPs) were tested for association with BSR in a set of 4735 accessions from four diversity panels evaluated for resistance from 1989 to 2003. Using a unified mixed linear model and stepwise model selection, we refined the signals within the *Rbs* interval on chromosome 16 by finding associations that explain a substantial proportion of the total variation of BSR resistance. In combination with significant GWAS signals found elsewhere in the genome, our study will aid efforts to improve BSR resistance by providing new targets for MAS.

## Core Ideas

- GWAS narrowed chromosome 16 interval containing a BSR resistance QTL to 0.3 Mb
- SNPs identified provide new targets for marker-assisted selection of BSR resistance
- GWAS of public BSR resistance data failed to identify three separate resistance genes

**B**ROWN STEM ROT OF SOYBEAN, caused by the soil-borne fungus *C. gregata*, affects soybean production in the northern United States, Canada, Brazil, and also has a minor impact in China (Wrather et al., 2010). Yield losses of up to 38% have been reported (Gray, 1972), and annual damage to the US soybean crop was estimated to average 422,000 Mg (15.5 million bushels) from 2006 to 2009 (Koenning and Wrather, 2010). Management of this disease is best achieved with host genetic resistance (Bachman et al., 1997b). Multiple screens of the USDA Soybean Germplasm Collection for resistance have identified almost 400 accessions with resistance similar to current resistant cultivars (Bachman et al., 1997a; Bachman and Nickell, 2000a; Chamberlain and Bernard, 1968; Hughes et al., 2004; Nelson et al., 1989; Patzoldt et al., 2003). However, introgression of disease resistance into cultivars has only been achieved with the accessions PI 84946-2 and PI 88788.

K. Rincker, A.E. Lipka, and B.W. Diers, Dep. of Crop Sciences, Univ. of Illinois, Urbana, IL 61801. Received 5 Aug. 2015. Accepted 15 Dec. 2015. \*Corresponding author (bdiers@illinois.edu).

Published in Plant Genome  
Volume 9. doi: 10.3835/plantgenome2015.08.0064

© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Abbreviations:** BIC, Bayesian information criterion; BSR, brown stem rot; FDR, false discovery rate; GRIN, USDA Germplasm Resources Information Network; GWAS, genome-wide association study; LD, linkage disequilibrium; MAS, marker-assisted selection; MG, maturity group; MLM, mixed linear model; MLM, multilocus mixed linear model; NBS-LRR, nucleotide binding site-leucine rich repeat; QTL, quantitative trait loci; SNP, single-nucleotide polymorphism.

Three genes conferring resistance to BSR have been identified through genetic studies: *Rbs1*, *Rbs2*, and *Rbs3* (Hanson et al., 1988; Willmont and Nickell, 1989). These original studies showed the genes were unlinked; however, subsequent genetic mapping placed all three resistance genes onto the same region on chromosome 16 (linkage group J) near the simple-sequence repeat markers Satt215 and Satt431 (Bachman et al., 2001; Lewers et al., 1999; Patzoldt et al., 2005b). Additional BSR resistance quantitative trait loci (QTL) in experimental crosses derived from eight Chinese soybean accessions have been mapped to the same region on chromosome 16 near Satt431, Satt547, or Satt244 (Patzoldt et al., 2005a; Perez et al., 2010). The efficiency of MAS for this locus has been limited because it was mapped to a large region of nearly 10.2 Mb of the Williams 82 genome. Thus, there has been a significant need to study this region and elucidate the key locus or loci underlying this QTL. Rincker et al. (2016) recently fine mapped *Rbs* genes from multiple sources and found intervals containing *Rbs1*, *Rbs2*, and *Rbs3* to be located between 32.7 and 33.1 Mb on chromosome 16. All intervals identified in this study included the same 0.04 Mb interval, suggesting the possibility that only one resistance locus is present among the sources tested.

Although additional loci have been proposed as interacting with the *Rbs* locus or acting as modifiers (Bachman and Nickell, 2000b; Sebastian et al., 1985; Waller et al., 1991), their identification using linkage mapping approaches has been difficult. Some common drawbacks of linkage mapping include the time required to develop populations, limitations in genomic resolution arising from high linkage disequilibrium (LD), and limitations in statistical power arising from small population sizes, all of which usually lead to low precision of QTL mapping (Cardon and Bell, 2001). The GWAS offers an alternative to linkage mapping for identifying QTL. Using plant accessions that capture many generations of recombination, the GWAS typically has greater resolution for locating QTL relative to linkage mapping (Cardon and Bell, 2001). Furthermore, these accessions typically include more segregating loci than what is found in a biparental cross used in a linkage mapping study.

Recent advances in genotyping and phenotyping technologies are making it possible to dissect the genetic underpinnings of important soybean traits and disease resistance with unprecedented resolution. For example, the genotyping of 19,652 accessions in the USDA Soybean Germplasm Collection with an Illumina Infinium chip, SoySNP50K, containing 52,041 SNPs (Song et al., 2013) is now complete. Phenotypic data on many of these accessions for BSR resistance and other traits are also available through the USDA Germplasm Resources Information Network (GRIN) and from previous accession screening studies (Nelson et al., 1989; Bachman et al., 1997a; Bachman and Nickell, 2000a; Patzoldt et al., 2003).

Genome-wide association studies have been conducted in soybean to identify QTL controlling a wide variety of soybean traits including disease resistance.

Hwang et al. (2014) and Vaughn et al. (2014) each used the SoySNP50K array to identify QTL associated with soybean seed protein and oil concentration. In addition, Vaughn et al. (2014) relied on historical phenotypic data available in GRIN. Both studies identified previously reported QTL and refined their genomic locations. Wen et al. (2014) also used the SoySNP50K array to study sudden death syndrome [*Fusarium virguliforme* (syn. *F. solani* f. sp. *glycines*)] resistance, which resulted in the identification of novel loci and further refinement of the genomic regions already known to contain QTL. In addition, associations with resistance to sclerotinia stem rot of soybean [*Sclerotinia sclerotiorum* (Lib.) de Bary] were identified by Bastien et al. (2014). These studies demonstrate the adaptability of GWAS to soybean traits including disease resistance, the use of GRIN phenotypic data, and the use of the SoySNP50K array as a source of genotypic data. To our knowledge, GWAS of soybean BSR resistance has not been reported.

A more precise location of previously identified resistance genes underlying the QTL on chromosome 16 (Hanson et al., 1988; Willmont and Nickell, 1989) and the elucidation of additional novel loci should increase the efficiency and effectiveness of MAS and aid in the determination of whether this QTL consists of separate genes or a common resistance gene. Genome-wide association studies with diverse soybean accessions and a dense set of SNP markers should contribute to this endeavor by refining QTL to a narrow genomic interval. Therefore, the objectives of this study were to use GWAS to (i) rigorously study the interval containing the known BSR resistance QTL on chromosome 16 and (ii) map novel resistance QTL in a diverse set of accessions previously screened for BSR resistance.

