



# Genome-wide association study for resistance to the *Meloidogyne javanica* causing root-knot nematode in soybean

Jean Carlos Alekcevetch<sup>1</sup> · André Luiz de Lima Passianotto<sup>2</sup> · Everton Geraldo Capote Ferreira<sup>1</sup> ·  
Adriana Brombini dos Santos<sup>3</sup> · Danielle Cristina Gregório da Silva<sup>3</sup> · Waldir Pereira Dias<sup>3</sup> · François Belzile<sup>4</sup> ·  
Ricardo Vilela Abdelnoor<sup>3</sup> · Francismar Correa Marcelino-Guimarães<sup>3</sup>

Received: 25 July 2020 / Accepted: 3 November 2020 / Published online: 19 January 2021  
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

**Key message** A locus on chromosome 13, containing multiple TIR-NB-LRR genes and SNPs associated with *M. javanica* resistance, was identified using a combination of GWAS, resequencing, genetic mapping and expression profiling.

**Abstract** *Meloidogyne javanica*, a root-knot nematode, is an important problem in soybean-growing areas, leading to severe yield losses. Some accessions have been identified carrying resistance loci to this nematode. In this study, a set of 317 soybean accessions was characterized for resistance to *M. javanica*. A genome-wide association study was performed using SNPs from genotyping-by-sequencing, and a region of 29.2 kb on chromosome 13 was identified. An analysis of haplotypes showed that SNPs were able to discriminate between susceptible and resistant accessions, with 25 accessions sharing the haplotype associated with resistance. Furthermore, five accessions that exhibited resistance without carrying this haplotype may carry different loci conferring resistance to *M. javanica*. We also conducted the screening of the SNPs in the USDA soybean germplasm, revealing that several soybean accessions previously reported as resistant to other nematodes also shared the resistance haplotype on chromosome 13. Two SNP-based TaqMan® assays were developed and validated in two panels of soybean cultivars and in biparental populations. In silico analysis of the region associated with resistance identified the occurrence of genes with structural similarity with classical major resistance genes (NBS-LRR genes). Specifically, several nonsynonymous SNPs were observed in *Glyma.13g194800* and *Glyma.13g194900*. The expression profile of these candidate genes demonstrated that the two gene models were up-regulated in the resistance source PI 505,099 after nematode infection. Overall, the SNPs associated with resistance and the genes identified constitute an important tool for introgression of resistance to the root-knot nematode by marker-assisted selection in soybean breeding programs.

Communicated by Henry T. Nguyen.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00122-020-03723-9>) contains supplementary material, which is available to authorized users.

✉ Francismar Correa Marcelino-Guimarães  
francismar.marcelino@embrapa.br

<sup>1</sup> Londrina State University (UEL), Celso Garcia Cid Road, km 380, Londrina, PR, Brazil

<sup>2</sup> Department of Plant Agriculture, University of Guelph, Guelph, ON N1G 2V7, Canada

<sup>3</sup> Brazilian Agricultural Research Corporation – National Soybean Research Center (Embrapa Soja), Carlos João Strass road, Warta County, PR, Brazil

<sup>4</sup> Department of Plant Sciences and Institute of Integrative Biology and Systems (IBIS), Université Laval, Quebec City, Quebec G1V 0A6, Canada

## Introduction

Soybean [*Glycine max* (L.) Merrill] is one of the most important oil and protein crops in the world. Currently, Brazil is the largest producer with 127.6 million metric tons, followed by USA with 110.18 million tons (USDA 2019). Despite the success of the soybean crop both in the USA and Brazil, the crop is frequently challenged by biotic stresses. In soybean, among all phytosanitary problems, nematode parasitism deserves special attention. It is estimated that nematodes can cause annual losses of 10% to 15%, representing almost US\$78 billion worldwide (Lima et al. 2017). The *Meloidogyne* genus is composed of more than 90 species, but basically *Meloidogyne javanica* and *Meloidogyne incognita* are the members of this genus with the largest economic impact on worldwide soybean production (Jones et al. 2013; Lima et al. 2017). In Brazil, losses due to these two

nematode species have been reported in all the main regions where soybean is cultivated, including Rio Grande do Sul, Paraná, Mato Grosso and Mato Grosso do Sul (Silva et al. 2001a; Wrather et al. 2010). The use of resistant cultivars and crop rotation are the best ways to decrease and control nematode populations, thus minimizing production losses and allowing cultivation in infected areas (Ferraz 2001).

Studies have been conducted in order to identify resistance sources in soybean accessions. For example, through the phenotypic screening of 2370 soybean accessions it was possible to identify PI 230,977 as a source of resistance to *M. javanica*, as well as PI 200,538 as source of resistance to *Meloidogyne arenaria* (Luzzi et al. 1987). In another study, soybean breeding lines and reported resistance sources were evaluated using different nematode species. It was observed that PI 595,099 and PI 230,977 showed resistance against both *M. javanica* and *M. arenaria*, and against some races of *Heterodera glycines* (Davis et al. 1998). Inheritance of *M. javanica* was identified as quantitative and showing high heritability in PI 595,099 and cv. CD 201 (Silva et al. 2001a). Currently, in the Brazilian market, more than 80 cultivars resistant or moderately resistant to nematodes are available; however, there are few varieties with strong resistance to *M. javanica*. The resistance present in Brazilian soybean varieties is derived from only one source, the North American cultivar Bragg, in which resistance is reported to be quantitative (Silva et al. 2001a; Dias et al. 2010).

Genetic mapping studies have been performed to identify loci conferring *M. javanica* resistance. Initially, an F<sub>2</sub> progeny derived from the cross between the cv. CNS and PI 230,977 identified RFLP markers linked to resistance. Two QTLs contributing to resistance were identified; a QTL located on chromosome 13 accounted for 46% of the variation in the number of galls and a QTL on chromosome 01 accounted for 13% (Tamulonis et al. 1997a). Interestingly, resistance to *M. arenaria* mapped to the same region of chromosome 13 in a cross between PI 200,538 and the susceptible cv. CNS (Tamulonis et al. 1997b). In a third population [cv. Gazelle (resistant to *M. javanica*) and cv. Prima (susceptible)], resistance was mapped to chromosome 13 as well. Finally, populations derived from crosses between CD 201 and BRS 133, and between PI 595,099 and BRS 133 identified SSR markers associated with resistance to *M. javanica* on chromosome 13 (Silva et al. 2001b; Fuganti et al. 2004).

All these previous studies were carried out using biparental populations, which access only the alleles segregating between the two parents and provide limited resolution due to the small number of recombination events captured in such populations (Korte and Farlow 2013). Nowadays, different high-throughput genotyping methodologies, for example genotyping-by-sequencing (GBS) (Elshire et al. 2011), SoySNP50K chip (Song et al. 2015) and whole-genome resequencing (dos Santos et al. 2016), have been

used to obtain SNPs for Genome-Wide Association Studies (GWASs). Association mapping has been receiving unprecedented attention because it overcomes several limitations of QTL mapping and easily expands the knowledge about the occurrence of new sources of resistance. It provides high resolution, cost efficiency and does not require the production of mapping populations.

GWASs have been performed for the discovery of genomic regions underlying important diseases in soybean. For example, genomic regions for *Sclerotinia* stem rot resistance (white mold) (Bastien et al. 2014; Iquira et al. 2015; Boudhrioua et al. 2020), brown stem rot (BSR) (Rincker et al. 2016) and soybean stem canker (Maldonado dos Santos et al. 2019) were identified by GWAS. However, there is a lack of association studies regarding resistance against *Meloidogyne* sp. nematodes, especially in soybean. As far we know, only a GWAS for resistance to *M. incognita* has been published for soybean (Passianotto et al. 2017) and another for *Arabidopsis* (Warmerdam et al. 2018). Indeed, almost all the previous GWASs were focused on soybean cyst nematode (SCN) caused by *Heterodera glycines* (Vuong et al. 2015; Zhang et al. 2016, 2017; Tran et al. 2019). Further functional characterization of candidate genes in these regions has been conducted using different methodologies, such as RNA-seq, RT-qPCR and transgenic approaches, in order to better understand the molecular mechanisms involved in plant resistance. Fuganti et al. (2010), studying the resistance to *M. javanica* in soybean, observed that a polymorphism previously associated with resistance was located inside the promoter of the Gmhsp17.6-L gene. The comparison of the nucleotide sequences showed differences in the number of AT insertions between the resistant (PI 595,099) and susceptible (BRS 133) progenitors, which were hypothesized to be involved in the regulation of the expression levels of Gmhsp17.6-L. It was observed that the resistant individuals showed higher expression levels of the Gmhsp17.6-L when compared with the susceptible individuals. Transcriptomic studies have been also conducted in PI 595,099 in order to identify candidate genes involved in interaction between *M. javanica* and soybean. All the resulting data suggested the key role of glycosyltransferases, auxins and components of gibberellin signal transduction, biosynthesis and deactivation pathways in the resistance reaction (de Sá et al. 2012; Beneventi et al. 2013).

The objective of this study was to describe the genetic architecture underlying *M. javanica* resistance in a diverse set of 317 soybean accessions, via a GWAS approach, and to develop SNP markers useful in breeding for resistance. In order to further comprehend the role of the genomic regions identified by GWAS, resistance gene candidates present in the associated LD block were selected for study of their expression levels by RT-qPCR after pathogen infection, using both susceptible and resistant soybean accessions.

## Materials and methods

### Plant materials

A set of 317 accessions composed of cultivars and PIs (Supplementary Table S1) were obtained from the Soybean Germplasm Bank, located at Embrapa Soja in Londrina, PR, Brazil. Initially, one seed for each soybean accession was sown and grown under green-house conditions, and the leaf sample for each accession was individually collected, frozen in liquid nitrogen and stored in a  $-80^{\circ}\text{C}$  freezer. The leaf samples were ground to a fine powder and stored until DNA isolation. The resulting seeds for each plant corresponding for each accession were used for the phenotypic approach.

Six seeds of each accession were pre-germinated in plastic pots of 0.25 L filled with sterilized sand. Five days after germination, seedlings were transferred individually to plastic tubes of 0.5 L filled with substrate (sterilized by autoclaving) composed of sand and soil (3:1), totalizing six biological replications (one seedling per plastic tube). The plants were kept in a greenhouse under 16 h of daylight supplemented with 600 W high-pressure sodium lamps (Light Systems PL). All six samples were used for nematode resistance evaluation.

One population was derived by crossing the resistant soybean cultivar CD 224 and the susceptible cultivar BRS 133, while the other one derived by crossing the resistance source PI 595,099 and cultivar BRS 133.  $F_1$  plants from these crosses were self-pollinated to produce the  $F_2$  progeny. A total of 135 and 122  $F_2$  seeds were sown in greenhouse conditions and evaluated for *M. javanica* resistance in the CD 224 and PI 595,099 populations, respectively.

