JEAN CARLOS ALEKCEVETCH

MAPEAMENTO ASSOCIATIVO AMPLO EM SOJA PARA RESISTÊNCIA A *Meloidogyne javanica*

Londrina
2018

JEAN CARLOS ALEKCEVETCH

MAPEAMENTO ASSOCIATIVO AMPLO EM SOJA PARA RESISTÊNCIA A *Meloidogyne javanica*

Tese apresentada à Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Doutor pelo Programa de Pós-graduação em Genética e Biologia Molecular

Orientador: Prof. Dr. Ricardo Vilela Abdelnoor

Coorientadora: Prof.ª. Dra. Francismar Correa Marcelino Guimarães

Londrina
2018

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Alekcevetch, Jean Carlos.
MAPEAMENTO ASSOCIATIVO AMPLO EM SOJA PARA RESISTÊNCIA A Meloidogyne javanica / Jean Carlos Alekcevetch. - Londrina, 2018.
91 f. : il.

Orientador: Ricardo Vilela Abdelnoor.
Coorientador: Francismar Correa Marcelino-Guimarães.
Tese (Doutorado em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Genética e Biologia Molecular, 2018. Inclui bibliografia.

1. Mapeamento Associativo - Tese. 2. GWAS - Tese. 3. Meloidogyne javanica - Tese. I. Vilela Abdelnoor, Ricardo . II. Correa Marcelino-Guimarães, Francismar. III. Universidade Estadual de Londrina. Centro de Ciências Biológicas. Programa de Pós-Graduação em Genética e Biologia Molecular.

JEAN CARLOS ALEKCEVETCH

MAPEAMENTO ASSOCIATIVO AMPLO EM SOJA PARA RESISTÊNCIA A *Meloidogyne javanica*

Tese apresentada à Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Doutor pelo Programa de Pós-graduação em Genética e Biologia Molecular

BANCA EXAMINADORA

_____

Orientador: Prof. Dr. Ricardo Vilela Abdelnoor
Empresa Brasileira de Pesquisa Agropecuária – Embrapa - UEL

_____

Prof. Dr. François Belzile
Universitè Laval - Quebec

_____

Prof. Dr. Luiz Filipe Protássio Pereira
Empresa Brasileira de Pesquisa Agropecuária - Embrapa

_____

Dr. Gaspar Malone
GDM Genética do Brasil – Cambé

_____

Dr. Carlos Arrabal Arias
Empresa Brasileira de Pesquisa Agropecuária – Embrapa

Londrina, 19 de abril de 2018

DEDICO

À todos cientistas que tanto se dedicam para um mundo melhor, dedico.

AGRADECIMENTOS

À Deus, pelo milagre da vida e todas as oportunidades a mim oferecidas.

À Universidade Estadual de Londrina e ao programa de pós-graduação em Genética e Biologia Molecular.

Á CAPES pelo apoio financeiro em forma de bolsa de estudo.

À EMBRAPA Soja, pela disponibilidade da estrutura física, e provimento de reagentes.

Ao orientador, Prof. Dr. Ricardo Vilela Abdelnoor, pela oportunidade, confiança, atenção despendida, revisões e correções.

À co-orientadora, Prof.ª Dra. Francismar Correa Marcelino-Guimarãoes, pelo convívio, troca de ideias, revisões, correções e incentivo.

Aos professores do programa de pós-graduação em Genética e Biologia Molecular, pelos conhecimentos repassados.

À secretaria do programa de pós-graduação em Genética e Biologia Molecular, pelo constante esforço em me orientar para que todos os prazos, regras e obrigações fossem compridos sem problemas.

Aos pesquisadores da Embrapa Soja, Dr. Carlos Alberto Arabal, Dr. Waldir Pereira Dias, Dr. Marcelo Fernandes de Oliveira e Dra. Ivani de Oliveira Negrão Lopes pelo apoio técnico.

Aos técnicos de laboratório da Embrapa Soja, Claudemir, Adriana, Danielle, Ana Maria, Silvana, Márcia, César, Robson e Vera Pierote.

Ao meu colega, grande amigo e compadre André Passianotto, pelos ensinamentos em bioinformática, discussões e repasse de conhecimentos nos momentos de trabalho e companheirismo nos momentos de descontrações.

À minha família, em especial pai (João Alekcevetch), mãe (Terezinha Maria Ficagna Alekcevetch), irmã (Ligia Alekcevetch) e cunhado (Renato Pereira) pelo apoio e incentivo incondicional. E a minha sobrinha Daniella, que mesmo um bebê deu muita força para o titio.

Aos meus sogros, Miguel Reis e Beatriz Ribeiro pelo incentivo, paciência e apoio.

Aos amigos e colegas de laboratório de biotecnologia vegetal e de genética molecular e seleção assistida.

À minha esposa e companheira de vida, Rafaela Ribeiro Reis, pelo amor incondicional, companheirismo, mesmo nas horas mais difíceis e estressantes, paciência, incentivo e força para o meu crescimento intelectual, profissional e pessoal. Pelas dicas, sugestões, correções, troca de informações e madrugadas de estudos. Meu muito obrigado pelo carinho e compreensão.

A todos que, de alguma forma, contribuíram para o desenvolvimento do meu trabalho. Gratidão.

ALEKCEVETCH, Jean Carlos. **Estudo genético de associação em soja para resistência a *Meloidogyne javanica*.** 2018. 81p Tese (Doutorado em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Londrina, 2018.

RESUMO

A soja (*Glycine max* (L.) Merrill) é uma leguminosa anual. Atualmente, considerada uma das principais *commodities* mundiais, teve na última safra (2016/2017) uma produção mundial de 351,32 milhões de toneladas, as quais foram produzidas em 120,30 milhões de hectares. A sojicultura, no entanto, pode ser acometida por diversos patógenos, entre eles os nematoides. Entre as principais espécies capazes de parasitar a soja, os nematoides de galhas (*Meloidogyne javanica* e *M. incognita*) ganham destaque, juntamente com os nematoides de cisto (*Heterodera glycines*) e das lesões radiculares (*Pratylenchys brachyurus*). Devido a ausência de um efetivo controle destes parasitas, a utilização de espécies vegetais com baixo índice de reprodução do nemaoide e/ou cultivares resistentes vem de demonstrando eficientes formas de manejo destes nematoides em áreas contaminadas. Desta forma, os programas de melhoramento genético, visando o desenvolvimento de novas cultivares com boa resistência aos nematoides, são de extrema importância. Estudos genômicos são importantes ferramentas de suporte ao melhoramento genético, provendo uma melhor compreensão dos mecanismos de resistência, identificação de genes responsáveis pela resistência, bem como o desenvolvimento de marcadores moleculares. O estudo de mapeamento genômico associativo (*GWAS – Genome Wide Association Mapping*) é uma metodologia que vem se demonstrando altamente eficaz em relacionar marcadores moleculares do tipo *Single Nucleotide Polymorphism* (SNP) à diferentes características. Assim, este trabalho teve como objetivo a identificação de marcadores moleculares do tipo SNP, associados à resistência ao *M. javanica*, utilizando um painel de 369 materiais, composto por cultivares e *Plant introductions* (PIs). Este estudo permitiu a identificação de sete SNPs com alto grau de associação em resposta ao nematoide de galhas. Com base nestes polimorfismos, foi possível desenvolver e validar ensaios baseados em TaqMan eficazes na identificação e discriminação de genótipos contrastantes.

Palavras-chave: Genotipagem por sequenciamento, Associação genômica ampla, *Meloidogyne javanica*, *Glycine max*, SNPs, resistência a nematoide

ALEKCEVETCH, Jean Carlos. **Genome wide Association study for *Meloidogyne javanica* resistance in soybean.** 2018. 56p Thesis (Ph.D in Genetics and Molecular Biology) – State University of Londrina, Londrina, 2018.

ABSTRACT

The soybean (*Glycine max* (L.) Merrill) is an annual leguminous. Currently, it is considered one of the main world-wide commodities, with a production of 351.32 million of ton in an area of 120.30 million of hectares in the last crop season (2016/2017). The soybean crop can be affected by different pathogen, including nematodes. The major nematode species able to infect soybean are the root-knot nematode (*Meloidogyne javanica* e *M. incognita*), cyst nematodes (*Heterodera glycines*) and lesion nematodes (*Pratylenchys brachyurus*). Currently there are no effective nematode control method, thus the crop management with the use of non-host or resistant cultivars are the best solutions to be used on infected area. Thus, soybean genetic breeding programs aiming development of new cultivars with good resistance level are extremely important. Genetics studies represent an imperative support tool for breeding programs, through better comprehension of resistance mechanisms, identification of gene responsible to resistance and molecular markers development. Genome Wide Association Studies (GWAS) is a methodology that has been demonstrated to be highly effective in correlate *Single Nucleotide Polymorphism* (SNP) molecular marker to different traits. Thus, the purpose of this study was, to identify SNPs related to *M. javanica* resistance in a panel containing 369 accessions (cultivars and PIs (*Plant Introduction*)), through GWAS. This study allowed the identification of seven SNPs with high degree of association to *M. javanica*. Based on that, three efficient TaqMan assays were designed and validated for identification and discrimination of contrasting genotypes.