## Materials and Methods

### Germplasm

Phenotypic data for BSR resistance of 4744 soybean accessions were obtained for four diversity panels designated N-1989, B-1997, B-2000, and P-2003. N-1989 included binary data (i.e., resistant vs. susceptible) based on a combination of foliar and stem observations for 2773 accessions ranging from maturity group (MG) 000 through IV and originating from primarily Asia (84%) and Europe (14%) (Nelson et al., 1989). The remaining accessions are from North America, North Africa, Australia, and unknown origins. Eighteen accessions were characterized by the USDA GRIN database as developed cultivars, including eight commercial cultivars used as susceptible checks. Because of a low number of resistant accessions, the four resistant commercial check cultivars that derived their resistance from PI84946-2 were excluded to prevent any bias from introgression of genomic content from a single source of BSR resistance. From 1981 to 1986, susceptible accessions from N-1989 were culled from nonreplicated field tests conducted in four Midwest states (Iowa, Illinois, Minnesota, and Wisconsin) and putative resistant

**Table 1. Characteristics of association panels analyzed with genome-wide association study and stepwise procedures.**

Panel	Data type	Symptoms measured	Accessions	SNP† markers	Box-Cox lambda	BSR Score‡		
						Mean	SD§	$h^2\parallel$
N-1989	Binary	Foliar and stem	2773	33,240	na	na	na	na#
B-1997	Proportion 0–1	Foliar	540	33,486	log	0.09	0.15	0.49
B-1997	Proportion 0–1	Stem	540	33,486	1	0.38	0.20	0.61
B-2000	Proportion 0–1	Foliar	825	32,150	0.25	0.33	0.29	0.93
P-2003	Proportion 0–1	Stem	606	29,815	0.75	0.39	0.25	0.68

† SNP, single-nucleotide polymorphism.

‡ BSR, brown stem rot; BSR score is measured by the percentage of plants or nodes of a plant showing foliar and stem symptoms.

§ SD, standard deviation.

¶  $h^2$ , narrow-sense heritability;  $h^2$  estimated within GAPIT by the use of SNP markers and genetic relatedness among individuals# Calculation of  $h^2$  by GAPIT is not appropriate for this binary trait

lines were evaluated in replicated field and greenhouse tests. Phenotypic data were obtained from GRIN (<https://npgsweb.ars-grin.gov/gringlobal/search.aspx>). B-1997 included quantitative data of separate foliar and stem measurements from 540 accessions originating from central China and ranging in MG II through IV (Bachman et al., 1997a). Evaluations were conducted in 1994 in a non-replicated field test at Urbana, IL. Phenotypic data were obtained from Michael Bachman (personal communication, 2014). Foliar data were expressed as the percentage of plants exhibiting symptoms, while stem data were measured as the proportion of nodes exhibiting brown pith and averaged over four plants. B-2000 included quantitative foliar measurement data from 825 accessions originating from central and southern China and ranging in MG IV through VIII (Bachman and Nickell, 2000a). Greenhouse data from evaluations with the Oh2 isolate of *C. gregata* were included in this analysis because no selection for BSR resistance was conducted before evaluating the 825 accessions. The phenotypic data for B-2000 were expressed as the percentage of the nodes exhibiting foliar symptoms and averaged over two to five plants tested within each replication. A total of 281 accessions with symptoms failing to progress past the first trifoliate in two plants were included in a second replication. For these accessions, phenotypic data were averaged across the two replicates and obtained from Bachman (1999). P-2003 included quantitative stem data from 606 accessions originating from south-central China and ranging in MG I through IX (Patzoldt et al., 2003). Evaluation of these accessions was conducted in a greenhouse with one replication using the Oh2 *C. gregata* isolate. Phenotypic data were obtained from GRIN and expressed as a percentage of the nodes exhibiting brown pith and averaged over five plants. Methods to evaluate resistance in greenhouse tests among the N-1989, B-2000, and P-2003 panels were similar with the exception that Nelson et al. (1989) inoculated with a different isolate. Separate analyses were conducted on each of the four diversity panels. Moreover, two separate analyses were conducted for stem and foliar BSR data in the B-1997 panel. The Box-Cox (Box and Cox, 1964) procedure was

conducted in SAS v9.3 (SAS Institute, 2011) to find the optimal transformations to correct for non-normality of the error terms and unequal variance of BSR resistance in the B-1997, B-2000, and P-2003 panels.

### Genome-Wide Association Study

Genotypic data of soybean accessions from the SoyS-NP50K BeadChip (Song et al., 2013) were obtained from Perry Cregan, USDA-ARS. Single-nucleotide polymorphisms with more than 10% missing data were discarded before the GWAS and accessions were removed if there were more than 10% heterozygous genotypes. After removal of SNPs with minor allele frequency <5%, between 29,815 and 33,486 SNPs (depending on the panel) were used in the GWAS (Table 1). Missing SNP genotypes were imputed with the major allele. To conduct GWAS, a unified mixed linear model (MLM; Yu et al., 2006) with population parameters previously determined (Zhang et al., 2010) was implemented in the GAPIT R package (Lipka et al., 2012). To control for population structure and familial relatedness, the unified MLM included principal components (Price et al., 2006) and a kinship matrix computed by the VanRaden method (VanRaden, 2008). For each GWAS scan, the Bayesian information criterion (BIC; Schwarz 1978) was used to determine the optimal number of principal components to include in the GWAS model. The variance component estimates from this model were used to estimate narrow-sense heritability; specifically the genetic variance component estimate was divided by the sum of the genetic and residual variance component estimates. The phenotypic variation explained by the model was assessed with a likelihood-ratio-based  $R^2$  statistic, denoted  $R^2_{LR}$  (Sun et al., 2010). The Benjamini and Hochberg (1995) procedure was implemented to the GWAS results of each panel to control for the multiple testing problem at a false discovery rate (FDR) of 10%. This rate was chosen because phenotypic data of the panels were based on limited replication and trends among panels would be identified. To further elucidate the genomic underpinnings of the peak GWAS signals,