### Phenotyping for RKN resistance

Nematode inoculation was performed during the summer. One day before transferring seedlings to plastic tubes, some infected soybean roots from a nematode stock were ground in water to obtain the suspension containing *M. javanica* eggs. The number of eggs was estimated in a Peters' chamber under microscopy, and the concentration of the suspension was adjusted to 1250 eggs/ml. The inoculation was performed by deposition of 4 mL of the nematode suspension in the same hole used to introduce the seedling. Thirty days after inoculation, all plants were individually removed from the tubes. Excess sand and soil around roots were carefully removed, and roots were washed in running water. The severity of the infestation was rated on a scale of 1–5 (adapted from Dall'Agnol and Antônio 1982), where 1 = < 10% of the root system is

infected with small galls; 2 = 10–25% of the root system galled, most being small galls; 3 = 26–50% of the root system with large galls; 4 = 51–90% of the root system with large galls; 5 = 91–100% of root system with large galls and necrotic roots. Accessions with a rating between 1 and 2.5 were considered resistant (*R*), between 2.6 and 3.5 were deemed moderately resistant (*MR*); and those with a score above 3.6 were rated as susceptible (*S*).

Two biparental populations were developed to evaluate the phenotypic distribution for resistance against *M. javanica* and to validate the co-segregation of SNP markers. The phenotypic evaluation was performed as described above for the GWAS panel. A Chi-square test was performed to test the phenotypic distribution for goodness of fit to the Mendelian ratio of 7:9 as previously described for *M. javanica* genetic inheritance (Silva et al. 2001a).

### DNA extraction, GBS library preparation and SNP calling

The DNA from each sample was extracted using the DNeasy® Plant Mini Kit (Qiagen) according to the manufacturer's instructions. DNA integrity was checked by electrophoresis on an agarose gel (1%), followed by quantification on a NanoDrop® ND-1000 spectrophotometer (Uniscience) and diluted with water to a concentration of 10 ng/μL. GBS libraries were produced using the protocol described by Elshire et al. (2011) and modified by Sonah et al. (2013). Thus, DNA was digested using *ApeKI*, followed by ligation of barcoded adapters and pooling of 96 samples per library. These 96-plex GBS libraries were sequenced on either Illumina HiSeq2000 (McGill University-Genome Quebec Innovation Centre, Montreal, QC, Canada) or Ion Torrent (Université Laval, Québec, QC, Canada) DNA sequencers. Using the Fast-GBS pipeline (Torkamaneh et al. 2017), variants were called using a minimal read depth of two reads and loci with more than 80% missing data were removed. Heterozygous genotypes were replaced with missing data, and any accessions with > 80% missing data were removed from the dataset. Finally, imputation of missing genotypic data was performed using fastPHASE 1.3 (Scheet and Stephens 2006). For the GWAS, only loci with a minor allele frequency (MAF)  $\geq 0.05$  were used.

### Association mapping and haplotype analysis

The Genomic Association and Prediction Integrated Tool–GAPIT (Lipka et al. 2012)—was used to conduct GWAS using a compressed mixed linear model (cMLM) that takes into account both population structure and genetic relatedness between lines. The first three principal components (PC) from principal component analysis (PCA) were used to capture population structure and produce a P matrix.

A VanRaden kinship matrix ( $K$ ) was used to capture genetic relatedness. Marker–trait associations were declared significant using FDR-adjusted  $p$  values with the threshold set at 0.0001.

In order to verify the alleles of the SNPs identified by GWAS in a wider panel, the USDA soybean germplasm collection previously genotyped with the SoySNP50K array was used. The USDA germplasm genotype dataset was imputed with the soybean haplotype map (GmHapmap) constructed using WGS data for 1007 *G. max* accessions (<https://www.soybase.org/projects/SoyBase.C2020.01.php>).

All SNPs from GBS with  $MAF \geq 0.05$  and located on chromosome Chr13 were loaded into PLINK version 1.9 (Purcell et al. 2007), and the correlation coefficient ( $r^2$ ) was calculated to determine pairwise linkage disequilibrium (LD). Haplotype blocks were identified and visualized using Haploview (Barrett et al. 2005) using default Gabriel's rules (Gabriel et al. 2002), confidence interval minimal for strong LD, upper = 0.98, lower = 0.7— $D' > 0.8$  and fraction of strong LD  $\geq 0.95$ . Nucleotide variation in a WGS dataset obtained from sequencing 12 Brazilian soybean cultivars (dos Santos et al. 2016) as well as 14 additional Brazilian cultivars and two other accessions (Torkamaneh et al. 2020) was also examined to identify additional SNPs in the region of chromosome 13 identified through GWAS. First, all the SNPs in the region from 30.7 Mb to 30.9 Mb were extracted, and only SNPs present in at least two accessions and showing no heterozygous alleles in the set were kept. Finally, pairwise LD values ( $r^2$ ) between the SNPs from the WGS data and the peak SNP (Chr13\_30,804,961) were calculated using PLINK, and LD heatmaps were visualized using the “LDheatmap” R package (Shin et al. 2006). Only SNPs showing high levels of LD with the peak SNP ( $r^2 \geq 0.9$ ) were used in haplotype analysis. Predictions of SNP effects were performed using SnpEff version 4.3i (Cingolani et al. 2012), using version 2 of the soybean reference genome and physical gene locations (GFF3 file) ([https://phytozome-next.jgi.doe.gov/info/Gmax\\_Wm82\\_a2\\_v1](https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1)).

### Expression levels obtained by RT-qPCR

To investigate the transcriptional regulation and polymorphism in the genes located in the same LD block as the marker showing the greatest association with resistance to *M. javanica*, the expression profile of four gene models annotated as resistance genes containing TIR-NBS-LRR domains was quantified by RT-qPCR at three timepoints after pathogen infection (1, 2 and 6 days post-inoculation or DPI). Gene expression was measured on both susceptible (BRS 133) and resistant (PI 595,099) accessions, under inoculated and mock conditions. The experiment was conducted in a completely randomized block design with five replicates. Inoculation was done as previously described.

After 1, 2 and 6 DPI, roots from mock-inoculated and inoculated plants were collected, frozen in liquid nitrogen, and stored 80 °C. Using Trizol reagent (Life Technologies), total RNA was extracted from samples. First-strand cDNAs were generated using the SuperScript III First-Strand Synthesis SuperMix (Invitrogen). Primers for each of the four candidate gene models in the mapped interval were designed using Primer3Plus (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>). The cDNA samples were amplified with primers specific to each gene and also for the endogenous control ( $\beta$ -actin gene), at a final concentration of 0.1–0.25  $\mu$ M, with the 1X SYBR Green Master Mix Kit (Applied Biosystems) in a final volume of 12.5  $\mu$ L. All PCR reactions were conducted in three technical replicates. The  $E = [10 - 1/\text{slope}] - 1$  formula was employed to calculate the reaction efficiency and to adjust the final primer concentration. PCR efficiency was determined using standard curves for each primer pair constructed with serial dilutions (1:10, 1:100 and 1:1000) of the cDNA preparation. The reactions were performed in a StepOnePlus Real-Time PCR System (Applied Biosystems) following the manufacturer's instructions. After initial steps at 50 °C for 2 min and at 95 °C for 10 min, a two-step program of 95 °C for 15 s and 62 °C for 1 min was run for 40 cycles. The final relative quantification of each gene compared with the control conditions (mock-inoculated samples) for each timepoint was estimated considering the RQ obtained in each biological replicate, represented by each independent experiment, with three replicates each. Significant differences were determined based on estimates of the standard deviation (SD) and with REST software version 2.0.7 ( $p < 0.05$ ).

### SNP genotyping using TaqMan assays

The 200-bp flanking sequence for each SNP identified in the GWAS was extracted from the soybean reference genome and was used to design TaqMan® MGB allelic discrimination assays (Applied Biosystems, Foster City, CA, USA) in Primer Express Software v3.0.1 using default parameters for MGB assays. The primers and probe sequences generated were aligned against the reference genome (<https://phytozome.jgi.doe.gov/pz/>) in order to check their specificity. The primers and probe sequences for two of the SNPs (Chr13\_30,804,961 and Chr13\_30,792,409) aligned in the desired region on chromosome 13 and therefore were synthesized. TaqMan® qPCR end-point amplifications were performed according to manufacturer's instructions on an ABI 7900 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Genotypes were acquired and clustered as plots by the TaqMan® Genotyper v1.6 Software (Applied Biosystems, Foster City, CA, USA).

The M700 (Chr13\_30,792,409) and M800 (Chr13\_30,804,961) TaqMan® assays were used to genotype



both  $F_2$  populations (CD 224 and PI 595,099 populations), as well as for genotyping a set of 69 soybean accessions, part of them from the GWAS panel and some selected based on their reported phenotype (*R* or *S*). The Chi-square for each SNP marker was conducted to evaluate Mendelian segregation in Genes Software (Cruz, 1998).

## Results

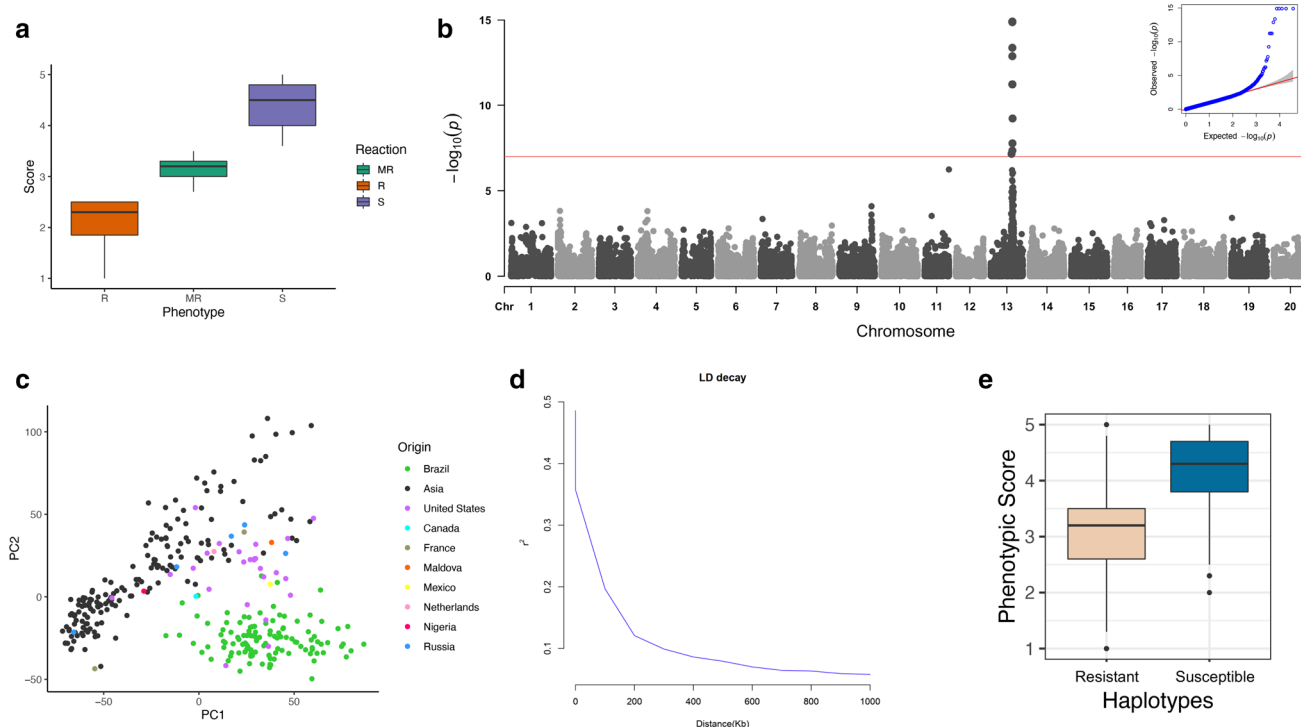
### Evaluation of root-knot nematode resistance in the soybean accessions

