the plant journal

SEB Society for Experimental Biology

The Plant Journal (2015) **84,** 1124–1136 doi: 10.1111/tpj.13069

Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean

Jiaoping Zhang[1], Arti Singh[1], Daren S. Mueller[2,*] and Asheesh K. Singh[1,*]
[1]*Department of Agronomy, Iowa State University, Ames, IA 50011, USA, and*
[2]*Department of Plant Pathology, Iowa State University, Ames, IA 50011, USA*

Received 10 August 2015; revised 20 October 2015; accepted 29 October 2015; published online 12 November 2015.
*For correspondence (e-mails dsmuelle@iastate.edu; singhak@iastate.edu).

SUMMARY

Soybean [*Glycine max* (L.) Merr.] is an economically important crop that is grown worldwide. Sudden death syndrome (SDS), caused by *Fusarium virguliforme*, is one of the top yield-limiting diseases in soybean. However, the genetic basis of SDS resistance, especially with respect to epistatic interactions, is still unclear. To better understand the genetic architecture of soybean SDS resistance, genome-wide association and epistasis studies were performed using a population of 214 germplasm accessions and 31 914 SNPs from the SoySNP50K Illumina Infinium BeadChip. Twelve loci and 12 SNP–SNP interactions associated with SDS resistance were identified at various time points after inoculation. These additive and epistatic loci together explained 24–52% of the phenotypic variance. Disease-resistant, pathogenesis-related and chitin- and wound-responsive genes were identified in the proximity of peak SNPs, including *stress-induced receptor-like kinase gene 1* (*SIK1*), which is pinpointed by a trait-associated SNP and encodes a leucine-rich repeat-containing protein. We report that the proportion of phenotypic variance explained by identified loci may be considerably improved by taking epistatic effects into account. This study shows the necessity of considering epistatic effects in soybean SDS resistance breeding using marker-assisted and genomic selection approaches. Based on our findings, we propose a model for soybean root defense against the SDS pathogen. Our results facilitate identification of the molecular mechanism underlying SDS resistance in soybean, and provide a genetic basis for improvement of soybean SDS resistance through breeding strategies based on additive and epistatic effects.

Keywords: *Glycine max* (L.) Merr., sudden death syndrome, genome-wide association study, SNP–SNP interaction, epistatic interaction, linkage disequilibrium, plant defense.

INTRODUCTION

Sudden death syndrome (SDS) of soybean [*Glycine max* (L.) Merr.] is caused by *Fusarium virguliforme*, a soil-borne fungal pathogen that infects and colonizes the soybean root and causes root rot (Roy *et al.*, 1997). Toxins produced by the fungus translocate from the root to the foliage through the vascular system, and lead to vascular discoloration and characteristic foliar symptoms including inter-veinal chlorosis and necrosis on the upper leaves, defoliation, pod abortion and premature plant death (Mueller *et al.*, 2002).

Sudden death syndrome spread rapidly and widely across the major soybean-producing regions of the world. Since the first observation of SDS on soybean in Arkansas in 1971, occurrence of the disease has been reported in most soybean-producing states throughout the United States and in other soybean-growing countries in South America, Africa and Asia (Roy *et al.*, 1997; Hartman *et al.*, 2015). Severe yield losses have been observed in soybean fields with SDS (Farm, 1995; Brzostowski *et al.*, 2014). In 2009, the yield losses caused by SDS in the United States were equivalent to an economic cost of $0.33 billion (Koenning and Wrather, 2010). Due to the widespread areas of infection and significant yield losses, SDS is considered one of the most important fungal diseases in soybean, requiring research efforts to develop control measures.

Developing resistant cultivars is the most effective method for controlling SDS. Over recent decades, significant efforts have been made to identify new resistance sources. A large number of soybean plant introductions (PIs) (>6800) and publicly and privately developed soybean

entries (>2300) have been evaluated for SDS resistance in greenhouse or field tests (Rupe *et al.*, 1991; Hartman *et al.*, 1997; Njiti *et al.*, 1997; Mueller *et al.*, 2002, 2003). Although abundant genetic variation exists, no source of complete resistance has been discovered. Soybean breeders mainly rely on sources with partial resistance for cultivar improvement. For example, Cianzio *et al.* (2014) described an SDS-tolerant soybean germplasm line that inherited three resistance loci for SDS from the partially resistant donor parent Ripley. This study demonstrates that combining multiple resistant loci is an effective way to increase SDS resistance in soybean (Njiti *et al.*, 2002). A recent study identified a receptor-like kinase (RLK) gene *GmRLK18-1* at the *Rhg1/Rfs2* locus, conferring resistance to both SDS and soybean cyst nematode (Srour *et al.*, 2012). Given the ability of *GmRLK18-1* to bind the CLAVATA3/EMBRYO-SURROUNDING REGION-related peptide found in nematode secretions (Srour *et al.*, 2012), a gene-for-gene resistance mechanism may be also important for soybean SDS resistance.

Linkage mapping using biparental populations revealed 56 quantitative trait loci (QTL) associated with SDS resistance across 12 of the 20 soybean chromosomes (http://www.soybase.org/). These explained between 1.4 and 47.0% of the phenotypic variation, suggesting both minor and major loci contribute resistance to SDS. However, *GmRLK18-1*, coincident with *Rhg1/Rfs2*, is the only gene that has been cloned and functionally characterized (Srour *et al.*, 2012). In addition to the small effect of most loci, the limited allelic segregation between the two parents and the limited recombination that occurs during creation of biparental populations in linkage mapping studies also hinder identification of SDS resistance genes. Genome-wide association studies (GWAS) are a powerful tool that overcomes the limitations of biparental populations and dissects complex traits at high mapping resolution. In soybean, GWAS have been performed with respect to seed composition (Hwang *et al.*, 2014), various agronomic traits (Sonah *et al.*, 2015; Zhang *et al.*, 2015), and responses to biotic stresses such as white mold caused by *Sclerotinia sclerotiorum* (Bastien *et al.*, 2014) and abiotic stresses such as iron deficiency chlorosis (Mamidi *et al.*, 2014). A recent GWAS for SDS using elite soybean cultivars identified diversity of gene candidates, including disease resistance-related genes (Wen *et al.*, 2014), indicating the advantages of GWAS in dissecting the genetic basis of complex traits in soybean.

Although GWAS is a powerful tool, it only explains a limited proportion of the heritability for complex traits (Manolio *et al.*, 2009). Interaction between genetic variants is an important source of the missing heritability in GWAS. Genome-wide epistasis studies (GWES) are a state-of-the-art technique in the search for epistatic interactions with genome-wide dense markers. They have been used for human disease research (Hu *et al.*, 2010; Prabhu and Pe'er, 2012), but have not been adequately utilized in plants. In soybean, a large genetic variation in SDS resistance was observed in germplasm accessions (Mueller *et al.*, 2002). However, our knowledge about the genetic basis of the natural variation in soybean SDS resistance is limited. Previous studies mainly focused on the additive effect, and information on epistatic interactions associated with SDS resistance in soybean, although critical, is still lacking.

To further accelerate improvement in soybean SDS resistance, new resistance sources and the underlying genetic basis of such resistance need to be identified. The objectives of this study were to better understand the natural genetic variants in both additive and epistatic effects associated with SDS resistance in soybean, to provide additional genetic sources of SDS resistance and new breeding strategies for soybean SDS resistance, and to provide insights into the mechanisms of plant root defense against soil-borne pathogens. We performed GWAS and GWES for resistance to SDS in a soybean germplasm panel containing 214 accessions saturated with high-density SNPs. The results suggest the necessity of using GWES as a complement to GWAS to develop a comprehensive understanding of the genetic basis of the traits and to facilitate development of new breeding strategies. Causal genes that putatively function in plant defense are proposed. A model of soybean defense to the soil-borne SDS pathogen is presented.

RESULTS AND DISCUSSION

Marker distribution and linkage disequilibrium

A total of 31 914 markers from the SoySNP50K Illumina Infinium BeadChip had a minor allele frequency (MAF) ≥5% in our population of 214 PIs, which were selected from successive screenings of over 6000 PIs for SDS resistance (Table S1). The mean inter-marker distance was 29.7 kb. However, the SNPs were unevenly distributed across chromosomes and within each chromosome (Figure S1). Chromosomes 18 and 20 (Gm18 and Gm20) harbored the largest number of SNPs (2504) and the smallest number of SNPs (1121), respectively. The chromosome-wide densities varied from 45.3 kb per SNP on Gm01 to 22.4 kb per SNP on Gm13. The majority of SNPs (76.7%) were located in euchromatic regions. This was reasonable because 78% of the putative genes were found in euchromatin where over 80% of recombination events occur (Schmutz *et al.*, 2010).