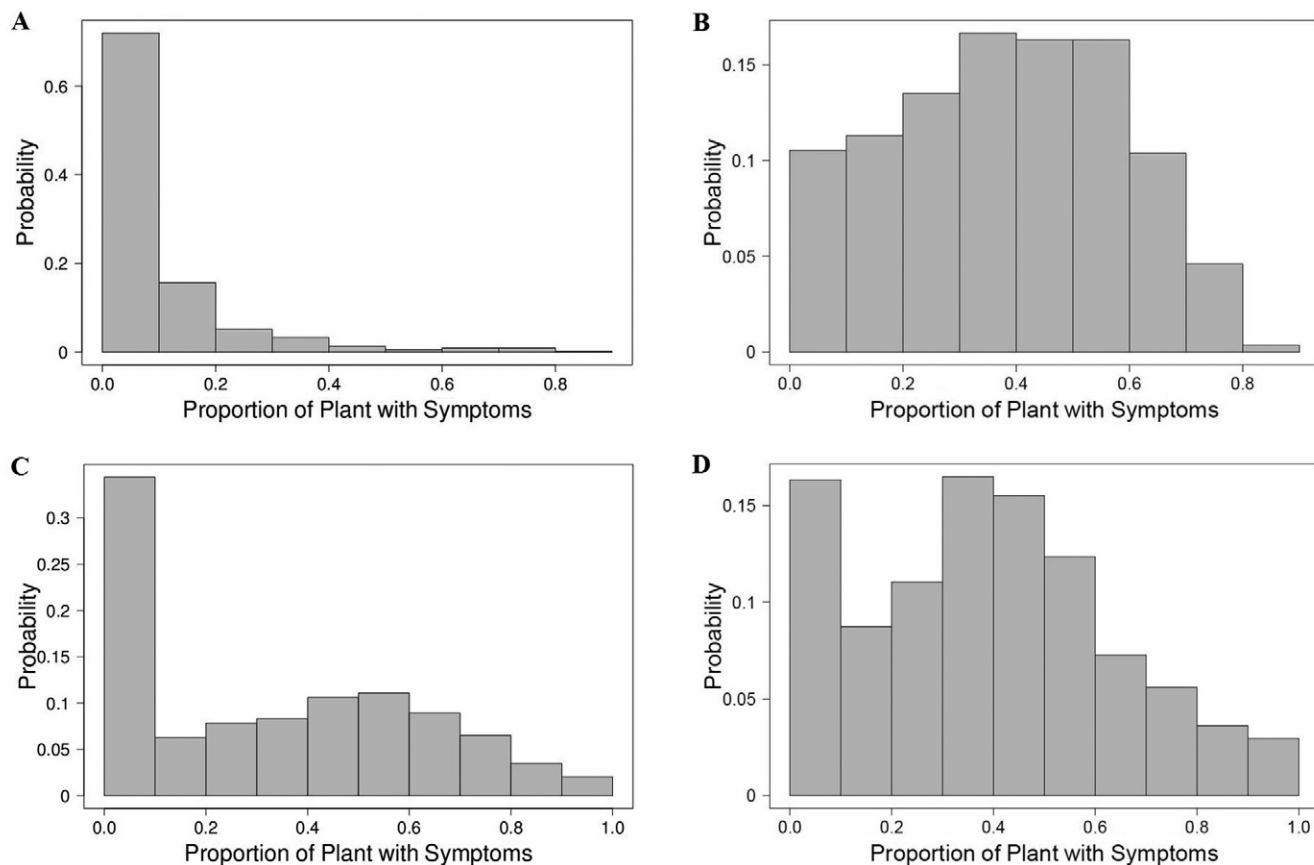


Fig. 1. Histograms of phenotypic data in genome-wide association studies are shown in the following panels: (A) foliar symptoms from the B-1997 study, (B) stem symptoms from the B-1997 study, (C) foliar symptoms from the B-2000 study, and (D) stem symptoms from the P-2003 study.

a multilocus mixed linear model (MLMM) (Segura et al., 2012) in R software was implemented using all SNPs across the genome. The final MLMM was selected using the extended BIC (Chen and Chen, 2008). All SNPs that were identified in the final MLMM were then entered as covariates in a second GWAS.

Because the unified MLM was developed to analyze quantitative traits, some of its statistical assumptions (e.g., normality of error terms) are violated when fitted to the binary BSR data in the N-1989 panel. Thus, it was imperative that the GWAS signals identified using the unified MLM were confirmed using an alternate approach specifically designed to analyze binary data, namely the logistic regression model (Agresti 2013). All statistically significant markers from the original GWAS scan of the N-1989 panel were considered for inclusion into a logistic regression model using SAS PROC LOGISTIC (SAS Institute, 2011), where the optimal model was determined using the BIC (Schwarz 1978) in a stepwise model selection procedure. Unlike the unified MLM, these logistic regression models did not include individuals as random effects. Phenotypic variation explained by the model was assessed with a likelihood-ratio-based  $R^2$  statistic, denoted  $R^2_{\text{McF}}$  (McFadden, 1974).

## Linkage Disequilibrium Analysis

Linkage disequilibrium was assessed by computing the squared allele frequency correlations ( $r^2$ ) between unimputed marker pairs using the site-by-all option in TASSEL version 5.0 (Bradbury et al., 2007). Only SNPs with a minor allele frequency >5% and <10% missing were used to estimate LD.

## Results

### Phenotypic Data

The N-1989 panel contains 2738 susceptible and 35 resistant accessions. The remaining panels measured either foliar or stem symptoms, which were quantified as the proportion of the plant expressing symptoms. For these remaining panels, the proportion of individual plants exhibiting BSR symptoms averaged between 0.33 and 0.39 (Table 1). Interestingly, the B-1997 panel had, on average, threefold more plants showing no signs of foliar BSR development than stem BSR development in the B-1997 panel and BSR development in the B-2000 and P-2003 panels (Fig. 1). The narrow-sense heritabilities of BSR resistance in the B-1997, B-2000, and P-2003 panels ranged from 0.49 to 0.93 (Table 1), suggesting that genetic variability might play a substantial role in BSR resistance.



## Genome-Wide Association Study and Stepwise Procedures

Consistent with previous studies, our GWAS detected significant associations between BSR resistance and marker loci on chromosome 16. Using the unified MLM, significant SNPs were identified on chromosomes 2, 16, and 17 in the N-1989 panel (Supplemental Table S1; Supplemental Fig. S1). On chromosome 16, statistically significant GWAS signals were found in the vicinity of the previously reported BSR QTL (Fig. 2). In contrast to the N-1989 panel, no statistically significant associations were identified in either the B-1997 foliar or stem panel (Supplemental Fig. S1). The GWAS of the B-2000 panel found significant peaks on chromosome 16, 5, and 8 (Fig. 3; Supplemental Table S1; Supplemental Fig. S1). The most significant SNP ( $P$ -value  $4.79 \times 10^{-35}$ ) was located 21 kb from the interval identified by Rincker et al. (2016) as potentially containing *Rbs1*, *Rbs2*, and *Rbs3* and explained 17% of the variation for symptoms (Supplemental Table S1; Fig. 3). In agreement with the N-1989 and B-2000 panel GWAS results, the peak signals from the GWAS of the P-2003 panel were in the vicinity of the putative BSR QTL on Chromosome 16 (Supplemental Table S1; Fig. 4; Supplemental Fig. S1). Moreover, the marker with the strongest association with BSR resistance ( $P$ -value  $1.59 \times 10^{-7}$ ) was the same as that identified in the GWAS of the B-2000 panel.

Within the N-1989 panel, the MLM analysis identified no significant associations; however, stepwise logistic regression of significantly associated SNPs identified in the GWAS revealed that three SNPs on three different chromosomes (i.e., chromosomes 2, 16, and 17) best described the binary resistant and susceptible phenotypes (Table 2). When the GWAS was reconducted with these three SNPs included as covariates in the model, no statistically significant signals were detected at 10% FDR (Fig. 2), suggesting that these three SNPs sufficiently account for BSR resistance variability in this panel. Within the B-1997 panel, no significant associations were found using the MLM, which is consistent with the original GWAS scan. The optimum model obtained from the MLM analysis for the B-2000 panel included two SNPs on chromosome 16 (Table 2). When these two SNPs were included as covariates in the GWAS model, an additional four SNPs (three of which are proximal to the putative BSR QTL on chromosome 16 and another on chromosome 8) were significant at 10% FDR (Fig. 3). This could indicate that more variability exists than can be explained by the two MLM-identified SNPs. Finally, the optimum model obtained with MLM for the P-2003 panel identified two SNPs on chromosome 16 (Table 2) explaining 10% of the variation. Subsequently, the GWAS rescan that included these two SNPs as covariates in the model identified no statistically significant signals at 10% FDR (Fig. 4), which suggests that these two SNPs sufficiently account for BSR resistance variability in this panel.

## Discussion and Conclusions

Before this study, genetic mapping of BSR resistance had been conducted primarily through the use of biparental populations. Given recent advances in genotyping and the diversity present in soybean accessions, GWAS is now a viable approach for identifying new genetic resistance loci and pinpointing the location of *Rbs* genes on chromosome 16. We conducted GWAS with a 50K SNP array in four panels of diverse soybean germplasm that include all publically available BSR resistance data. While the scope of this study is limited by the pathological screening and inoculation methods, this work constitutes one of the most extensive analyses of natural variation for BSR resistance to date.