Among the 317 soybean accessions comprising the association panel (Supplementary Table 1), 30 accessions received scores between 1.0 and 2.5 (indicating a high level of resistance), 82 accessions received scores ranging from 2.6 to 3.5 (moderate resistance), while the remaining 205 accessions exhibited scores higher than 3.6 (classifying as susceptible materials) (Fig. 1a). Considering the population size and the number of accessions in each phenotypic class, it was assumed that the genomic regions that are responsible for

resistance and susceptibility were well sampled. The known *M. javanica* resistance sources PI 595,099, Bragg and also PI 437,127 B showed the lowest score (1.0), whereas PI 230,977 showed a score of 2.5 and was also classified as resistant (Supplementary Table 1). Interestingly, another previously described resistance source to *M. arenaria*, PI 200,538, showed a score of 2.7, and was thus classified as moderately resistant. As expected, BRS 133, a cultivar known to be susceptible to *M. javanica*, showed a score of 4.5. It can be observed that only approximately 10% of the accessions were resistant to *M. javanica*, but only PI 595,099, PI 437,127 B and Bragg showed the lowest score (1.0). Around 26% were classified as moderately resistant, while the majority of accessions (~65%) were classified as susceptible to *M. javanica*.

### Genome-wide association study for *M. javanica* resistance

A total of 44,992 SNPs was identified by GBS along the 20 soybean chromosomes, providing an extensive coverage of the genome (Supplementary Fig. 1). This represented



**Fig. 1** Genome-wide association study for soybean resistance to *Meloidogyne javanica*. **a** Distribution of the soybean reaction to *Meloidogyne javanica* expressed as score values across all of the 317 accessions, *R* represents resistant phenotype, *MR* represents moderately resistant and *S* represents susceptible phenotype. **b** GWAS result as Manhattan plot, in which are presented by negative  $\log_{10} p$  values against position on each of 20 soybean chromosomes and quantile–quantile plot (upper right), the red line indicates the signifi-

cant threshold. **c** Heatmap of the VanRaden kinship matrix between the soybean accessions. **c** Population structure of soybean accessions used in GWAS reflected by the first two principal components. **d** Linkage disequilibrium (LD) decay rate estimated as squared correlation coefficient ( $r^2$ ) in the soybean genome. **e** Boxplot of *Meloidogyne javanica* phenotypic variation between the different haplotypes formed by SNPs identified by GWAS in 317 soybean accessions (colour figure online)

an average of 2250 SNPs per chromosome or one variation every 21 kb approximately. The largest number of variants (8.5% of total SNPs) was found, as expected, on the largest chromosome (Chr18, 58 Mb). On the other hand, the smallest number of variants was not found on the smallest chromosome (Chr11–34 Mb; 1382 variants), but rather on Chr 12 (40 Mb) (only 1255 variants) (Supplementary Table 2). Regarding the marker distribution within coding versus non-coding regions, SNPs were identified in all genic segments as well as in the intergenic regions, with the intergenic regions presenting the largest number of SNPs, followed by downstream regions, intronic regions, upstream regions, exons, 3'UTRs and 5'UTRs (Supplementary Table 2). Based on the results obtained in the linkage disequilibrium analysis performed using the full set of 44,992 SNPs, the regression curve fitted to the LD decayed below  $r^2=0.2$  at  $\sim 0.15$  Mb (Fig. 1d). Finally, after removing markers showing a MAF lower than 0.05, a total of 37,784 SNPs were kept for subsequent analysis.

A principal component analysis (PCA) was performed in order to capture the population structure in the panel (Fig. 1c, Supplementary Fig 2). The PC1 explained approximately 9% of the observed genetic variance, PC2 approximately 5% and PC3 approximately 4%; together, the first three PCs explained about 18% of the total genetic variance. The GWAS was conducted using a cMLM taking into account both genetic relatedness ( $K$  matrix) and population structure ( $P$  matrix). The quantile–quantile plot showed that observed  $p$  values strongly deviated from the expected  $p$  values only for the markers showing a very high degree of association (Fig. 1b), which suggests that the cMLM model performed well in limiting confounding effects (Fig. 1b).

All the significantly associated SNPs were located in the same region of chromosome 13 (Fig. 1b) and exhibited a very high degree of association (Table 1). The strongest degree of association (FDR-adjusted  $p$  value =  $1.00\text{E}^{-11}$ )

was shared by five SNPs. These five 5 peak SNPs explained approximately 34% of the phenotypic variation. The physical interval delimited by the peak SNPs was only 29.2 kb, extending between position 30,776,090–30,805,508 on chromosome 13, and all of these markers are located in intergenic regions.

Based on the peak SNPs identified by GWAS, individually or jointly, we could define only two haplotypes among the association panel, with accessions carrying Haplotype 1 showing a much lower disease rating (mean = 3.06) than accessions bearing Haplotype 2 (mean = 4.20) (Fig. 1e, Supplementary Table 1). In other words, it was possible to appropriately separate the resistant and moderately resistant accessions from susceptible accessions. For the SNP located at Chr13:30,804,961, the C allele was present in 25 of 30 resistant accessions (including the resistant checks PI 595,099, PI 230,977, Bragg, CD 201 and PI 200,538), and in approximately 66% of the moderately resistant accessions and in only 10% of susceptible accessions. In contrast, the alternate allele (T) was present in only 5 of the resistant accessions, and in approximately 34% of the moderately resistant accessions and 184 of the 205 (90%) susceptible accessions (Table 2).

Among the SoySNP50K data for the USDA soybean germplasm collection, two of the peak SNPs identified here were genotyped (Chr13:30,804,961 and Chr13:30,805,508). We identified a total of 19,978 accessions for which homozygous calls had been made at these two SNP loci and were able to classify these accessions as bearing Haplotype 1 or Haplotype 2 (Supplementary Table 1). In total, 2150 accessions were found to carry Haplotype 1 and 17,828 accessions bear Haplotype 2. This suggests that approximately 11% of the whole USDA collection may possess the same *M. javanica* resistance QTL on the chromosome 13.

**Table 1** SNPs significantly associated with resistance to *M. javanica* identified by GWAS

SNP ID	Chr <sup>a</sup>	Pos (bp) <sup>b</sup>	<i>P</i> -value	MAF <sup>c</sup>	<i>R</i> <sup>bd</sup>	FDR adjusted <i>p</i> -values <sup>e</sup>
GBSRmj961	13	30,804,961	$1.23 \times 10^{-15}$	0.31	0.34	$1.00 \times 10^{-11}$
GBSRmj499	13	30,805,499	$1.23 \times 10^{-15}$	0.31	0.34	$1.00 \times 10^{-11}$
GBSRmj500	13	30,805,500	$1.23 \times 10^{-15}$	0.31	0.34	$1.00 \times 10^{-11}$
GBSRmj508	13	30,805,508	$1.23 \times 10^{-15}$	0.31	0.34	$1.00 \times 10^{-11}$
GBSRmj090	13	30,776,090	$1.33 \times 10^{-15}$	0.29	0.34	$1.00 \times 10^{-11}$
GBSRmj409	13	30,792,409	$4.20 \times 10^{-14}$	0.31	0.32	$2.64 \times 10^{-10}$
GBSRmj474	13	30,792,474	$1.30 \times 10^{-13}$	0.31	0.31	$7.03 \times 10^{-10}$
GBSRmj686	13	30,804,686	$5.76 \times 10^{-12}$	0.27	0.29	$2.18 \times 10^{-8}$
GBSRmj726	13	30,804,726	$5.76 \times 10^{-12}$	0.27	0.29	$2.18 \times 10^{-8}$
GBSRmj752	13	30,804,752	$5.76 \times 10^{-12}$	0.27	0.29	$2.18 \times 10^{-8}$

<sup>a</sup>Chromosome, <sup>b</sup>physical position of the SNPs based on soybean genome (W82.a2.v1), <sup>c</sup>minor allele frequency, <sup>d</sup>*R* squared value (%) of the model with the SNP, which corresponds to the percentage of phenotypic variation explained by the SNP. <sup>e</sup>False-discovery rate (FDR)

**Table 2** Haplotypes obtained using the most significant SNPs from GWAS

Haplotype ID	Positions in the soybean genome—chromosome 13				<i>M. javanica</i> reaction <sup>b</sup>		
	30.804.961 <sup>a</sup>	30.805.499	30.805.500	30.805.508	<i>R</i>	<i>MR</i>	<i>S</i>
Hap-resistant	C	C	A	T	25	54	21
Hap-susceptible	T	A	G	G	5	28	184
Total = 317					30	82	205

<sup>a</sup>Physical position of the SNPs based on soybean genome (W82.a2.v1); <sup>b</sup>resistant—*R*; moderately resistant—*MR* and susceptible *S*

