Association mapping of a locus that confers Southern stem canker resistance in soybean

João Vitor Maldonado dos Santos
Embrapa Soybean
Everton Geraldo Capote Ferreira
Universidade Estadual de Londrina
André Luiz de Lima Passianotto
Universidade Estadual de Londrina
Bruna Bley Brumer
Universidade Estadual de Londrina
Adriana Brombini Dos Santos
Embrapa Soybean
Rafael Moreira Soares
Embrapa Soybean
Davoud Torkamaneh
Universite Laval
Carlos Alberto Arrabal Arias
Embrapa Soybean
François Belzile
Universite Laval
Ricardo Vilela Abdelnoor
Embrapa Soybean
Francismar Marcelino-Guimaraes (francismar.marcelino@embrapa.br)
Embrapa Soybean https://orcid.org/0000-0002-5828-7249

Research article

Keywords: Diaporthe aspalathi, GWAS, haplotype analysis, marker assisted selection (MAS)

Posted Date: July 11th, 2019

DOI: https://doi.org/10.21203/rs.2.11243/v1

Version of Record: A version of this preprint was published on October 31st, 2019. See the published version at https://doi.org/10.1186/s12864-019-6139-6.

Abstract

Background Southern stem canker (SSC), caused by Diaporthe aspalathi (E. Jansen, Castl. & Crous) is an important soybean disease, which has been responsible for severe losses in the past. The main strategy to control this fungus is through the introgression of resistance genes. So far, five main loci have been associated with resistance to Southern stem canker. However, there is a lack of information about useful allelic variation at these loci. In this work, a genome-wide association study (GWAS) was performed to identify allelic variation associated with resistance against Diaporthe aspalathi and to provide molecular markers useful in breeding programs. Results We characterized the response to Southern stem canker infection in a panel of 295 accessions from different regions of the world including important Brazilian elite cultivars. Using a GBS approach, the panel was genotyped and we identified marker loci associated with Diaporthe aspalathi resistance using GWAS analysis. We identified 19 SNPs associated with Southern stem canker resistance, all on chromosome 14. The peak SNP showed an extremely high degree of association (p-value = 6.35E-27) and explained a high level of the phenotypic variance (R2 = 70%). This strongly suggests that a single major gene is responsible for resistance to D. aspalathi present inn most of the lines comprising this panel. We also identified in resequenced soybean materials other SNPs in the region identified by GWAS in the same LD block that clearly differentiate resistance and susceptible accessions. The peak SNP was selected and used to develop a cost-effective molecular marker assay, which was validated in a subset of the initial panel. In an accuracy test, this SNP assay demonstrated 98% of selection efficiency. Conclusions Our results suggest a relevant importance of this locus in SSC resistance in soybean cultivars and accessions from different countries and the SNP marker assay developed in this study can be directly applied in MAS studies in breeding programs to select resistance materials against this pathogen and support its introgression.

Background

The cultivated soybean [*Glycine max* (L.) Merrill] is one of the most important crops worldwide. It has been estimated that wild soybean (*Glycine soja*) was domesticated to cultivated soybean around 7,000–9,000 years ago in Asia, but reached the Americas only on 18th century [1]. Nowadays, the Americas are responsible for 90% of the world production. In Brazil, soybean is a major agricultural commodity with a production of 119 M tons from 35 M hectares of cultivated area in the 2017/18 growing season [2]. Due its major importance for the Brazilian economy, a large number of studies have been undertaken to better understand genetic variation in the soybean genome and its relationship to traits of interest [3].

An important barrier to increased soybean production and seed quality is the large number of biotic factors that affect soybean production. Among the main pathogens responsible for meaningful losses in soybean fields is Southern stem canker (SSC). It is caused by the fungus *Diaporthe aspalathi*, anamorph *Phomopsis aspalathi* (Cooke & Ellis), belonging to the *Diaporthe/Phomopsis* complex, one which is associated with other diseases in soybean such as seed decay as well as pod and stem blight. Historically, two causal agents of SSC have been described: *Diaporthe phaseolorum* var *. meridionalis* (*Dpm*) F.A. Fernández, *Diaporthe phaseolorum* var. *caulivora* (*Dpc*) K. L. Athow & R. M. Caldwell. Recently, the names of these

species (*Dpm* and *Dpc*) have been changed to *Diaporthe aspalathi* (E. Jansen, Castl. & Crous) (*Da*) and *Diaporthe caulivora* (Athow & Caldwell) J.M. Santos, Vrandecic & A.J.L. Phillips (*Dc*), respectively [4–6].

The *Da* fungus was reported for the first time in Brazil during the 1989/90 soybean cropping season in the states of Paraná and Mato Grosso, and in the following cropping season, SSC was observed in almost all soybean production areas in the country [7, 8]. In 1994, SSC was responsible for losses of 1.8 million metric tons in Brazil, making it the most serious disease for the Brazilian soybean crop at that time [9].

Currently, the genetic resistance is the main method of control and most of the cultivars being cropped carry resistance genes to SSC. To date, five major dominant, non-allelic SSC resistance loci (*Rdc1*, *Rdc2*, *Rdc3*, *Rdc4* and *Rdc5*) have been reported [10, 11]. Moreover, another source of resistance, distinct from *Rdc1-4*, was identified in PI 398469 and has provisionally been named *Rdc?* [12]. However, these loci were identified using *Da* isolates from the southern United States, and according to other studies, genes that confer resistance to one pathogen do not confer resistance to another [12,13]. Due to this fact, it was proposed to rename the major loci related to *Da* resistance as *Rdm1*, *Rdm2*, *Rdm3*, *Rdm4*, and *Rdm5* [13, 14]. Recently, *Rdm4* and *Rdm5* were mapped closely on chromosome 08 in the cultivar (cv.) Hutcheson [15]. The knowledge associated with the accurate localization of major genes responsible for host plant resistance to a pathogen is an important step in the identification of molecular markers that may be helpful on the development of resistant cultivars to SSC. In this respect, genome-wide association studies (GWAS) may offer a great opportunity to identify and to find these resistance genes as well markers associated with resistance, thereby representing an important tool for breeding programs.

The advent of new platforms for large-scale sequencing associated with the complete sequencing of the soybean genome [16] allowed the genome-wide identification of a great number of variations that can be used both to characterize nucleotide and structural diversity in collections of soybean accessions and to perform GWAS. A large number of GWAS are already available for soybean. Hwang et al. [17] identified 40 Single Nucleotide Polymorphisms (SNPs) associated with protein content in 17 different genomic regions. Moreover, in this study, 25 SNPs were related to the control of the oil content in 13 genomic regions. Two different studies identified QTLs associated with resistance to *Sclerotinia sclerotiorum* [18, 19]. Mamidi et al. [20, 21] generated two studies on iron deficiency chlorosis (IDC). Finally, Contreras-Soto [22] using a panel of 169 soybean cultivars, identified 17, 59 and 11 SNPs associated with 100-seed weight, plant height and seed yield, respectively.

Despite the emergence of a large number of GWAS analyses, many of these studies were carried out using SNPs obtained from Genotyping By Sequencing (GBS) approach and, due to this fact, may not have ensured a full coverage of the soybean genome. Improved marker coverage can be achieved using whole-genome sequencing (WGS) data, and such exhaustive data can be useful to identify and refine regions identified by GWAS performed with SNPs from GBS. For example, Zhou et al. [23] found associations for 10 selected regions and 13 previously uncharacterized agronomic loci, which included pubescence form, plant height, and oil content. Maldonado dos Santos et al. [24] identified 5.8 million SNPs, and 1.3 million InDels, in 28 Brazilian soybean cvs. that could be used as a complementary source of information in GWAS studies. Valliyodan et al. [24] detected over 10 million SNPs in 106 soybean genomes, some of

which some were associated with oil and protein content, salinity, and domestication traits. Recently, a genome-wide study was developed and it was responsible for the identification of two genes with a relevant association with a soybean seed permeability trait in *Glycine max* and *Glycine soja* [25]. These studies highlighted the great power of whole-genome sequencing technologies for GWAS.