KeyWords: GBS, Genome Wide Association Mapping, *Meloidogyne javanica*, *Glycine max*, SNPs, nematode resistance

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

M. javanica	*Meloidogyne javanica*
M. incognita	*Meloidogyne incognita*
H. glycines	*Heterodera glycines*
P. brachyurus	*Pratylenchys brachyurus*
GWAS	*Genome-Wide Association Studies*
SNP	*Single Nucleotide Polymorphism*
PI	*Plant introduction*
USDA	*United States Department of Agriculture*
J1	Juvenil 1
J2	Juvenil 2
J3	Juvenil 3
J4	Juvenil 4
PAMPs	P*athogen-associated molecular patterns*
MAMPs	*Microbe-associated molecular patterns*
Avr	*Avirulence*
Genes R	Genes de resistência
HR	*Hypersensitivity response*
Ca^{2+}	Íons de cálcio
NBS-LRR	*Nucleotide-Binding Site Leucine-Rich-Repeat*
TNL	TIR-NB-LRR
IL-1R	*Interleukin-1 Receptor*
CNL	Classe não-TIR
CC	*Coiled-coil*
L	Leucina
DNA	Ácido desoxirribonucleico
SAM	Seleção assistida por marcadores moleculares
RFLP	*Restriction Fragment Length Polymorphism*
LG	*Linkage Group*
Chr	*Chromosome*
AFLP	*Amplified Fragment Length Polymorphism*
SCARs	*Sequence Characterized Amplified Regions*
PCR	*Polymerase Chain Reaction*
SSR	*Single Sequence Repeat*
R	Resistente
S	Suscetível
χ^2	chi-quadrado
QTL	*Quantitative Trait Locus*
SNPs	*Single Nucleotide Polymorphism*
GBS	*Genotyping-by-Sequencing*
Mb	Megabase
RKN	*Root-knot nematode*

GAPIT	*Genomic association and prediction integrated tool*
cMLM	*Compressed Mixed Linear Model*
PC	*Principal Components*
G. max	*Glycine max*
MAF	*Minor Allele Frequency*
K	*VanRaden Kinship matrix*
FDR	*False discovery rate*
P	*Population structure*
QQplot	*Quantile-Quantile Plot*
AE	*Allelic effect*
AV	*Allelic Variation*
CVs	*Cultivars*
ASE	*Asymptotic Square Error*
FPKM	*Fragments Per Kilobase Million*
ABA	Ácido Abicisico
JA	Ácido Jasmônico
ET	Etileno
SA	Ácido Salicílico
GB	Giberelina
CK	Citocininas
SAR	*Systemic Acquired Resistance*

SUMÁRIO

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

INTRODUÇÃO E FUNDAMENTAÇÃO TEÓRICA

A soja

A soja (*Glycine max* (L.) Merrill) é uma leguminosa anual, pertencente à divisão Magnoliophyta, classe Magnoliopsida, subclasse Rosidae, ordem Fabales, família Fabaceae, subfamília Faboideae, gênero *Glycine* L., subgênero Glycine subg. soja (Moench). Estima-se que sua ancestral (*Glycine soja*) tenha sido domesticada na Ásia entre 7000 e 9000 anos (LEE et al., 2011). A introdução da cultura no Brasil, aconteceu na Bahia, por volta de 1882, onde Gustavo Dutra, um professor da Escola de Agronomia da Bahia, começou a conduzir os primeiros estudos com a planta visando utilização como forrageira. Embora o primeiro relato de produção comercial seja datado de 1914, somente no final da década de 60, a cultura começou a ganhar destaque na agricultura brasileira, por ser uma boa cultura de verão, principalmente para o Rio Grande do Sul (Sindmilho & Soja, 2017).

Atualmente a soja é uma das principais *commodities*, com produção mundial (safra 2016/2017) de 351,32 milhões de toneladas produzidas em uma área de 120,30 milhões de hectares. Os Estados Unidos destacam-se como maior produtor do grão (119,92 milhões de toneladas), enquanto que o Brasil assume a posição de maior exportador (65 milhões de toneladas) e segundo produtor (114,10 milhões de toneladas) (USDA 2017a; USDA 2017b).

Nematoides de galhas na cultura da soja

Mais de 100 espécies de nematoides, pertencentes a 50 gêneros, já foram relacionados a cultura da soja. Dentro destas, cinco espécies são as principais responsáveis por maiores perdas: Os nematoides de cisto (*Heterodera glycines* Ichinohe), nematoides de galhas (*Meloidogyne incognita* e *M. javanica*), nematoides renifores (*Rotylenchulus reniformis*) e os nematoides das lesões (*Pratylenchus brachyururs*) (Veloso da Silva et al. 2006; Lima et al. 2017).

Estima-se que, entre todas as culturas economicamente importantes, as perdas agrícolas mundiais causadas por nematoides sejam entre 10% e 15%, o que representa em torno de $78 bilhões de dólares anuais. A soja, assim como várias outras culturas, também está sujeita ao ataque destes vermes (Yorinori 2000). A agricultura brasileira sofre anualmente, perdas estimadas em R$ 35 bilhões, sendo R$ 16 bilhões somente na sojicultura nacional (SBN, 2015).

Dentre os nematoides que atacam a cultura da soja, os nematoides de galhas se destacam pelas perdas potenciais e pela ampla gama de hospedeiros, dificultando assim o seu controle. Os nematoides de galhas são endoparasitas sedentários capazes de colonizar mais de 3000 espécies de plantas. A formação das galhas nas raízes é a principal característica destes nematoides. Esse processo é devido a alterações morfológicas induzidas pelos nematoides. Os principais sintomas observados nas lavouras comprometidas pelos nematoides de galhas é a presença de plantas subdesenvolvidas e amareladas em disposição de reboleira. Esses sintomas são devidos a deficiência na absorção de água e nutrientes e são fortemente influenciados pela densidade populacional dos nematoides bem como outros fatores ambientais como estresse hídrico e outros (Gheysen e Fenoll 2002; Abad et al 2003; Dias et al 2010).

Os nematoides de galhas, durante o ciclo de vida, passam pela fase de embriogênese que é seguida por outras quatro etapas de desenvolvimento (mudas). A primeira muda (embriogênese para J1) ocorre ainda dentro dos ovos. Após a primeira muda, já fora dos ovos, os vermes passam para o estádio J2, o qual os indivíduos apresentam motilidade e capacidade de infectar as raízes das plantas. As duas outras mudas ocorrem no interior das raízes do hospedeiro (Abad et al, 2003).

Com a utilização do aparelho bucal em formato de estilete retrátil, os nematoides invadem as raízes na zona de alongamento e deslocam-se de maneira intercelular até a ponta do cilindro vascular. Ao atingir este tecido, o verme irá percorre-lo de forma ascendente até selecionar algumas células para iniciar a formação do sítio de alimentação (células gigantes), estrutura a qual proverá água e nutrientes para que o mesmo possa permanecer até sofrer mais duas mudas (J3 e J4) completando o ciclo de vida e reprodução (Abad et al, 2009). Com algumas exceções, os nematoides de galhas reproduzem-se por partenogênese, assim as fêmeas assumem formato de pêra e depositam os ovos (em média 400-500 ovos) em uma massa gelatinosa rica em glicoproteínas, enquanto que os machos deixam as raízes e voltam para o solo (Gheysen et al, 2002; Castagnone-Sereno, 2006; Caillaud et al 2008, Abad et al, 2003).

As células gigantes têm início a partir de algumas células binucleadas (primeira característica da estrutura) e desenvolvem-se por vários ciclos de endoreduplicação (Williamson e Hussey, 1996). Em outras palavras, a replicação celular ocorre normalmente, porém a divisão celular (citocinese) é inibida. Desta forma, seu tamanho final, pode chegar a

aproximadamente 400 vezes o tamanho de uma célula do cilindro vascular não infectada e conter mais de 100 núcleos (Wiggers et al, 1990).

Fisiologicamente hiperativas, as células gigantes apresentam citoplasma denso, complexo de Golgi bastante desenvolvido, grande número de mitocôndrias, ribossomos e plastídios, por outro lado, o retículo endoplasmático é reduzido (Jones e Payne, 1978, Engler et al 2005).

Interação patógeno/hospedeiro

As plantas respondem aos mais variados patógenos através de um sistema de defesa intrínseco e dinâmico. O mecanismo de defesa tem sido classificado como uma resposta inata e sistêmica em vegetais (Kiraly et al., 2007; Jones and Dangl, 2006). Em plantas parasitadas por nematoides, por exemplo, logo após a penetração dos nematoides nas raízes, bem como durante o estabelecimento do sítio de alimentação, as plantas sofrem uma complexa modulação da expressão gênica envolvendo genes de defesa, da formação de parede, do ciclo celular, de fatores de transcrição, de hormônios, de transporte e do uso da água (Gheysen and Fenoll, 2002). Respostas de defesa incluem barreiras morfológicas e estruturais (parede celular, tricomas, espinhos, camada de epiderme entre outros), alterações bioquímicas (terpenóides, metabólitos, esteróides, compostos fenólicos, compostos nitrogenados, saponinas e glucosinolatos), além da modificação no complexo e interação proteico e enzimático (Dahal, et al., 2009).