### Development of SNP genotyping assays to facilitate introgression of *M. javanica* resistance

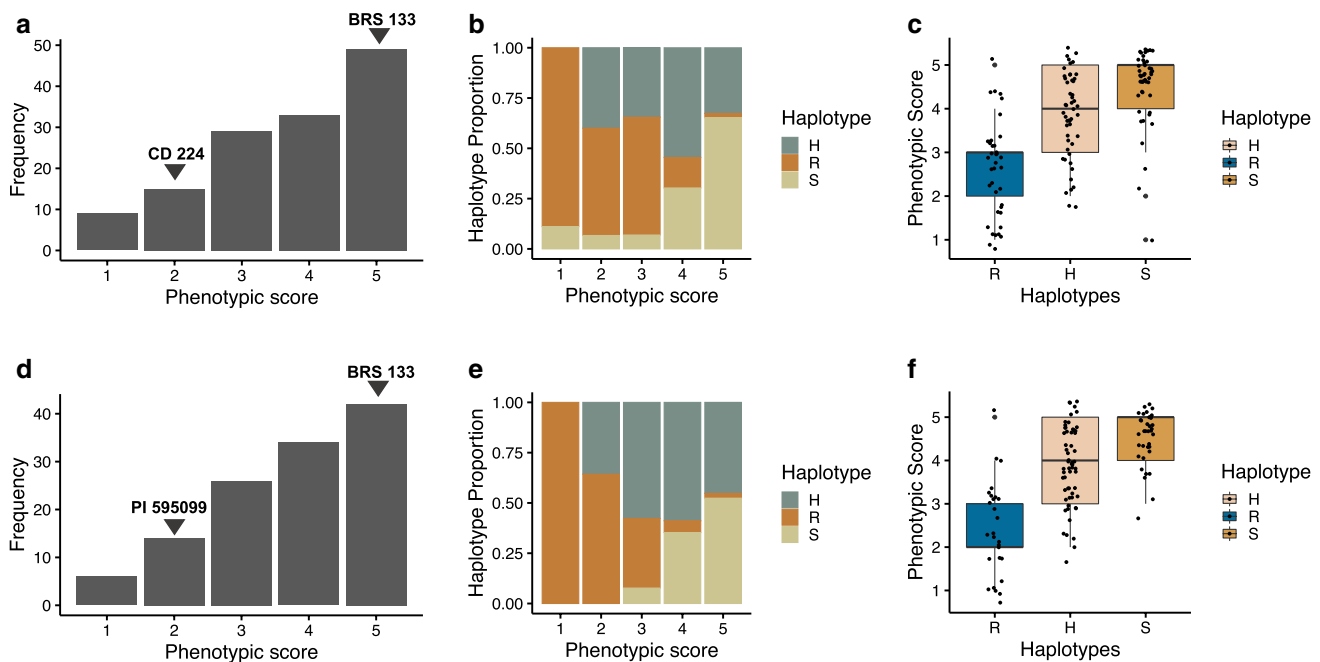
Two TaqMan® assays (M700 and M800) were designed to enable high-throughput genotyping of two of the peak SNPs. Both assays precisely discriminated between the two alleles in a set of 69 soybean accessions (Supplementary Fig. 3). In the group of resistant soybean accessions, 38 of 41 accessions carried the alleles associated with resistance in both assays. Thus, about 93% of concordance between phenotype and genotype (Supplementary Table 3) was observed. Regarding the incongruences, three cultivars classified as being resistant to *M. javanica* (BRS Pétala, BRSGO 204 and BRSGO 8661 RR) nonetheless carried the alleles associated with susceptibility. Of the 26 accessions classified as susceptible, 92.3% carried the alleles associated with susceptibility. Only BRS 285 and PI 398,887 had the alleles associated with resistance while being classified as susceptible.

To assess the degree of co-segregation between these SNP markers and *M. javanica* resistance, two F<sub>2</sub> populations derived from crossing resistant accessions known to carry Haplotype 1, PI 595,099 (disease score = 1) and CD 224 (score = 1.30) with the susceptible parent BRS 133 (score = 4.5) carrying Haplotype 2 were examined. The frequency distribution of the reaction to *M. javanica* in 135 F<sub>2</sub> individuals from the CD 224 × BRS 133 population (expressed as score values) is shown in Fig. 2a. The score means for the resistant parental CD 224 (2.76) and for the susceptible parent BRS 133 (4.72) were significantly different ( $P < 0.05$ ). The phenotypic distribution among the F<sub>2</sub> progeny spanned a wide range with a skewed distribution (Shapiro–Wilk's test;  $W = 0.853$ ,  $p < 0.01$ ). A majority of individuals were rated as susceptible, and few individuals showed ratings above or below the parental lines. The frequency distribution in the PI 595,099 × BRS 133 F<sub>2</sub> progeny was similar to the one described above, as shown in Fig. 2d. The average phenotypic scores of the parents were significantly different ( $p < 0.05$ ), since PI 595,099 and BRS 133 showed scores of 2.18 and 4.72, respectively. The F<sub>2</sub> phenotypic distribution was again skewed toward susceptibility ( $W = 0.860$ ,  $p < 0.01$ ) (Fig. 2a, d). In order to clarify the genetic basis of inheritance to *M. javanica*

resistance, Chi-square tests were performed in both populations (CD 224 × BRS 133 and PI 595,099 × BRS 133) to determine the goodness of fit of observed F<sub>2</sub> phenotypic segregation patterns to genetic models of one dominant gene (3:1 ratio) and two recessive genes (7:9 ratio). As expected for nematode resistance, segregation of resistant (*R*) and susceptible (*S*) progenies did not fit a 3R:1S ratio in F<sub>2</sub> populations, showing significant  $\chi^2$  values (Supplementary File 4). However, the Chi-square test for segregation of both F<sub>2</sub> populations showed a good fit of a 7 resistant: 9 susceptible (7R:9S) model, showing non-significant  $\chi^2$  values (Table 3).

In these two populations, both SNP markers segregated in the expected Mendelian ratio in the PI 595,099 × BRS 133 population, but not in the CD 224 × BRS 133 population (Supplementary Table 4). In the latter case, surprisingly, an excess of homozygotes and a deficit of heterozygotes were observed with both SNP markers. In the PI 595,099 × BRS 133 population, 58 of 63 individuals (92%) scored as homozygous (for either allele) exhibited a disease rating that was consistent with the allele at these two associated markers. In the CD 224 × BRS 133 population, 75 of 85 individuals (88%) scored as homozygous (for either allele) exhibited a disease rating that was consistent. In both crosses, a majority of individuals scored as heterozygous produced a susceptible disease rating.

To gain a further understanding of the relationship between the scored haplotype and the observed disease rating, we examined the proportion of individuals bearing each haplotype within each phenotypic class (Fig. 2b, d). In progeny with an extreme disease rating (either high or low score), the individuals mostly carried the expected haplotype. For example, in the CD 224 population, among progeny with a score of 1, eight were homozygous for Haplotype 1 (*R* haplotype) and one was homozygous for Haplotype 2 (*S* haplotype), while in the PI 595,099 population, individuals with score values of 1 and 2 showed only the *R* haplotype (Fig. 2b, Supplementary File 5). Finally, in progenies with susceptible reactions (score of 5), most of the individuals shared Haplotype 2 in both populations, and in progenies with intermediate disease scores (2, 3 and 4 values) all three haplotypes were observed.



**Fig. 2** Frequency distribution for *Meloidogyne javanica* resistance expressed as score values in CD 224×BRS 133 and PI 595,099×BRS 133  $F_2$  populations and haplotypes proportion constructed by genotyping results from M700/M800 assays. **a** Frequency distribution in CD 224  $F_2$  population. **b** Haplotype proportion in each phenotypic class (CD 224 population). **c** Box plot for *Meloidogyne javanica* plotted as different haplotypes (CD 224 population). **d** Frequency distribution in PI 595,099 population. **e** Haplotype proportion in each phenotypic class (PI 595,099 population). **f** Box plot for

*Meloidogyne javanica* plotted as different haplotypes (PI 595,099 population). R represents the haplotype containing homozygous resistant alleles, H represents the haplotype containing heterozygous alleles and S is the haplotype showing the susceptible alleles for M700/M800 assays. The middle line is the median and the outer dots are outliers in the boxplots. Black triangles represent the average values of resistant (CD 224 and PI 5,950,909) and susceptible (BRS 133) parents

**Table 3** Genetic of resistance to *Meloidogyne javanica* in  $F_2$  individuals derived from the soybean crosses CD 224×BRS 133 and PI 595,099×BRS 133

Parents/ $F_2$ populations	Number of plants			Chi-square		
	R <sup>a</sup>	S	Total	Model	$\chi^2$ value	<i>p</i> value
CD 224	17	—	17			
BRS 133	—	25	25			
PI 595,099	22	—	22			
CD 224×BRS 133	53	82	135	7:9	1.12 NS	0.29
PI 595,099×BRS 133	46	76	122	7:9	1.71 NS	0.19

<sup>a</sup>Plants showing up to a score of 3 were classified as resistant to *M. javanica*, S susceptible plants. NS: non-significance of the Chi-square value ( $P=0.05$ )

To characterize the effect of the locus on chromosome 13, all individuals from each  $F_2$  progeny set were grouped into the different three allelic states (Fig. 2c, f). Significant differences in *M. javanica* resistance among these genotypic classes were observed (Kruskal–Wallis test;  $p < 0.001$ ) (Supplementary File 4). Then, further analysis, including the Wilcoxon signed rank test for pairwise comparisons using Bonferroni correction, showed that QTL effects in each haplotype were significantly different. Indeed, in progeny homozygous for the resistant haplotype, accessions showed a mean score of 2.56 and 2.37 in the CD 224×BRS 133

and PI 595,099×BRS 133 populations, respectively. The progeny scored as heterozygous showed a mean score of 3.88 (CD 224×BRS 133) and 3.90 (PI 595,099×BRS 133). Among progeny homozygous for the allele associated with susceptibility, mean scores of 4.54 (CD 224×BRS 133) and 4.56 (PI 595,099×BRS 133) were observed. These stark phenotypic contrasts between the different genotypic classes in both populations confirm that marker-assisted selection among populations segregating for the locus on chromosome 13 would facilitate the selection of progeny with improved resistance.



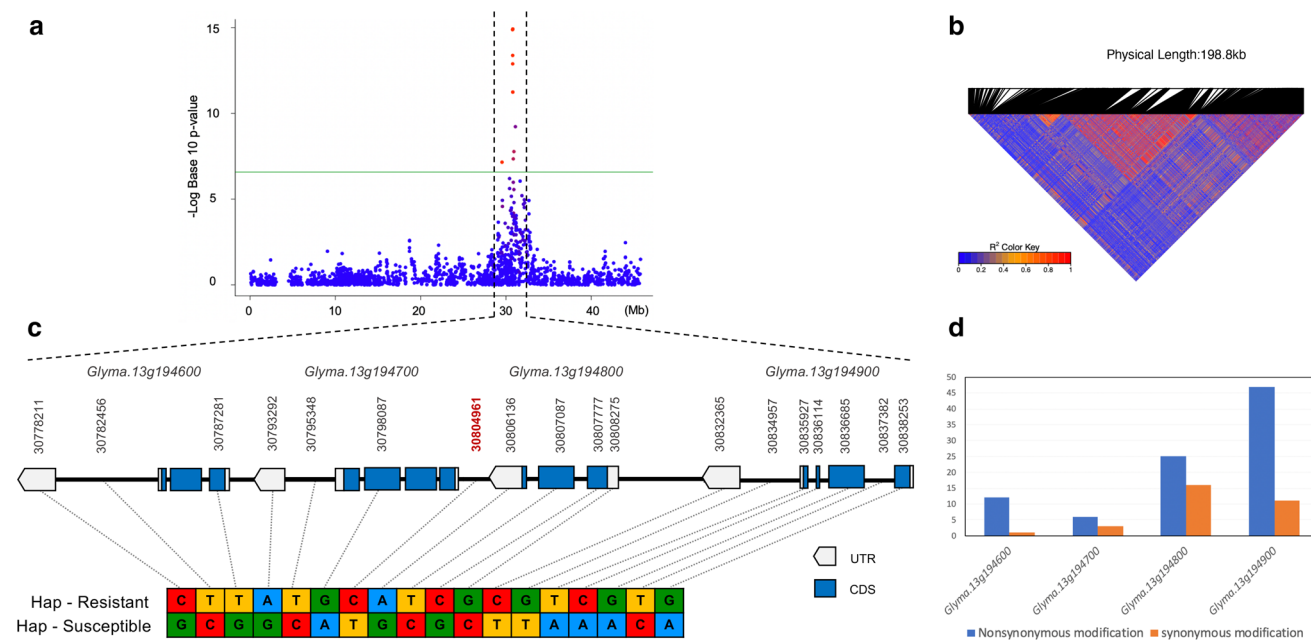
## Identification of candidate genes and their expression

According to the analysis of LD blocks, nine of the ten significant SNPs (SNP 30,776,090 was the exception) were in a single LD block of 117 kb (Supplementary Fig. 4) along with four other SNPs that were not significantly associated with *M. javanica* resistance. To help gain insights into possible candidate genes and causal variants not detected by the GBS approach, we also examined this LD block using all SNPs located in this region of chromosome 13 within a WGS dataset for 28 accessions of known reaction to *M. javanica* (5 resistant, 7 moderately resistant and 16 susceptible). We examined the LD in the mapped region (Fig. 3b). Within this set, 187 SNPs were in complete LD ( $r^2=1$ ) with the peak SNPs identified via GWAS (Supplementary Table 5). Considering only markers in complete LD, these encompassed a region of 118.9 kb, containing nine gene models (Table 4). This LD region mostly overlapped with the LD region identified using SNPs from GBS, thus demonstrating the accuracy of prediction of LD size.