### Linkage Disequilibrium and Population Structure

The genomic region on chromosome 16 from 32 to 34 Mb that contains significant SNPs in two of the MLM analyses is gene rich with rapid LD decay. The N-1989 panel is composed of more accessions ranging from wider geographic origins, which we would expect to have less LD. However, the greater LD present in the N-1989 panel than other panels (Fig. 2, 3, 4) suggests this group of accessions could have been subject to greater selection pressure, founding events, or the presence of population structure compared with the germplasm that comprised the other three association panels. Hwang et al. (2014) also used the SoySNP50K array and reported an approximate coverage of one SNP every 17 kb in euchromatic regions and 100 kb in heterochromatic regions. This coverage is sufficient for SNPs to be in LD with most QTL; however, a gap of 700 kb exists between the two significant GWAS peaks on chromosome 16 (Fig. 2, 3, 4) where none of the SNPs tested in GWAS were called.

For each of the four panels, the BIC was used to determine that none of the principal components needed to be in the optimum GWAS model. Given that the ancestry of Asian soybean accessions can be traced back to a domestication event that reduced diversity from a wild relative (*Glycine soja* Siebold & Zucc.) that already had low sequence diversity (Hyten et al., 2006), this is not particularly surprising. However, two explanations for this result are that the kinship matrix is explaining both population structure and familial relatedness, or BSR resistance is not associated with population structure. To explore this possibility further, the correlation of BSR resistance and principal components was calculated, and no meaningful correlations were found (data not shown).

### Genome-Wide Association Study and Stepwise Procedures

In three of the four tested panels, statistically significant GWAS signals were identified in the chromosome 16 region where three *Rbs* genes were previously mapped. Interestingly, one SNP in this region (ss715624573) had the strongest association with BSR resistance in two of the panels. Although this SNP was not identified in

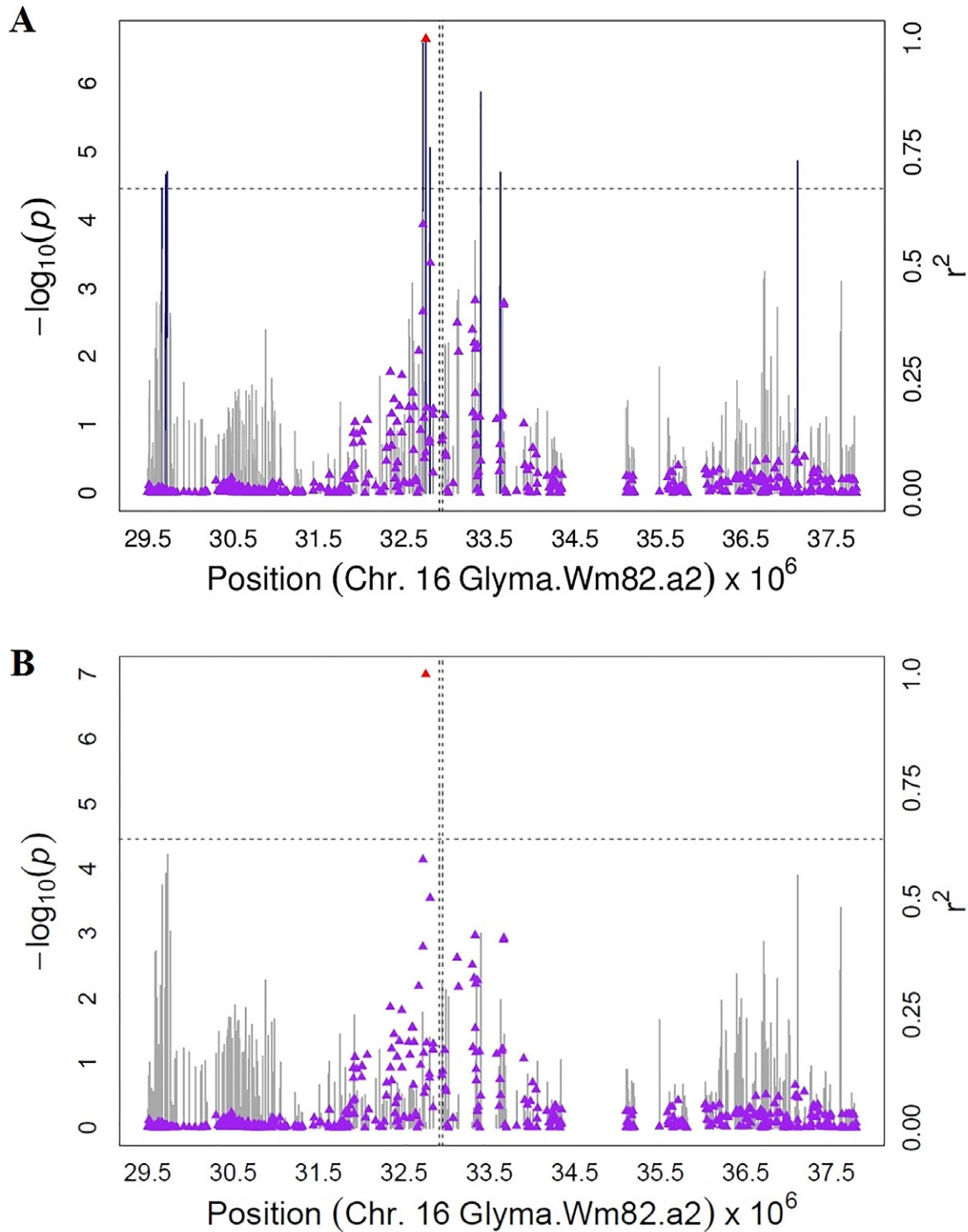


Fig. 2. Genome-wide association study (GWAS) results for brown stem rot (BSR) resistance of soybean in the N-1989 panel. (Panel A) Association results from a unified mixed model analysis of BSR and linkage disequilibrium (LD) estimates ( $r^2$ ) across a chromosome 16 region. Negative  $\log_{10}$ -transformed  $P$ -values (left, y-axis) from a GWAS for BSR and  $r^2$  values (right, y-axis) are plotted against physical position (Glyma.W82.a2) for a 8-Mb region on chromosome 16. The blue vertical lines are  $-\log_{10}$   $P$ -values for single-nucleotide polymorphisms (SNPs) that are statistically significant for BSR at 10% false discovery rate (FDR), whereas the gray vertical lines are  $-\log_{10}$   $P$ -values for SNPs that are nonsignificant at 10% FDR. Triangles are the  $r^2$  values of each SNP relative to the peak SNP (indicated in red) at 32,796,708 bp of chromosome 16. The black horizontal dashed line indicates the  $-\log_{10}$   $P$ -value of the least statistically significant SNP at 10% FDR. The black vertical dashed lines indicate the positions of BARCSOYSSR\_16\_1114 and BARCSOYSSR\_16\_1115 identified by Rincker et al. (2016), which fine map an interval containing an *Rbs* gene. (Panel B) Association results from a conditional unified mixed model analysis of BSR and LD estimates ( $r^2$ ) across a chromosome 16 region, as in (A). The three SNPs (ss715627222, ss715624549, and ss715582351) from the step-wise logistic regression model were included as covariates in the unified mixed model to control for the *Rbs* effect.