SSC is mainly controlled by the introgression of resistance genes in elite cultivars and these genes are present in most cultivars released over the last 20 years in Brazil. However, the potential for considerable damage remains if the current resistance genes are overcome by the pathogen. Thus, molecular characterization of SSC resistance loci in a diverse set of soybean germplasm is essential to understand the genetic base of the resistance. Therefore, the objective of this study was to identify allelic variation associated with resistance against *Da* in a diverse panel containing a broad distribution of soybean cultivars and plant introductions from different regions of the world.

Results

Phenotypic evaluation for Southern stem canker resistance in soybean accessions

All accessions were inoculated with mycelium from CMES 480 isolate using the toothpick method under greenhouse conditions [26, 27]. The results of the inoculation were expressed as percentage of dead plants (%DPs) and all the differential genotypes showed a small lesion at the point of the stem where the toothpick penetrated, indicating that an infection had successfully occurred in all the inoculated plants. The resistance sources of SSC resistance, the cultivars Tracy-M (*Rdm1*/*Rdm2*), Crockett (*Rdm3*) and Hutcheson (*Rdm5*) showed complete resistance against the *D. aspalathi* isolate CMES 480, while PI 398469 (*Rdm*?) also showed a high degree of resistance, but we still observed 3% of DPs. On the other hand, the interactions between CMES 480 and the accessions with the *Rdm1* (D85-10404), *Rdm2* (D85-10412) and *Rdm4* (cv. Dowling) genes were all compatible, such that these accessions were all highly susceptible (Table 1). The isolate CMES 480 was able to be recognized by multiple R genes, thus, leading the possibility to identify different resistant loci if they are distributed in the GWAS-panel.

Southern stem canker symptoms were evaluated at 60 days after inoculation and, as expected, known resistant (cv. Tracy-M) and susceptible (cv. BR 23) accessions showed a highly contrasting result (Figure 1a). The resistant plants showed only a small necrosis in the stem tissue around the toothpick, presence of callus at the toothpick insertion point and no damage to plant development. On the other hand, susceptible accessions presented both infected and dead plants, the infected plants being identified on the basis of the absence of callus, a reduction in the development of the aerial part of the plant, a large necrosis at the point of inoculation, as well as chlorotic and withered plants. Another parameter that easily distinguished resistant and susceptible plants was the length of the internal lesion; resistant plants usually showed a lesion length less than 1 cm, unlike susceptible plants, which presented lesions greater than 1 cm (Figure 1b).

The pathogenicity test was carried out for all the 295 accessions used on the GBS panel, and 205 were considered resistant and 90 susceptible. Just to highlight the diversity of the panel, among the resistant plants, 26% of the accessions came from China, 22% from Brazil, 20% from Japan and 12% from the USA. In the susceptible group, Brazil contributed with 33% of the susceptible accessions, the USA with 20%, China with 18% and South Korea with 17%. Based on the year of release/cataloging of the materials, in the resistant group, China and North Korea contains the oldest resistant accessions from the panel (1930's), while cv. Tropical and cv. Doko are the oldest resistant Brazilian materials (1980's). In the susceptible group, PI 090763 from China (1930's), PI 196170 (South Korea), accessions from Japan (1950's), cv. Santa Rosa (1957), and the American cultivars Bragg and Davis (1960's) are examples of the oldest susceptible materials composing this panel.

Identification and mapping of the Southern stem canker resistance locus

The Fast-GBS pipeline produced about 50,000 high-quality SNPs from GBS data. Using a MAF of ≥0.05 as cut-off, it was selected a total of 32,836 polymorphic SNP markers that we used in GWAS. The resulting SNPs were distributed over the whole genome. These SNPs covered proportionally all soybean chromosomes, with a mean SNP density of one SNP every 29.1 Kbp and a mean of 1,642 SNP markers per chromosome. The greatest number of SNPs was detected on chromosome 18 (2,845 SNPs), followed by chromosome 4 (2,145 SNPs), and the lowest was observed on chromosomes 12 (951 SNPs) and 11 (959 SNPs) (Additional File 1). Regarding population structure, Principal Component Analysis (PCA) were performed, which PC1 explained about 9% of variance, PC2 about 7% and the PC3 about 4%; together, the three PCs explained about 20% of the total genetic variance (Figure 2a and 2b). The GWAS was performed with the compressed mixed linear model (cMLM), which accounts for population structure (PCA) and relatedness by kinship matrix (K matrix). The quantile-quantile plot showed that observed p-values strongly deviated from the expected p-values in few SNPs, which means that the cMLM model was appropriated for GWAS performed (Figure 2c). We identified a single locus on chromosome 14, in which a total of 19 SNPs showed significant associations (FDR < 0.001) with SSC resistance (Figure 2d). Among these significant SNPs, the FDR-adjusted *p*-value ranged between 6.35E-27 to 4.13E-09, with SNPs explaining approximately 50% to 70% of the total phenotypic variation (Table 2).

The interval delimited by the significant SNPs extends just over 400 kbp, although the three most significant SNPs were located within a span of 34 kbp, thus identifying a very specific region. Within this region, the most significant SNP resided inside *Glyma.14g024300* (a DEA(D/H)-box RNA helicase family protein), the second most significant SNP resided inside *Glyma.14g024100* (a Rho GTPase-activating protein), andthe third most significant SNP was located inside *Glyma.14g23900* (a methionine sulfoxide reductase).

Based on the results, the peak SNP by itself was enough to separate the resistant and susceptible accessions with a high level of concordance. At the peak SNP (1,744,370 – SNP1), the C allele was detected in 194 resistant accessions, while four resistant accessions were heterozygous and the remaining

seven resistant accessions showed the T allele. Similarly, an elevated concordance between the phenotype and genotype was observed among susceptible materials. Among 90 susceptible accessions, 71 showed the T allele. Of the 19 apparent discrepancies, 16 accessions were heterozygous and the remaining three carried the C allele. A comprehensive description of SNP genotypes (at all 19 significant positions) and phenotype for each accession is provided in Additional File 2.

Among the differential accessions, in the peak SNP, the C allele was detected in all accessions that showed resistance to isolate CMES 480, as well as in the susceptible D85-10404, which is a line derived from cv. Tracy-M. In another hand, cv. Dowling and D85-10412 line showed both the susceptible phenotype and the T allele (Additional File 3).

We performed a haplotype analysis for the 295 accessions using SNPs associated with SSC resistance. First, from the initial 19 SNPs significant associated, we eliminated the redundant SNPs, which means, SNPs associated with SSC but bringing the same information. After that, we obtained four haplotypes containing the combination of four SNPs able to discriminate the main SSC resistance sources and grouped the accessions presented in the panel (Table 3). The haplotype 1 was present in the majority of resistant materials, shared by cv. Hutcheson and the PI 398469 and just one susceptible accession. The haplotype 2 was shared only by cv. Crockett and 35 resistant accessions. The haplotype 3, shared by the cv. Tracy-M and line D85-10404 was also present in 22 resistant and two susceptible accessions. Finally, the haplotype 4 was distributed in 70 susceptible accessions and by Dowling and line D85-10412, and other 5 resistant accessions.

Whole-genome sequencing at the resistance locus interval reveals additional allelic variation

Analysis of the region associated with resistance against *Da* was performed by examining allelic variation 278 kb upstream and 200 kb downstream of the first peak SNP of the GWAS in the resequencing soybean dataset. This specific interval was based on SNPs with D' values higher than 0.7, according to the LD analysis. (Additional File 4). We observed a total of 4,440 SNPs and 1,105 InDels in this interval (Table 4). Of these, 3,375 SNPs were identified in non-coding regions, 421 in intronic regions, 247 in UTR regions, and 397 in exons. Among the latter, 248 non-synonymous SNPs were also observed in 39 different genes. Moreover, there were 69 InDels in UTR regions, 98 InDels in introns, and 37 InDels in exons. Twenty-three InDels were responsible for a frameshift modification in nine different genes.