All nine genes showed SNPs in both their non-coding and coding regions. However, among the 334 SNPs linked to the peak SNP from GWAS, most of these variants (186 SNPs) were located inside only four gene models (Fig. 3c,

Supplementary Table 6). As expected, due to the  $r^2$  values  $> 0.9$ , the haplotype analysis showed that these additional SNPs identified in WGS dataset are completely redundant, therefore bringing the same power of discrimination between resistance and susceptible in the 28 accessions with *M. javanica* reaction (Supplementary Table 1). Summarizing, all the five resistant soybean accessions shared the resistant alleles; in the moderately resistant group, six accessions shared the resistant alleles and one accession shared the susceptible haplotype. Among the susceptible group, 14 accessions shared the susceptible haplotype and only two accessions shared the resistant haplotype.

Regarding the functional annotation of the genes in the region, five gene models are predicted as disease resistance TIR-NBS-LRR class protein-coding genes (*Glyma.13g194600*, *Glyma.13g194700*, *Glyma.13g194800*, *Glyma.13g194900* and *Glyma.13g195100*). Two genes showed unknown domains (*Glyma.13g195000* and *Glyma.13g195400*), one gene domain coding for serine threonine/protein kinase/ankyrin repeat domain (*Glyma.13g195200*) and one gene predicted as a hydrolase (*Glyma.13g195300*). Among the TIR-NBS-LRR class genes, *Glyma.13g194600* contains a Toll/interleukin-1 receptor (TIR) domain. *Glyma.13g194800* and *Glyma.13g194900* genes were predicted as containing TIR and NB-ARC



**Fig. 3** Haplotypes analyses using additional SNPs identified on 31 resequenced soybean accessions and allelic effects. **a** Manhattan plot showing the negative  $\log_{10} p$  values from GWAS for soybean resistance to *M. javanica* with the significant SNPs plotted against base pair positions (Mb) on soybean chromosome 13. **b** Heatmap representing the  $R^2$  values between 1048 SNPs inside of GWAS region (SNPs identified as heterozygous as well as SNPs showing values of MAF  $< 0.05$  were not used). **c** Schematic graph shows the position of

the additional SNPs identified in the 4 genes models, the SNP position highlighted in red represents the peak SNP identified by GWAS and located between the *Glyma.13g194700* and *Glyma.13g194800*, for clear visualization, only some SNPs were used in the haplotypes construction, since 186 SNPs showed  $R^2$  values of  $> 0.9$  with the peak SNP were identified. **d** Predicted impact of variants in the four gene models, representing synonymous and non-synonymous SNPs (colour figure online)

**Table 4** Candidate resistance genes located in the GWAS region on the chromosome 13

Candidate gene <sup>a</sup>	Functional annotation	Physical location <sup>b</sup>	Ortholog number <sup>c</sup>
<i>Glyma.13g194600</i>	Disease resistance protein (TIR-NBS-LRR class) family	30,777,645–30,787,581 (–) <sup>d</sup>	AT4G14370.1
<i>Glyma.13g194700</i>	Disease resistance protein (TIR-NBS-LRR class) family	30,793,151–30,800,421 (–)	AT5G17680.1
<i>Glyma.13g194800</i>	Disease resistance protein (TIR-NBS-LRR class) family	30,806,038–30,808,277 (–)	AT1G64070.1
<i>Glyma.13g194900</i>	Disease resistance protein (TIR-NBS-LRR class) family	30,832,071–30,838,537 (–)	AT5G36930.2
<i>Glyma.13g195000</i>	No domain predicted	30,833,034–30,833,451 (+)	–
<i>Glyma.13g195100</i>	Disease resistance protein (TIR-NBS-LRR class) family	30,850,314–30,856,196 (+)	AT5G17680.1
<i>Glyma.13g195200</i>	Protein kinases; ubiquitin-protein ligases	30,862,185–30,877,782 (+)	AT5G13530.1
<i>Glyma.13g195300</i>	Nudix hydrolase homolog	30,878,334–30,882,722 (–)	AT3G12600.1
<i>Glyma.13g195400</i>	No domain predicted	30,895,968–30,896,400 (+)	–

<sup>a</sup>Genes ID from Williams 82 genome, assembly version 2.0—Wm82.a2.v1 (<https://phytozome.jgi.doe.gov/>)

<sup>b</sup>Physical location in chromosome 13 (Wm82.a2.v1)

<sup>c</sup>Accession number of orthologs obtained from *Arabidopsis* Information Resource (TAIR) database

<sup>d</sup>Represents the antisense strand and + represents the sense strand

domains, while *Glyma.13g194700* and *Glyma.13g195100* genes were predicted to contain a complete TIR-NBS-LRR, with TIR and NB-ARC domains and LRR-regions (Table 4).

The functional impact of the SNPs was predicted in the nine gene models (Supplementary Table 6), especially regarding nonsynonymous modifications. The highest values were predicted in the TIR-NBS-LRR class genes (*Glyma.13g194900*, *Glyma.13g194800* and *Glyma.13g194600*). The other two TIR-NBS-LRR genes showed similar values of nonsynonymous modifications, six and seven for *Glyma.13g194700* and *Glyma.13g195100*, respectively. Therefore, we selected the region containing the four gene models (*Glyma.13g194600*, *Glyma.13g194700*, *Glyma.13g194800* and *Glyma.13g194900*) in which most of the SNPs identified by GWAS are located. Summarizing, while the *Glyma.13g194600* and *Glyma.13g194700* gene models showed lower nonsynonymous SNPs with functional impact, several nonsynonymous SNPs with moderate effect were identified in TIR and NB-ARC domains coding regions for *Glyma.13g194800* and *Glyma.13g194900*. Therefore, these results suggested that these genes might be under strong selective pressure to modification.

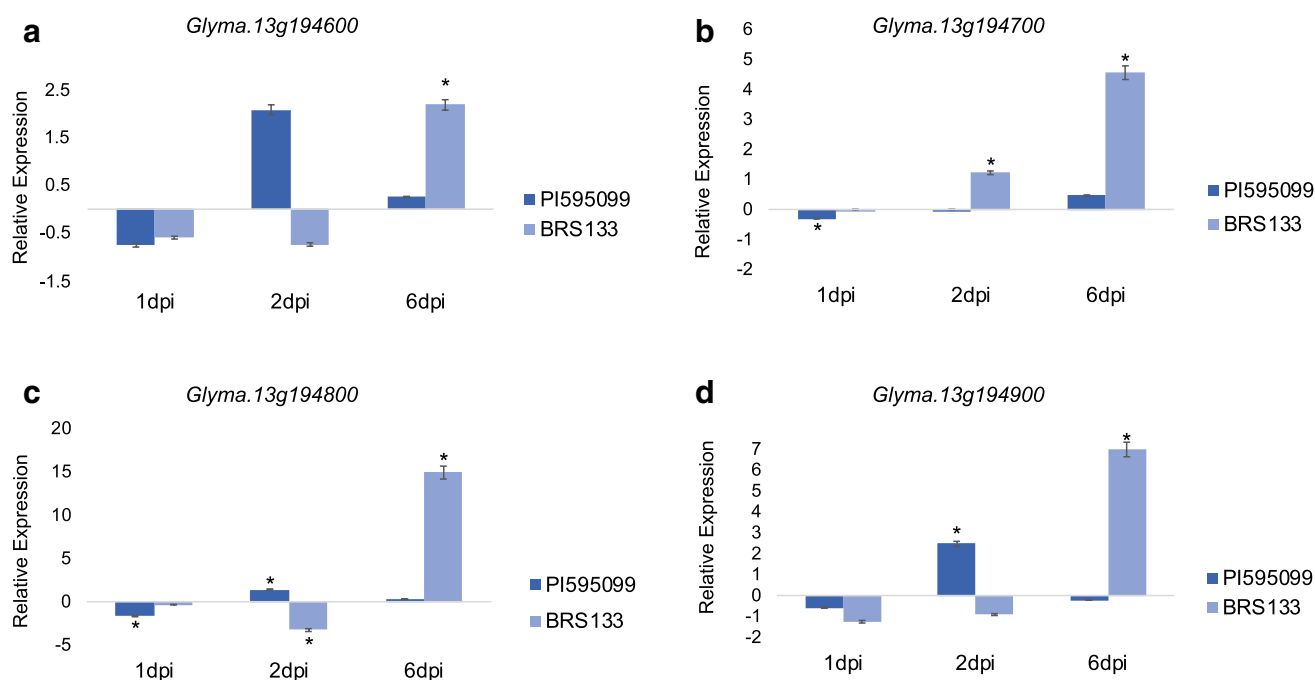
Finally, we investigated the gene expression levels of these four TIR-NBS-LRR class genes in PI 595,099 and BRS 133 soybean roots inoculated with *M. javanica* by RT-qPCR (Fig. 4, Supplementary Table 7). After inoculation, the gene model *Glyma.13g194600* showed a significant difference in expression (~2.5-fold) only in BRS 133 after 6 dpi (Fig. 4a). Interestingly, *Glyma.13g194700* was down-regulated in PI 595,099 at 1 dpi; however, it showed increased expression levels in BRS 133 from ~1.5-fold to 4.5-fold at 2 and 6 dpi, respectively (Fig. 4b). For the *Glyma.13g194800* and *Glyma.13g194900* gene models, a significant difference between resistant and susceptible parents was observed. Initially, *Glyma.13g194800* was down-regulated in PI 595,099

at 1 dpi, and then, at 2 dpi, the gene expression level was significantly up-regulated in the resistant accession PI 595,099 (~3.5-fold) and down-regulated in BRS 133 (Fig. 4c). At 6 dpi, gene expression levels significantly increased only in the susceptible accession BRS 133 (~15.0-fold). A similar expression profile was observed for *Glyma.13g194900*, in which no induction for both accessions was observed at 1 dpi (Fig. 4d). Otherwise, at 2dpi, gene expression levels significantly increased in PI 595,099 (~3.0-fold) and, at 6 dpi, gene expression levels were significantly increased in susceptible accession BRS 133 (~7.0-fold). Altogether, results showed that *Glyma.13g194800* and *Glyma.13g194900* genes were differentially expressed between resistant and susceptible soybean accessions in the presence of *M. javanica* and may play roles in plant defense responses to nematode infection in the initial stages of infection.