Due to the substantial difference in recombination rate, the linkage disequilibrium (LD) decay rates for euchromatin and heterochromatin were estimated separately at 208 and 1502 kb, respectively (Figure 1). We observed a faster LD decay than in previous association studies of soybean where larger populations were used (Zhang *et al.*, 2015), suggesting high genetic diversity in our association panel.

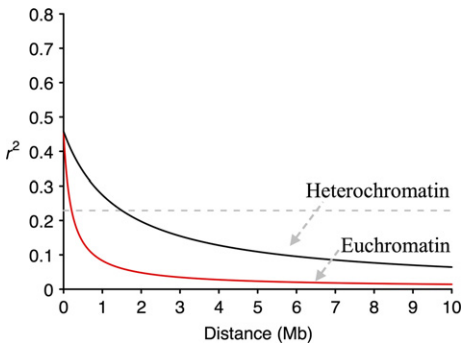

Figure 1. Mean linkage disequilibrium (LD) decay rate in euchromatic and heterochromatic regions of the soybean genome.
The mean LD decay rate was estimated as r^2 using all pairs of SNPs located within 10 Mb of physical distance in euchromatic and heterochromatic regions in 214 soybean germplasm accessions. The dashed line indicates the position where r^2 has decreased to half its maximum value.

Given that the marker densities of euchromatic and heterochromatic regions were 17.5 and 70.0 kb per SNP, respectively, the SNP set was dense enough to capture the genetic variation in the association panel.

Genetic diversity and phenotypic variation

The heterozygous genotype rate was low (0.5%), consistent with the highly inbred nature of the soybean. The SNP set had a mean nucleotide diversity (polymorphism information content) of 0.27. This was similar to that in a previous study (0.28) (Wen *et al.*, 2014) in which diverse elite cultivars were used, but lower than the genetic diversity (0.35) in a broad sample of wild (*G. soja*) and domesticated (*G. max*) soybean (Li *et al.*, 2010). This indicates that the current association panel is a diverse population. However, at some chromosomal regions on Gm02, Gm08, Gm14 and Gm19, the nucleotide diversity remained consistent (Figure S1). Further examination revealed that these regions were comprised extensive LD blocks in heterochromatin (Figure S2).

The disease severity was scored at 20, 23, 26 and 29 days after inoculation (DAI), and the area under the disease progress curve (AUDPC) was calculated for each PI. The observed phenotypic distribution was wide at each time point of SDS rating and AUPDC in the soybean association panel (Table S2 and Figure S3). The individual disease ratings and AUDPC values showed a continuous distribution, reflecting the quantitative nature of SDS resistance in soybean (Figure S3). Disease severity progressed with each time point, and high correlations between time points were observed (Table S3).

Population structure

Because population structure is one of the major sources for spurious association, we investigated the population stratification in our panel. A neighbor-joining analysis using the whole set of SNPs classified the population into three clusters (Figure 2a). Using the same SNP set, principal component analysis was also performed. We found that the first two principal components, which explained 17% of the total variation, defined the three clusters well (Figure 2b), and therefore were used to account for the population structure in this study.

The geographical origin and photoperiod response are primary factors shaping population stratification in soybean, which is a photoperiod-sensitive crop. A high correlation was observed between the country of origin and the population structure of the test accessions (Figure 2c). Maturity groups, determined by photoperiod response, were spread across the sub-groups (Figure 2d), suggesting that geographical origin, rather than photoperiod, was the main factor driving the population structure in the association panel. A possible explanation may be that the geographical origins of the PIs are at similar latitude, and

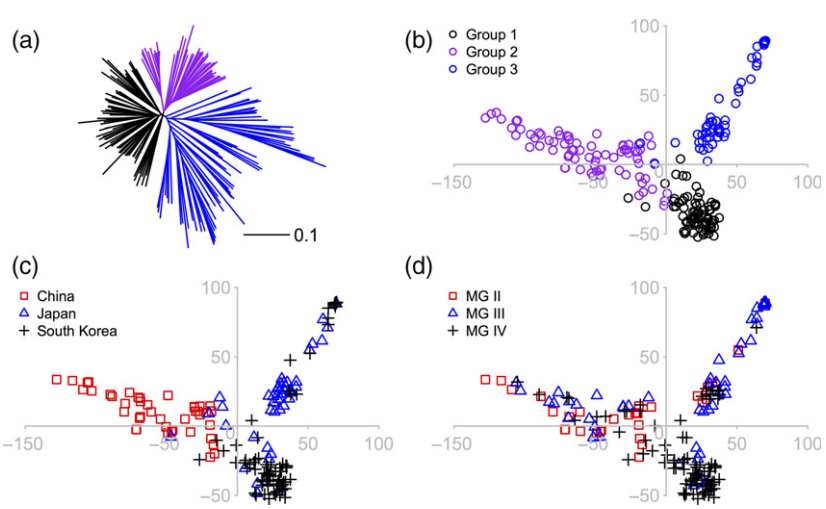

Figure 2. Population structure of the soybean association panel.
(a) Neighbor-joining tree of 214 soybean accessions in the panel. The sub-groups are color-coded.
(b–d) Plots of the first two principal components of the panel including color-coded sub-groups as defined by the neighbor-joining analysis, country of origin and maturity group (MG) of the accessions, respectively.

the lines with different origins may undergo different artificial selections to meet local needs during long-term cultivation.

Genome-wide association and epistatic interaction analyses

Genome-wide scans were performed using the naive model (without correction for kinship and population stratification), the mixed linear model (MLM) with kinship (*K* model), and the MLM with both kinship and population structure (*K* + *P* model) for each measurement. The quantile/quantile plots indicated that the MLMs were consistently better than the naive model in terms of genomic control across five disease severity measurements. Similar control of genomic inflation was observed between the *K* model and the *K* + *P* model (Figure S4). Therefore, only the results from the analysis using the *K* model are presented. This was consistent with the results of a model fitness test (Table S4).

The GWAS revealed 14, 4, 5, 15 and 7 SNPs associated with DAI20, DAI23, DAI26, DAI29 and AUDPC, respectively (Figure S5). The SNPs were grouped at $r^2 > 0.70$, and only the strongest trait-associated SNP (or peak SNP) within each LD block was retained (Zhang *et al.*, 2015). After imposing this condition, seven, four, three, two and five loci associated with SDS resistance were identified for DAI20, DAI23, DAI26, DAI29 and AUDPC, respectively (Table 1). These loci alone contribute 12–31% of the phenotypic variance (Figure 3). Six of them were detected multiple times (Table 1). The loci *ss715584189_T_C* (MAF = 0.15) and *ss715634180_G_A* (MAF = 0.50) were detected at DAI20, DAI23 and AUDPC. The locus *ss715631747_C_T* (MAF = 0.11) at the 52.0 Mb position on Gm18 was associated with DAI23, DAI26 and AUDPC. Another locus, *ss715582444_T_G* (MAF = 0.22) at 39.6 Mb on Gm02, was observed at DAI20 and DAI23. No common locus was found between DAI29 and other ratings. A pathogen's effect on yield largely depends on the timing of disease development and spread, so selection of the desired alleles at loci associated with early response to SDS may postpone development of the disease and reduce yield losses. This may be an important strategy for improvement of soybean SDS resistance through marker-assisted selection when no complete resistance source is available.

For epistatic tests, three, five, one, one and three SNP–SNP interactions associated with DAI20, DAI23, DAI26, DAI29 and AUDPC, respectively, were identified after clustering SNPs as described above (Table 2 and Figure S6). The addition of these epistatic loci to the GWAS loci increased the explained proportion of phenotypic variance up to 24–52% (Figure 3), indicated that genomic selection for SDS resistance based on both additive and non-additive effects may be more efficient than using additive effects alone (Wang *et al.*, 2012a). All of the interactions detected were between SNPs located on different chromosomes. The interaction between *ss715586656_A_G* (MAF = 0.39) on Gm03 and *ss715594736_T_C* (MAF = 0.17) on Gm06 associated with SDS resistance at DAI20 represented the strongest epistatic effect ($P = 8.8 \times 10^{-13}$) in the present study (Figure 4a and Table 2). The disease severity for the most resistant genotype combination (*G*C*) of the epistatic loci was 23% less than that for the most susceptible one (*A*C*) (Figure 4b). Further examination of all four genotype combinations for the epistatic loci revealed that they were significantly different from each other, implying that selection based on the epistatic effect is still effective even though one of the two loci is fixed in a breeding population. However, selection using *ss715586656_A_G* or *ss715594736_T_C* alone may have no effect on SDS resistance if both are segregating in a population. The GWAS and GWES results revealed that PI 82278, one of the most resistant lines in the association panel, contained the desired alleles at all 12 additive loci and 11 of the 12 epistatic loci (Table S5). Therefore, PI 82278 may be a good donor parent for breeding to improve soybean SDS resistance.