The most significant SNP was a non-synonymous modification located at exon 6 of the gene *Glyma.14G024300* (a DEAD/DEAH box RNA helicase). We also identified another three non-synonymous SNPs associated with this gene (Figure 3), which are in perfect LD with the first peak SNP, and could not be detected by GBS strategy, due the lower coverage of the technique compared to the whole genome sequencing. Unsurprisingly given the large size of the haplotype block comprising the peak SNP, we observed 203 SNPs and 46 InDels in perfect LD with the first peak SNP of the GWAS. Some of these allelic variations are distributed inside genes in the interval, which present structural domains commonly found in

resistance genes, revealing other potential candidate genes to SSC resistance. Fifteen non-synonymous SNPs were observed in eight genes, including two Leucine-rich-repeat receptor-like protein kinases (LRR-RPK) (*Glyma.14G026300*, and *Glyma.14G026500*), a serine-threonine protein kinase (PRSTK) (*Glyma.14G026700*), a PH domain LRR-containing protein phosphatase 1 (*Glyma.14G024400*), a methyltransferase (*Glyma.14G026600*), an acid phosphatase related (*Glyma.14G024700*), and a gene involved in DNA repair (*Glyma.14G026900*) (Table 5).Finally, an insertion of two nucleotides responsible for a frameshift modification in the exon of LRR-RPK (*Glyma.14G026500*) was observed only in susceptible cvs. based on our analysis. To confirm the association of these allelic variations, and the role of the potential candidate genes to the resistance against SSC, functional validation should be conducted in future studies.

Allelic discrimination using *Rdm* SNP KASP assay

The peak SNP (1,744,370) was selected to develop a KASP assay in order to confirm the alleles obtained by GBS and to apply this assay in future MAS. Thus, a subset of 146 accessions from the GWAS panel was analyzed with this assay and, as expected, all the alleles/genotypes obtained by GBS were the same using the KASP assay (Additional File 5). Furthermore, the assay developed was able to correct the heterozygous genotypes obtained by GBS (Figure 4). Among the accessions showed as heterozygote in the peak SNP, 15 accessions were present in the subset analysed with the assay and all were found to be homozygous.

Therefore, the efficiency of the SNP marker, and type I/II error rates were calculated and shown in Table 6. The SNP1 marker was present in 98% of the accessions phenotyped as resistant, resulting in a low type I error rate (2.4%), which suggests a low probability of erroneously selecting a susceptible line based on the marker genotype. In addition, the marker also presented a low type II error rate or false negative rate of 1.19%.

Discussion

Southern stem canker reaction in the GWAS panel

Resistance to Southern stem canker is an important trait to release a new soybean cultivar, considering that this disease presents a high potential to cause losses of up to 100% in soybean fields [8]. Nowadays, almost all soybean cultivars registered in Brazil and in other countries are resistant to Southern stem canker. However, few genetic studies had documented the main sources of resistance present in soybean cultivars. Regarding the Brazilian cultivars, there are not genetic studies showing the main SSC resistance sources present in Brazilian germplasm.

Considering the importance of SSC for Brazil, recently, Brumer et al. characterized a Brazilian collection of isolates of the pathogen comprising samples collected in different regions and years, and demonstrated

the occurrence of at least three different races in Brazil [28]. Only the sources Tracy-M (*Rdm1*/ *Rdm2*) and cv. Crockett (*Rdm3*) showed resistance reaction for all isolates in that study, thus these genes becoming target for the plant breeding programs. Given that fact we did not know about the main sources in our GWAS-panel, the isolate CMES 480 was selected to our phenotyping approach due to show incompatible reactions when inoculated in the main SSC resistance sources, the cultivars Tracy-M, Crockett, Hutcheson and the PI 398469.

In the present study, the method carried out was the toothpick inoculation, which has been used successfully in the evaluation of soybean materials since the onset of the disease in the late 1980s [8, 13, 26, 28]. In our panel, 205 accessions were classified as resistant by this inoculation method, including the differential genotypes like cv. Tracy-M, cv. Crockett, cv. Hutcheson and PI 398469, confirming their resistance obtained in other studies [8, 10–12, 29–33]. Therefore, demonstrating the good reproducibility of this approach to assess the correct SSC phenotype in the accessions, which is a crucial step to obtaining good results in GWAS.

Genomic wide association study for Southern stem canker disease

Using a MAF of 5%, we filtered from initial SNPs data about 36 K SNPs, which were used in the GWAS analysis. The SNPs were distributed in all the soybean chromosomes and, as expected, larger number of SNPs were detected on the largest chromosomes, as we can see on chromosome 18. In another hand, a smaller number of SNPs were detected on the smallest chromosomes, as chromosome 11. A very similar pattern of distribution of the SNPs was obtained in recent GWAS for resistance to *Sclerotinia sclerotiorum* [19] and *Meloidogyne incognita* [34]

The GWAS conducted in the present work revealed a highly significant association of resistance to SSC with a 478 kbp region on chromosome 14. Therefore, we may assume that the main SSC resistance presenting in our panel is related to this region, although previous studies of genetic mapping have detected other loci involved in SSC resistance, and even we have used an isolate able to select to different R genes. In the present study, we used the CMES 480, which selects for different R genes, thus, we may not completely assume that the peak SNP in the chromosome 14 is associated with resistance locus in all the accessions, indeed, probably, some accessions showed resistance derived by other R genes located in other genomic regions.

A similar region on chromosome 14 was recently identified by a GWAS analysis conducted with SNPs from SoySNP50K array and using the USDA germplasm bank phenotype information [35]. In that study, it was also identified two SNPs associated with resistance to SSC caused by *D. aspalathi and D. caulivora* on chromosome 14 in a region spanning about 400 kb. However, it was previously demonstrated that the *Rdm1-Rdm5* genes that confer resistance to *D. aspalathi* do not confer resistance to *D. caulivora* [13], leading an assumption that the region might contains different R genes for both *D. aspalathi* and *D. caulivora*. In our study, all accessions were screened for SSC resistance in a same experiment with the pure

isolated of *D. aspalathi,* previously characterized both, morphological and molecularly [28]. The SNP (ss715617869) previously identified related to SSCresistance[35] is located in the 1,731,256 pb in chromosome 14, while the three peak SNPs detected in our association analysis are located in the interval between 1,710,287-1,744,370. So, our SNPs overlapped the region identified by Chang et al. [35], suggesting that the region identified in both studies are related by SSC caused by *D. aspalhati*.

Interestingly, although the peak SNP was present in almost all the SSC sources, that haplotypes described was able to differentiate the main resistance sources, leading to infer about the origin of R gene conferring the resistance present in the accessions. Most of the resistant materials in the panel shared the cvs. Hutcheson and PI 398469 haplotype (Additional File 2). Therefore, we may assume that the SSC resistance in the panel is the same from these sources. In contrast, Chiesa et al. [15], using *D. aspalhati* isolates and $F_{2:3}$ populations derived from cv. Hutcheson, reported the genetic mapping of *Rdm4 and Rdm5* on the chromosome 8, indicating different regions conferring resistance present in this source. The use of different isolates in each study, which means, isolates selecting to different R genes, and the panel composition are the main explanations, once it has a direct consequence in the region identified in the mapping studies. Similarly, the others sources as cv. Crockett and cv. Tracy-M showed specified haplotypes each one and a considerable part of the resistant accessions were grouped in these haplotypes, leading an assumption that these accessions probably contains the same resistance source shared by the these sources.

Other studies have been showed the success of haplotype analysis in order to discriminate resistance sources in soybean. Pham et al. [36] performed the fine mapping of the resistance to *Cercospora sojina* K. Hara in two accessions, and constructed the haplotype using 11 SoySNP50K SNPs in the known resistance source (cv. Davis) and 45 lines and cultivars, obtained unique haplotype to this two resistance accessions. Furthermore, they analyzed haplotype allele variation in *Rcs3* locus (a resistance gene to *C. sojina*) in the same accessions panel. It was observed that Davis haplotype was shared with only four cultivars and not for the two resistance accessions, which suggested that all the cultivars with Davis haplotype may have the same resistance sources and confirming the unique resistance haplotype to the other two accessions. In other recent study, King et al. [37] mapped the *Rpp4-b* locus in PI 423971 and using five SoySNP50K SNPs made the *Rpp4-b* haplotype, which was unique to the PI 423971 and by just four lines, while the all other *Rpp* sources genotypes and the 32 susceptible soybean ancestors did not have this haplotype. Then, the authors suggested that these lines may possess *Rpp4-b* locus. Altogether, these studies and our results demonstrate the applicability of the haplotype analysis in order to have initial information about the resistance sources and the success of discriminate these sources.