## Discussion

### Characterization of soybean resistance against *M. javanica* in 317 accessions

The root-knot nematode (*M. javanica*) is an important pest problem for soybean growers in Brazil and around the world (Wrather et al. 1997; Dias et al. 2010; Oyekanmi and Fawole 2010). The development of cultivars showing genetic resistance is one of most efficient control strategies for the management of this nematode problem. An important step in the development of resistant cultivars is (1) to identify loci conferring partial or complete resistance, (2) the accessions sharing the same or different resistance loci and (3) SNP markers closely linked to these loci to enable marker-assisted selection.



**Fig. 4** Gene expression profile of candidate genes for resistance to *Meloidogyne javanica*. **a** relative expression levels of *Glyma.13g194600*. **b** relative expression levels of *Glyma.13g194700*. **c** relative expression levels of *Glyma.13g194800*, **d** relative expres-

sion levels *Glyma.13g194900* (**d**). \*Significant expression level at 95% probability. Relative fold expression levels between soybean roots inoculated with *Meloidogyne javanica* eggs at 1, 2 and 6 dpi (days post-inoculation)

In the present study, we successfully performed the characterization of the reaction against *M. javanica* in a collection of 317 soybean accessions from different origins. Overall, among 317 accessions, 31 were classified as resistant, 82 accessions scored as moderately resistant and a larger proportion as susceptible, 204 accessions. Our phenotypic characterization identified a small number of 12 accessions with high levels of resistance (disease score  $\leq 2$ ) that are from different origins (Brazil, USA, Japan and China) and most of these are supported by previous studies. For example, the resistance sources PI 595,099 (Davis et al. 1998; Silva et al. 2001a; Beneventi et al. 2013), cv. Bragg and CD 201 (Silva et al. 2001a, b) had been previously described. Interestingly, other resistant accessions had been described as resistant to other nematodes, such as PI 506,862 and PI 089,772, both resistant to *M. incognita* and races of *H. glycines* (Anand and Luedders 1989; Lee et al. 2015); PI 437,679 and PI 608,357 also described as resistant to *H. glycines* (Anand and Luedders 1989; Kilen and Young 2000). The known R source, PI 230,977 (Luzzi et al. 1987), which is described as donor of resistance to PI 595,099 (Davis et al. 1998), surprisingly, showed a lower level of resistance (score of 2.50) compared to PI 595,099 (score of 1). As expected, the susceptible soybean accessions cv. BRS 133 (Silva et al. 2001a), cv. BRSMT Pintado (Mattos et al. 2016) and cv. BRSMG 250 (Nobreza) (Mattos et al. 2016) showed also susceptible reactions in our results.

Finally, we provided an in-depth characterization of a diverse soybean set and novel information regarding several soybean accessions identified as resistant or moderately resistant is reported for the first time. Among those accessions, some were highlighted for their high levels of resistance, such as PI 437127B, cv. CD 224, CD 219 RR, and BRSMG 850G RR. All these data will be useful in different breeding programs, since we provided a diverse set of resistant accessions from several origins.

### GWAS highlights a locus on chromosome 13 associated with resistance against *M. javanica*

We genotyped a diverse set of 317 accessions with approximately 45 K SNPs obtained from GBS, allowing us to identify SNPs associated with a region on chromosome 13 conferring resistance to *M. javanica*. Our results provide advantages in mainly two aspects. First, given the nature of the GWAS approach, we narrowed down the genomic region previously identified in biparental mapping studies, and also providing SNP markers tightly associated with the R locus, which are more compatible and useful in marker-assisted selection. Indeed, the SNPs identified by GWAS (30,776,090–30,805,508 bp) define a region of ~ 29 kb, or the extent of the LD region (~ 119 kb), which contains only nine genes, and the genomic region was significantly reduced. In prior studies, the SSR markers SOYHSP

176, Sat133 and Satt114 flanked a region of ~4 Mb (24,949,516–29,041,673 bp) in progenies of biparental crosses derived from both PI 595,099 and CD 201 {Formatting Citation}. The RFLP marker B212-1 (~29,000,000 bp on chr. 13) was also associated with resistance to *M. javanica* in progenies derived from PI 230,977 (Tamulonis et al. 1997a) and cv. Gazelle (Mienie et al. 2002) and to resistance against *M. arenaria* in PI 200,538 (Tamulonis et al. 1997b). The phenotypic variance explained by SNPs identified by our GWAS was around 34%, which was similar to phenotypic variance explained by SNPs from GWAS for resistance to *M. incognita* in soybean (25–40%). These values of phenotypic variance are indicative of other rare alleles (MAF < 0.05) and small-effect QTLs contributing to resistance, but not captured by our GWAS, as pointed out in the GWAS for susceptibility to *M. incognita* in *Arabidopsis thaliana*, in which the phenotypic variance was only 22%.

In addition to the already known sources of resistance to *M. javanica*, 21 soybean accessions sharing the resistant haplotype were identified in this study, including adapted cultivars that can be used for introgression of resistance instead of ill-adapted PIs. This fact suggests that the locus on chromosome 13 is widely spread in resistant soybean accessions. Interestingly, we also identified six accessions (BRS 239, CD 225 RR, UFV10, PI 253651D, PI 424,588 and PI 438,190) showing a resistance phenotype, but susceptible haplotype (Haplotype 2) on chromosome 13. The resistance in these accessions is likely due to other *loci* than the one of chromosome 13. Because of their small number (only 1.6% of the 317 accessions), even if they shared the same resistance locus or not, markers associated with this locus would likely have been removed when filtering for minor allele frequency ( $\geq 5\%$  or  $1\%$ ). Such accessions would, however, be extremely useful as targets to perform QTL mapping in biparental crosses with a susceptible accession to explore whether or not their resistance is conferred by another locus. If this were the case, with marker-assisted selection, it could be possible to pyramid these different QTLs to develop improved lines with a much more robust resistance to *M. javanica*.

Interestingly, the screening of accessions belonging to the USDA soybean collection, using our peak SNPs linked to *M. javanica* resistance, identified many accessions showing the resistant alleles that also have already been described as containing resistance to other nematodes species, for example the cv. Peking, and PI 339,868 B, described as resistant to various races of *H. glycines* and *Rotylenchulus reniformis*, as well as cv. Forrest described as resistant to *H. glycines*, *R. reniformis* and *M. incognita* (Klepadlo et al. 2018). In our results, PI 200,538 (known source of resistance to *M. arenaria*) was classified as moderately resistant to *M. javanica* and shared the resistant alleles on chromosome 13. Similarly, PI 595,099 was already described as

resistant to *M. arenaria* (Luzzi et al. 1997). Based on this, we propose that the R locus on chromosome 13 may confer resistance to both *Meloidogyne* species; however, further studies should be done to confirm this hypothesis. For plant breeding purposes, the information might be useful to select soybean sources to introgression of multiple resistance for nematodes species.

Finally, our GWAS results are important due to the lack of prior association studies aimed at resistance against *Meloidogyne* sp. nematodes, especially in soybean. As far we know, only a single GWAS for resistance to *M. incognita* has been published for soybean (Passianotto et al. 2017) and for *Arabidopsis* (Warmerdam et al. 2018). Indeed, almost all the previous GWAS were focused on soybean cyst nematode (SCN) caused by *Heterodera glycines* (Vuong et al. 2015; Zhang et al. 2016, 2017; Tran et al. 2019).

### Validation of SNP assays

In our view, one important further step of any GWAS is to validate the QTLs associated with the SNPs identified. Most of the GWAS published for soybean only identified SNPs associated with genomic regions without any validation step or development of marker assays to directly apply in marker-assisted selection. Therefore, we also developed marker assays for two of the peak SNPs from the GWAS and analyzed the agreement between these SNPs and the *M. javanica* phenotype in two  $F_2$  populations derived from known R sources.

The phenotypic distribution in both progenies, one for PI 595,099 and other for cv. CD 224, indicates that the *M. javanica* resistance was quantitatively inherited, following a non-normal distribution, skewed toward the susceptible phenotypes. Such phenotypic distribution patterns are supported by studies with *Meloidogyne* species and other nematodes in different legumes, as *M. incognita* (Oliveira et al. 2015), *M. javanica* (Silva et al. 2001a) and *Rotylenchulus reniformis* (Cardoso et al. 2014) in soybean, as well as *M. javanica* in cowpea (Huynh et al. 2016) and *M. arenaria* in peanut (Burow et al. 2014; Ballén-Taborda et al. 2019). Results of the two TaqMan® assays evaluated in both populations showed that in general,  $F_2$  plants showing high levels of resistance (score  $\leq 2$ ) were homozygous for the resistance allele (Haplotype 1). Otherwise, most of the lines showing moderate resistance were heterozygous, while a large proportion of the lines showing high levels of susceptibility showed the susceptible alleles (Haplotype 2).

As resistance to *M. javanica* is widely described as a quantitative trait, as expected, we were not able to explain all the genetic control of resistance in PI 595,099 and CD 224. The inheritance of resistance to *M. javanica* in soybean was described as quantitative, showing a moderate to high heritability ( $h^2 = 0.48$ – $0.76$ ) (Tamulonis et al. 1997a). The



authors proposed that different accessions might possess resistance genes at different *loci* or different alleles at the same *locus*. Overall, the genetic resistance to nematodes in soybean is a complex trait (Mitchum 2016). For example, cyst nematode resistance was early described as qualitative, with dominant and recessive *Rhg* genes (Cook et al. 2012; Liu et al. 2012; Bao et al. 2014) and later as quantitative (QTLs) in soybean (Anand and Luedders 1989; Vuong et al. 2015), with 31 putative QTL mapped to 17 of the 20 soybean chromosomes (Concibido et al. 2004). The complex nature of resistance to nematode contributes to explain the failure in explaining all the phenotypic variation by the SNP markers. The validation of a codominant SCAR marker in a F<sub>2</sub> soybean population derived from a cross between a resistant cultivar (Gazelle) and a highly susceptible variety (Prima) explained 42% of gall-index variation in the progenies inoculated with *M. javanica* (Mienie et al. 2002), not being able to explain all the variation as well. By validating SNP markers linked to two different QTLs in a F<sub>2</sub> cowpea population, high resistance to *M. javanica* was observed in the lines with both QTLs in the homozygous resistant condition compared to lines harboring only one of the QTL. Probably additional minor QTLs not detected in GWAS are contributing to minor effect to the resistance to *M. javanica* in PI 595,099 and CD 224 progenies. QTL mapping studies should be done in order to identify these other *loci* and to clarify the resistance to *M. javanica* in PI 595,099 and CD 224 accessions.

To further characterize the use of these marker assays, we genotyped a subset of the GWAS panel composed mostly of Brazilian cultivars. Most breeding efforts to introduce resistance to *M. javanica* in Brazilian soybean cultivars have used cv. Bragg (Dias et al. 2010) as a resistant donor. As this accession carries the resistant haplotype (Haplotype 1) on chromosome 13, this can explain the good concordance (~93%) observed between the genotype determined using the TaqMan® assays and the resistance phenotype of these cultivars. In the few cases where there was a lack of concordance, it could be due to the resistance having been derived from another resistance source or errors in phenotyping or genotyping.