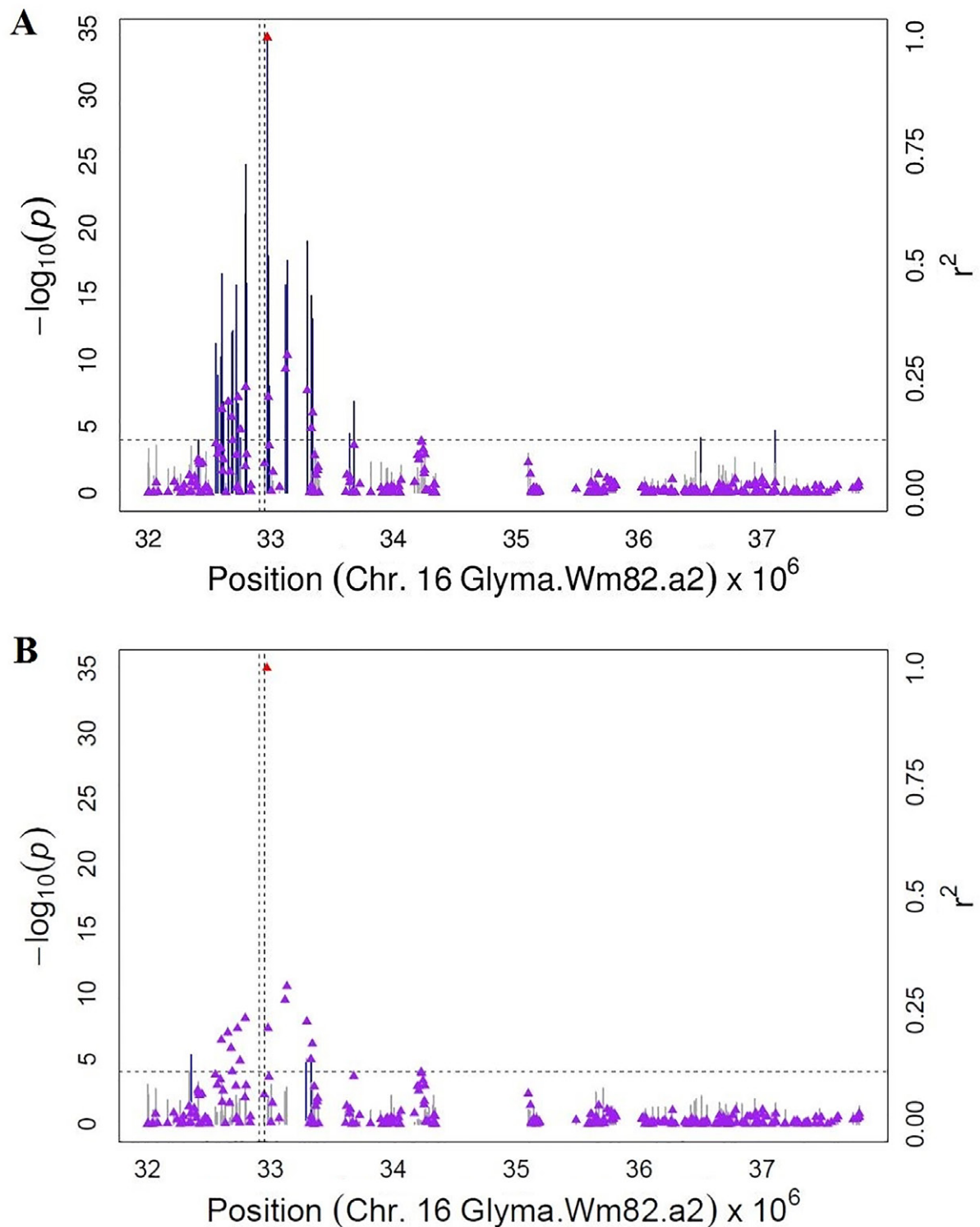


Fig. 3. Genome-wide association study (GWAS) results for brown stem rot (BSR) resistance of soybean in the B-2000 panel. (Panel A) Association results from a unified mixed model analysis of BSR and linkage disequilibrium (LD) estimates ( $r^2$ ) across a chromosome 16 region. Negative  $\log_{10}$ -transformed  $P$ -values (left, y-axis) from a GWAS for BSR and  $r^2$  values (right, y-axis) are plotted against physical position (Glyma.W82.a2) for a 6-Mb region on chromosome 16. The blue vertical lines are  $-\log_{10}$   $P$ -values for single-nucleotide polymorphisms (SNPs) that are statistically significant for BSR at 10% false discovery rate (FDR), whereas the gray vertical lines are  $-\log_{10}$   $P$ -values for SNPs that are nonsignificant at 10% FDR. Triangles are the  $r^2$  values of each SNP relative to the peak SNP (indicated in red) at 33,018,083 bp of chromosome 16. The black horizontal dashed line indicates the  $-\log_{10}$   $P$ -value of the least statistically significant SNP at 10% FDR. The black vertical dashed lines indicate the positions of BARCSOYSSR\_16\_1114 and BARCSOYSSR\_16\_1115 identified by Rincker et al. (2016), which fine map an interval containing an *Rbs* gene. (Panel B) Association results from a conditional unified mixed model analysis of BSR and LD estimates ( $r^2$ ) across a chromosome 16 region, as in (A). The two SNPs (ss715624557 and ss715624573) from the optimal multilocus mixed model were included as covariates in the unified mixed model to control for the *Rbs* effect.