Three SNPs were involved in at least two interactions detected in the same disease resistance measurement. The SNP *ss715580815_A_G* (MAF = 0.05) on Gm01 interacted with SNPs on both Gm02 and Gm09 in DAI20 (Table 2). Another SNP, *ss715611120_C_T* (MAF = 0.15) on Gm11, interacted with SNPs on Gm02 and Gm17 in DAI23. Notably, all three interactions associated with AUDPC shared a common locus at the 36.2 Mb position on Gm12, which was mapped by two SNPs (*ss715612644_G_T*, MAF = 0.07; *ss715612645_C_A*, MAF = 0.08) in high LD ($r^2 = 0.81$). One of them, the interaction with *ss715599552_T_C* on Gm08, was repeatedly detected for DAI26 and AUDPC (Table 2). Interestingly, none of the epistatic loci were detected in GWAS, implying that a search for epistatic effects exclusively between main effects loci may be insufficient. The above results suggested a complex network of epistatic effects independent from main effects for soybean SDS resistance. Similar observations have been made in other species such as barley (*Hordeum vulgare*; Xu and Jia, 2007) and wheat (*Triticum aestivum*; Singh *et al.*, 2013).

SNP locus confirmation and candidate genes

The loci for SDS resistance identified in this study had considerable overlap with previously reported loci, adding support to our findings. Four of the 12 unique loci identified via GWAS, including the locus *ss715584189_T_C* (MAF = 0.15) that was detected in multiple ratings, were mapped to regions where SDS resistance QTLs had been previously reported (Table 1). Two of the eight novel loci overlapped with QTLs for resistance to other soybean fungal pathogens. The locus *ss715614656_G_A* at 28.5 Mb on Gm13 mapped to a similar region associated with

Table 1 SNPs significantly associated with soybean sudden death syndrome resistance and the candidate genes

Trait	SNP[a]	Chromosome[b]	Position (bp)[b]	MAF	*P*	Allelic effect[c]	R^2	QTL[d]	Candidate genes[e]	Distance (kb)[f]	Annotation
DAI20	*ss715584189_T_C*	02	9450450	0.15	3.61E-04	0.15	0.06	Abdelmajid *et al.* (2012)	*SIK1*	0	LRR-RLK
	ss715582444_T_G	02	39634936	0.22	3.92E-04	−0.11	0.06	N	*Glyma.02g211000*	0	Integral membrane protein
	ss715599474_G_A	08	1373179	0.34	4.84E-05	−0.11	0.07	Abdelmajid *et al.* (2012)	*Glyma.0g017000*	0	B-1,3-glucosyltransferase
	ss715614656_G_A	13	28548247	0.20	4.00E-04	−0.13	0.06	N	Unknown		
	ss715627896_C_A	17	4085459	0.20	4.19E-04	0.13	0.06	Arahana *et al.* (2001)	Unknown		
	ss715634180_G_A	19	3489366	0.50	3.46E-04	0.10	0.06	N	*Glyma.19g027900*	0	Casein lytic proteinase related to heat shock
	ss715637220_A_G	20	30662956	0.08	1.86E-04	−0.20	0.07		Unknown		
DAI23	*ss715584189_T_C*	02	9450450	0.15	7.85E-05	0.14	0.07	Abdelmajid *et al.* (2012)	*SIK1*	0	LRR-RLK
	ss715582444_T_G	02	39634936	0.22	1.15E-04	−0.10	0.07	N	*Glyma.02g211000*	0	Integral membrane protein
	ss715631747_C_T	18	51968926	0.11	5.16E-05	0.17	0.08		*Glyma.18g231000*	+4.2	LRR-containing protein; disease resistance RPP13-like protein
	ss715634180_G_A	19	3489366	0.50	9.13E-05	0.09	0.07	N	*Glyma.19g027900*	0	Casein lytic proteinase related to heat shock
DAI26	*ss715631747_C_T*	18	51968926	0.11	9.26E-05	0.14	0.07		*Glyma.18g231000*	+4.2	LRR-containing protein; disease resistance RPP13-like protein
	ss715634431_C_T	19	36112662	0.27	5.00E-05	−0.09	0.08		*Glyma.19g108500*	0	Apyrase
	ss715634448_A_G	19	36199222	0.12	1.15E-04	−0.12	0.07	Abdelmajid *et al.* (2012)	*Glyma.19g109300*	+56.2[g]	NBS-LRR domain-containing disease resistance protein
DAI29	*ss715586494_C_T*	03	44251912	0.47	2.39E-05	−0.14	0.09	Chang *et al.* (1996), Hnetkovsky *et al.* (1996)	*Glyma.03g245200*	0	Unknown function
	ss715633619_C_T	19	2901793	0.27	1.67E-04	−0.13	0.07	N	Unknown		
AUDPC	*ss715584189_T_C*	02	9450450	0.15	1.25E-04	6.69	0.07	Abdelmajid *et al.* (2012)	*SIK1*	0	LRR-RLK
	ss715627896_C_A	17	4085459	0.20	2.33E-04	5.39	0.06	Arahana *et al.* (2001)	Unknown		
	ss715631747_C_T	18	51968926	0.11	5.97E-05	8.30	0.08		*Glyma.18g231000*	4.2	LRR-containing protein; disease resistance RPP13-like protein
	ss715634180_G_A	19	3489366	0.50	7.54E-05	4.54	0.07	N	*Glyma.19g027900*	0	Casein lytic proteinase related to heat shock
	ss715634431_C_T	19	36112662	0.27	3.02E-04	−4.53	0.06		*Glyma.19g108500*	0	Apyrase

[a]SNP name followed by the major and minor allele of the marker.
[b]Physical position of the SNP based on soybean reference genome Glyma.Wm82.a2 (Gmax2.0) in SoyBase (www.soybase.org).
[c]With respect to the minor allele.
[d]QTL previously reported or newly identified in the present study; N, QTLs have not been reported previously.
[e]Genes annotated in Glyma1.1, Glyma1.0 and NCBI RefSeq gene models in SoyBase (www.soybase.org) were used as the source of candidate genes.
[f]Distance from the SNP to the candidate gene transcript based on soybean reference genome Gmax2.0 in SoyBase (www.soybase.org). '0' indicates that the SNP is within the candidate gene; '+' indicates that it is upstream of the gene.
[g]The candidate gene is within the LD block ($r^2 > 0.70$) tagged by the lead SNP.

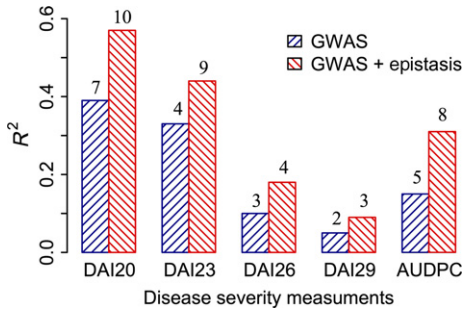

Figure 3. Contributions of identified sudden death syndrome loci via genome-wide association studies (GWAS) and epistatic analysis to the phenotypic variance of each disease severity measurement.
The numbers of loci or pairs of loci used to estimate contributions are indicated above the bars.

resistance to *Phytophthora sojae*, a soybean root fungus causing root and stem rot, and *S. sclerotiorum*, the causal agent of soybean white mold (Arahana *et al.*, 2001; Wang *et al.*, 2010, 2012b; Lee *et al.*, 2013). Another locus, *ss715627896_C_A* on Gm17, overlapped with a QTL associated with soybean resistance to *S. sclerotiorum* (Arahana *et al.*, 2001). These results imply a cluster arrangement of disease resistance loci in soybean as suggested by Graham *et al.* (2002) or a common genetic basis for defense against various pathogens in soybean (Srour *et al.*, 2012).

We also investigated the allelic distribution of both the additive and epistatic loci in a sample of 96 elite soybean cultivars (Song *et al.*, 2013). At least 80% of the elite cultivars carried the unfavorable allele at eight of the 12 identified main effect loci (Figure S7). For nine of the 12 epistatic effect loci, at least 90% of the tested elite cultivars carried the unfavorable alleles (Figure S8). These results confirm the breeding potential of the SDS resistance loci identified in this study.