### Functional annotation and gene expression levels supporting the TIR-NBS-LRR genes underlying resistance to *M. javanica*

In our results, the SNPs associated with the resistance to *M. javanica* in soybean are located in a region that is rich in genes presenting the TIR-NBS-LRR domains. Indeed, *R* genes to nematodes have been also identified and cloned in different crops, and the majority of them are described as LRR-NBS-type genes, which supports our results, for example genes for *M. incognita* in tomato (*Mi-1* and *Mi-9* genes)

(Jablonska et al. 2007), pepper (*CaMi*) (Chen et al. 2007), for *Meloidogyne floridensis* in *Prunus* species (*Ma*) (Clavierie et al. 2011), for *M. incognita* and *Heterodera schachtii* in *Arabidopsis* (*NILRI*) (Mendy et al. 2017; Warmerdam et al. 2020), and for in cowpea (*QRk-vu9.1*) (Santos et al. 2018). The *M. javanica* resistance in soybean in the biparental populations did not show dominant inheritance; however, two recessive genes were indicated as underlying resistance to *M. javanica*. Examples of recessive genes controlling resistance to many pathogens, including to nematodes, have been described, for example, a single recessive gene controlling resistance against *M. javanica* in cucumber (Walters et al. 1997; Devran et al. 2011) and *M. arenaria* in peanut (Church et al. 2005). Most of the LRR-NBS-types genes are described as dominant genes, conferring race-specific resistance based on the gene-for-gene theory (Flor, 1956). However, this model does not provide a clear explanation for all types of disease resistance in plants. At least one recessive *R* gene being the LRR-NBS type has been described, the *RRS1-R* resistance gene to *Ralstonia solanacearum* in *Arabidopsis* (Deslandes et al. 2002). Additional studies are necessary, but we cannot exclude the possibility of the LRR-NBS-type gene involved in a recessive resistance at least in the PI595099.

While the functional impact predictions identified few nonsynonymous on *Glyma13.g194600* and *Glyma13.g194700*, several nonsynonymous SNPs inside of *Glyma13.g194800* and *Glyma13.g194900* genes were identified, pointing out these genes are strong resistance genes candidates. Indeed, the effect of nonsynonymous SNPs modifying the gene function on soybean resistance to nematode was previously reported in locus *Rgh4*, where changes in two nucleotides alter the role of this protein in resistance to cyst nematode (Liu et al. 2012).

Also, the expression levels of *Glyma13g.194800* and *Glyma13g.194900* were up-regulated in the resistance accession PI 595,099 and down-regulated in cv. BRS 133 at 2 dpi, giving the assumption that the timing of induction might be an important factor in the resistance response to *M. javanica* in soybean. As expected, the initial recognition of the pathogen by the resistance genes, followed by signal transduction cascade, might contribute to the global gene activation toward the nematode resistance on soybean. On the other hand, *Glyma13g.194700* was only significantly induced in the susceptible cv. BRS 133 in 2 and 6 dpi. It is possible that the times evaluated in our study may not be enough to capture all the gene expression profile. Indeed, the up-regulated genes in soybean plants inoculated with *M. incognita* showed high levels of expression in 12 dpi (Pham et al. 2013), many hours before the first time we have evaluated in our study.

Our study reported that the TIR and NB domains encoding-genes (*Glyma.13g194800* and *Glyma.13g194900*)

possessed several nonsynonymous variants in their sequences and were differentially expressed in the resistance soybean source compared to susceptible one. Therefore, our gene expression results and findings reported in previous studies suggested that LRR-NBS-type genes may be the mainly R genes related to nematodes resistance in several plant species.

## Conclusion

The present study reported the identification of a locus conferring partial resistance to *M. javanica* on chromosome 13 in soybean, and this region contains a LRR-NBS cluster. As far as we know, our GWAS is the first one reporting SNPs markers associated with the resistance to this important nematode. We validated two SNPs-based TaqMan® assays on two biparental populations and in Brazilian cultivars, and we observed high correlation levels between the SNPs and the phenotype. Surprisingly, several soybean accessions reported as resistant to other important nematodes species, as *M. incognita*, *M. arenaria* and *H. glycine* shared the haplotype on chromosome 13 in our panel, leading us the assumption that these soybeans accessions may possess multiple-resistance to nematodes, which are helpful for plant breeding programs. Several additional nonsynonymous SNPs on *Glyma.13g194800* and *Glyma.13g194900* were identified in the mapped region using WGS data, and expression levels of these genes were only detected in PI 595,099, suggesting them as strong candidates to resistance genes for future functional validation. Finally, the SNPs assays identified here can be used in the marker-assisted selection for screening for resistance to *M. javanica* in soybean breeding programs, as well as the resistance sources.

**Acknowledgements** This work was supported by INCT Plant Stress Biotech Grant from MCT of the Brazilian Government and CNPq, EMBRAPA Soybean and the National Council for the Improvement of Higher Education (CAPES)—Finance code 001 for a scholarship. Thanks also to the team of professionals in Plant Biotechnology Laboratory and Marker-Assisted and Molecular Genetics Laboratory from Embrapa-Soja, Londrina, Brazil.

**Author contribution statement** RVA and FCMG designed and planned the project. JCA and WPD conducted the phenotypic approach. JCA, ALLP and EGCF performed GWAS analysis. ABS and DCGS performed the genotyping in germplasm and biparental populations. JCA performed the RT-qPCR studies. EGCF conducted the prediction of effects, WGS and statistical analysis. JCA and EGCF drafted the manuscript. FCMG, RVA, EGCF and FB supervised the study and provided assistance for manuscript preparation.

**Funding** This work was supported by INCT Plant Stress Biotech Grant from MCT of the Brazilian Government and CNPq, EMBRAPA Soybean and the National Council for the Improvement of Higher Education (CAPES)—Finance code 001 for a scholarship.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no competing interest.

**Availability of data and material** All phenotypic data are provided in supplementary data as well genotypic data (SNPs on chromosome 13) for 317 soybean accessions. GBS data and GAPIT code for running GWAS are available through direct contact to the corresponding author Dra. Francismar C. Marcelino-Guimarães by email: francismar.marcelino@embrapa.br.

## References

- Anand SC, Luedders VD (1989) Use of soybean cyst nematode inbreds to determine genetic diversity among resistant soybeans. *Crop Prot* 8:380–382. [https://doi.org/10.1016/0261-2194\(89\)90059-8](https://doi.org/10.1016/0261-2194(89)90059-8)
- Ballén-Taborda C, Chu Y, Ozias-Akins P et al (2019) A new source of root-knot nematode resistance from *Arachis stenosperma* incorporated into allotetraploid peanut (*Arachis hypogaea*). *Sci Rep* 9:1–13. <https://doi.org/10.1038/s41598-019-54183-1>
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bth457>
- Bao Y, Vuong T, Meinhardt C et al (2014) Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. *Plant Genome*. <https://doi.org/10.3835/plantgenome2013.11.0039>
- Bastien M, Sonah H, Belzile F (2014) Genome Wide Association Mapping of Resistance in Soybean with a Genotyping-by-Sequencing Approach. *Plant Genome* 7:1–13. <https://doi.org/10.3835/plantgenome2013.10.0030>
- Beneventi MA, Bonfim O, Eugênia M, et al (2013) Transcription profile of soybean-root-knot nematode interaction reveals a key role of phytohormones in the resistance reaction. *BMC Genomics* 14:322. <https://doi.org/10.1186/1471-2164-14-322>
- Boudhrioua C, Bastien M, Torkamaneh D, Belzile F (2020) Genome-wide association mapping of Sclerotinia sclerotiorum resistance in soybean using whole-genome resequencing data. *BMC Plant Biol* 20(1):195. <https://doi.org/10.1186/s12870-020-02401-8>
- Burow MD, Starr JL, Park CH et al (2014) Introgression of homeologous quantitative trait loci (QTLs) for resistance to the root-knot nematode [*Meloidogyne arenaria* (Neal) Chitwood] in an advanced backcross-QTL population of peanut (*Arachis hypogaea* L.). *Mol Breed* 34:393–406. <https://doi.org/10.1007/s11032-014-0042-2>
- Cardoso PC, Asmus GL, Gonçalves MC et al (2014) Inheritance of soybean resistance to *Rotylenchulus reniformis*. *Trop Plant Pathol* 39:251–258. <https://doi.org/10.1590/S1982-56762014000300009>
- Chen R, Li H, Zhang L et al (2007) CaMi, a root-knot nematode resistance gene from hot pepper (*Capsium annuum* L.) confers nematode resistance in tomato. *Plant Cell Rep* 26:895–905. <https://doi.org/10.1007/s00299-007-0304-0>
- Church GT, Starr JL, Simpson CE (2005) A recessive gene for resistance to *Meloidogyne arenaria* in interspecific *Arachis* spp. hybrids. *J Nematol* 37:178–184
- Cingolani P, Platts A, Wang LL et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff Fly (Austin). <https://doi.org/10.4161/fly.19695>
- Claverie M, Dirlwanger E, Bosselut N et al (2011) The ma gene for complete-spectrum resistance to *meloidogyne* species in prunus is a TNL with a huge repeated c-terminal post-LRR region. *Plant Physiol* 156:779–792. <https://doi.org/10.1104/pp.111.176230>