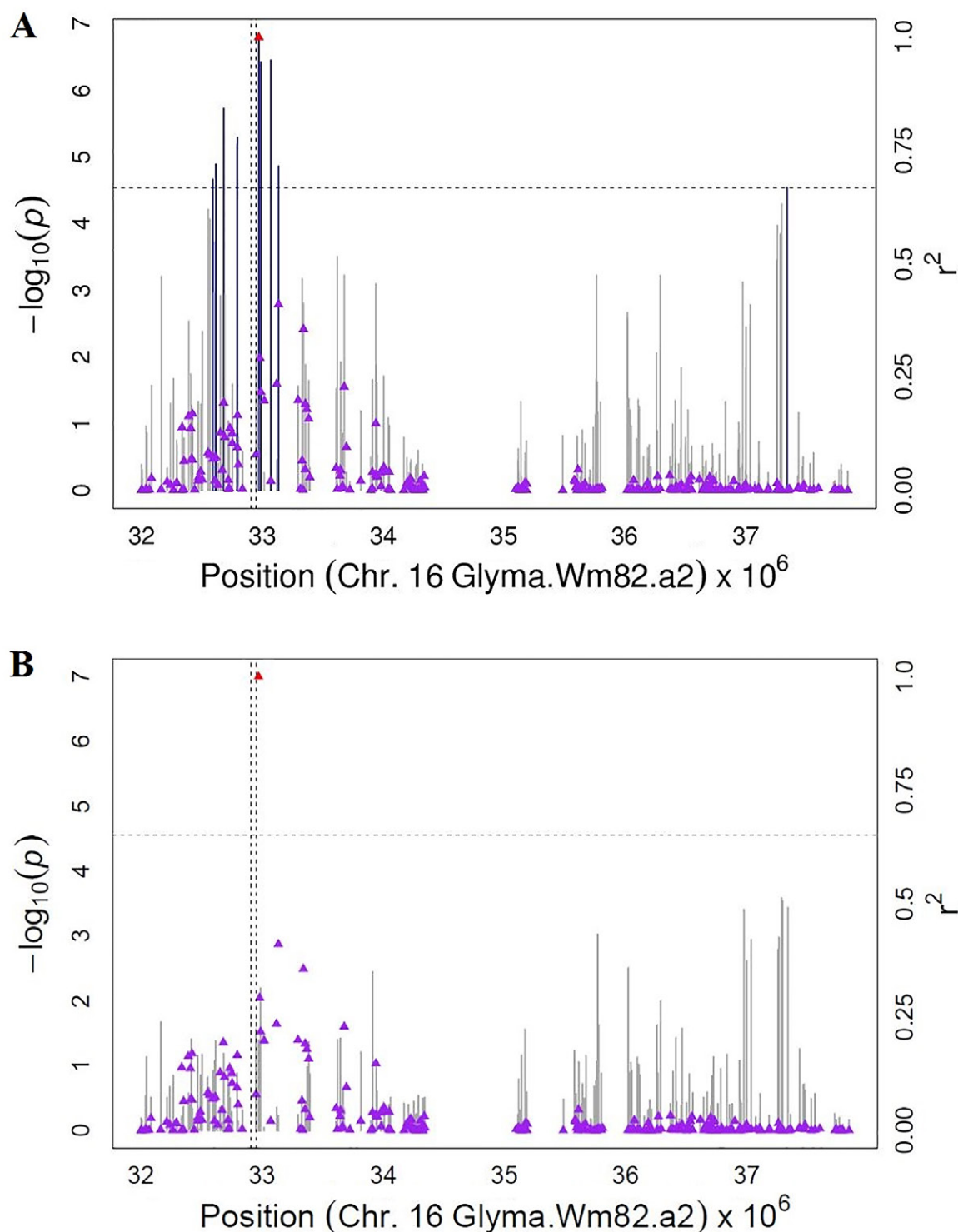


Fig. 4. Genome-wide association study (GWAS) results for brown stem rot (BSR) resistance of soybean in the P-2003 panel. (Panel A) Association results from a unified mixed model analysis of BSR and linkage disequilibrium (LD) estimates ( $r^2$ ) across a chromosome 16 region. Negative  $\log_{10}$ -transformed  $P$ -values (left, y axis) from a GWAS for BSR and  $r^2$  values (right, y axis) are plotted against physical position (Glyma.W82.a2) for a 6-Mb region on chromosome 16. The blue vertical lines are  $-\log_{10}$   $P$ -values for single-nucleotide polymorphisms (SNPs) that are statistically significant for BSR at 10% false discovery rate (FDR), whereas the gray vertical lines are  $-\log_{10}$   $P$ -values for SNPs that are nonsignificant at 10% FDR. Triangles are the  $r^2$  values of each SNP relative to the peak SNP (indicated in red) at 33,018,083 bp of chromosome 16. The black horizontal dashed line indicates the  $-\log_{10}$   $P$ -value of the least statistically significant SNP at 10% FDR. The black vertical dashed lines indicate the positions of BARCSOYSSR\_16\_1114 and BARCSOYSSR\_16\_1115 identified by Rincker et al. (2016) which fine map an interval containing an *Rbs* gene. (Panel B) Association results from a conditional unified mixed model analysis of BSR and LD estimates ( $r^2$ ) across a chromosome 16 region, as in Panel A. The two SNPs (ss715624583 and ss715624558) from the optimal multilocus mixed model were included as covariates in the unified mixed model to control for the *Rbs* effect.



**Table 2. Significant single-nucleotide polymorphism (SNPs) identified in stepwise procedures for the foliar or stem symptoms of brown stem rot.**

Stepwise logistic regression									
Panel	Symptoms	SNP added to model		Chromosome	Position, Glyma2.0 assembly	P-value	Bayesian information criterion (BIC)	$R^2_{MCF}$	$R^2_{MCF}$ with all SNPs identified in stepwise logistic regression†
N-1989	Foliar and stem	Gm17_37574384_T_C	ss715627222	17	37,284,864	$5.09 \times 10^{-11}$	368.181	0.07	0.21
N-1989	Foliar and stem	Gm16_32298597_T_G	ss715624549	16	32,796,708	$2.09 \times 10^{-8}$	339.544	0.10	
N-1989	Foliar and stem	Gm02_4208733_C_T	ss715582351	2	4260,493	$4.09 \times 10^{-5}$	330.345	0.06	
Multilocus mixed model									
Panel	Symptoms	SNP added to model		Chromosome	Position, Glyma2.0 assembly	P-value	extBIC§	$R^2_{LR}$ †	$R^2_{LR}$ with all SNPs identified in MLM§
B-2000	Foliar	Gm16_32340079_G_A	ss715624557	16	32,838,190	$1.04 \times 10^{-44}$	-767.80526	0.10	0.22
B-2000	Foliar	Gm16_32517934_C_T	ss715624573	16	33,018,083	$2.93 \times 10^{-15}$	-804.055686	0.17	
P-2003	Stem	Gm16_32617666_T_C	ss715624583	16	33,119,116	$2.09 \times 10^{-6}$	-19.7654342	0.04	0.10
P-2003	Stem	Gm16_32342381_C_T	ss715624558	16	32,840,492	$1.89 \times 10^{-2}$	-17.4209656	0.03	

† Defined as the likelihood-ratio based  $R^2$  statistic that measures the increase in phenotypic variation explained by adding the indicated SNP into the model.

‡ Defined as the likelihood-ratio based  $R^2$  statistic that measures the increase in phenotypic variation explained by adding the three SNPs identified in the “SNP added to model” column compared to the intercept-only model.

§ ExtBIC, extended Bayesian information criterion (Chen and Chen, 2008).

¶ Defined as the likelihood-ratio based  $R^2$  statistic that compares the variation explained by the final multilocus mixed models (MLMMs) fitted to (i) the foliar symptoms in the B-2000 panel and (ii) the stem symptoms in the P-2003 panel relative to the intercept-only model. For each of these panels, the respective SNPs included in the final MLMM are indicated in the “SNP added to model” column.

the N-1989 or B-1997 panels, it is only 0.22 Mb from the chromosome 16 SNP identified in N-1989 as having the peak signal. Based on the position of the significant markers in the MLMM analysis, our results have narrowed the region containing BSR resistance QTL on chromosome 16 to between 32.8 and 33.1 Mb based on the Glyma2.0 assembly (Table 2), which is a substantial refinement compared with the 10.2 Mb interval identified in Patzoldt et al. (2005b). These results also agree with fine mapping results of Rincker et al. (2016) that map the position of *Rbs1*, *Rbs2*, and *Rbs3* to between the positions of 32.7 and 33.1 Mb. We therefore expect these results to directly benefit MAS breeding efforts by providing a focused region on chromosome 16 from which to make selections for improved BSR resistance.

Although three *Rbs* genes have been previously mapped to the same region of chromosome 16 (Bachman et al., 2001; Lewers et al., 1999; Patzoldt et al., 2005b), the results of our study failed to confirm the presence of three distinct genes. It is possible that the three previously identified *Rbs* genes are in fact one gene and that environmental effects or errors in the original allelism tests could have resulted in the incorrect conclusion that different resistance genes were present in resistance sources. Although our study identified multiple significant SNPs located in a narrow interval, it is possible that a single QTL is present, and the different patterns of LD within the four panels result in unique sets of SNPs with significant associations (Platt et al., 2010). A second possibility is that the multiple significant SNPs identified in our study arise from multiple QTL within the region. Nucleotide binding site–leucine rich repeat (NBS-LRR) genes have been found to be

the largest class of disease resistance genes in flowering plants (Dangl and Jones, 2001), and many NBS-LRR genes have been identified in this defined region of the Williams 82 genome sequence. Furthermore, duplicate NBS-LRR genes have been found to confer disease resistance in plants (Dixon et al., 1996). In soybean, Kang et al. (2012) found that locations of disease resistance QTL were correlated with the number of NBS-LRR genes. In fact, resistance to multiple soybean pests such as *Phytophthora* root rot (*Rps1k*), soybean mosaic virus (*Rsv1* and *Rsv3*), and Asian soybean rust (*Rpp4*) have been mapped or cloned to NBS-LRR genes (Gao et al., 2005; Hayes et al., 2004; Meyer et al., 2009; Suh et al., 2011).