O sistema de defesa em plantas contra ataque de patógenos é complexa e envolve muitas vias de transdução de sinal, as quais são mediadas por uma rede de fitohormônios. Esses, desempenham um papel importante na regulação do crescimento e desenvolvimento vegetal em condições de cultivo ideais. O ácido jasmônico (JA), etileno (ET) e ácido salicílico (SA) são os três principais fitohormônios de resposta à estresses bióticos (Bari e Jones, 2009). O ácido salicílico, um derivado do ácido benzóico, é um marcante agente na regulação da resposta de defesa das plantas, sendo as rotas que o envolve, potenciais ferramentas de estudo para entendimento e indução de resistência (War et al., 2011). Outros fitohormônios, como auxinas, ácido abscísico (ABA), brassinosteróides, giberelinas (GBs) e citocininas (CKs), foram detectados recentemente como reguladores de defesa (Robert-Seilaniantzet al., 2011). ABA, um composto sesquiterpeno, derivado da clivagem do γ-caroteno, regula diversos processos de desenvolvimento e respostas adaptativas ao estresse em plantas. O ABA pode

atuar regulando positivamente a defesa das plantas nos estágios iniciais da infecção, mediando o fechamento dos estômatos contra os invasores ou induzindo a deposição de calosidade, caso o patógeno desvie dos mecanismos decorrentes da primeira linha de defesa (Asselbergh et al., 2008). Se induzido em estágios posteriores, o ABA pode suprimir a indução de espécies reativas de oxigênio (EROs) e a transdução de sinal SA ou JA, dispensando assim, as respostas controladas por essas duas vias (Ton et al., 2009).

Estratégias de defesa envolvendo estes compostos, podem ser classificadas nas categorias, inata ou *Systemic Acquired Resistance* (SAR). Embora a imunidade inata tenha maior eficiência e seja a forma mais comum de resistência (em plantas) a patógenos, ambos os mecanismos de defesa dependem da capacidade da planta de distinguir entre moléculas próprias e exógenas.

A resposta das plantas ao ataque de patógenos pode ocorrer de forma basal, onde o sistema imune geral da planta é ativado a partir de estimulos genéricos (elicitores). Esse mecanismo é conhecido como *pathogen-associated molecular patterns* (PAMPs) ou *microbe-associated molecular patterns* (MAMPs) (Albersheim e Anderson-Prouty, 1975). As plantas ainda podem responder ao ataque de patógenos de uma forma mais específica, para tal, há o envolvimento de genes de resistência (R). Estes fazem o reconhecimento de proteínas específicas codificadas pelos genes Avr (avirulence) do patógeno. Este método de resistência também é conhecido como resistência gene-a-gene e é frequentemente associada à morte celular dos locais de infecção dos patógenos também conhecido como resposta de hipersensibilidade (HR) (Flor, H. 1947; Flor, H. 1971, Bent, 1996, Jones e Dangl, 2006).

Além da HR, os genes R podem atuar no fluxo de íons Ca^{2+}, fosforilação de proteínas, geração de compostos nocivos aos patógenos (espécies reativas de oxigênio, ácido salicílico quitinases, glucanases e metabólitos com atividade antimicrobiana (fitoalexinas)) (Lamb, 1994; Dixon et al., 1994), conferindo resistência em plantas à vírus, bactérias, oomicetos, pulgões e nematoides (Strange e Scott, 2005, McHale, et al. 2006). Os genes R podem ser classificados em função da organização dos motifs dos aminoácidos ou domínios de membranas. A maior parte dos genes de resistência conhecidos em plantas pertencem à família proteica *Nucleotide-Binding Site Leucine-Rich-Repeat* (NBS-LRR) (Ooijen, et al. 2007), porém, Gururani et al. (2012), evidenciou em seu trabalho de revisão, a existência de sete outras classes.

Família NBS-LRR

A família NBS-LRR atua de forma intercelular e pode ser dividida em duas classes com base em sua região N-terminal. A classe TIR-NB-LRR (TNL) caracteriza-se pela similaridade com o receptor IL-1R (Interleukin-1 Receptor), encontrado tanto em humanos como em *Drosophila melanogaster*. A outra classe é a não-TIR (CNL), cuja proteína contém estruturas "coiled-coil" (CC) na porção N-terminal do domínio, por esse motivo são chamadas de CC-NB-LRR (Ooijen, et al. 2007).

O domínio característico da família proteica LRR, geralmente apresenta segmentos de 20 a 29 aminoácidos e possuem um segmento conservado de 11 aminoácidos característico: LxxLxLxxNxL, onde "L" representa o aminoácido Leucina, podendo eventualmente ser substituído por Valina, Isoleucina ou Fenilalanina, "N" representa Asparagina, Treonina, Serina ou Cisteína e "x" pode representar qualquer aminoácido (Kobe and Kajava 2001, Matsushima et al. 2007).

Os genes NBS-LRR possuem grande dispersão no reino vegetal, são bastante numerosos e frequentemente estudados. Diversas plantas já tiveram o número de genes dessa família estimado: *Arabidopsis thaliana,* 149 genes (Meyer, et al. 2003; Guo et al. 2011), *A. lyrata*, 185 genes (Guo et al. 2011), *Brachypodium distachyon,* 126 genes (Tan e Wu, 2012), *Medicago truncatula,* 617 genes (Song e Nan 2014), *Populus trichocarpa,* 402 genes (Kohler et al. 2008), *Lotus japonicus*, 158 genes (Li, et al. 2010), *Oryza sativa L.* spp. *indica* e *O. sativa L.* spp. *japônica,* 653 e 553 genes respectivamente (Shang et al.2009), *Vitis vinífera* e *V. poplar*, 459 e 330 genes respectivamente (Yang, et al. 2008), *Carica papaya*, 54 genes (Porter, et al. 2009), *Glycine max*, 319 genes (Kang, et al. 2012), *Solanum tuberosum*, 435 genes (Lozano, et al. 2012), *Phaseolus vulgaris* L., 178 genes (Wu, et al. 2017), *Cucumis sativus*, 57 genes (Wan et al. 2013), *Cicer arietinum*, 104 genes (Sharma, et al. 2017), *Actinidia chinensis*, 96 genes (Li, et al. 2016), *Zea mays*, 129 genes (Li, et al. 2010).

Diversos genes, isolados em diferentes espécies, já foram idenficados e relacionados a resposta à diferentes nematoides, incluindo RKN como por exemplo: o GHNTR1, relatado conferindo resistência a *M. incognita* raça 1 em *Nicotiana benthamiana* e *N. tabacum* (Zhang et al. 2015), o gene *Mi*-1 conferindo resistência a *M. incognita* (Vos et al. 1998, Milligan et al. 1998) e o gene *Mi*-9, que confere resistência ao *M. incognita, M. javanica* e *M. arenaria* em plantas de tomates, o gene *PsoRPM2* oriundo de *Prunus sogdiana*, promovendo resistência a

M. incognita em plantas de tabaco (Zhu et al 2017), o gene *Hero* de tomate, que confere resistência ao nematoide de cisto de batatas *Globodera rostochiensis* (Ernst et al 2002).

Melhoramento genético e seleção assistida por marcadores moleculares

O melhoramento genético consiste basicamente na seleção de materiais que apresentam fenótipo desejado, porém, frequentemente são necessárias a obtenção de populações segregantes, as quais são originadas por cruzamentos entre indivíduos contrastantes. Essa metodologia tem possibilitado grande adaptabilidade, aumento de produtividade e obtenção de características de interesse agronômico. Porém, sua eficácia é reduzida quando a característica avaliada sofre forte influência do ambiente e/ou é multigênica (como caracteres quantitativos) dificultando a seleção de indivíduos geneticamente superiores (Bered et al. 1997, Schuster 2011).

A partir da década de 80, o processo de seleção foi facilitado com o desenvolvimento dos marcadores moleculares. Estes são marcadores genéticos a nível de DNA que não são influenciados pelo ambiente e que devem possuir herança simples, ou seja, são herdados de acordo com as leis de Mendel (Bered et al. 1997). Assim, podem ser empregados na seleção indireta de indivíduos de interesse agronômico e/ou geneticamente superiores em um procedimento conhecido como *seleção assistida por marcadores* (SAM) (Jannink et al. 2010, Schuster 2011).

O desenvolvimento da SAM vem acelerando consideravelmente o ganho genético, uma vez que as seleções, tanto dos genitores quanto dos descendentes, tornam-se mais eficientes, o número de indivíduos avaliados pode ser escalonado e com um preço relativamente reduzido (Bered et al. 1997, Schuster 2011). A SAM ainda apresenta as vantagens de poder ser aplicadas em qualquer estádio de desenvolvimento da planta, permitindo adiantar a seleção de características expressas somente em estádios avançados de desenvolvimento, reduzindo o tamanho das populações mantidas a campo ou casa de vegetação, o fato de não ser uma metodologia destrutiva auxilia programas de melhoramento na obtenção de descendentes destes materiais, mesmo após avaliação.

Resistencia genética a nematoides de galhas

A identificação de genes associados a resistência aos nematoides é de grande importância para os programas de melhoramento genético na obtenção de variedades

resistentes (Wendland, 2004). Ao longo dos anos diversas metodologias moleculares buscaram elucidar os mecanismos moleculares que controlavam a resistência. Tamulonis et al (1997) desenvolveu uma população F2 composta por 84 indivíduos (obtidos pelo cruzamento da PI230977 (R) e outra cultivar suscetível à *M. javanica*) e submeteu-a à avaliação com marcadores do tipo *Restriction Fragment Length Polymorphism* (RFLP) utilizando cinco enzimas de restrição (*Dra*I, *Eco*RI, *Eco*RV, *Hind*III e *Taq*I). Oito marcadores puderam ser associados à resistência ao nematoide (P<0,001). O marcador A725-2 localizado no grupo de ligação (LG) D1a (Cromossomo 1 = chr 1) apresentou a menor contribuição no número de galhas (13%). Marcadores localizados no Chr 13 foram os que apresentaram maior contribuição na redução do número de galhas. O B212-1 apresentou maior contribuição (46%), seguido pelo R045-1, com 42% de contribuição.