- Concibido VC, Diers BW, Arelli PR (2004) A decade of QTL mapping for cyst nematode resistance in soybean. *Crop Sci*
- Cook DE, Lee TG, Guo X et al (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* 338:1206–1209. <https://doi.org/10.1126/science.1228746>
- Cruz CD (1998) Programa GENES: Aplicativo Computacional em Estatística Aplicada à Genética (GENES - Software for Experimental Statistics in Genetics). *Genet Mol Biol* 21. <https://doi.org/10.1590/s1415-47571998000100022>
- Dall'Agnol A, Antônio H (1982) Reação de genótipos de soja aos nematóides formadores de galhas *Meloidogyne incognita* e *M. javanica*. *Soc Bras Nematol* 51–77
- Davis EL, Meyers DM, Burton JW, Barker KR (1998) Resistance to root-knot, reniform, and soybean cyst nematodes in selected soybean breeding lines. *J Nematol* 30(4s):530–541
- Deslandes L, Olivier J, Theulières F et al (2002) Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes. *Proc Natl Acad Sci U S A* 99:2404–2409. <https://doi.org/10.1073/pnas.032485099>
- Devran Z, Firat AF, Tör M et al (2011) AFLP and SRAP markers linked to the mj gene for root-knot nematode resistance in cucumber. *Sci Agric* 68:115–119. <https://doi.org/10.1590/s0103-90162011000100017>
- Dias WP, Garcia A, Silva JFV, Cameiro GEDS (2010) Nematóides em soja: identificação e controle. *Embrapa Soja Circ* 76:1–8
- dos Santos JVM, Valliyodan B, Joshi T et al (2016) Evaluation of genetic variation among Brazilian soybean cultivars through genome resequencing. *BMC Genom* 17:1–18. <https://doi.org/10.1186/s12864-016-2431-x>
- Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:1–10. <https://doi.org/10.1371/journal.pone.0019379>
- Ferraz LCCB (2001) As Meloidoginoses da Soja: Passado, Presente e Futuro. *Relações Parasito-Hospedeiro nas Meloidoginoses da Soja*
- Flor HH (1956) The Complementary Genic Systems in Flax and Flax Rust. *Adv Genet* 8:29–54. [https://doi.org/10.1016/S0065-2660\(08\)60498-8](https://doi.org/10.1016/S0065-2660(08)60498-8)
- Fuganti R, Fuganti R, Beneventi MA et al (2004) Identificação de Marcadores Moleculares de Microssatélites para Seleção de Genótipos de Soja Resistentes a *Meloidogyne javanica*. *Nematol Bras* 28:125–130
- Fuganti R, Machado M de FP da S, Lopes VS et al (2010) Size of AT(n) insertions in promoter region modulates Gmhsp17.6-L mRNA transcript levels. *J Biomed Biotechnol* 2010:1–9. <https://doi.org/10.1155/2010/847673>
- Gabriel SB, Schaffner SF, Nguyen H et al (2002) The structure of haplotype blocks in the human genome. *Science*. <https://doi.org/10.1126/science.1069424>
- Huynh BL, Matthews WC, Ehlers JD et al (2016) A major QTL corresponding to the Rk locus for resistance to root-knot nematodes in cowpea (*Vigna unguiculata* L. Walp.). *Theor Appl Genet* 129:87–95. <https://doi.org/10.1007/s00122-015-2611-0>
- Iquira E, Humira S, François B (2015) 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* 15:5. <https://doi.org/10.1186/s12870-014-0408-y>
- Jablonska B, Ammiraju JSS, Bhattarai KK et al (2007) The Mi-9 gene from *Solanum arcanum* conferring heat-stable resistance to root-knot nematodes is a homolog of Mi-1. *Plant Physiol* 143:1044–1054. <https://doi.org/10.1104/pp.106.089615>
- Jones JT, Haegeman A, Danchin EGJ et al (2013) Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol. Plant Pathol* 14(9):946–961. <https://doi.org/10.1111/mpp.12057>
- Kilen TC, Young LD (2000) Registration of D95–5246 soybean germplasm line resistant to *Phytophthora* Rot and soybean cyst nematode races 3 and 14. *Crop Sci* 40:304–304. <https://doi.org/10.2135/cropsci2000.0010orgp>
- Klepadlo M, Meinhardt CG, Vuong TD et al (2018) Evaluation of soybean germplasm for resistance to multiple nematode species: *Heterodera glycines*, *Meloidogyne incognita*, and *Rotylenchulus reniformis*. *Crop Sci* 58:2511–2522. <https://doi.org/10.2135/cropsci2018.05.0327>
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* 9:29. <https://doi.org/10.1186/1746-4811-9-29>
- Lee J-D, Kim H-J, Robbins RT et al (2015) Reaction of soybean cyst nematode resistant plant introductions to root-knot and reniform nematodes. *Plant Breed Biotechnol* 3:346–354. <https://doi.org/10.9787/pbb.2015.3.4.346>
- Lima FSO, Correa VR, Nogueira SR, Santos PRR (2017) Nematodes Affecting Soybean and Sustainable Practices for Their Management. In: *Soybean - The Basis of Yield, Biomass and Productivity*. InTech, pp 95–110
- Lipka AE, Tian F, Wang Q et al (2012) GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Liu S, Kandoth PK, Warren SD et al (2012) A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492:256–260. <https://doi.org/10.1038/nature11651>
- Luzzi BM, Boerma HR, Hussey RS (1987) Resistance to three species of root-knot nematode in soybean I. *Crop Sci* 27:258–262. <https://doi.org/10.2135/cropsci1987.0011183X002700020027x>
- Luzzi BM, Boerma HR, Hussey RS, Wood ED (1997) Registration of javanese root-knot nematode resistant soybean germplasm line G93–9223. *Crop Sci* 37:1035–1036
- Maldonado dos Santos JV, Ferreira EGC, Passianotto A et al (2019) Association mapping of a locus that confers southern stem canker resistance in soybean and SNP marker development. *BMC Genomics* 20(1):798. <https://doi.org/10.1186/s12864-019-6139-6>
- Mattos VS, Furlanetto C, Silva JGP et al (2016) *Meloidogyne* spp. populations from native Cerrado and soybean cultivated areas: genetic variability and aggressiveness. *Nematology* 18:505–515. <https://doi.org/10.1163/15685411-00002973>
- Mendy B, Wang'ombe MW, Radakovic ZS et al (2017) Arabidopsis leucine-rich repeat receptor-like kinase NILR1 is required for induction of innate immunity to parasitic nematodes. *PLoS Pathog* 13:1–22. <https://doi.org/10.1371/journal.ppat.1006284>
- Mienie CMS, Fourie H, Smit MA et al (2002) Identification of AFLP markers in soybean linked to resistance to *Meloidogyne javanica* and conversion to Sequence Characterized Amplified Regions (SCARs). *Plant Growth Regul* 37:157–166. <https://doi.org/10.1023/A:1020585023976>
- Mitchum MG (2016) Soybean resistance to the soybean cyst nematode *Heterodera glycines*: an update. *Phytopathology* 106:1444–1450. <https://doi.org/10.1094/PHYTO-06-16-0227-RVW>
- Oliveira L, Vinholes P, Montecelli T et al (2015) Inheritance of resistance of soybean for *Meloidogyne incognita* and identification of molecular marker for marker assisted selection. *J Sci Res Rep* 8:1–8. <https://doi.org/10.9734/jsrr/2015/18648>
- Oyekanmi EO, Fawole B (2010) Nematodes of soybean and their management. *Soybean Bot Prod Uses*. <https://doi.org/10.1079/9781845936440.0325>
- Passianotto A, Sonah H, Dias WP et al (2017) Genome-wide association study for resistance to the southern root-knot nematode (*Meloidogyne incognita*) in soybean. *Mol Breed* 37:148. <https://doi.org/10.1007/s11032-017-0744-3>
- Pham AT, McNally K, Abdel-Haleem H et al (2013) Fine mapping and identification of candidate genes controlling the resistance to southern root-knot nematode in PI 96354. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-013-2095-8>

- 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*. <https://doi.org/10.1086/519795>
- Rincker K, Hartman GL, Diers BW (2016) Fine Mapping of Resistance Genes from Five Brown Stem Rot Resistance Sources in Soybean. *Plant Genome* 9(1). <https://doi.org/10.3835/plantgenom2015.08.0063>
- Santos JRP, Ndeve AD, Huynh BL et al (2018) QTL mapping and transcriptome analysis of cowpea reveals candidate genes for root-knot nematode resistance. *PLoS ONE* 13:1–22. <https://doi.org/10.1371/journal.pone.0189185>
- Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet*. <https://doi.org/10.1086/502802>
- Shin J-H, Blay S, Graham J, McNeney B (2006) LDheatmap : an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw*. <https://doi.org/10.18637/jss.v016.c03>
- Silva JFV, Ferraz LCCB, Arias CA (2001) Herança da resistência a *Meloidogyne javanica* em soja. *Nematropica* 31:209–217
- Silva JFV, Ferraz LCBC, Arias CAA, Abdelnoor RV (2001) Identificação de Marcadores Moleculares de Microssatélites Associados a Resistência de Genótipos de Soja a *Meloidogyne javanica*. *Nematol Bras* 25:79–83
- Sonah H, Bastien M, Iqura E et al (2013) An improved Genotyping by Sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* 8:1–9. <https://doi.org/10.1371/journal.pone.0054603>
- Song Q, Hyten DL, Jia G et al (2015) Fingerprinting soybean germplasm and its utility in genomic research. *G3 Genes, Genomes, Genet* 5:1999–2006. <https://doi.org/10.1534/g3.115.019000>
- Tamulonis JP, Luzzi BM, Hussey RS et al (1997a) DNA markers associated with resistance to Javanese root-knot nematode in soybean. *Crop Sci*. <https://doi.org/10.2135/cropsci1997.0011183X003700030015x>
- Tamulonis JP, Luzzi BM, Hussey RS et al (1997b) DNA marker analysis of loci conferring resistance to peanut root-knot nematode in soybean. *Theor Appl Genet* 95:664–670. <https://doi.org/10.1007/s001220050610>
- Torkamaneh D, Laroche J, Bastien M et al (2017) Fast-GBS: a new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinform* 18:1–7. <https://doi.org/10.1186/s12859-016-1431-9>
- Torkamaneh D, Laroche J, Valliyodan B et al (2020) Soybean (*Glycine max*) haplotype map (GmHapMap): a universal resource for soybean translational and functional genomics. *Plant Biotechnol J*. <https://doi.org/10.1111/pbi.13466>
- Tran DT, Steketee CJ, Boehm JD et al (2019) Genome-Wide Association Analysis Pinpoints Additional Major Genomic Regions Conferring Resistance to Soybean Cyst Nematode (*Heterodera glycines* Ichinohe). *Front Plant Sci* 10:1–13. <https://doi.org/10.3389/fpls.2019.00401>
- USDA (2019) World agricultural production. Foreign Agric Serv. pp 1–31. <https://doi.org/10.32317/2221-1055.201907059>
- Vuong TD, Sonah H, Meinhardt CG et al (2015) Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genom* 16:1–13. <https://doi.org/10.1186/s12864-015-1811-y>
- Walters SA, Wehner TC, Barker KR (1997) A single recessive gene for resistance to the root-knot nematode (*Meloidogyne javanica*) in *Cucumis sativus* var. hardwickii. *J Hered* 88:66–69
- Warmerdam S, Sterken MG, van Schaik C et al (2018) Genome-wide association mapping of the architecture of susceptibility to the root-knot nematode *Meloidogyne incognita* in *Arabidopsis thaliana*. *New Phytol* 218(2):724–737. <https://doi.org/10.1111/nph.15034>
- Warmerdam S, Sterken MG, Sukarta OCA et al (2020) The TIR-NB-LRR pair DSC1 and WRKY19 contributes to basal immunity of *Arabidopsis* to the root-knot nematode *Meloidogyne incognita*. *BMC Plant Biol* 20:73. <https://doi.org/10.1186/s12870-020-2285-x>
- Wrather JA, Anderson TR, Arsyad DM et al (1997) Soybean disease loss estimates for the top 10 soybean producing countries in 1994. *Plant Dis* 81:107–110. <https://doi.org/10.1094/PDIS.1997.81.1.107>
- Wrather A, Shannon G, Balardin R et al (2010) Effect of Diseases on Soybean Yield in the Top Eight Producing Countries in 2006. *Plant Heal Prog* 11:29. <https://doi.org/10.1094/php-2010-0102-01-rs>
- Zhang J, Song Q, Cregan PB, Jiang GL (2016) Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor Appl Genet* 129:117–130. <https://doi.org/10.1007/s00122-015-2614-x>
- Zhang S, Zhang Z, Bales C et al (2017) Mapping novel aphid resistance QTL from wild soybean, *Glycine soja* 85–32. *Theor Appl Genet* 130:1941–1952. <https://doi.org/10.1007/s00122-017-2935-z>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.