Due to the relatively high mapping resolution, candidate genes for eight of the 12 unique main-effect loci associated with SDS resistance were predicted (Table 1). Six of the eight gene candidates contained the peak SNPs. The SNP locus *ss715584189_T_C* (MAF = 0.15) at 9.5 Mb on Gm02 was repeatedly detected in this study. In comparison with the alternative allele, the desired allele of this locus may reduce the disease severity by 10% (Figure 5a). An SDS resistance QTL has been previously reported at this genomic region (Abdelmajid *et al.*, 2012). Further analysis pinpointed *ss715584189_T_C* to the coding region of *stress-induced receptor-like kinase gene 1* (*SIK1*) (Figure 5b and Table 1). *ss715584189_T_C* is a non-synonymous SNP according to the information on SoyBase (http://soybase.org/) and Phytozome (http://phytozome.jgi.doe.gov/). It altered the amino acid coding from 'AAA' (Lys388) to 'AGA' (Arg388). Additionally, the LD decayed rapidly, and *SIK1* was the only gene candidate identified at this locus (Figure 5b), indicating that *ss715584189_T_C* may be the causal genetic variant associated with soybean SDS resistance. *SIK1* encodes a leucine-rich repeat receptor-like protein kinase (LRR-RLK), and is mainly expressed in root and flower in soybean (SoyBase, http://soybase.org/) (Woody *et al.*, 2011). Many disease resistance (*R*) genes in plants, including the recently identified soybean SDS *R* gene *GmRLK18-1* at the *Rhg1*/*Rfs2* locus (Srour *et al.*, 2012), encode LRR-RLKs. These proteins contain extracellular LRRs and an intracellular kinase domain, and serve as cell surface-localized pattern recognition receptors in pathogen recognition and innate immunity signaling processes (Liu *et al.*, 2014). Research in rice (*Oryza sativa* L.) demonstrated that OsSIK1 participates in salt and drought stress tolerance through regulation of the antioxidative system (Ouyang *et al.*, 2010), indicating pleiotropism of *SIK1*. This is consistent with a previous study showing that LRR-RLK proteins may be involved in plant responses to both biotic and abiotic stresses (Gou *et al.*, 2010). The above evidence strongly suggests that *SIK1* is the candidate *R* gene of this SDS-resistant locus on Gm02. In this study, LRR-encoding disease resistance genes were also identified at the other two loci tagged by *ss715631747_C_T* on Gm18 and *ss715634448_A_G* on Gm19 (Table 1). In soybean, disease resistance QTLs have been found to co-localize with genes encoding LRR-containing proteins (Hayes *et al.*, 2004; Kang *et al.*, 2012). The LRR domain is the major determinant of pathogen recognition (Jones and Jones, 1997), indicating the importance of gene-for-gene mechanism for SDS resistance in soybean.

Although none of the 12 unique SNP–SNP interactions included SNPs identified by GWAS, we found that three of them included at least one SNP located in a previously identified QTL associated with SDS resistance in soybean (Table 2). Two SNPs, *ss715586656_A_G* (MAF = 0.39) and *ss715594736_T_C* (MAF = 0.17), conferring the interaction between Gm03 and Gm06, were located in known SDS resistance loci in soybean. Both loci were detected in two independent studies as main effects (Chang *et al.*, 1996; Hnetkovsky *et al.*, 1996). Further examination of the gene candidates for these loci revealed that the putative genes *Glyma.03G261600* containing the SNP on Gm03 and *Glyma.06G287300* at 167 bp downstream of the SNP on Gm06 encode an oxidoreductase and a zinc finger FYVE domain-containing lipase, respectively (Figure 4c,d). Oxidoreductases are crucial for the generation of the reactive oxygen species O_2^- and H_2O_2 that fulfill a central role in programmed cell death (Pennell and Lamb, 1997), which is one of the major outcomes of the hypersensitive response triggered by recognition of the plant pathogen at the early stage of plant defense (Staskawicz *et al.*, 1995). The information available on the small family of FYVE-containing proteins in plants is limited. However,

Table 2 SNP–SNP interactions significantly associated with soybean resistance to SDS and candidate genes

Trait	SNP_1[a]	Chromosome position[b]	MAF_1	QTL_1[c]	Candidate $gene_1$[d]	$Annotation_1$	SNP_2[a]	Chromosome position[b]	MAF_2	QTL_2[c]	Candidate $gene_2$[d]	$Annotation_2$	*P*
DAI20	*ss715580815_A_G*	01_801884	0.05	N			*ss715581582_T_C*	02-2447382	0.47	N			5.02E-11
	ss715580815_A_G	01_801884	0.05	N			*ss715603837_T_C*	09-3733550	0.42	N	*Glyma.09G043900*	F-box and WD40 domain protein	6.02E-11
	ss715586656_A_G	03_45496371	0.39	Chang *et al.* (1996), Hnetkovsky *et al.* (1996)	*Glyma.03G261600*	Oxidoreductase	*ss715594736_T_C*	06-47618313	0.17	Chang *et al.* (1996), Hnetkovsky *et al.* (1996)	*Glyma.06G287300*	Zinc finger FYVE domain-containing GDSL-like lipase	8.83E-13
DAI23	*ss715583725_C_A*	02_6380008	0.48	Abdelmajid *et al.* (2012)			*ss715597201_C_T*	07-3337039	0.23	N			2.51E-11
	ss715599599_C_A	08_14644929	0.44	N			*ss715621673_T_C*	15-37387010	0.45	N	*Glyma.15g218900*	Pathogenesis-related protein Bet v I family	6.65E-11
	ss715602039_A_G	08_42977728	0.25	N	*Glyma.08g310300*	Tam3 transposase	*ss715615423_C_T*	13-33466062	0.16	N			5.39E-11
	ss715611120_C_T	11_7454653	0.15	N			*ss715581188_T_C*	02-1357769	0.20	N	*Glyma.02g015300*	Galacturonosyl transferase	3.07E-11
	ss715611120_C_T	11_7454653	0.15	N			*ss715626639_T_C*	17-2906735	0.19	N	*Glyma.17g039200*	*Pleiotropic drug resistance proteins, ABC transporter*	2.69E-11
DAI26	*ss715612645_C_A*[e]	12_36207711	0.08	N	*Glyma.12g200300 – Glyma.12g200800 Glyma.12g201000*	Wound-responsive genes	*ss715599552_T_C*	08-14251580	0.43	N	*Glyma.08g178000*	Phosphate-induced protein	4.91E-11
DAI29	*ss715594216_T_C*	06_3459326	0.39	Abdelmajid *et al.* (2012)	*Glyma.06g045400*	Salt tolerance zinc finger protein, responsive to chitin oligomers	*ss715618025_A_G*	14-2201645	0.30	N	*Glyma.14g030300*	Auxin responsive protein	3.42E-11
AUDPC	*ss715612644_G_T*[f]	12_36185445	0.07	N	*Glyma.12g200300 – Glyma.12g200800Glyma.12g201000*	Wound-responsive genes	*ss715592910_A_G*	06-12456134	0.48	N	*Glyma.06g152800*	Phosphatidylinositol 4-kinase, root hair development	4.64E-11
	ss715612644_G_T[f]	12_36185445	0.07	N	*Glyma.12g200300 – Glyma.12g200800Glyma.12g201000*	Wound-responsive genes	*ss715634880_T_G*	19-39791264	0.36	N			8.61E-11
	ss715612645_C_A[e]	12_36207711	0.08	N	*Glyma.12g200300 – Glyma.12g200800Glyma.12g201000*	Wound-responsive genes	*ss715599552_T_C*	08-14251580	0.43	N	*Glyma.08g178000*	Phosphate-induced protein	6.17E-11

[a]SNP name followed by the major and minor allele of the marker.
[b]Chromosome and physical position (bp) on the chromosome of the SNP based on the soybean reference genome Glyma.Wm82.a2 (Gmax2.0) in SoyBase (www.soybase.org).
[c]QTL previously reported or newly identified in the present study; N, QTLs have not been reported previously.
[d]Genes annotated in Glyma1.1, Glyma1.0, and NCBI RefSeq gene models in SoyBase (www.soybase.org) were used as the source of candidate genes.
[e,f]In high linkage disequilibrium ($r^2 = 0.81$).
[g]The gene is 79.4 kb away from the SNP, within the LD block ($r^2 > 0.70$) tagged by the lead SNP.