Considering that some *D. aspalathi* Brazilian isolates were able to cause disease in the cv. Hutcheson and PI 398469 [28], but not in the cv. Crocket and cv. Tracy-M, is possible to that the SNPs associated with SSC on chromosome 14 might be linked with one or more *Rdm* genes in the region, however, in order to confirm this hypothesis, a further study of fine mapping need to be conducted in a bi-parental population obtained from independently crosses with these resistance sources. Therefore, we chose to name this locus as just a common locus of resistance to Southern steam canker present in many different soybean accessions

evaluated in this study. Furthermore, based on our results, the KASP assay using the most significant SNP associated with SSC in soybean can be considered useful to breeding programs in marker-assisted selection for SSC resistance.

New allelic variations based on resequencing analysis of soybean genomes

To confirm our results, we examined the nucleotide variation on a whole-genome resequencing data in a collection of 51 accessions that were also characterized for their reaction to SSC isolates. In the latter case, the SNP haplotypes in the vicinity of the SNPs shown to be significantly associated with *Da* resistance in the GWAS analysis were again clearly associated with the disease reaction.

The most significant SNP associated with SSC resistance based on GWAS was identified in *Glyma.14G024300*, a DEAD/DEAH box RNA helicase, described as involved in important biological processes, such as transcription, translation initiation, mRNA splicing and export, and ribosome biogenesis [38–41]. There are a large number of studies which have associated DEAD-box RNA helicases to soybean stress, such as salt stress [38, 42], cold tolerance [38, 43], and resistance to a fungal pathogen [44].

Moreover, we identified allelic variations in perfect LD with SNP1 in LRR-RPK genes (*Glyma.14G026300*, and *Glyma.14G026500*). In *Arabidopsis thaliana*, several studies associated LRR-RPK with defense mechanisms. It was described a LRR-RPK gene as a positive regulator of ABA response during stress response and plant development [45]. Another study in *Arabidopsis* identified that the ERECTA gene, previously described associated with development pathways, were also related to the resistance against bacterial blight[46]. In soybean, there are some studies linked LRR-RPK genes associated with stress. It was observed in *Glycine soja* that the overexpression of GsLRPK gene contribute to an increase of tolerance to cold [47]. Finally, a RNA-seq study of the *Rbs3* locus aid to identify some candidate genes associated with the resistance against brown stem root, which included some LRR-RPK genes [48]. Besides LRR-RPK genes, allelic variations were also observed in PRSTK (*Glyma.14G026700)*. Plant-receptor-like serine/threonine kinase was one of the first genes cloned and associated to defense mechanisms, which plays a key role in signal transduction pathway in plants [49, 50]. It was described the presence of PRSTK involved in a defense response due the interaction plant-pathogen in some organisms, such as rice [51], *Arabidopsis thaliana* [52], and soybean [53, 54]. The existence of a non-synonymous SNP and InDels in coding region of these genes associated with plant stress studies could clarify the plant defense mechanisms to SSC resistance. Thus, the DEAD-box RNA helicases (*Glyma.14G024300*), LRR-RPK (*Glyma.14G026300*, and *Glyma.14G026500*), and PRSTK (*Glyma.14G026700*) genes might be interesting targets for future functional studies to determinate the effect of these genes in soybean during *Da* infection.

Conclusion

In this study, we identified and confirmed the location of an important locus related to SSC resistance in soybean. At least three important sources of resistance to SSC, PI 398469, and the cv. Hutcheson and Crocket presented the locus mapped on chromosome 14. The peak SNP identified was able to correctly define the resistant accessions in the panel with high precision. The marker assay associated with the *Rdm* locus developed is a useful tool in breeding programs for marker-assisted selection to identify accessions carrying the allele conferring resistance against infection by *D. aspalathi* and to follow its introgression. Our results demonstrated for the first time the relevance of *Rdm* locus on chromosome 14 for the resistance to SSC in Brazilian cvs and other alternatives sources of resistance. Additionally, we characterize a significant number of plant accessions and cvs sharing different resistant haplotypes, which can be exploited by breeders.

Materials And Methods

Plant materials

The source material for the analysis comprises a set of 295 soybean accessions (Additional File 6) representing different maturity groups and various regions of origin such as China, Japan, North and South Korea, Russia, the United States, India and Brazil among others. The panel included accessions carrying previously described resistance genes (in parentheses): cv. Tracy-M (*Rdm1/Rdm2*), D84-10404 (*Rdm1*), D84-10412 (*Rdm2*), cv. Crockett (*Rdm3*), cv. Dowling (*Rdm4*), cv. Hutcheson (*Rdm4/Rdm5*) and PI 398469 (*Rdm?*), while the Cultivar BR23 served as a susceptible control. The seeds were obtained from the Embrapa Soybean germplasm bank.

Phenotypic evaluation for stem canker

The soybean accessions composing the GWAS panel and the accessions subjected to WGS were infected with isolate CMES 480 of *D. aspalathi* (collected in Rio Verde (GO) in 2001) and evaluated in a greenhouse at Embrapa Soybean, in Londrina (PR, Brazil) in 2015. The phenotyping was conducted using the toothpick method with colonized mycelium as described by Keeling [26], and modified by Yorinori [27]. The experimental design was completely randomized with two replicates containing 10 plants in each pot. In both phenotyping trials, all inoculations were carried out on 10- to 15-day-old seedlings that were kept under high humidity (45-second nebulization every hour throughout the day), and average temperatures of 26 ± 4 ºC (day) and 17 ± 3 ºC (night). As a negative control, cv. BR 23 was inoculated with sterile toothpicks without mycelium. The evaluation of each genotype was performed 60 days after inoculation by counting the number of dead plants (DPs). The percentage of DPs (% DP) was calculated according to the method described by Yorinori [27]: %DP = {[DP + (IP/2)]/TP}*100, PI being the total number of infected plants and TP the number of inoculated plants.

The accessions were classified based on plant-fungus interaction reactions described by Yorinori [27] and modified by Pioli et al. [13] in two categories: i) incompatible or avirulent (0-14.9% DP), which means,

accession was considered resistant to the isolate and ii) compatible (> 15% DPs), plants classified as susceptible to SSC.

DNA extraction and GBS library preparation

The DNA was extracted with the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) according the manufacturer's instructions, and subsequently quantified using a Nanodrop 8000 spectrophotometer (Thermo Fischer Scientific Inc., Waltham, MA, USA). Then, the samples were diluted to 10 ng/ul. The GBS libraries were constructed according to the protocol described by Elshire et al. [55], as modified by Sonah et al. [56]. A part of the resulting libraries was sequenced on Illumina HiSeq2000 (McGill University-Genome Quebec Innovation Centre, Montreal, QC, Canada) and another part via Ion Torrent (IBIS - Institute of Integrative Biology and Systems, Université Laval, Quebec City, QC, Canada) sequencers.

SNP identification and GWAS analysis

Illumina and Ion Torrent read processing, mapping, SNP/indel calling and genotyping were performed using the Fast-GBS pipeline [56]. Any heterozygous calls were replaced with missing data and only SNPs with less than 80% missing data were kept. Indels were not used in the downstream analyses. Imputation of missing data was performed using Beagle [57]. Marker-trait associations were calculated with the GAPIT R package [58] using a compressed mixed linear model (cMLM). In order to control for population structure and relatedness between individuals, we used the first three principal components (PCs) obtained from principal component analysis (PCA) and the VanRaden kinship matrix in the GWAS model. We declared SNPs as being significant at an FDR-adjusted *p*-value of less than 0.001.

Haplotype analysis and linkage disequilibrium detection

First, we performed haplotype analysis on the GWAS panel using the set of 19 SNPs most highly associated with SSC resistance in the GWAS. Then, we removed the redundant SNPs, and the haplotype in the differential lines was created and haplotypes selecting most of resistant accessions were obtained. We carried out an analysis of linkage disequilibrium (LD) decay using the GBS-derived SNP dataset on the GWAS panel, with the PopLDdecay 3.30 software package and the LD was measured using the squared allele frequency correlations (r^2).