Mienie et al, (2002), com objetivo de desenvolver novos marcadores moleculares economicamente viáveis para a seleção de plantas resistentes a *M. javanica*, submeteu sessenta plantas F2 (obtidas a partir do cruzamento entre as cultivares Gazelle (R) e Prima (S)) à avaliação por marcadores do tipo *Amplified Fragment Length Polymorphism* (AFLP) e o marcador RFLP B212-1. Este último explicou em torno de 62% da variação no número de galhas. Além disso, foi possível converter sete AFLP em marcadores *Sequence Characterized Amplified Regions* (SCARs) cujos são baseados em PCR.

Silva et al. (2001b) desenvolveu dois cruzamentos entre indivíduos contrastantes, também para a resistência a *M. javanica* (BRS133 (S) x PI595099 (R) e BRS133 x CD201 (R)). Cada cruzamento gerou uma população F2 com 120 indivíduos os quais tiveram vinte locus SSR (*Single Sequence Repeat*) avaliados. Esses lócus foram selecionados em função da proximidade com os marcadores do chr 1 e chr13 previamente avaliados por Tamulonis et. al. (1997a). No cruzamento BRS133 x CD201 foi obtida significância na análise de variância do número de galhas para o Satt266 (Chr1) e Sat133 (Chr13). Nenhuma correspondência estatística foi observada no cruzamento BRS133 x PI595099, porém, ao considerar P=13 e R^2=1,98, o microssatélite SOYHSP176 também no Chr13 foi aceito.

Com o objetivo de identificar microssatélites associados a resistência a *M. javanica* e eficientes para utilização em SAM, Fuganti et al. (2004) gerou e submeteu à análise molecular 51 linhas de soja (25R e 26S) e as cultivares genitoras (BRS133 e PI595099) destas. Em um *screening* inicial, foram utilizados 97 SSR. Destes, 21 marcadores foram capazes de evidenciar polimorfismos nos pais (Satt266, Satt290 e Satt041 (chr 2), Satt584 (chr3), Satt155

e Satt236 (chr 5), Satt202 (chr 6), Sat 132 (chr 10), Sat128 e Satt509 (chr11), Satt434 e Satt192 (chr12), Satt423, Satt554, Satt114 e SOYHSP176 (chr13), Satt543 (chr17), Satt418 (chr19), Satt571, Satt419 e Satt367 (chr 20)). Destes, sete (Sat128, Sat 132, Satt571, Satt419 e Satt367 com baixa probabilidade, e SOYHSP176 (previamente avaliado por Silva et al. 2001) e Satt114 com alto valor de significância ($P<0,01$) no teste de chi-quadrado (x^2)) foram eficientes na descriminação dos genótipos R e S.

Mesmo utilizando SSR, que possui maior poder informativo que os demais marcadores até então utilizados, os mapeamentos de QTLs em populações biparentais apresentam resolução limitada em função da frequência de eventos de recombinação ser relativamente baixa nessas populações (Korte e Farlow, 2013). Além disso, os SSRs baseam-se apenas no tamanho do fragmento amplificado na PCR, não sendo possível a identificação de origens evolucionarias distintas em um mesmo locus. Em outras palavras, eventos de homoplasia não são identificados via SSR (Viard et al. 1998, Estoup et al. 1995).

Os *Single Nucleotide Polymorphism* (SNPs) estão presentes no genoma em um número virtualmente ilimitado e podem ser utilizados como marcadores moleculares. Compreendem em variações alélicas de uma única base que distribuem-se ao longo dos genomas (regiões codantes e não codantes), (Frazer et al. 2007). O acesso aos SNPs pode ser realizado por diferentes metodologias (Ganal et al. 2009). Resequenciamento de genomas completos vem se demonstrando uma efetiva e importante opção metodológica para a identificação de SNPs, para as mais variadas finalidades (Lam et al.2010, Zhou et al.2015, Maldonado dos Santos et al. 2016), uma vez que os valores de sequenciamento estão mais acessíveis. O GBS (*Genotyping-by-Sequencing*), metodologia proposta por Elshire et al. 2011, também é uma opção altamente eficiente na identificação de SNPs (Sonah et al, 2013, Sonah et al. 2014, Iquira et al. 2015, Li et al. 2015, Passianotto et al 2017). Essa metodologia baseia-se na redução da complexidade do genoma, utilizando enzimas de restrição. É uma técnica que em função da característica da endoenzima, regiões genômicas mais ativas são priorizadas.

Mapeamento associativo amplo - GWAS

O estudo genomico associativo amplo, do inglês *Genome-Wide Association Study* (GWAS), é uma poderosa metodolgoia que foi desenvolvida para detectar variações genéticas naturais que estão relacionadas a alguma característica fenotípica. O desenvolvimento do GWAS só foi possível com o advento de tecnologias capazes de identificar (sequenciamento, chips de sequenciamento, GBS) grandes quantidade de marcadores SNPs permitiram o desenvolvimento de estudos genômicos amplos que buscam a identificação de SNPs correlacionados com o fenótipos de interesse (Brachi et al. 2011). Essas análises não requerem a separação de fragmentos por tamanho, e ainda podem ter a eficiência maximizadas com a utilização de outros SNPs próximos, formando haplótipos de interesse (Rafalski 2002).

O GWAS foi desenvolvido inicialmente para estudos genéticos em humanos, que tinham como objetivo a idenficação de genes relacionados à doenças (Altshuler et al. 2008). Porém, observou-se que essa metodologia também poderia ser utilizada em plantas e com maior eficiência que em humanos, em função dos tamanhos dos grupos de ligação, que em plantas, de modo geral, são menores que em humanos (Brachi et al. 2011, Huang et al. 2013). O GWAS apresenta algumas vantagens frente ao mapeamento de QTLs, como por exemplo, pode ser realizado em populações não correlacionadas geneticamente e que não tenham sido previamente estudada, provê maior resolução e a quantidade de marcadores avaliados em cada indivíduo é maior.

Diferentes espécies vegetais como *Arabidopsis thaliana* (Nemri et al. 2010), Arroz (Huang et al 2010, Huang et al 2012), milheto (Jia et al.2013), Milho (Kump et al. 2011), Sorgo (Morris et al. 2012) já foram estudadas por GWAS com diferentes objetivos, inclusive em soja, objetivando a identificação de marcadores associados a resistência à nematoides de cisto (Vuongh et al.2015, Zhang et al2016) e de galhas - *M. incognita* (Passianotto et al. 2017).

REFERÊNCIAS

ABAD, P. et al. Root-knot nematode parasitism and host response: molecular basis of a sophisticated interaction. **Molecular Plant Pathology**. V.4, No.4, 217-224p. 2003.

NEMRI A, ATWELL S, TARONE AM, HUANG YS, ZHAO K, STUDHOLME DJ, NORDBORG M, JONES JDG: Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. Proceedings of the National Academy of Sciences. V.107, No.22, 2010.

ABAD. P. Invasion, feeding and development. In: Roland N. Perry.; Maurice Moens.; James L. Starr. **Root-knot Nematodes**. 2009.

ALBERSHEIM, P.; ANDERSON-PROUTY, A.J. Carbohydrates, proteins, cell surfaces and biochemistry of pathogenesis. **Annual Review Plant Physiology**. No.26, 31–52p. 1975 In BENT, A.F. E MACKEY, D. Elicitors, effectors, and R genes: The new paradigm and a lifetime supply of questions. **Annual Review Plant Physiology**, No.43, 399-436p. 2007.

ALTSHULER, D.; DALY, M.J.; LANDER, E.S. Genetic mapping in human disease. **Science**. V.7, No.322, 881-888p. 2008.

ASSELBERGH B, DE VLEESSCHAUWER D, HOFTE M. Global switches and fine-tuning-ABA modulates plant pathogen defense. **Molecular Plant Microbe Interaction**. No.21, 709–719p. 2008.

HUANG X,WEI X, SANG T, ZHAO Q, FENG Q, et al.. Genome-wide association studies of 14 agronomic traits in rice landraces. **Nature Genetics**. V.42, 961–967p, 2010

HUANG X, ZHAO Y, WEI X, LI C, WANG A, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. **Nature Genetics**. V.44: 32–39p. 2012

BARI R, JONES JD. Role of plant hormones in plant defence responses. **Plant Molecular Biology**. 69:473–488p. 2009.

JIA G, HUANG X, ZHI H, ZHAO Y, ZHAO Q, et al. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). **Nature Genetics**. V.45, 957–961p, 2013.

BENT, A.F. Plant disease resistance genes: Function meets structure. **The plant cell**. V.8, 1757-1771p. 1996.

MORRIS GP, RAMU P, DESHPANDE SP, HASH CT, SHAH T, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. **Proceeding of the National Academy of Sciences**. 453–458p. 2012.

KUMP KL, BRADBURY PJ, WISSER RJ, BUCKLER ES, BELCHER AR, et al. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. **Nature Genetics**. N.43, 163–168p. 2011.

BERED, F.; NETO, J..F.B.; de CARVALHO, F.I.F. Marcadores moleculares e sua aplicação no melhoramento genético de plantas. **Ciências Rurais**. V.27, N.3, 513-520p. 1997.

BRACHI, B; MORRIS, G.P.; BOREVITZ, J.O. Genome-wide association studies in plants: the missing heritability is in the field. **Genome Biology.** V.12, N.232. 2011.