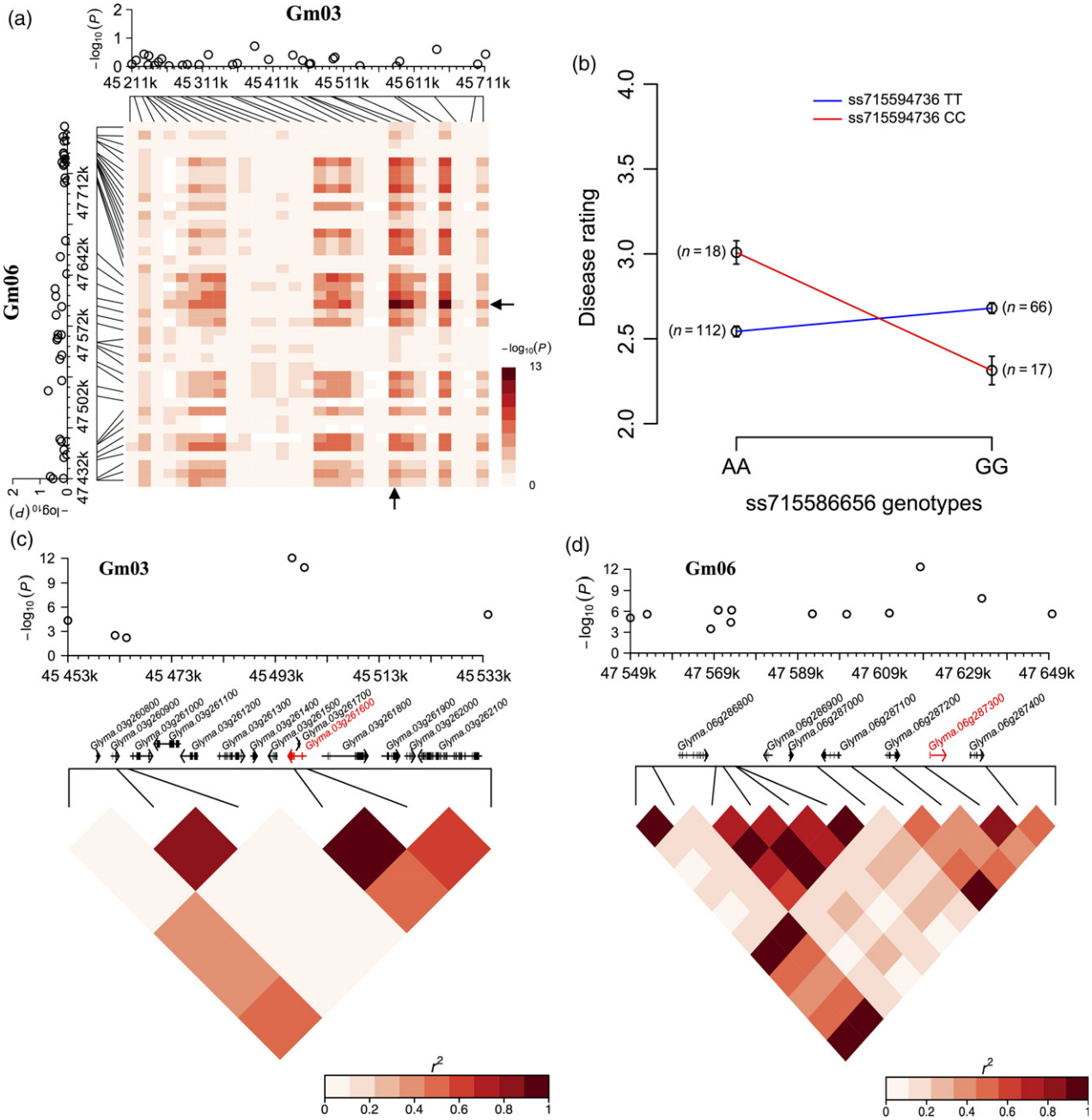

Figure 4. Epistatic interaction between *ss715586656_A_G* and *ss715594736_T_C* associated with DIA20, and candidate gene for each SNP locus.
(a) Heatmap of the SNP–SNP interactions at regions adjacent to the peak SNPs on Gm03 and Gm06. The color intensity indicates the $-\log_{10}$-transformed *P* value of each interaction as indicated in the color key. Black arrows indicate SNPs involved in the identified interaction. The transformed *P* values of the SNP/trait associations for each chromosomal region are plotted against positions on the chromosome as indicated at the top and left.
(b) Phenotypic differences between genotype combinations of the two SNP loci. The standard error of the disease rating for each combination is indicated, and the frequencies are shown in parentheses.
(c,d) Candidate genes for *ss715586656_A_G* and *ss715594736_T_C* loci, respectively. The $-\log_{10}$-transformed *P* values of the SNP–SNP interactions within the adjacent region of the peak SNPs are plotted against positions on the chromosome. The middle panels show all putative genes in the region. The proposed causal genes are indicated in red. The bottom panel depicts the extent of linkage disequilibrium in the regions based on pairwise r^2 values. The r^2 values are indicated using the color intensity index shown.

the FYVE domains are highly conserved across various organisms (Wywial and Singh, 2010), and research on animals showed that the zinc finger FYVE motif mediates cell signaling to phosphatidylinositol 3-phosphate-containing membranes (Abouzeid *et al.*, 2011). Notably, phosphatidylinositol 3-phosphate is involved in abscisic

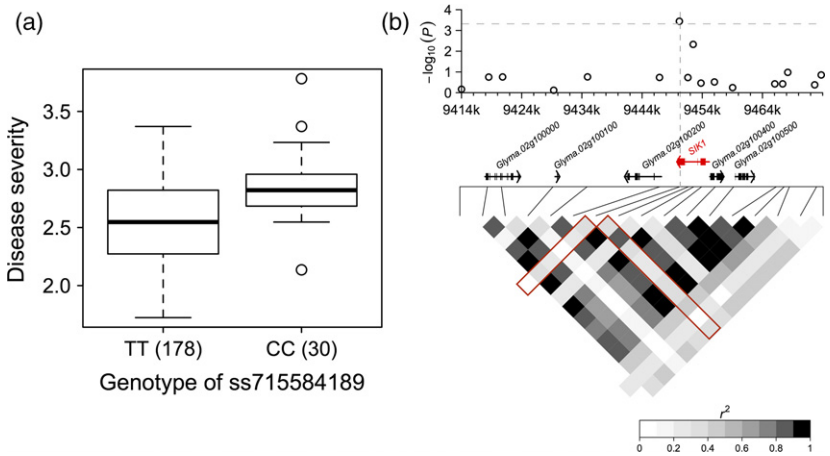

Figure 5. Phenotypic differences between lines carrying alternative alleles and candidate genes at the *ss715584189_T_C* locus associated with DAI20, DAI23 and AUDPC.

(a) Box plot showing differences in disease severity at DAI20 (averaged over four replications) between lines segregating at the peak SNP. The box shows the first, second (median) and third quartile. The whiskers extend to 1.5 times the interquartile range or the data extreme, whichever is smaller. The number of individuals for each allele is shown in parentheses.

(b) The top panel shows a 60 kb region on Gm02 harboring the peak SNP, whose position is indicated by a vertical gray dashed line. Negative $\log_{10}$-transformed *P* values from the mixed linear model are plotted on the vertical axis. The significance threshold is indicated as the horizontal grey dashed line ($P < 4.8 \times 10^{-4}$). The middle panel shows all putative genes within this region. The proposed causal gene *SIK1* is highlighted in red. The bottom panel depicts the extent of linkage disequilibrium in the regions based on pairwise r^2 values. The r^2 values are indicated using the color intensity index. The red boxes indicate the pairwise r^2 values related to the lead SNP.

acid-induced generation of reactive oxygen species in guard cells (Park *et al.*, 2003).

The SNP *ss715621673_T_C* (MAF = 0.45) on Gm15 involved in the SNP–SNP interaction in DAI23 tagged to an LD block harboring the predicted gene *Glyma.15g218900*. This encodes a pathogenesis-related protein and is highly expressed in root (SoyBase, http://soybase.org/). Pathogenesis-related proteins produced locally or systematically in the host plant are induced by pathogen attack, and are part of systemic acquired resistance (Van Loon and Van Strien, 1999). A recent study identified pathogenesis-related proteins in xylem sap from SDS-infected soybean (Abeysekara and Bhattacharyya, 2014), indicating that root-sourced pathogenesis-related proteins may play important roles in soybean defense against the SDS fungus. Previous studies showed that infection of plants by phytopathogenic fungi is usually accompanied by damage to the plant cell wall, and may induce wound and defense signaling in host plants (Leon *et al.*, 2001; Vorwerk *et al.*, 2004). In this study, the locus at the 36.2 Mb position on Gm12, which was detected in DAI26 and AUDPC and was involved in three epistatic interactions, harbors seven wound-responsive genes spanning a 30 kb region (Table 2). In addition, the SNP *ss715612644_G_T* is located in the wound-responsive gene *Glyma.12g201000*. Additionally, chitin-responsive genes were also identified in the epistatic analysis (Table 2).