Further, we investigated the allelic variation present in a subset of 51 accessions comprising 27 Brazilian soybean cvs. [59] and 23 other accessions from the center of origin [24], as well as PI 595099, and Williams 82 (reference genome) for the putative resistance locus mapped in this study using WGS data (Additional File 7). We used an LD analysis to identify SNPs associated with the peak SNP identified by GWAS. We used TASSEL software to generate r^2 values, and to determine which SNPs were in LD with the peak SNP. Finally, we used SnpEff [60] as a way to detect SNPs associated with candidate genes in the

soybean genome. The focus of this analysis was allelic variation within genes located within the region identified based on GWAS analysis. Graphical genotype visualization was performed using Flapjack [61].

SNP assay design and genotyping

For the development of markers to be used for high throughput genotyping, the peak SNP identified in the GWAS was selected, and a Kompetitive Allele Specific PCR (KASP) assay was designed. For SNP marker validation, a subset of the GWAS panel, comprising 146 resistant and susceptible accessions were selected, including the seven differential lines [Tracy-M (*Rdm1*/ *Rdm2*), D85-10404 (*Rdm1*), D85-10412 (*Rdm2*), Crockett (*Rdm3*), Dowling (*Rdm4*), Hutcheson (*Rdm4*/ *Rdm5*) and PI 398469 (*Rdm?*)], (Additional File 5). DNA extraction was carried out using the DNeasy Plant Mini Kit. Briefly, for the KASP assay, the final volume of the reaction was 5.07 µL, containing 2.5 µL of diluted DNA (10 ng/ul), 1x KASP master mix, and 0.0014x KASP assay mix. The SNP genotyping was performed using an ABI7900 equipment, following a touchdown thermal cycling protocol described by the manufacturer. Genotypes were acquired and clustered using the Taqman Genotyper Software v2.1 (Life Technologies, Applied Biosystems Inc.; Foster City, CA, USA).

List Of Abbreviations Used

SSC: Southern Stem Canker

Dpm: Diaporthe phaseolorum var *. meridionalis*

Dpc: Diaporthe phaseolorum var. *caulivora*

Da: Diaporthe aspalathi

Dc: Diaporthe caulivora

cv.: cultivar

GWAS: Genome-Wide Association Analysis

SNPs: Single Nucleotide Polymorphisms

QTLs: Quantitative Trait Loci

CNVs: Copy-Number Variations

DP: Dead Plants.

IP: Infected Plants.

TP: Total Plants.

bp: base pair,

Mbp: Megabase pair

cMLM: compressed Mixed Linear Model

mAF: minor Allele Frequency

LD: Linkage Disequilibrium

kbp: kilobase pair

LRR-RPK: Leucine-rich-repeat receptor-like protein kinase

PRSTK: Plant-receptor-like serine/threonine kinase

MAS: Marker-Assisted Selection

Declarations

Competing Interests

The authors declare that they have no competing interest.

Availability of data and materials

The supporting data are available in the additional files.

Authors' contribution

FCMG, CAA, ALLP, and RVA conceived and planned the study; BBB, EGCF, and RMS were responsible for the phenotyping; ALLP, BBB, EGCF and ABS performed DNA extraction and sent the samples for GBS sequencing; JVMS, EGCF, DT, and ALLP performed the bioinformatics analyses and data interpretation. BBB, ABS and EGCF executed the SNP assay standardization and samples genotyping; JVMS, EGCF, ALLP and BBB wrote the manuscript; FB, JVMS, FCMG, and RVA edited the intellectual content of the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

We would like to thank the Coordination for the Improvement of Higher Level for Education (CAPES) program and the National Council of Technological and Scientific Development (CNPq) for financial support of this study. Moreover, we greatly appreciate the members of the

Plant Biotechnology and Bioinformatics Laboratory, and the Plant Pathology Laboratory at Embrapa Soja, Brazil, and Université Laval, Canada for supporting this study.

Reference

1. Lee G-A, Crawford GW, Liu L, Sasaki Y, Chen X. Archaeological soybean (*Glycine max*) in East Asia: does size matter? PLoS One. 2011;6:e26720.

2. Companhia Nacional de Abastecimento. Séries Históricas de Área Plantada, Produtividade e Produção, Relativas às Safras 1976/77 a 2017/18 de Grãos, 2001 a 2017 de Café, 2005/06 a 2017/18 de Cana-de-Açúcar. 2018. http://www.conab.gov.br/conteudos.php?a=1252&. Accessed 2 Feb 2017.

3. Benko-iseppon AM, Nepomuceno AL, Abdelnoor R V. GENOSOJA – The Brazilian Soybean Genome Consortium : High throughput omics and beyond. Genet Mol Biol. 2012;35:i–iv.

4. Janse Van Rensburg JC, Lamprecht SC, Groenewald JZ, Castlebury LA, Crous PW. Characterisation of *Phomopsis* spp. associated with die-back of rooibos (*Aspalathus linearis*) in South Africa. Stud Mycol. 2006;55:65–74.

5. Santos JM, Vrandečić K, Ćosić J, Duvnjak T, Phillips AJL. Resolving the diaporthe species occurring on soybean in Croatia. Persoonia Mol Phylogeny Evol Fungi. 2011;27:9–19.

6. Udayanga D, Liu X, Crous PW, McKenzie EHC, Chukeatirote E, Hyde KD. A multi-locus phylogenetic evaluation of *Diaporthe* (*Phomopsis*). Fungal Divers. 2012;56:157–71.

7. Yorinori JT, Almeida AMR, Homechin M, Miranda LC, Kiihl RAS, Pola JN. Epifitia do cancro da haste da soja nos municípios de Castro, Palmeira, Ponta Grossa e Tibagi no Paraná e Rondonópolis, no Mato Grosso, na safra de 1988/89. In: 5. Seminário Nacional de Pesquisa de Soja, Campo Grande MS. Londrina: Embrapa Soja; 1989. p. 22–3.

8. Yorinori JT. Cancro da haste da soja: epidemiologia e controle. Londrina; 1996.

9. Wrather JA, Anderson TR, Arsyad DM, Gai J, Ploper LD, Porta-Puglia A, et al. Special Report Soybean Disease Loss Estimates for the Top 10 Soybean Producing Countries in 1994. Plant Dis. 1997;81:107–10.

10. Kilen TC, Hartwig EE. Identification of Single Genes Controlling Resistance to Stem Canker in Soybean. Crop Sci. 1987;27:863.

11. Bowers GR, Ngeleka K, Smith O. Inheritance of Stem Canker Resistance in Soybean Cultivars Crockett and Dowling. Crop Sci. 1993;33:67.

12. Tyler JM. Additional sources of stem canker resistance in soybean plant introductions. Crop Sci. 1995;35:376–7.

13. Pioli RN, Morandi EN, Martínez MC, Lucca F, Tozzini A, Bisaro V, et al. Morphologic, Molecular, and Pathogenic Characterization of *Diaporthe phaseolorum* Variability in the Core Soybean-Producing Area of Argentina. Phytopathology. 2003;93:136–46.

14. Chiesa M a., Pioli RN, Cambursano M V., Morandi EN. Differential expression of distinct soybean resistance genes interacting with Argentinean isolates of *Diaporthe phaseolorum* var. *meridionalis*. Eur J Plant Pathol. 2013;135:351–62.

15. Chiesa MA, Cambursano M V, Pioli RN, Morandi EN. Molecular mapping of the genomic region conferring resistance to soybean stem canker in Hutcheson soybean. Mol Breed. 2017;37.

16. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010;463:178–83.

17. Hwang E-Y, Song Q, Jia G, Specht JE, Hyten DL, Costa J, et al. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics. 2014;15:1.

18. Iquira E, Humira S, François B. Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. BMC Plant Biol. 2015;:1–12.

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

20. Mamidi S, Lee RK, Goos RJ, McClean PE. Genome-Wide Association Studies Identifies Seven Major Regions Responsible for Iron Deficiency Chlorosis in Soybean (*Glycine max*). PLoS One. 2014;9.

21. Mamidi S, Chikara S, Goos RJ, Hyten DL, Annam D, Moghaddam SM, et al. Genome-Wide Association Analysis Identifies Candidate Genes Associated with Iron Deficiency Chlorosis in Soybean. Plant Genome J. 2011;4:154.

22. Contreras-Soto RI, Mora F, Oliveira MAR de, Higashi W, Scapim CA, Schuster I. A Genome-Wide Association Study for Agronomic Traits in Soybean Using SNP Markers and SNP-Based Haplotype Analysis. PLoS One. 2017;12:1–22.

23. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol. 2015; April 2014.

24. Valliyodan B, Dan Q, Patil G, Zeng P, Huang J, Dai L, et al. Landscape of genomic diversity and trait discovery in soybean. Sci Rep. 2016;6 October 2015:23598.

25. Ramakrishna G, Kaur P, Nigam D, Chaduvula PK, Yadav S, Talukdar A, et al. Genome-wide identification and characterization of InDels and SNPs in *Glycine max* and *Glycine soja* for contrasting seed permeability

traits. BMC Plant Biol. 2018;18:1–15.

26. Keeling BL. A Seedling Test for Resistance to Soybean Stem Canker Caused by *Diaporthe phaseolorum* var. *caulivora*. Phytopathology. 1982;72:807–9.

27. Yorinori JT. Metodologia para avaliação da resistência ao cancro da haste da soja. Fitopatol Bras. 1991;16:58.

28. Brumer BB, Lopes-Caitar VS, Chicowski AS, Beloti JD, Castanho FM, Gregório da Silva DC, et al. Morphological and molecular characterization of *Diaporthe* (anamorph *Phomopsis*) complex and pathogenicity of *Diaporthe aspalathi* isolates causing stem canker in soybean. Eur J Plant Pathol. 2018;151:1009–25.

29. Chiesa M a., Pioli RN, Morandi EN. Specific resistance to soybean stem canker conferred by the *Rdm4* locus. Plant Pathol. 2009;58:1032–8.

30. Siviero A, Menten JOM. Uso de método do palito para inoculação de *Diaporthe* phaseolorum f. sp. *meridionalis* em soja. Summa Phytopathol. 1995;21:259–60.

31. Gonçalves ECP, Centurion MAP da C, Di Mauro AO. Avaliação de características agronômicas e de reação ao cancro-da-haste a ao oídio em linhagens de soja. Rev Ceres. 2006;53:197–203.

32. Hillen T, Juliatti FC, Polizel AC, Hamawaki OT, Brito CH de. Reação de genótipos de soja quanto à resistência ao cancro da haste. Biosci J. 2006;22:27–35.

33. Olliveira AL de, Lam-Sanchez A. Herança da resistência ao cancro da haste (*Diaporthe phaseolorum* (Cke .& Ell). Sacc . F . sp . *meridionalis* Morgan Jones) em soja (*Glycine max* (L.) Merrill). Rev Nucl. 2005;3:113–8.

34. Passianotto ALDL, Sonah H, Dias WP, Abdelnoor R V. Genome-wide association study for resistance to the southern root-knot nematode (*Meloidogyne incognita*) in soybean. Mol Breed. 2017;37:1–11.

35. Chang H, Lipka AE, Domier LL, Hartman GL. Characterization of Disease Resistance Loci in the USDA Soybean Germplasm Collection Using Genome-Wide Association Studies. Genet Resist. 2016;106:1139–51.

36. Pham AT, Harris DK, Buck J, Hoskins A, Serrano J, Abdel-Haleem H, et al. Fine mapping and characterization of candidate genes that control resistance to *Cercospora sojina K. Hara* in two soybean germplasm accessions. PLoS One. 2015;10.

37. King ZR, Childs SP, Harris DK, Pedley KF, Buck JW, Boerma HR, et al. A new soybean rust resistance allele from PI 423972 at the *Rpp4* locus. Mol Breed. 2017;37.

38. Chung E, Cho C, Yun B, Choi H, So H, Lee S, et al. Molecular cloning and characterization of the soybean DEAD-box RNA helicase gene induced by low temperature and high salinity stress. Gene.

2009;443:91–9.

39. Rocak S, Linder P. DEAD-box proteins: the driving forces behind RNA metabolism. Nat Rev Mol Cell Biol. 2004;5:232–41.

40. Cordin O, Banroques J, Tanner NK, Linder P. The DEAD-box protein family of RNA helicases. Gene. 2006;16:17–37.

41. Linder P. Dead-box proteins: a family affair--active and passive players in RNP-remodeling. Nucleic Acids Res. 2006;34:4168–80.

42. Sanan-mishra N, Pham XH, Sopory SK, Tuteja N. Pea DNA helicase 45 overexpression in tobacco confers high salinity tolerance without affecting yield. Proc Natl Acad Sci. 2005;102:509–14.

43. Gong Z, Dong C, Lee H, Zhu J, Xiong L, Gong D. A DEAD Box RNA Helicase Is Essential for mRNA Export and Important for Development and Stress Responses in Arabidopsis. plan. 2005;17:256–67.

44. Li D, Liu H, Zhang H, Wang X, Song F. OsBIRH1 , a DEAD-box RNA helicase with functions in modulating defence responses against pathogen infection and oxidative stress. J Exp Bot. 2008;59:2133–46.

45. Osakabe Y, Maruyama K, Seki M, Satou M, Shinozaki K, Yamaguchi-shinozaki K. Leucine-Rich Repeat Receptor-Like Kinase1 Is a Key Membrane-Bound Regulator of Abscisic Acid Early Signaling in Arabidopsis. Plant Cell. 2005;17 April:1105–19.

46. Godiard L, Sauviac L, Torii KU, Grenon O, Mangin B, Grimsley NH, et al. ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. Plant J. 2003;36:353–65.

47. Yang L, Wu K, Gao P, Liu X, Li G, Wu Z. GsLRPK, a novel cold-activated leucine-rich repeat receptor-like protein kinase from *Glycine soja*, is a positive regulator to cold stress tolerance. Plant Sci. 2014;215–216:19–28.

48. McCabe CE, Cianzio SR, O'Rourke JA, Graham MA. Leveraging RNA-Seq to Characterize Resistance to Brown Stem Rot and the Rbs3 Locus in Soybean. Mol Plant Microbe Interact. 2018;31:1083–94.

49. Martin GB, Brommonschenkel SH, Chunwongse J, Frary a, Ganal MW, Spivey R, et al. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science (New York, N.Y.). 1993;262:1432–6.

50. Zhou J, Loh YT, Bressan R a., Martin GB. The tomato gene Pti1 encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response. Cell. 1995;83:925–35.

51. Song W-Y, Wang G-L, Chen L-L, Kim H-S, Pi L-Y, Holsten T, et al. A Receptor Kinase-Like Protein Encoded by the Rice Disease Resistance Gene, Xa21. Science (80-). 1995;270:1804–6.

52. Gómez-Gómez L, Boller T. FLS2: An LRR Receptor-like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in *Arabidopsis*. Mol Cell. 2000;5:1003–11.

53. Li Y, Sun S, Zhong C, Wang X, Wu X, Zhu Z. Genetic mapping and development of co-segregating markers of RpsQ, which provides resistance to *Phytophthora sojae* in soybean. Theor Appl Genet. 2017;0:1–11.

54. Ithal N, Recknor J, Nettleton D, Maier T, Baum TJ, Mitchum MG. Developmental transcript profiling of cyst nematode feeding cells in soybean roots. Mol Plant Microbe Interact. 2007;20:510–25.

55. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;6:1–10.

56. Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, et al. An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. PLoS One. 2013;8:1–9.

57. Scheet P, Stephens M, Tier B, Werf JH van der, Cleveland MA, Scheet P, et al. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. Am J Hum Genet. 2006;78:629–44.

58. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: Genome association and prediction integrated tool. Bioinformatics. 2012;28:2397–9.

59. Maldonado dos Santos JV, Valliyodan B, Joshi T, Khan SM, Liu Y, Wang J, et al. Evaluation of genetic variation among Brazilian soybean cultivars through genome resequencing. BMC Genomics. 2016;17:110.

60. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Landes Biosci. 2012;6:80–92.

61. Milne I, Shaw P, Stephen G, Bayer M, Cardle L, Thomas WTB, et al. Flapjack-graphical genotype visualization. Bioinformatics. 2010;26:3133–4.