CAILLAUD, M.C et al. Root-knot nematodes manipulate plant cell functions during a compatible interaction. **Journal of Plant Physiology**. V.165, 104-113p. 2008.

CASTAGNONE-SERENO P. Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. **Heredity**, V. 96, 282–289p, 2006.

DAHAL, D., HEINTZ., D., VAN DORSSELAER, A., BRAUN H.-P., WYDRA, K.. Pathogenesis and stress related, as well as metabolic proteins are regulated in tomato stems infected with Ralstonia solanacearum. **Plant Physiology and Biochemistry** No.47: 838–846p, 2009.

DIAS, W.P. et al. Nematoides em soja: Identificação e controle. **Circular Técnica**, 76. 2010.

DIXON, R.A.; HARRISON, M.J.; AND LAMB, C.J. Early events in the activation of plant defense responses. **Annual Review Phytopathology**. No.32, 479-501p, 1994.

ELSHIRE, R. J. et al. A robust, simple genotyping by sequencing (GBS) approach for high diversity species. **Plos ONE** 6, 2011.

ENGLER, J. de A. et al. Loss of susceptibility as an alternative for nematode resistance. **Current Opinion in Biotechnology**, No. 16, 112-117p. 2005.

ERNST, K. et al. The broad-spectrum potato cyst nematode resistance gene (*Hero*) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. **The Plant Journal**. V.31, No.2, 127-136p. 2002.

ESTOUP, A. et al. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). **Molecular Biology and Evolution**, No.12, 1074-1084p. 1995.

FLOR, H.H. Current status of the gene-for-gene concept. **Annual Review phytopathology**. No.9, 275-296p. 1971 In: MARONE, D. et al. Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: Active guardians in host defense responses. **International Journal of Molecular Sciences**. No.14, 7302-7306p. 2013.

FLOR, H.H. Host-parasite interactions in flax rust-Its genetics and other implications. **Phytopathology** No.45, 680-685p. 1947 In. BENT, A.F. Plant disease resistance genes: Function meets structure. **The plant cell**. V.8, 1757-1771p. 1996.

FRAZER, K.A. et al. International HapMap Consortium, A second generation human haplotype map of over 3.1 million SNPs. **Nature**, No. 449, 851-861p. 2007.

FUGANTI, R. et al. Identificação de marcadores moleculares de microssatélites para seleção de genótipos de soja resistentes a *Meloidogyne javanica*. **Nematologia Brasileira**. V.28, No.2, 125-130p. 2004.

GANAL, M.W.; ALTMANN, T.; RÖDER, M. SNP identification in crop plants. **Current Opinion in Plant Biology**. No.12, 211-217p. 2009.

GHEYSEN, G.; FENOLL, C. Gene expression in nematode feeding sites. **Annual Review Phytopathology**. V.40. 191-219p. 2002.

GUO, Y.L. et al. Genome-wide comparison of nucleotide-binding site-leucine-rich-repeat-encoding genes in *Arabidopsis*. **Plant Physiology**. No.157, 757–769p. 2011.

GURURANI, M.A. et al. Plant disease resistance genes: Current status and future directions. **Physiological and Molecular Plant Pathology**. No.78, 51-65p. 2012.

HUANG, X. HAN, B. Natural variations and genome-wide association studies in crop plants. **Annual Review Plant Biology**. V. 64, No.4, 4,1-4,21p. 2014.

IQUIRA, E.; SONAH, H.; BELZILE F. Association mapping of QTLs for *sclerotinia* stem rot resistance in collection of soybean plant introduction using a genotyping by sequencing (GBS) approach. **BMC Plant Biology**. 2015.

JANNINK J-L.; LORENZ, A.J.; IWATA.; H. Genomic selection in plant breeding: from theory to practice. **Briefing Functional Genomics**. V.9, No2, 166–177p. 2010.

JONES, J.D.G. and DANGL, J.L. The plant immune system. **Nature**, V.444, 323-329p. 2006.

JONES, M.G.K.; PAYNE, H.L. Early stages of nematode-induced giant cell formation in roots of *Impatiens balsamina*. **The Journal of Nematology**. No.10, 70-84p. 1978.

KANG, Y.J.et al. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. **BMC Plant Biology**. No.12, V.139, 2012.

KHAN, R. et al. Microarray analysis of gene expression in soybean roots susceptible to the soybean cyst nematode two days post invasion. **Journal of nematology**. No.36, V.3, 241-248p. 2004.

KIRALY L, BARNAZ B, KIRALYZ Z. Plant resistance to pathogen infection: forms and mechanisms of innate and acquired resistance. **Journal of Phytopathology**. No.155, 385–396p. 2007

KOBE, B. and KAJAVA, A.V. The leucine-rich repeat as a protein recognition motif. **Current Opinion in Structural Biology**. V.11, 725-735p, 2001p.

KOHLER, A. et al. Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. **Plant Molecular Biology**. No.66, 619–636p. 2008.

KORTE, A. and Farlow, A. The advantages and limitations of trait analysis with GWAS: a review. **Plant Methods**. V.9, No.20. 2013.

KUMARI, C. et al. Comparing the defense-related gene expression changes upon root-knot nematode attack in susceptible versus resistant cultivars of rice. **Scientific reports**. V.6, No.22846. 2016.

LAM, HON-M. et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. **Nature Genetics**V.42, No.12. 2010.

LAMB, C.J. Plant disease resistance genes in signal perception and transduction. **Cell**, V76, 419-422p. 1994.

LEE, G. et al. Archaeological soybean (*Glycine max*) in East Asia: does size matter? **Plos one**, v. 6, n. 11, p. 2011.

LI, H. et al. A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. **BMC Genomics**. V. 16, No.216. 2015.

LI, J. et al. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. **Molecular Genetics and Genomics**. V. 283, No.5, 427-438p. 2010.

LI, X. et al. Identification and characterization of NBS-encoding disease resistance genes in *Lotus japonicus*. **Plant Systematic and Evolution**. No.289, 101–110p. 2010.

LI, Y. et al. Genome wide analysis of NBS-encoding genes in kiwi fruit (*Actinidia chinensis*). **Journal of genetics**. V.95, No.4, 2016.

LIMA, F.S.O. et al. Nematodes affecting soybean and sustainable practices for their management. In. KASAI, M. Soybean – The basis of yield, biomass and productivity. Available in: https://www.intechopen.com/books/soybean-the-basis-of-yield-biomass-and-productivity. *InTech*, 2017.

LOPES, M.J.C. Aspectos histopatológicos e mudanças na expressão de genes em genótipos de soja resistente, durante a interação com *Meloidogyne javanica.* UFLA – Universidade Federal de Lavras. **Tese de doutorado.** 2009.

LOZANO, R. et al. Genome-wide identification and mapping of NBS-encoding resistance genes in *Solanum tuberosum* group phureja. **Plos one**, V.7, No.4, 2012.

MALDONADOS Dos SANTOS, J.V. et al. Evaluation of genetic variation among Brazilian soybean cultivars through genome resequencing. **BMC Genomics**. V.17, No.110. 2016.

MATSUSHIMA, N. et al. Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate tool-like receptors. **BMC Genomics**. V8, No.124, 2007.

MCHALE, L. et al. Plant NBS-LRR proteins: adaptable guards. **Genome Biology**. No.7, Vol.212. 2006.

MEYERS, B.C. et al. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. **Plant Cell**, No.15, 809–834p. 2003.

MIENIE, C.M.S. et al. Identification of AFLP markers in soybean linked to resistance to *Meloidogyne javanica* and conversion to sequence characterized amplified regions (SCARS). **Plant Growth Regulation**. No.35, 157-166p. 2002.

MILLIGAN, S.B, et al. The root knot nematode resistance gene *Mi* from tomato is a member of the Leucine Zipper, Nucleotide Binding, Leucine-Rich Repeat family of plant genes. **The Plant Cell**. V.10, 1307-1319p. 1998.

OOIJEN, G. Van. et al. Structure and function of resistance proteins in solanaceous plants. **Annual Review Phytopathology**. No. 45, 43-72p. 2007.

PASSIANOTTO, A.L.de L. et al. Genome-wide association study for resistance to the southern root-knot nematode (*Meloidogyne incognita*) in soybean. **Molecular Breeding**. V.37, No.148, 2017.

PORTER, B.W. et al. Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. **Molecular Genetics and Genomics**, No.281, 609–626p. 2009.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**. No.5, 94-100p. 2002.

ROBERT-SEILANIANTZ A, GRANT M, JONES JDG. Hormone crosstalk in plant disease and defense: more than just JASMONATE-SALICYLATE antagonism. **Annual Reviews Phytopathology**. No.49, 317–343p. 2011.

SBN – Sociedade Brasileira de Nematologia, 2015. In: RIVAS, L. **Agrolink**. Disponível em: < https://www.agrolink.com.br/noticias/por-ano--nematoides-causam-prejuizos-de-r--35-bilhoes-ao-agronegocio-nacional_343212.html>

SCHUSTER, I. Marker-assisted selection for quantitative traits. **Crop Breeding and Applied Biotechnology**. S1. 50-55p, 2011.

SHANG, J. et al. Identification of a new rice blast resistance gene, *Pid3*, by genome wide comparison of paired nucleotide-binding site leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. **Genetics**, No.182, 1303–1311p. 2009.

SHARMA, R.; RAWAT, V.; SURESH, C.G. Genome-wide identification and tissue-specific expression analysis of nucleotide binding site-leucine rich repeat gene family in *Cicer arietinum* (kabuli chickpea). **Genomics data**. V.14, 24-31p. 2017.