In contrast to foliar disease, relatively little is known about the mechanism of root defenses against soil-borne pathogens (De Coninck *et al.*, 2015). Our results provide insight into the nature of soybean resistance to the root fungal pathogen *F. virguliforme*. A model of soybean root defense against the SDS pathogen, similar to the foliar defense model, is outlined in Figure 6 based on the function of identified candidate genes in this study. First, penetration of the cell walls of soybean root cells by the pathogen leads to cell-wall damage and release of elicitors that trigger wound responses, which are important for induction of jasmonic acid synthesis (Heil and Bostock, 2002). Then elicitors such as chitin oligomers produced by the pathogen are recognized by *R* gene products (e.g. LRR-RLK proteins). Subsequently, the hypersensitive response is induced by recognition of the pathogen in the host plant, resulting in programmed cell death at the site of infection. Concurrently, *R* gene-modulated pathogen-specific resistance is also activated. Finally, the products of chitin-responsive genes and pathogenesis-related proteins are translocated to aboveground organs through the xylem, leading to systemic acquired resistance of the host plant. However, further research is required to validate the predicted causal genes and clarify their direct or indirect interaction network.

We noted that almost all of the loci identified had a minor effect, which is similar to the quantitative resistance to Northern/Southern leaf blight in maize (*Zea mays*; Kump *et al.*, 2011; Poland *et al.*, 2011). A possible explanation is that the association panel used in this study was generated through screening for SDS resistance from over 6000 germplasm accessions (Mueller *et al.*, 2002). The screen for resistant genotypes sacrificed the genetic diversity of

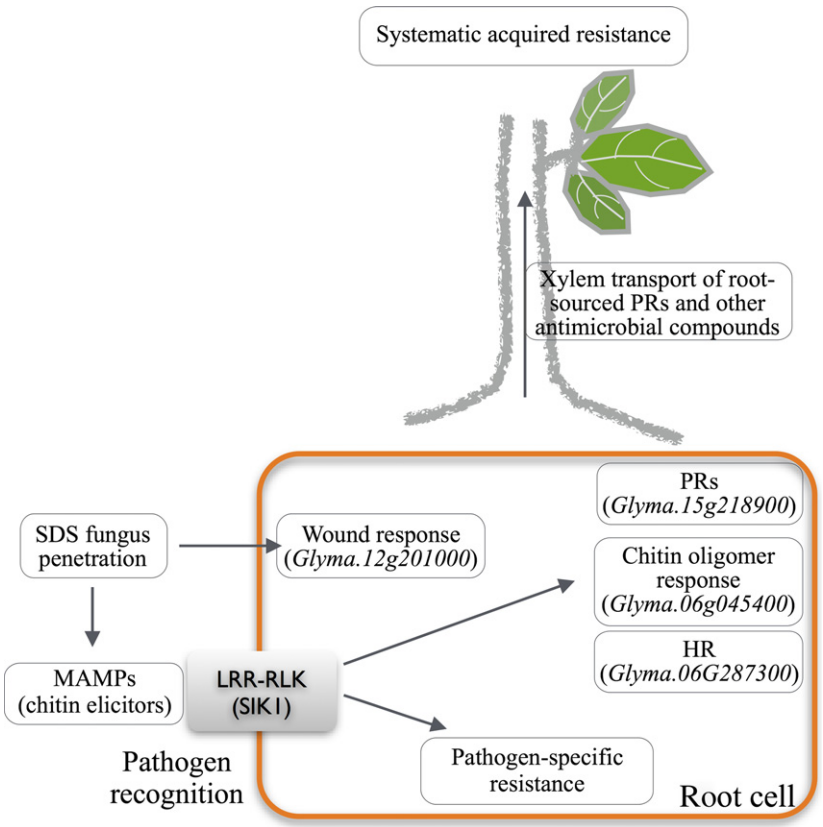

Figure 6. Putative model for soybean defense against sudden death syndrome (SDS) based on the results of genome-wide association and epistasis studies.
MAMPs, microbe-associated molecular patterns; LRR-RLK, leucine-rich repeat receptor-like protein kinase; SIK1, stress-induced receptor-like kinase; HR, hypersensitive response; PRs, pathogenesis-related proteins.

the population by eliminating or decreasing highly susceptible alleles while fixing the major resistant ones. Consequently, valuable resistance alleles that are necessary for resistance may be undetectable because of the way in which this population was selected, and the genetic variation of SDS resistance of the association panel is therefore contributed by numerous minor effect loci. It is also possible that the susceptible alleles fixed in North American germplasm collections may occur at low frequency in the association population and are thus undetectable, or there may be no major resistant alleles in the original germplasm panel. However, the screening process potentially increased the frequencies of some rare SDS-resistant alleles and made them detectable in a population of moderate size.

CONCLUSIONS

The present study identified 12 additive loci and 12 epistatic interactions associated with soybean resistance to SDS at various disease development stages with a high mapping resolution. *SIK1*, a putative *R* gene pinpointed by a trait-associated SNP, and other genes involved in plant defense were identified. The germplasm accession PI 82278 was identified as a valuable donor parent of SDS resistance alleles for improvement of soybean cultivars. More importantly, our results demonstrated that epistatic effects are a substantial complement to additive effects in contributing resistance to SDS in soybean, which provides a foundation for improvement in SDS resistance through marker-assisted selection and genomic selection. This study also suggested the existence of a similar defense network in plant roots as in leaves. Further studies, including expression profiling and functional analyses of the promising candidate genes, will facilitate elucidation of the molecular mechanisms underlying soybean defense to the SDS fungus and other soil-borne pathogens.

EXPERIMENTAL PROCEDURES

Plant materials and greenhouse experiments

Two hundred and fourteen soybean germplasm accessions were selected for this study by successive screening of 6037 PIs for SDS resistance (Mueller *et al.*, 2002). These PI lines originated in China, South Korea and Japan, and correspond to maturity groups II–IV that adapted to the major soybean-producing areas of the Midwest United States (Tables S1 and S6).

Phenotypic data for disease severity on the selected PIs were obtained in a greenhouse experiment using a randomized complete block design with four replications (Mueller *et al.*, 2002). Briefly, SDS-infested plants were rated for SDS foliar symptoms at 20, 23, 26 and 29 DAI. The disease severity rating was based on a scale of 1–9, where 1 = no symptoms, 2–4 = light symptom development (1–5%, 5–10% and 10–20% of foliage affected, respectively), 5 and 6 = moderate symptom development (20–35% and

35–50% of foliage affected, respectively), 7 and 8 = heavy symptom development (50–65% and 65–80% of foliage affected, respectively), and 9 = severe symptom development (80–100% foliage affected). The ratings were then converted to percentage midpoint values to calculate the AUDPC values (Mueller *et al.*, 2002).

Genotyping and quality control

The SNP dataset for the PIs involved in this study was genotyped with the Illumina (https://www.illumina.com) Infinium SoySNP50K BeadChip as described by Song *et al.* (2013) and was downloaded from SoyBase (http://www.soybase.org/). A total of 42 180 SNPs were identified in the association panel. Of them, 284 SNPs that were presented in unanchored sequence scaffolds were excluded from further analyses. The dataset had a missing rate of 0.3%. Markers with a missing rate larger than 10% were ruled out, and the remaining missing data were imputed using BEAGLE version 3.3.1 with default parameter settings (Browning and Browning, 2007, 2009). SNPs with a MAF < 5% after imputation were also excluded from further analyses. Finally, a total of 31 914 SNPs were used for GWAS and GWES.

Linkage disequilibrium and genetic diversity

Pairwise LD between markers was calculated as the squared correlation coefficient (r^2) of alleles using the R package synbreed (Wimmer *et al.*, 2012). Due to the substantial difference in recombination rate between euchromatic and heterochromatic regions, r^2 was calculated separately for the two chromosomal regions. The physical length of euchromatin and heterochromatin on each chromosome is defined in SoyBase (www.soybase.org). Only r^2 values for SNPs with a pairwise distance less than 10 Mb in either the euchromatic or heterochromatic region of each chromosome were used to calculate the mean LD decay using the equation described previously (Remington *et al.*, 2001). The LD decay rate of the population was measured as the chromosomal distance at which the mean r^2 dropped to half of its maximum value (Huang *et al.*, 2010). The nucleotide diversity polymorphism information content was calculated as described previously (Nagy *et al.*, 2012).

Genome-wide association and epistatic interaction analyses

Best linear unbiased predictors of each rating for individual lines were calculated using the R package lme4 (Bates *et al.*, 2012). They were then used to fit the one-way ANOVA model for the naive test (without correction of population structure and familial relatedness) in R 2.15.3 (www.R-project.org) and MLMs in GAPIT software (Zhang *et al.*, 2010; Lipka *et al.*, 2012) for association analyses. The latter takes into account both familial relatedness and population structure (depending on the model fitness).