Tables

Table 1. Reaction of soybean differentials against *D. aspalathi* isolate CMES480

Accession	Resistance Gene	%DP	SSC Phenotype
Tracy-M	*Rdm1*/*Rdm2*	0%	Resistant
D8510404	*Rdm1*	65%	Susceptible
D8510412	*Rdm2*	72%	Susceptible
Crockett	*Rdm3*	0%	Resistant
Dowling	*Rdm4*	56%	Susceptible
Hutcheson	*Rdm5*	0%	Resistant
PI398469	*Rdm?*	2%	Resistant

%DPs: The values for the percentage of dead plants were obtained according to the formula proposed by Yorinori (1991).

Table 2. The most significant SNPs associated with SSC resistance identified in this study.

Marker ID	Chrom.	Pos (bp)	MAF	*p*.value	R²	FDR Adjusted *p*-values
GBSRdm370	14	1,744,370	0.30	1.93E-31	0.70	6.35E-27
GBSRdm556	14	1,725,556	0.24	2.97E-28	0.64	4.88E-24
GBSRdm287	14	1,710,287	0.28	5.61E-28	0.63	5.66E-24
GBSRdm224	14	1,986,224	0.27	6.89E-28	0.63	5.66E-24
GBSRdm562	14	1,740,562	0.25	2.87E-27	0.62	1.89E-23
GBSRdm793	14	1,768,793	0.42	3.66E-25	0.59	2.00E-21
GBSRdm339	14	1,921,339	0.28	4.17E-22	0.54	1.71E-18
GBSRdm374	14	1,921,374	0.28	4.17E-22	0.54	1.71E-18
GBSRdm219	14	1,795,219	0.45	8.48E-21	0.52	3.09E-17
GBSRdm204	14	1,751,204	0.21	1.60E-19	0.50	5.27E-16
GBSRdm516	14	1,612,516	0.27	2.26E-17	0.47	6.75E-14
GBSRdm964	14	1,850,964	0.40	2.57E-17	0.47	7.02E-14
GBSRdm114	14	1,851,114	0.40	4.80E-17	0.46	1.21E-13
GBSRdm450	14	1,612,450	0.26	9.59E-16	0.45	2.25E-12
GBSRdm397	14	1,612,397	0.23	1.19E-14	0.43	2.60E-11
GBSRdm518	14	1,744,518	0.46	1.52E-14	0.43	3.13E-11
GBSRdm120	14	1,741,120	0.45	4.36E-14	0.42	8.43E-11
GBSRdm712	14	1,581,712	0.23	4.26E-13	0.41	7.77E-10
GBSRdm875	14	1,581,875	0.32	2.39E-12	0.40	4.13E-09

Chrom.: Chromosome; Pos(bp): physical position of the allelic variation; **MAF:** minor allele frequency; $R^{2\dagger}$: R square of model with SNP.

Table 3. Haplotypes obtained using SNPs from GWAS for the accessions

Haplotype ID	*SNPs positions in the soybean genome*				*SSC phenotype*	
	Differential Genotypes	*1,744,370*	*1,768,793*	*1,744,518*	*R*	*S*	*Total*
Hap1	Hutcheson/PI 398469	C	C	C	124	1	125
Hap2	Crockett	C	C	A	36	0	36
Hap3	Tracy-M/D85-10404	C	G	A	23	2	25
Hap4	Dowling/D85-10412	T	G	A	5	70	75

SNPs positions: physical position of the allelic variation are from chromosome 14 of the soybean genome (Wm82.a2); **R:** SSC resistant accessions; **S:** SSC Susceptible accessions. In the haplotypes analyses were just considered accessions showing homozygous alleles for all the three SNPs.

Table 4. Summary of the allelic variation observed in the putative Rdm locus region.

Region		Modification	SNPs	InDels
Non-Coding		Intergenic region	143	14
		Upstream 5k	2,537	683
		Downstream 5k	695	204
Coding	UTR	5' UTR	88	23
		5' UTR premature start gained	8	--
		3' UTR	151	46
	Intron	Intron	399	96
		Splice region	19	2
		Splice acceptor site	2	--
		Splice donor site	1	--
	Exon	Disruptive + Inframe Deletion	--	3
		Disruptive + Inframe Insertion	--	2
		Frameshift	--	23
		Inframe Deletion	--	5
		Inframe Insertion	--	4
		Non-synonymous modification	248	--
		Start lost	1	--
		Stop retained	1	--
		Stop gained	5	--
		Synonymous modification	142	--
Total			**4,440**	**1,105**

Intergenic region: the variant is in an intergenic region; **Upstream 5k:** SNPs detected up to 5 kb upstream of the coding region; **Downstream 5k:** SNPs detected up to 5 kb downstream of the coding region; **5' UTR:** hits 5'UTR region; **5' UTR premature start gained:** a variant in 5'UTR region produces a three base sequence that can be a START codon; 3' UTR: variant hits 3'UTR region; **Intron:** SNPs detected inside an intron; **Splice region:** a sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron; **Splice acceptor site:** the variant hits a splice acceptor site; **Disruptive + In frame Deletion:** one codon is changed and one or more codons are deleted; **Disruptive + In frame Insertion:** one codon is changed and one or many codons are inserted; **Frameshift:** insertion or deletion causes a frameshift; **In frame Deletion:** one or

many codons are deleted; **In frame Insertion:** one or many codons are inserted; **Start lost:** variant causes start codon to be mutated into a non-start codon; **Stop G.:** variant causes a STOP codon; **Non-synonymous modification:** SNP variants causing a codon that produces a different amino acid; inside coding region; **Synonymous modification**: variant causes a codon that produces the same amino acid.

Table 5. Fifteen non-synonymous mutations with similar pattern of GBSRdm370 detected in haplotype analysis.

Pos	REF	ALT	Ncl	AA	Gene	Exon
1,747,030	G	A	Gag/Aag	p.Glu8Lys/c.22G>A	Glyma.14G024400	1
1,747,238	C	G	gCt/gGt	p.Ala77Gly/c.230C>G	Glyma.14G024400	1
1,747,,247	C	T	tCg/tTg	p.Ser80Leu/c.239C>T	Glyma.14G024400	1
1,747,286	G	A	aGg/aAg	p.Arg93Lys/c.278G>A	Glyma.14G024400	1
1,747,288	G	A	Gaa/Aaa	p.Glu94Lys/c.280G>A	Glyma.14G024400	1
1,783,495	A	G	Acc/Gcc	p.Thr417Ala/c.1249A>G	Glyma.14G024700	5
1,887,464	G	A	aGc/aAc	p.Ser4Asn/c.11G>A	Glyma.14G026300	1
1,908,059	G	A	Ggt/Agt	p.Gly285Ser/c.853G>A	Glyma.14G026500	3
1,908,548	T	G	tgT/tgG	p.Cys336Trp/c.1008T>G	Glyma.14G026500	4
1,909,357	T	C	Atg/Gtg	p.Met274Val/c.820A>G	Glyma.14G026600	9
1,909,360	T	A	Atc/Ttc	p.Ile273Phe/c.817A>T	Glyma.14G026600	9
1,912,650	A	G	Tct/Cct	p.Ser44Pro/c.130T>C	Glyma.14G026600	1
1,915,601	G	A	Gcc/Acc	p.Ala31Thr/c.91G>A	Glyma.14G026700	1
1,917,088	T	G	Tat/Gat	p.Tyr193Asp/c.577T>G	Glyma.14G026700	5
1,942,282	C	T	cCt/cTt	p.Pro558Leu/c.1673C>T	Glyma.14G026900	3

Pos: position of the identified SNP in base pair (bp); **REF:** allele correspond to the reference genome (Williams 82); **ALT:** alternative allele observed in this position; **Ncl:** nucleotide modification observed due this SNP; **AA:** amino acid modification observed due this SNP; **Exon:** this illustrate in which exon of the gene the SNP were identified

Table 6. Agreement analysis between genotyping and phenotyping using the CMES 480 isolate

SNP	Pos(bp)	Accuracy (%)	Recall (%)	Type I error rate (%)	Type I error rate (%)
GBSRdm?370	1,744,370	98.0	98.0	2.4	1.18

Pos: position of the identified SNP in base pair (bp); **Accuracy (%):** accuracy is the percentage of resistant and susceptible genotypes correctly classified by the marker; **Recall (%):** a value given by the number of true positive divided by the number of true positives plus the number of false negatives; **Type I error rate (%):** false positive rate; **Type II error rate (%):** false negative rate.