SILVA, J.F.V; FERRAZ, L.C.B.C; ARIAS, C.A.A; ABDELNOOR, R.V. Identificação de marcadores moleculares de microssatélites associados à resistência de genótipos de soja a *Meloidogyne javanica*. **Nematologia Brasileira**, V.25, 79-83, 2001b.

SINDMILHO & SOJA - Sindicato da indústria do milho, soja e seus derivados no estado de São Paulo – **Disponível em**: http://www.fiesp.com.br/sindimilho/sobre-o-sindmilho/curiosidades/soja-e-suas-riquezas-historia/. Acesso em Janeiro de 2018.

SONAH, H. et al. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and Genotyping. **Plos One** V.8, No.1. 2013

SONAH, H. et al. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. **Plant Biotechnology Journal.** 1-11p. 2014

SONG, H. and NAN, Z. Genome-wide analysis of nucleotide-binding site disease resistance genes in *Medicago truncatula*. **Chinese Science Bulletin**. Vol. 59, No.11, 2014.

STRANGE, R.N.; SCOTT, P.R. Plant disease: a threat to global food security. **Annual Review Phytopathology**. No.43, 83–116p. 2005.

TAMULONIS. J.P. et al. DNA marker associated with resistance to javaneses root-knot nematode in soybean. V.37. 783-788p. 1997.

TAN, S.; WU, S. Genome wide analysis of nucleotide-binding site disease resistance genes in *Brachypodium distachyon*. **Comparative and Functional Genomics**. V.2012. 2012.

TON J, FLORS V, MAUCH-MANI B. The multifaceted role of ABA in disease resistance. **Trends in Plant Science**. No.14, 310–317p. 2009.

UNITED STATES DEPARTMENT OF AGRICULTURE – USDA. Oilseeds:World market and trade. **December**. 2017b.

UNITED STATES DEPARTMENT OF AGRICULTURE – USDA. World Agricultural Production. **December**. 2017a.

VELOSO DA SILVA, J.F.; DIAS, W.P.; GARCIA, A.; CARNERO, G.E.S. Perdas por nematoides chegam a 10.6% da soja mundial. **Visão agrícola**. No.5, 2006.

VIARD, F. et al. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. **Journal Molecular Evolution**. No.47, 42-51, 1998.

VINHOLES, P. DA SILVA. Associação genômica para resistência da soja a *Meloidogyne javanica* e *Macrophomina phaseolina*, **Tese de doutorado**, UFV – Universidade Federal de Viçosa, 2014.

VOS, P. et al. The tomato *Mi-1* gene confers resistance to both root-knot nematodes and potato aphids. **Nature Biotechnology**. V.16, 1365-1369p. 1998.

VUONGH, T.D. et al. Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. **BMC Genomics**. V16, No.593, 2015.

WAN, H. et al. Genome-wide analysis of NBS-encoding disease resistance in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. **BMC Genomics**. No.14, V.109. 2013.

WAN, J. et al. Whole-genome gene expression profiling revealed genes and pathways potentially involved in regulating interaction of soybean with cyst nematode (*Heterodera glycines* Ichinohe). **BMC Genomics**. No. 16, V.148. 2015.

WAR AR, PAULRAJ MG, WAR MY, IGNACIMUTHU S. Herbivore and elicitor-induced resistance in groundnut to Asian armyworm, Spodoptera litura (Fab.) (Lepidoptera: Noctuidae). **Plant Signaling & Behavior**, No.6, V.11, 1769–1777p. 2011.

WENDLAND, A. et al. Genetic expression of soybeans: Meloidogyne javanica interaction revealed by analyses of microarrays. In: **World Soybean Research Conference,** VII, Foz do Iguaçu (PR). Anais, p. 320. 2004.

WIGGERS, R.G.; STARR, J.L.; PRICE, H.J. DNA content and variation in chromosome number in plant cells affected by *Meloidogyne incognita* and *M. arenaria*. **Phytopathology**. No. 80. 1391-1395p. 1990.

WILLIAMSON, V.M.; HUSSEY, R.S.; Nematode pathogenesis and resistance in plants. **Plant Cell**. No.8, 1735-1745p. 1996.

WU, J. et al. Genome-wide association study identifies NBS-LRR-Encoding genes related with Anthracnose and common Bacterial Blight in the common bean. **Frontiers in Plant Science**. V.8, 2017.

YANG, S. et al. Recent duplications dominate NBS-encoding gene expansion in two woody species. **Molecular Genetics and Genomics**. No.280, 187–198p. 2008.

YORINORI, J. T. Ameaças para a soja. **Cultivar**, v. 22, p. 26-30, 2000.

ZHANG, B. et al. A CC-NBS-LRR type gene *GHNTR1* confers resistance to southern root-knot nematode in *Nicotiana benthamiana* and *Nicotiana tabacum*. **European Plant Pathology**. V.142, No.4, 2015.

ZHANG, H. et al. Genome-wide association study of resistance to soybean cyst nematode (*Hererodera glycines*) HG type 2.5.7 in wild soybean (*Glycine soja*). **Frontiers in Plant Science**. V.7. 2016.

ZHOU, Z. et al. Resequencing 302Wild and cultivated accessions identifies genes related to domestication and improvement in soybean. **Nature Biotechnology**. V.33, No.4. 2015.

OBJETIVO GERAL

O presente trabalho tem como objetivo o mapeamento associativo amplo para a resistência ao nematoide de galhas, *Meloidogyne javanica* com base em uma população não estruturada, bem como desenvolver ensaios baseados em marcadores SNPs capazes de discriminar fenótipos contrastantes para tal caraterística.

OBJETIVO ESPECÍFICO

a) Identificar e validar SNPs em uma população de soja não relacionadas;
b) Associar SNPs à resistência a *M. javanica* através de mapeamento associativo*;*
c) Desenvolver e verificar a eficiência de sondas de hidrólise a serem utilizadas na discriminação de indivíduos com genótipos contrastantes para reação ao *M. javanica*.

CAPÍTULO 2 - Genome wide association study of resistance to *Meloidogyne javanica* in soybean

Jean Carlos Alekcevetch[1,2], André Luiz de Lima Passianotto[2§], Adriana Brombini dos Santos[2], Waldir Pereira Dias[2], François Belzile[3], Francismar Correa Marcelino-Guimarães[2], Ricardo Vilela Abdelnoor[2*]

[1]UEL - Londrina State University, Londrina, PR, Brazil;

[2]Brazilian Agricultural Research Corporation, National Soybean Research Center (Embrapa – Soja), P.O. Box 231, Londrina, PR, Brazil;

[3]Department of Plant Sciences and Institute of Integrative Biology and Systems (IBIS), Université Laval, Quebec City, Quebec, Canada G1V 0A6. [§]Currently address: University of Guelph, Guelph, Ontario, Canada N1G 2W1. *Corresponding author: ricardo.abdelnoor@embrapa.br

ABSTRACT

Soybean [*Glycine max* (L.) Merrill] is a major commodity worldwide and an important source of vegetable oil and protein meal. Despite its good adaptation in several regions in the world, it is constantly affected by abiotic and biotic factors, among which nematodes are responsible for extensive losses in many regions. The identification and study of nematode resistant accessions and genes able to confer this phenotype is very important and desirable in breeding programs. Therefore, the goal of this study was to detect *Single Nucleotide Polymorphism* (SNP) markers associated with resistance against the root-knot nematode (RKN), *M. javanica*, and to develop an efficient assay for use in marker-assisted selection.

In this study, the reaction to *M. javanica* was evaluated in a non-structured population of 369 accessions comprising cultivars and *Plant Introduced* (PIs). The genotypic evaluation was performed via a genotyping-by-sequencing (GBS) approach. It leads to identification and validation of 44,100 SNPs. Among them, seven markers were significantly associated to RKN

reaction. Six of these were located in a unique linkage disequilibrium block (LD), comprehending a intergenic region of 13kb, in chromosome 13, flanking a single gene coding to a NBS-LRR-TIR domain gene. Three TaqMan assays were developed

A group of PIs sharing the R haplotype could be identified. These may constitute a new source of *M. javanica* resistance. Those markers might be explored by marker assisted selection, representing a useful option to help breeders select resistant materials in a fast and economic way. Furthermore, the developed assays presented a satisfactory level of confidence and allelic discrimination power, representing an efficient, faster, less dispendious and laborious alternative to access genetically the nematode resistance.

KeyWords

Genotyping by Sequencing, Genome Wide Association Mapping, *Meloidogyne javanica*, *Glycine max*, SNPs, nematode resistance

Introduction

Soybean [*Glycine max* (L.) Merrill] is one of the most important oil and protein crops in the world. Currently, the United States is the largest producer with 116.92 million metric tons followed by Brazil with 114.1 million tons [1]. Despite the success of the soybean crop both in the United States and Brazil, the crop is frequently challenged by biotic and abiotic 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 U$78 billion worldwide [2].

The *Meloidogyne* genus is composed of more than 90 species, but basically *M. javanica* and *M. incognita* are the main members of the genus that are of economic importance affecting worldwide soybean production [2,3]. These species have been frequently detected in Argentina (15% and 40% of samples, composed by soil and roots, respectively) [4], in 91% of soils and roots sampled in South Africa [5], and in 24% of samples from the USA [6]. 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, São Paulo, Minas Gerais, Mato Grosso, Mato Grosso do Sul and Bahia [7,8].

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 [9,10]. The continuous development and introduction of new soybean varieties with increased resistance are therefore important. 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 main source of the resistance present in soybean varieties is derived from only one source, the North American cultivar Bragg in which resistance is reported to be quantitative [8,11].