For the naive test, the equation was

$$y = \mu + X\alpha + e$$

For the MLM analysis, the equation was

$$y = \mu + X\alpha + P\beta + Zu + e$$

where y is the phenotype best linear unbiased predictor of each line, μ is the total mean, X is the incidence matrix relating the individuals to the fixed marker effects α, P is the incidence matrix relating the individuals to the fixed principal component effects β, and Z is the incidence matrix relating the individuals to the random group effects (u) obtained from the compression algorithm. The random group effects follow a multivariate normal distribution with mean 0 and variance/covariance matrix $2KV_g$, where K is the kinship matrix, and V_g is the polygenic variance. The random error term e follows a multivariate normal distribution with mean 0 and variance/covariance matrix IV_e, where I is the identity matrix and V_e is the error variance component.

We performed 1000 permutations to access the empiric significance values for each association as described previously (Zhang *et al.*, 2015), as both the false discovery rate and the Bonferroni adjustment were too stringent. Briefly, for each iteration, we shuffled the rows randomly but kept the order of row names unchanged in a genotypic data file where each column represented one SNP and each row represented one germplasm accession. GAPIT software was run with the same parameter setting and kinship as the original test for each disease rating. However, the threshold for a significant association was determined as the lowest P value of the SNP-trait association that was not significant at an empiric value $P < 0.001$.

The genome-wide epistatic interaction test was implemented in PLINK version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell *et al.*, 2007). The equation was

$$y = b_0 + b_1A + b_2B + b_3AB + e$$

where b_0 is the overall mean, b_1 and b_2 are the additive effects of markers A and B, b_3 is the interaction effect between A and B, and e is the random error following $N(0, \sigma^2_e)$. A Bonferroni threshold ($\alpha = 0.05$) was used to correct the multiple comparisons.

The proportions of phenotypic variance explained by the additive loci identified via GWAS or by the additive loci plus the epistatic loci were calculated as the R^2 of the simple linear regression model by regressing the additive loci or the additive loci plus the epistatic loci on the mean performance of each trait.

Prediction of candidate genes

Genes annotated by the Glyma1.1, Glyma1.0 and NCBI RefSeq gene models on the Glyma.Wm82.a2 (Gmax2.0) reference genome (SoyBase, www.soybase.org) were used as the source of candidate genes. The search for candidate genes was confined to the region defined by clustering of trait-associated SNPs at LD $r^2 > 0.70$ or in a region of 50 kb each side of the peak SNP, whichever was smaller, unless specifically noted. Then the following preferences were applied: (i) genes of known function in soybean related to the trait under study, (ii) genes with known function orthologs in Arabidopsis related to the trait under study, and (iii) genes pinpointed by the peak SNPs.

ACKNOWLEDGEMENTS

The authors are thankful to Iowa State University, the Iowa Soybean Association and the RF Baker Center for Plant Breeding for research support, and to Yuba Raj Kandel (Department of Plant Pathology, Iowa State University) and Jennifer Hicks (Department of Agronomy, Iowa State University) for reviewing the manuscript. The authors thank Perry B. Cregan and Qijian Song (Soybean Genomics and Improvement Laboratory, Agricultural Research Service, US Department of Agriculture) for providing the SNP data for the elite soybean cultivars.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Distribution of SNPs and nucleotide diversity across the soybean genome.

Figure S2. Extensive linkage disequilibrium found on Gm02, Gm08, Gm14 and Gm19.

Figure S3. Frequency distribution of SDS disease severity at 20, 23, 26 and 29 days after inoculation, and the area under the disease progress curve.

Figure S4. Quantile/quantile plots for SDS resistance using various models.

Figure S5. Manhattan plots of GWAS for each trait.

Figure S6. Manhattan plots of GWES for each trait.

Figure S7. Allelic distribution of the 12 identified additive loci associated with SDS resistance in a sample of 96 elite soybean cultivars.

Figure S8. Allelic distribution of the 12 identified epistatic loci associated with SDS resistance in a sample of 96 elite soybean cultivars.

Table S1. List of the PIs in the soybean association panel, their country of origin and maturity groups.

Table S2. Phenotypic variation for SDS disease expression.

Table S3. Correlation coefficients between time-point measurements of SDS disease resistance.

Table S4. Bayesian information criterion (BIC) values for the mixed linear model with various numbers of principal components used for association analyses of each measurement of SDS resistance.

Table S5. Allelic analysis of the SDS resistance-associated loci in PI 82278.

Table S6. Observations of SDS resistance of the 214 PIs in a greenhouse test.

REFERENCES

Abdelmajid, K.M., Ramos, L., Leandro, L., Mbofung, G., Hyten, D.L., Kantartzi, S.K., Grier, R.L., Njiti, V.N., Cianzio, S. and Meksem, K. (2012) The 'PI 438489B' by 'Hamilton' SNP-based genetic linkage map of soybean [*Glycine max* (L.) Merr.] identified quantitative trait loci that underlie seedling SDS resistance. *J. Plant Genome Sci.* **1**, 18–30.

Abeysekara, N.S. and Bhattacharyya, M.K. (2014) Analyses of the xylem sap proteomes identified candidate *Fusarium virguliforme* proteinacious toxins. *PLoS ONE*, **9**, e93667.

Abouzeid, H., Boisset, G., Favez, T. *et al.* (2011) Mutations in the SPARC-related modular calcium-binding protein 1 gene, *SMOC1*, cause Waardenburg anophthalmia syndrome. *Am. J. Hum. Genet.* **88**, 92–98.

Arahana, V.S., Graef, G.L., Specht, J.E., Steadman, J.R. and Eskridge, K.M. (2001) Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in soybean. *Crop Sci.* **41**, 180–188.

Bastien, M., Sonah, H. and Belzile, F. (2014) Genome-wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. *Plant Genome*, **7**(10), 3835.

Bates, D., Maechler, M. and Bolker, B. (2012) lme4: linear mixed-effects models using using 'Eigen' and S4 [WWW document]. URL http://CRAN.R-project.org/package=lme4 [accessed on 13 November 2015].

Browning, B.L. and Browning, S.R. (2007) Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. *Genet. Epidemiol.* **31**, 365–375.

Browning, B.L. and Browning, S.R. (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223.

Brzostowski, L., Schapaugh, W., Rzodkiewicz, P., Todd, T. and Little, C. (2014) Effect of host resistance to *Fusarium virguliforme* and *Heterodera glycines* on sudden death syndrome disease severity and soybean yield. *Plant Health Prog.* **15**, 1.

Chang, S.J.C., Doubler, T.W., Kilo, V., Suttner, R., Klein, J., Schmidt, M.E., Gibson, P.T. and Lightfoot, D.A. (1996) Two additional loci underlying durable field resistance to soybean sudden death syndrome (SDS). *Crop Sci.* **36**, 1684–1688.

Cianzio, S.R., Bhattacharyya, M.K., Swaminathan, S., Westgate, G.M., Gebhart, G., Rivera-Velez, N., Lundeen, P., Van der Molen, K. and Pruski, T.I. (2014) Registration of AR10SDS soybean germplasm partially resistant to sudden death syndrome and resistant to soybean cyst nematode. *J. Plant Regist.* **8**, 200–210.

De Coninck, B., Timmermans, P., Vos, C., Cammue, B.P.A. and Kazan, K. (2015) What lies beneath: belowground defense strategies in plants. *Trends Plant Sci.* **20**, 91–101.

Farm, P.P.S. (1995) Occurrence of soybean sudden death syndrome in east-central Illinois and associated yield losses. *Plant Dis.* **79**, 315.

Gou, X., He, K., Yang, H., Yuan, T., Lin, H., Clouse, S.D. and Li, J. (2010) Genome-wide cloning and sequence analysis of leucine-rich repeat receptor-like protein kinase genes in *Arabidopsis thaliana*. *BMC Genom.* **11**, 19.

Graham, M.A., Marek, L.F. and Shoemaker, R.C. (2002) Organization, expression and evolution of a disease resistance gene cluster in soybean. *Genetics*, **162**, 1961–1977.

Hartman, G., Huang, Y., Nelson, R. and Noel, G. (1997) Germplasm evaluation of *Glycine max* for resistance to *Fusarium solani*, the causal organism of sudden death syndrome. *Plant Dis.* **81**, 515–518.

Hartman, G., Chang, H.-X. and Leandro, L. (2015) Research advances and management of soybean sudden death syndrome. *Crop Prot.* **73**, 60–66.

Hayes, A.J., Jeong, S.C., Gore, M.A., Yu, Y.G., Buss, G.R., Tolin, S.A. and Maroof, M.A.S. (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.

Heil, M. and Bostock, R.M. (2002) Induced systemic resistance (ISR) against pathogens in the context of induced plant defences. *Ann. Bot.* **89**, 503–512.

Hnetkovsky, N., Chang, S.J.C., Doubler, T.W., Gibson, P.T. and Lightfoot, D.A. (1996) Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). *Crop Sci.* **36**, 393–400.