The identification of BSR resistance QTL other than at the chromosome 16 *Rbs* interval has been elusive. The variation explained in each of our panels (Table 2) remains low, which suggests environmental effects are large, or more loci are responsible for resistance, or both. The most promising region identified in this study is a location distal of the *Rbs* interval at the positions of 36.6 to 37.4 Mb on chromosome 16. Three of the four GWAS panels contained significant SNPs from this location. However, no significant associations were found in this region when reconducting the GWAS with SNPs identified in MLMM as covariates. Lewers et al. (1999) identified two separate QTL on chromosome 16. Based on previous linkage maps, the restriction fragment-length polymorphic markers used in their study are not located near the positions found in our study. It is possible that the markers used by Lewers et al. (1999) might have been located near the significant GWAS signals identified in the N-1989 panel at 29.6 Mb. Regardless, neither of these locations identified in GWAS on each side of the *Rbs*

interval were identified in the MLM analysis or with stepwise logistic regression. In addition to these locations on chromosome 16, other significant SNPs identified within the N-1989 (chromosomes 2 and 17) and B-2000 (chromosome 8) panels could be involved in resistance. Because of the limited replication of phenotypic data obtained for GWAS and the nature of the binary data, more evidence is needed to substantiate these QTL. Increased replication of phenotypic data within these panels may identify additional QTL, but it is possible that the frequency of BSR resistance alleles at QTL other than the chromosome 16 *Rbs* locus is too low for a traditional GWAS to detect (Cardon and Bell, 2001). In this case a biparental or family-based association mapping population would be needed. Presumably, the significant association on chromosome 5 in the B-2000 panel is a false positive because it was not included in the optimal MLM and was not significant when reconducting the GWAS with SNPs identified in the MLM as covariates. The presence of another unknown QTL that interacts with the *Rbs* locus has been proposed (Bachman and Nickell, 2000b; Sebastian et al., 1985; Waller et al., 1991). If loci identified in our GWAS were substantiated, further research would be needed to determine if epistatic interactions with the *Rbs* locus exists.

The lack of significance in the B-1997 panel was unexpected because Patzoldt et al. (2005a) developed biparental populations with resistant accessions identified by Bachman et al. (1997a) and present in the B-1997 panel. Patzoldt et al. (2005a) then conducted linkage mapping with molecular markers on chromosome 16 and in all five of the resistant sources studied, QTL were mapped to the same region where *Rbs* was previously mapped. This suggests that the resistance QTL in the *Rbs* interval of chromosome 16 is present in the landrace population of the central China region where accessions included in the B-1997 GWAS were collected. It is possible that the lack of association is the result of a low frequency of resistance alleles found in the accessions of this panel, and a different design would be needed to identify an association on chromosome 16 (Cardon and Bell, 2001). Another explanation of the lack of GWAS signals for this panel is that many accessions did not have foliar symptoms and were possibly disease escapes. Yet another possibility is the inoculum for the 1994 Urbana, IL, field test where these accessions were evaluated. The *Rbs* gene may not have been effective against the population of *C. gregata* present in the field during that year. No interactions between *C. gregata* isolates and host genotypes have been reported but do exist in other soybean pathogenic fungi (Grau et al., 2004).

In this study, association mapping was used to refine the location of the known *Rbs* gene or genes and to identify new putative associations that should be tested in future research. This study demonstrates the ability to use data from the SoySNP50K array and historical BSR resistance test data to map disease resistance in soybean. The genomic positions identified will aid gene cloning

efforts. Furthermore, the markers identified in this study will improve MAS and development of resistant cultivars. Regardless, the prominence of only one major QTL for BSR resistance does present a challenge in the event that an isolate is able to overcome resistance from this locus. Further research is needed to validate additional QTL or distinguish the previously reported *Rbs* genes. In addition, research is needed to evaluate resistance to additional isolates and inoculation methods that can increase our overall resistance in germplasm.

### Acknowledgments

This work was supported by funding from Dow AgroScience. The authors thank Michael Bachman for providing BSR resistance results from tests that were not available on GRIN.

### References

- Agresti, A. 2013. Categorical data analysis. 3rd ed. John Wiley & Sons. Hoboken, NJ.
- Bachman, M.S. 1999. Characterization of brown stem rot resistance in soybean. Ph.D. diss. Univ. of Illinois. Urbana-Champaign.
- Bachman, M.S., and C.D. Nickell. 2000a. High frequency of brown stem rot resistance in soybean germplasm from central and southern China. *Plant Dis.* 84:694–699. doi:10.1094/PDIS.2000.84.6.694
- Bachman, M.S., and C.D. Nickell. 2000b. Investigating the genetic model for brown stem rot resistance in soybean. *J. Hered.* 91:316–321. doi:10.1093/jhered/91.4.316
- Bachman, M.S., C.D. Nickell, P.A. Stephens, and A.D. Nickell. 1997a. Brown stem rot resistance in soybean germ plasm from central China. *Plant Dis.* 81:953–956. doi:10.1094/PDIS.1997.81.8.953
- Bachman, M.S., C.D. Nickell, P.A. Stephens, A.D. Nickell, and L.E. Gray. 1997b. The effect of *Rbs2* on yield of soybean. *Crop Sci.* 37:1148–1151. doi:10.2135/cropsci1997.0011183X003700040019x
- Bachman, M.S., J.P. Tamulonis, C.D. Nickell, and A.F. Bent. 2001. Molecular markers linked to brown stem rot resistance genes *Rbs1* and *Rbs2* in soybean. *Crop Sci.* 41:527–535. doi:10.2135/cropsci2001.412527x
- Bastien, M., H. Sonah, and F. Belzile. 2014. Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. *Plant Genome* 7. doi:10.3835/plantgenome2013.10.0030
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Royal Stat. Soc.* 57:289–300.
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. doi:10.1093/bioinformatics/btm308
- Box, G.E.P., and D.R. Cox. 1964. An analysis of transformations. *J. Royal Stat. Soc. B. Met.* 26:211–234.
- Cardon, L.R., and J.I. Bell. 2001. Association study designs for complex diseases. *Nat. Rev. Genet.* 2:91–99. doi:10.1038/35052543
- Chamberlain, D.W., and R.L. Bernard. 1968. Resistance to brown stem rot in soybean. *Crop Sci.* 8:728–729. doi:10.2135/cropsci1968.0011183X00800060026x
- Chen, J., and Z. Chen. 2008. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* 95:759–771. doi:10.1093/biomet/asn034
- Dangl, J.L., and J.D. Jones. 2001. Plant pathogens and integrated defense responses to infection. *Nature* 411:826–833. doi:10.1038/35081161
- Dixon, M.S., D.A. Jones, J.S. Keddie, C.M. Thomas, K. Harrison, and J.D.G. Jones. 1996. The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* 84:451–459. doi:10.1016/S0092-8674(00)81290-8
- Gao, H., N.N. Narayanan, L. Ellison, and M.K. Battacharyya. 2005. Two classes of highly similar coiled coil-nucleotide binding-leucine rich repeat genes isolated from the *Rps1-k* locus encode *Phytophthora* resistance in soybean. *Mol. Plant Microbe Interact.* 18:1035–1045. doi:10.1094/MPMI-18-1035