Studies have been conducted with biparental populations to QTL mapping, using *Random Fragment Length Polymorphism* (RFLP) [12,13] *Amplified Fragment Length Polymorphism* (AFLPs) [13], *Sequence Characterized Amplified Regions* (SCAR) [13], and S*ingle Sequence Repeat* (SSR) [14,15]. Those works pointed the chromosome (Chr) 13 as the mains source of the resistance and, less frequently and with minor effect, the Chr 1 as carriers of resistance genes to RKN. Although, at moment, there are no studies to *M. javanica* reaction, SNPs are another type of molecular markers with potential to be used to that. SNPs can be accessed through different methodologies, as well as, *Genotyping by Sequencing* (GBS)[16], SoySNP50K chip [17], whole-genome resequencing [18] or another type of arrays [19,20].

Resistance to *M. javanica* is usually described as a complex trait and biparental populations have traditionally been used for QTL identification. In this type of study, it is possible to access only the allelic diversity segregating between the two parents of the population, so the genetic resolution is limited to few recombination events whose are typically captured in such populations [21]. On the other hand, the development of different methodologies highly efficient on SNPs discovery, allowed *genome-wide association study* (GWAS) identify not only individuals with common traits with their progenitor, but also expand the analysis by combining many genetic backgrounds, thus, linkage disequilibrium identified turn much less extended leading to a much-improved precision [22,23].

Initially, the GWAS was developed to genetic research aiming identify genes related to human diseases [24]. However, it also can be applied with success in plant genetic analysis, with more successful than in humans [25]. This new approach has been receiving unprecedented attention once it overcomes several limitations of QTL mapping. It containg high resolution, often to the gene level, cost efficiency, and it is non-requirement of pedigrees or crosses. Now, with development of sequencing technologies, GWAS is an useful and

655 powerful tool able to identify and dissect the genetic basis and variations that underlie many
656 important and complex traits, such as disease resistance, in different species, including
657 soybean [25, 26, 27]. To avoid spurious associations and perform correct association
658 mapping, a computational method, which account simultaneously population structure and
659 relativeness, must be used, as well as, stricter statistical tests (p-value cutoff) between marker
660 and trait and sample sizes sufficiently large [28, 26].

661 In soybean the GWAS have been used for a wide variety of purposes, including the
662 mapping of both simple and complex traits [29], seed weight [17], seed protein and oil
663 content [20], iron deficiency chlorosis [30], *M. incognita* resistance [31] and specific protein
664 domain mapping (NBS-LRR genes) [32].

665 Thus, the objective of this study was to identify genomic regions associated with *M.*
666 *javanica* resistance using a set a of 369 soybeans accessions and dense SNP marker coverage
667 via GWAS approach. We were able to identify a specific region in the chromosome 13
668 strongly associated with the resistance to *M. javanica* in our panel and to define a haplotype
669 shared by the 77,4% of the resistant accessions. In addition, at least seven accessions which
670 presented a resistant phenotype but not the haplotype might represent a new set of plant
671 introduction containing alternative sources of resistance. Finally, we developed a SNP assays
672 based on the most highly associated markers to discriminate between resistant and susceptible
673 genotypes for marker-assisted selection (MAS).

674 **Materials and Methods**

675 **Association panel**

676 A set of 369 accessions composed of cultivars and PIs (Supplementary Table 1), were
677 obtained from the Soybean Germplasm Bank, located at Embrapa Soja in Londrina, PR,
678 Brazil. Six seeds of each accession were pre-germinated in plastic pots of 0.25 L filled with

679 sterilized sand. Five days after germination, seedlings were transferred individually to plastic
680 tubes of 0.5 L filled with substrate (sterilized by autoclaving) composed by sand and soil
681 (3:1), in a greenhouse. The plants were kept under 16 hours of daylight and supplemented
682 with 600W high-pressure sodium lamps (Light Systems PL). All samples were used for
683 nematode resistance evaluation. Trifoliate leaves from two young plants from each genotype
684 were collected independently, frozen in liquid nitrogen and stored in a -80°C freezer. The leaf
685 samples were ground to a fine powder and stored until DNA isolation.

686 Additionally, to extend the number of material might have the QTL described in this
687 study, a new set of plants, composed by 22 fixed lines (cultivars and PIs), were phenotype to
688 *M. javanica* as describe previously and genotyped using the 3 SNPs selected in the assay
689 design (supplementary table 2).

690 **Phenotyping**

691 Nematode inoculation was performed during the summer. One day before transferring
692 seedlings to plastic tubes, some infected soybean roots from a nematode stock were ground in
693 water to obtain the suspension containing *M. javanica* eggs. The number of eggs was
694 estimated in a Peters' chamber under microscopy and the concentration of the suspension was
695 adjusted to 1250 eggs/ml. The inoculation was performed by deposition of 4mL of the
696 nematode suspension in the same hole used to introduce the seedling.

697 Thirty days after inoculation, all plants were individually removed from the tubes.
698 Excess sand and soil around roots was carefully removed and roots were washed in running
699 water. The severity of the infestation was rated on a scale of 1 to 5 adapted from [48], where
700 1 = <10% of the root system is infected with small galls; 2 = 10% - 25% of the root system
701 galled, most being small galls; 3 = 26% - 50% of the root system with large galls; 4 = 51% -
702 90% of the root system with large galls; 5 = 91% - 100% of root system with large galls and

703 necrotic roots. Accessions with a rating between 1 and 2.5 were considered resistant (R),
704 between 2.6 and 3.5 were deemed moderately resistant (MR); and those with a score above
705 3.6 were rated as susceptible (S). Ten accessions were used as phenotype check, thus, four
706 genotypes were used as resistant checks (PI595099, BRS Celeste, MG/BR 46 Conquista e
707 BRS282), three as moderately resistant controls (Santa Rosa, CD201 e BRS317) and two
708 other genotypes as susceptible checks (CD202, Embrapa 20 and BRSMTPintado).

709 **DNA extraction and GBS library preparation**

710 The DNA from each sample was extracted using the DNeasy® Plant Mini Kit
711 (Qiagen) according to the manufacturer's instructions. The DNA integrity was checked by
712 electrophoresis on an agarose gel (1%) followed by quantification on a NanoDrop® ND1000
713 spectrophotometer (Uniscience) and diluted with water to a concentration of 10 ng/µl. GBS
714 libraries were produced using the protocol described by [49] Elshire et al. (2011) and
715 modified by [16] Sonah et al. (2013). Thus, DNA was digested using *Ape*KI followed by
716 ligation of barcoded adapters and pooling of 96 samples per library. These 96-plex GBS
717 libraries were sequenced on either Illumina HiSeq2000 (McGill University-Genome Quebec
718 Innovation Centre, Montreal, QC, Canada) or Ion Torrent (Laval Université, Québec, QC,
719 Canada) DNA sequencers.

720 **Data analysis and SNP identification**

721 Using a custom-designed pipeline in perl language (IGST-GBS pipeline; J. Laroche,
722 Université Laval, data unpublished) short reads (≤100 bp) were first split into separate fastq
723 files based on the barcode and then trimmed to remove barcode and adapter sequences. The
724 resulting reads were mapped on the *G. max* reference genome (v.2) and variants were called
725 with SAM tools. Heterozygous genotypes were replaced with missing data and any accessions
726 with >80% missing data were removed from the dataset. Finally, imputation of missing
727 genotypic data was performed using fastPHASE 1.3 [50]. For the GWAS, only loci with a

minor allele frequency (MAF) ≥0.05 were used. SNP positions throughout the genome and the prediction of the functional impact of these variants was performed using SnpEff [51].

Association Mapping and Population Structure

The Genomic Association and Prediction Integrated Tool – GAPIT [52] 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 ten Principal Components (PC) from Principal Component Analysis (PCA) were used to capture and evaluate the 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 0.001. The minimum *p*-value to associated SNPs was 1.74E-08.

Linkage Disequilibrium Analysis

After filtering for a MAF ≥0.05, all SNPs located on chromosome Chr13 were loaded into PLINK and the correlation coefficient (r^2) was calculated to pairwise linkage disequilbrium (LD). Haplotype blocks were identified and visualized using Haploview [53]. using default Gabriel's rules [54], confidence interval minimal for strong LD, UPPER=0,98, LOWER=0,7 - D'>0,8 and fraction of strong LD ≥0,95. The most widely used method square allele frequency correlation (r^2) was used to assessment of LD decay.

***M. javanica* genotyping assays**

TaqMan assays were individually designed on three mapped SNPs (Supplementary table 3) using the Oligo Architect software. Assays were engineered by Sigma-Aldrich using two different fluorescent dyes. The 6-FAM dye was attached to the R-allele probe while the HEX dye was linked to the S-allele probe. Forward and reverse primers were also designed to amplify a small region around these probed sites.

Phenotyping was performed as previously described and the genotyping reactions were conducted in a final volume of 5 µL. The solution was composed of LuminoCt® SYBR® Green qPCR readyMix^{TM} (1x), forward and reverse primers (1uM each), R and S probes (0,20 uM each), DNA (6 to 15 ng in 1.5 µL), ultra-pure water (0,5 µL) and ROX dye (1x). Genotyping data were acquired using an ABI7900HT fast real-time PCR system (Applied Biosystems) under the following two-step PCR conditions: 50ºC for 2 min; followed by 95ºC for 10 min and another 50 cycles of 95ºC for 15 s and 60ºC for 1 min.