Hu, X., Liu, Q., Zhang, Z., Li, Z., Wang, S., He, L. and Shi, Y. (2010) SHEsisEpi, a GPU-enhanced genome-wide SNP–SNP interaction scanning algorithm, efficiently reveals the risk genetic epistasis in bipolar disorder. *Cell Res.* **20**, 854–857.

Huang, X.H., Wei, X.H., Sang, T. *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–976.

Hwang, E.Y., Song, Q., Jia, G., Specht, J.E., Hyten, D.L., Costa, J. and Cregan, P.B. (2014) A genome-wide association study of seed protein and oil content in soybean. *BMC Genom.* **15**, 1.

Jones, D.A. and Jones, J. (1997) The role of leucine-rich repeat proteins in plant defences. *Adv. Bot. Res.* **24**, 89–167.

Kang, Y., Kim, K., Shim, S., Yoon, M., Sun, S., Kim, M., Van, K. and Lee, S.-H. (2012) Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol.* **12**, 139.

Koenning, S.R. and Wrather, J.A. (2010) Suppression of soybean yield potential in the continental United States by plant diseases from 2006 to 2009. *Plant Health Prog.* doi:10.1094/PHP-2010-1122-01-RS.

Kump, K.L., Bradbury, P.J., Wisser, R.J. *et al.* (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* **43**, 163–168.

Lee, S., Mian, M.R., McHale, L.K., Wang, H., Wijeratne, A.J., Sneller, C.H. and Dorrance, A.E. (2013) Novel quantitative trait loci for partial resistance to *Phytophthora sojae* in soybean PI 398841. *Theor. Appl. Genet.* **126**, 1121–1132.

Leon, J., Rojo, E. and Sanchez-Serrano, J.J. (2001) Wound signalling in plants. *J. Exp. Bot.* **52**, 1–9.

Li, Y.H., Li, W., Zhang, C., Yang, L., Chang, R.Z., Gaut, B.S. and Qiu, L.J. (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytol.* **188**, 242–253.

Lipka, A.E., Tian, F., Wang, Q.S., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. and Zhang, Z.W. (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*, **28**, 2397–2399.

Liu, W.D., Liu, J.L., Triplett, L., Leach, J.E. and Wang, G.L. (2014) Novel insights into rice innate immunity against bacterial and fungal pathogens. *Annu. Rev. Phytopathol.* **52**, 213–241.

Mamidi, S., Lee, R.K., Goos, J.R. and McClean, P.E. (2014) Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). *PLoS ONE*, **9**, e107469.

Manolio, T.A., Collins, F.S., Cox, N.J. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

Mueller, D., Hartman, G., Nelson, R. and Pedersen, W. (2002) Evaluation of *Glycine max* germ plasm for resistance to *Fusarium solani* f. sp. *glycines*. *Plant Dis.* **86**, 741–746.

Mueller, D., Nelson, R., Hartman, G. and Pedersen, W. (2003) Response of commercially developed soybean cultivars and the ancestral soybean lines to *Fusarium solani* f. sp. *glycines*. *Plant Dis.* **87**, 827–831.

Nagy, S., Poczai, P., Cernak, I., Gorji, A.M., Hegedus, G. and Taller, J. (2012) PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. *Biochem. Genet.* **50**, 670–672.

Njiti, V., Suttner, R., Gray, L., Gibson, P. and Lightfoot, D. (1997) Rate-reducing resistance to *Fusarium solani* f. sp. *phaseoli* underlies field resistance to soybean sudden death syndrome. *Crop Sci.* **37**, 132–138.

Njiti, V.N., Meksem, K., Iqbal, M.J., Johnson, J.E., Kassem, M.A., Zobrist, K.F., Kilo, V.Y. and Lightfoot, D.A. (2002) Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. *Theor. Appl. Genet.* **104**, 294–300.

Ouyang, S.Q., Liu, Y.F., Liu, P., Lei, G., He, S.J., Ma, B., Zhang, W.K., Zhang, J.S. and Chen, S.Y. (2010) Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (*Oryza sativa*) plants. *Plant J.* **62**, 316–329.

Park, K.Y., Jung, J.Y., Park, J., Hwang, J.U., Kim, Y.W., Hwang, I. and Lee, Y. (2003) A role for phosphatidylinositol 3-phosphate in abscisic acid-induced reactive oxygen species generation in guard cells. *Plant Physiol.* **132**, 92–98.

Pennell, R.I. and Lamb, C. (1997) Programmed cell death in plants. *Plant Cell*, **9**, 1157–1168.

Poland, J.A., Bradbury, P.J., Buckler, E.S. and Nelson, R.J. (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc. Natl Acad. Sci. USA*, **108**, 6893–6898.

Prabhu, S. and Pe'er, I. (2012) Ultrafast genome-wide scan for SNP–SNP interactions in common complex disease. *Genome Res.* **22**, 2230–2240.

Purcell, S., Neale, B., Todd-Brown, K. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M. and Buckler, E.S.t. (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl Acad. Sci. USA*, **98**, 11479–11484.

Roy, K.W., Rupe, J.C., Hershman, D.E. and Abney, T.S. (1997) Sudden death syndrome of soybean. *Plant Dis.* **81**, 1100–1111.

Rupe, J., Gbur, E. and Marx, D. (1991) Cultivar responses to sudden death syndrome of soybean. *Plant Dis.* **75**, 47–50.

Schmutz, J., Cannon, S.B., Schlueter, J. *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Singh, A., Knox, R.E., DePauw, R.M., Singh, A.K., Cuthbert, R.D., Campbell, H.L., Singh, D., Bhavani, S., Fetch, T. and Clarke, F. (2013) Identification and mapping in spring wheat of genetic factors controlling stem rust resistance and the study of their epistatic interactions across multiple environments. *Theor. Appl. Genet.* **126**, 1951–1964.

Sonah, H., O'Donoughue, L., Cober, E., Rajcan, I. and Belzile, F. (2015) Identification of loci governing eight agronomic traits using a GBS–GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* **13**, 211–221.

Song, Q.J., Hyten, D.L., Jia, G.F., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE*, **8**, e54985.

Srour, A., Afzal, A.J., Blahut-Beatty, L. *et al.* (2012) The receptor like kinase at Rhg1-a/Rfs2 caused pleiotropic resistance to sudden death syndrome and soybean cyst nematode as a transgene by altering signaling responses. *BMC Genom.* **13**, 368.

Staskawicz, B.J., Ausubel, F.M., Baker, B.J., Ellis, J.G. and Jones, J.D.G. (1995) Molecular genetics of plant disease resistance. *Science*, **268**, 661–667.

Van Loon, L.C. and Van Strien, E.A. (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol. Mol. Plant Pathol.* **55**, 85–97.

Vorwerk, S., Somerville, S. and Somerville, C. (2004) The role of plant cell wall polysaccharide composition in disease resistance. *Trends Plant Sci.* **9**, 203–209.

Wang, H., Waller, L., Tripathy, S., St Martin, S.K., Zhou, L., Krampis, K., Tucker, D.M., Mao, Y., Hoeschele, I. and Maroof, S. (2010) Analysis of genes underlying soybean quantitative trait loci conferring partial resistance to *Phytophthora sojae*. *Plant Genome*, **3**, 23–40.

Wang, D., Salah El-Basyoni, I., Stephen Baenziger, P., Crossa, J., Eskridge, K.M. and Dweikat, I. (2012a) Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. *Heredity*, **109**, 313–319.

Wang, H., St Martin, S.K. and Dorrance, A.E. (2012b) Comparison of phenotypic methods and yield contributions of quantitative trait loci for partial resistance to *Phytophthora sojae* in soybean. *Crop Sci.* **52**, 609–622.

Wen, Z., Tan, R., Yuan, J. *et al.* (2014) Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genom.* **15**, 809.

Wimmer, V., Albrecht, T., Auinger, H.J. and Schon, C.C. (2012) synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics*, **28**, 2086–2087.

Woody, J.L., Severin, A.J., Bolon, Y.T. *et al.* (2011) Gene expression patterns are correlated with genomic and genic structure in soybean. *Genome*, **54**, 10–18.

Wywial, E. and Singh, S.M. (2010) Identification and structural characterization of FYVE domain-containing proteins of *Arabidopsis thaliana*. *BMC Plant Biol.* **10**, 157.

Xu, S.Z. and Jia, Z.Y. (2007) Genome-wide analysis of epistatic effects for quantitative traits in barley. *Genetics*, **175**, 1955–1963.

Zhang, Z.W., Ersoz, E., Lai, C.Q. *et al.* (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360.

Zhang, J., Song, Q., Cregan, P.B., Nelson, R.L., Wang, X., Wu, J. and Jiang, G.-L. (2015) Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genom.* **16**, 217.