Additional File Legend

Additional File 1. Single Nucleotide Polymorphism distribution in soybean chromosome.

Additional File 2. Haplotype analysis of the 295 accessions used in GWAS analysis.

Additional File 3. Haplotype analysis of the eight differential Rdm sources of resistance.

Additional File 4. Linkage Disequilibrium analysis for the *Rdm* loci region.

Additional File 5. List of soybean genotypes used for KASP assay validation.

Additional File 6. Basic description of the 295 soybean genotypes used in this study.

Additional File 7. The 51 soybean resequenced accessions used for haplotype analysis.

Figures

Figure 1

Phenotype response of the southern stem canker infection in soybean. A) Difference among the resistant (Tracy-M) and the susceptible (BR-23) cultivars. B) Lesion length in susceptible (left) and resistant (right) soybean accessions.

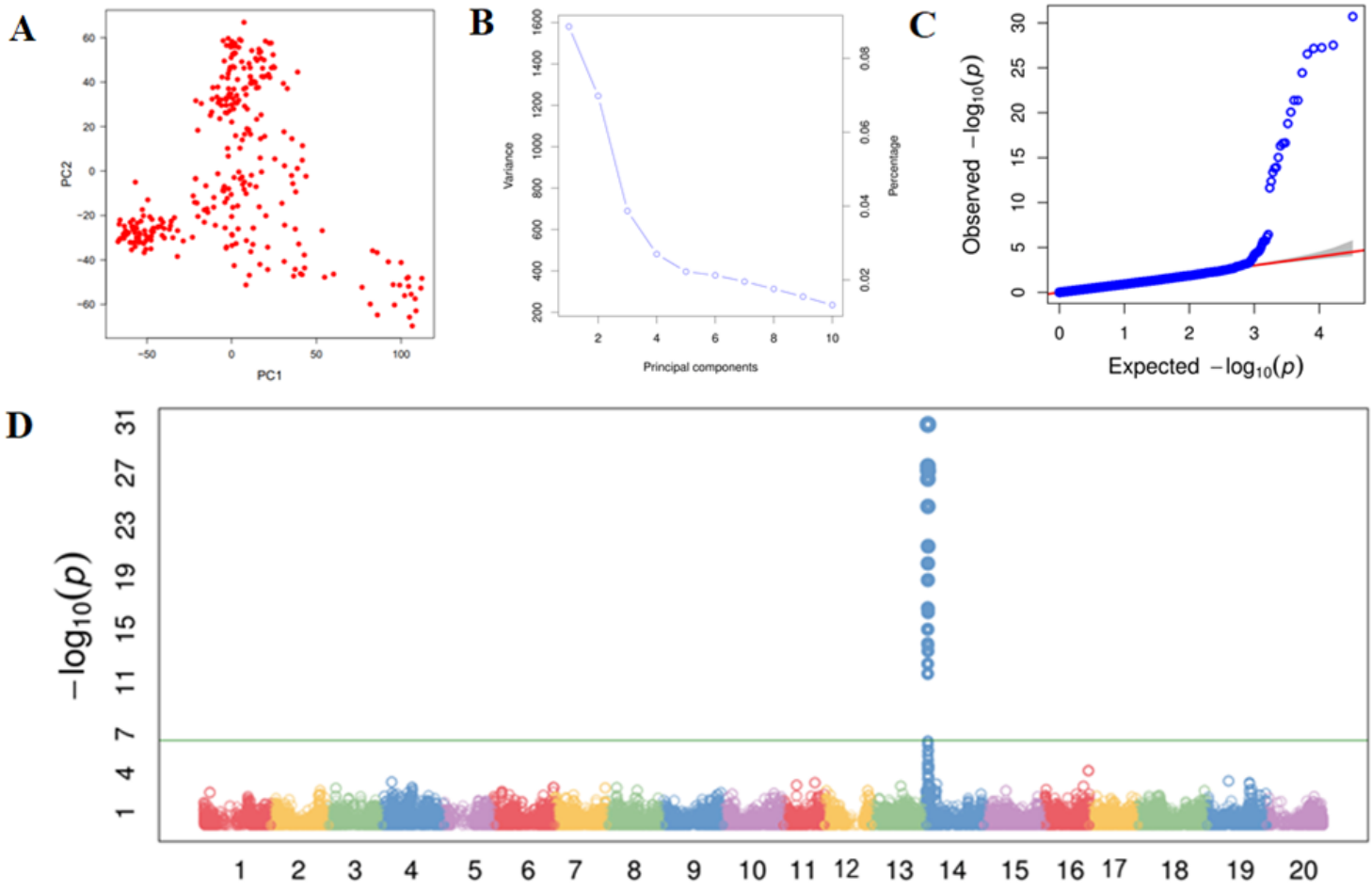

Figure 2

Manhattan-plot, Quantile-quantile (QQ) plots and PCA of population structure for southern stem canker. A Principal Component Analysis of the GBS panel. B The genetic variation explained using 3 PC's. C QQ-plot from this GWAS D Manhattan-plot obtained from GWAS.

Figure 3

The allelic variation observed in 51 resequenced soybean cultivar for GBSRdm370 used in this study. The soybean accessions on green square represent the resistance lines, while the soybean accessions on red square represent the susceptible lines.

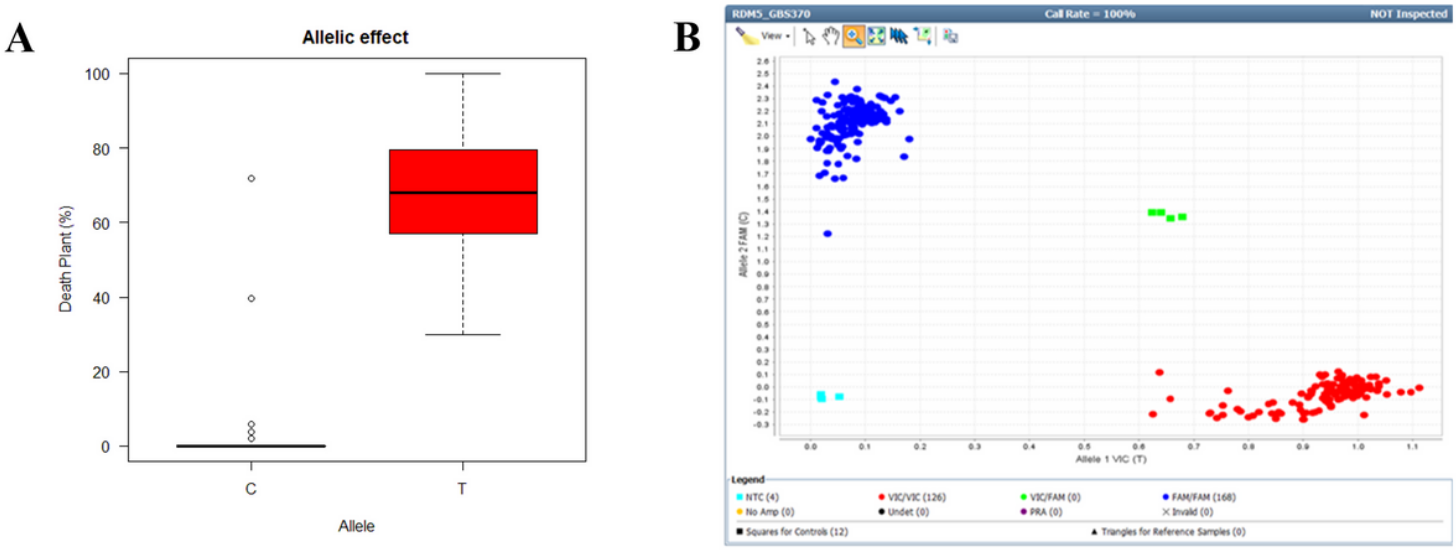

Figure 4

Box-Plot and Allelic discrimination of the SNP GBSRdm370. A) Box-Plot of the GBSRdm379 markers associated with the %DP. B) Allelic discrimination observed GBSRdm370. The blue dots represented the resistance allele, the red dots were the susceptible allele, and the green dots were the heterozygous samples.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.xlsx
- supplement2.xlsx
- supplement3.xlsx
- supplement4.xlsx
- supplement5.xlsx
- supplement6.xlsx
- supplement7.xlsx