- Grau, C.R., A.E. Dorrance, J. Bond, and J.S. Russin. 2004. Fungal diseases. In: H.R. Boerma and J.E. Specht, editors, Soybeans: Improvement, production, and uses. 3rd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI. p. 679-763.
- Gray, L.E. 1972. Effect of *Cephalosporium gregatum* on soybean yield. Plant Dis. Rep. 56:580-581.
- Hanson, P.M., C.D. Nickell, L.E. Gray, and S.A. Sebastian. 1988. Identification of two dominant genes conditioning brown stem rot resistance in soybean. Crop Sci. 28:41-43. doi:10.2135/cropsci1988.0011183X002800010010x
- Hayes, A.J., S.C. Jeong, M.A. Gore, Y.G. Yu, G.R. Buss, S.A. Tolin, and M.A. Saghai Maroof. 2004. Recombination within a nucleotide binding-site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to Soybean mosaic virus in soybeans. Genetics 166:493-503. doi:10.1534/genetics.166.1.493
- Hughes, T.J., N.C. Kurtzweil, B.W. Diers, and C.R. Grau. 2004. Resistance to brown stem rot in soybean germ plasm with resistance to the soybean cyst nematode. Plant Dis. 88:761-768. doi:10.1094/PDIS.2004.88.7.761
- Hwang, E., Q. Song, G. Jia, J.E. Specht, D.L. Hyten, J. Costa, and P.B. Cregan. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15:1. doi:10.1186/1471-2164-15-1
- Hyten, D.L., Q. Song, Y. Zhu, I.Y. Choi, R.L. Nelson, J.M. Costa, J.E. Specht, R.C. Shoemaker, and P.B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. USA 103:16666-16671. doi:10.1073/pnas.0604379103
- Kang, Y.J., K.H. Kim, S. Shim, M.Y. Yoon, S. Sun, M.Y. Kim, K. Van, and S. Lee. 2012. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol. 12:139. doi:10.1186/1471-2229-12-139
- Koenning, S.R., and J.A. Wrather. 2010. Suppression of soybean yield potential in the continental United States by plant diseases from 2006 to 2009. (online) Plant Health Prog. doi:10.1094/PHP-2010-1122-01-RS
- Lewers, K.S., E.H. Crane, C.R. Bronson, J.M. Schupp, P. Keim, and R.C. Shoemaker. 1999. Detection of linked QTL for soybean brown stem rot resistance in 'BSR 101' as expressed in a growth chamber environment. Mol. Breed. 5:33-42. doi:10.1023/A:1009634710039
- Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397-2399. doi:10.1093/bioinformatics/bts444
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In: P. Zarembka, editor, Frontiers in Econometrics. Academic Press, New York. p. 105-142.
- Meyer, J.D.F., D.C.G. Silva, C. Yang, K.F. Pedley, C. Zhang, M. Mortel, J.H. Hill, R.C. Shoemaker, R.V. Abdelnoor, S.A. Whitham, and M.A. Graham. 2009. Identification and analyses of candidate genes for *Rpp4*-mediated resistance to Asian soybean rust in soybean. Plant Physiol. 150:295-307. doi:10.1104/pp.108.134551
- Nelson, R.L., C.D. Nickell, J.H. Orf, H. Tachibana, E.T. Gritton, C.R. Grau, and B.W. Kennedy. 1989. Evaluating soybean germplasm for brown stem rot resistance. Plant Dis. 73:110-114. doi:10.1094/PD-73-0110
- Patzoldt, M.E., S.R. Carlson, and B.W. Diers. 2005a. Characterization of resistance to brown stem rot of soybean in five accessions from central China. Crop Sci. 45:1092-1095. doi:10.2135/cropsci2004.0393
- Patzoldt, M.E., W. Chen, and B.W. Diers. 2003. Evaluation of soybean plant introductions from China for resistance to brown stem rot. (online) Plant Health Prog. doi:10.1094/PHP-2003-0702-01-RS
- Patzoldt, M.E., C.R. Grau, P.A. Stephens, N.C. Kurtzweil, S.R. Carlson, and B.W. Diers. 2005b. Localization of a quantitative trait locus providing brown stem rot resistance in the soybean cultivar Bell. Crop Sci. 45:1241-1248. doi:10.2135/cropsci2003.0615
- Perez, P.T., B.W. Diers, P. Lundeen, G.M. Tabor, and S.R. Cianzio. 2010. Genetic analysis of new sources of soybean resistance to brown stem rot. Crop Sci. 50:2431-2439. doi:10.2135/cropsci2010.03.0159
- Platt, A., B.J. Vilhjálmsson, and M. Nordborg. 2010. Conditions under which genome-wide association studies will be positively misleading. Genetics 186:1045-1052. doi:10.1534/genetics.110.121665
- Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904-909. doi:10.1038/ng1847
- Rincker, K., G.L. Hartman, and B.W. Diers. 2016. Fine mapping of resistance genes from five brown stem rot resistance sources in soybean. Plant Genome 9. doi:10.3835/plantgenome2015.08.0063
- SAS Institute. 2011. The SAS system for Microsoft Windows. Release 9.3. SAS Inst. Inc., Cary, NC.
- Schwarz, G. 1978. Estimating the dimension of a model. Ann. Stat. 6:461-464. doi:10.1214/aos/1176344136
- Sebastian, S.A., C.D. Nickell, and L.E. Gray. 1985. Efficient selection for brown stem rot resistance in soybeans under greenhouse screening conditions. Crop Sci. 25:753-757. doi:10.2135/cropsci1985.0011183X002500050008x
- Segura, V., B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, and M. Nordborg. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44:825-830. doi:10.1038/ng.2314
- Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS ONE 8:E54985. doi:10.1371/journal.pone.0054985
- Suh, S.J., B.C. Bowman, N. Jeong, K. Yang, C. Kastl, S.A. Tolin, M.A. Saghai Maroof, and S. Jeong. 2011. The *Rsv3* locus conferring resistance to soybean mosaic virus is associated with a cluster of coiled-coil nucleotide-binding leucine-rich repeat genes. Plant Genome 4:55-64. doi:10.3835/plantgenome2010.11.0024
- Sun, G., C. Zhu, M.H. Kramer, S.S. Yang, W. Song, H.P. Piepho, and J. Yu. 2010. Variation explained in mixed-model association mapping. Heredity 105:333-340. doi:10.1038/hdy.2010.11
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423. doi:10.3168/jds.2007-0980
- Vaughn, J. N., R. L. Nelson, Q. Song, P. B. Cregan, and Z. Li. 2014. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3: Genes, Genomes, Genet. 4:2283-2294. doi:10.1534/g3.114.013433
- Waller, R.S., C.D. Nickell, D.L. Drzymalski, and J.E. Miller. 1991. Genetic analysis of the inheritance of brown stem rot resistance in the soybean cultivar Asgrow A3733. J. Hered. 82:412-417.
- Wen, Z., R. Tan, J. Yuan, C. Bales, W. Du, S. Zhang, M.I. Chilvers, C. Schmidt, Q. Song, P.B. Cregan, and D. Wang. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics 15:809. doi:10.1186/1471-2164-15-809
- Willmont, D.B., and C.D. Nickell. 1989. Genetic analysis of brown stem rot resistance in soybean. Crop Sci. 29:672-674. doi:10.2135/cropsci1989.0011183X002900030026x
- Wrather, A., G. Shannon, R. Balardin, L. Carregal, R. Escobar, G.K. Gupta, Z. Ma, W. Morel, D. Ploper, and A. Tenuta. 2010. Effect of diseases on soybean yield in the top eight producing countries in 2006. (online) Plant Health Prog. doi:10.1094/PHP-2010-0125-01-RS
- Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203-208. doi:10.1038/ng1702
- Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordoñez, and E.S. Buckler. 2010. Adaptation of mixed linear model for genome-wide association studies. Nat. Genet. 42:355-360. doi:10.1038/ng.546