The assays assertiveness was checked through genotyping of 43 accessions whose were also submitted to GWAS. Additionally, to increase the number of accessions molecularly characterized to resistance to RKN, twenty-one accessions (fifteen resistant and six susceptible to *M. javanica*) with previous literature information about phenotyping were evaluated using developed assays (Supplementary table 4).

Results

Nematode evaluation

To characterize the phenotypic response all ten reference phenotypes and 369 lines composing the association panel were inoculated with *M. javanica* and rated on a 1 to 5 scale. Three of the resistant checks (BRS Celeste, MG/BR 46 Conquista and BRS 282) received an identical score of 2.2, while PI595099 received the lowest score (1.3). Moderately resistant checks (Santa Rosa, CD 201 and BRS 317) presented scores ranging from 2.6 to 2.8, and susceptible checks CD 202 and Embrapa 20 also presents identical rating (4.3) while the BRSMT Pintado presents the highest susceptibility rate (5).

Among the 369 soybean accessions comprising the association panel, 32 accessions received scores between 1.0 and 2.5 (indicating a high level of resistance), 93 accessions

received scores ranging from 2.6 to 3.5, while the remaining 244 accessions exhibited scores higher than 3.6. (Figure 1). Considering the population size and the number of the accessions in each phenotypic class, it was assumed that the genomic regions that are responsible for resistance and susceptibility were well sampled.

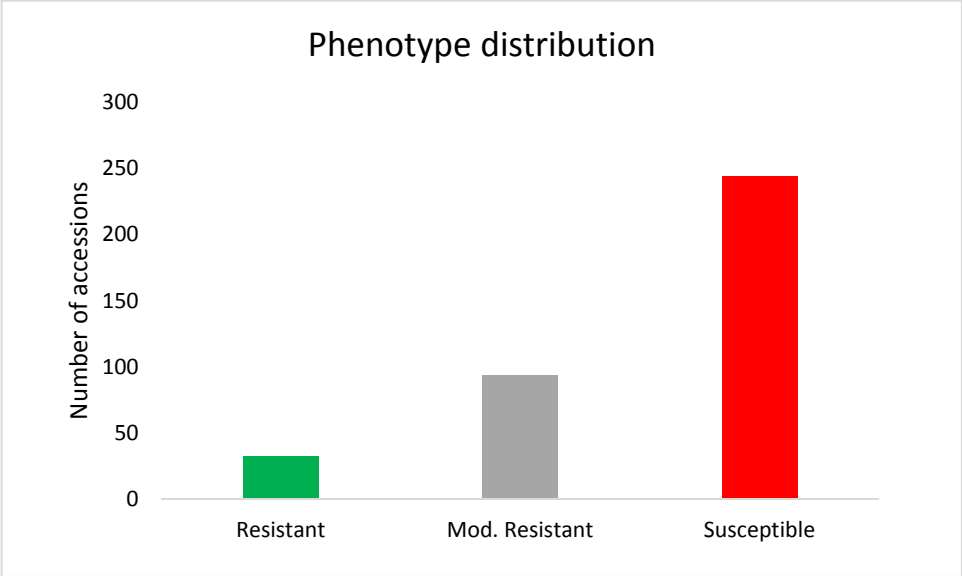
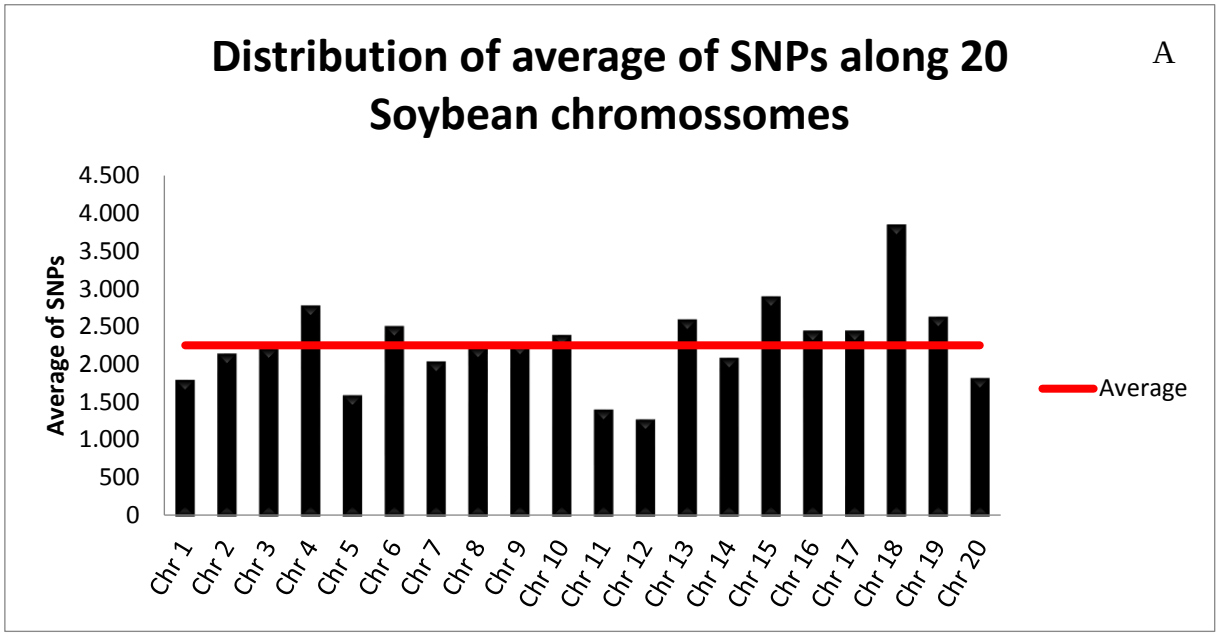


Figure 1 - Graphic distribution of resistant (R), moderately resistant (MR) and susceptible (S) soybean genotypes against the *M. javanica* pathogen.

Identification and distribution of SNP markers

Through GBS, a total of 44.992 SNPs was identified along all 20 chromosomes, providing an extensive coverage of the genome. This represented an average of 2,250 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, 58Mb). On the other hand, the smallest number of variants was not found on the smallest chromosome (Chr11 – 34Mb), but rather on Chr 12 (40 Mb) (only1,255 variants) (Figure 2a). Regarding

the marker distribution within coding vs non-coding regions, SNPs were identified in all segments of genes as well as in the intergenic regions, with the intergenic regions presenting the largest number of SNPs, followed by downstream regions, intronic region, upstream region, exons, 3'UTR and 5'UTR (Figure 2b).



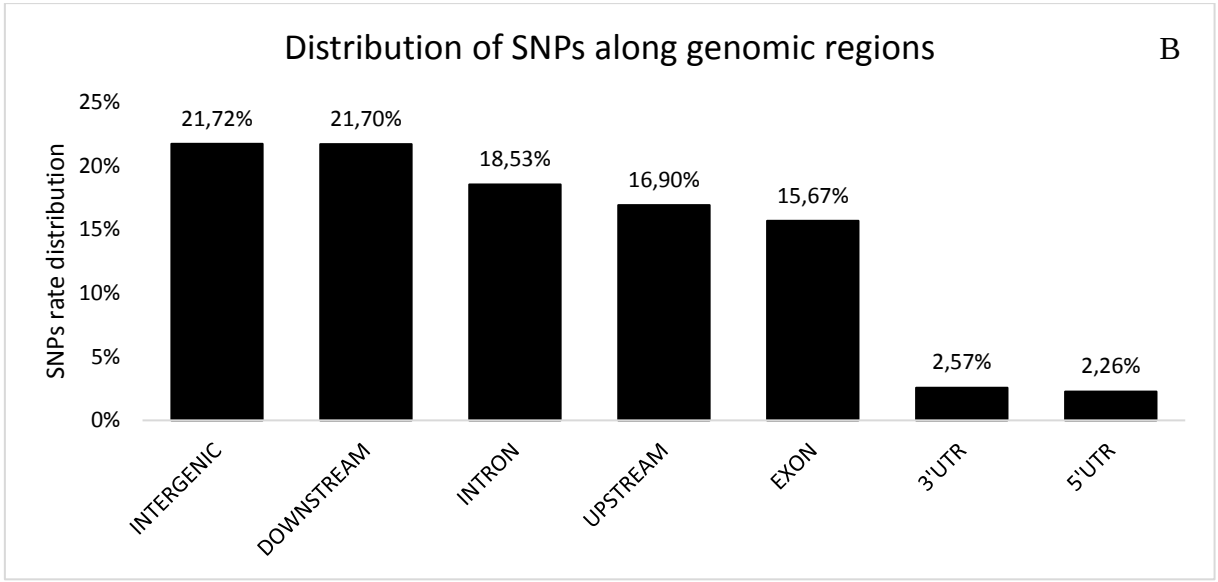


Figure 2 - A) Overview of SNP distribution along chromosomes. The number of SNPs with a MAF≥0.05 is shown on the Y axis, while the X axis represents individual soybean chromosomes. B) SNP distribution among the different genomic regions.

Population structure

The first ten principal components (PCs) were created using TASSEL and cumulatively, explained 31.4% of the variance, with the first PC alone explaining 8.2% of the variance. The first two PCs were able to separate the soybean accessions based on their genetic diversity. As expected, the accessions were spread broadly across the two dimensions of the plot, however, two groups were clearly formed: one with the commercial cultivars (less widely spread) and the other with the plant introduced lines (non-commercial). However, none clustering was observed considering only the reaction to the nematode (Figure 3).

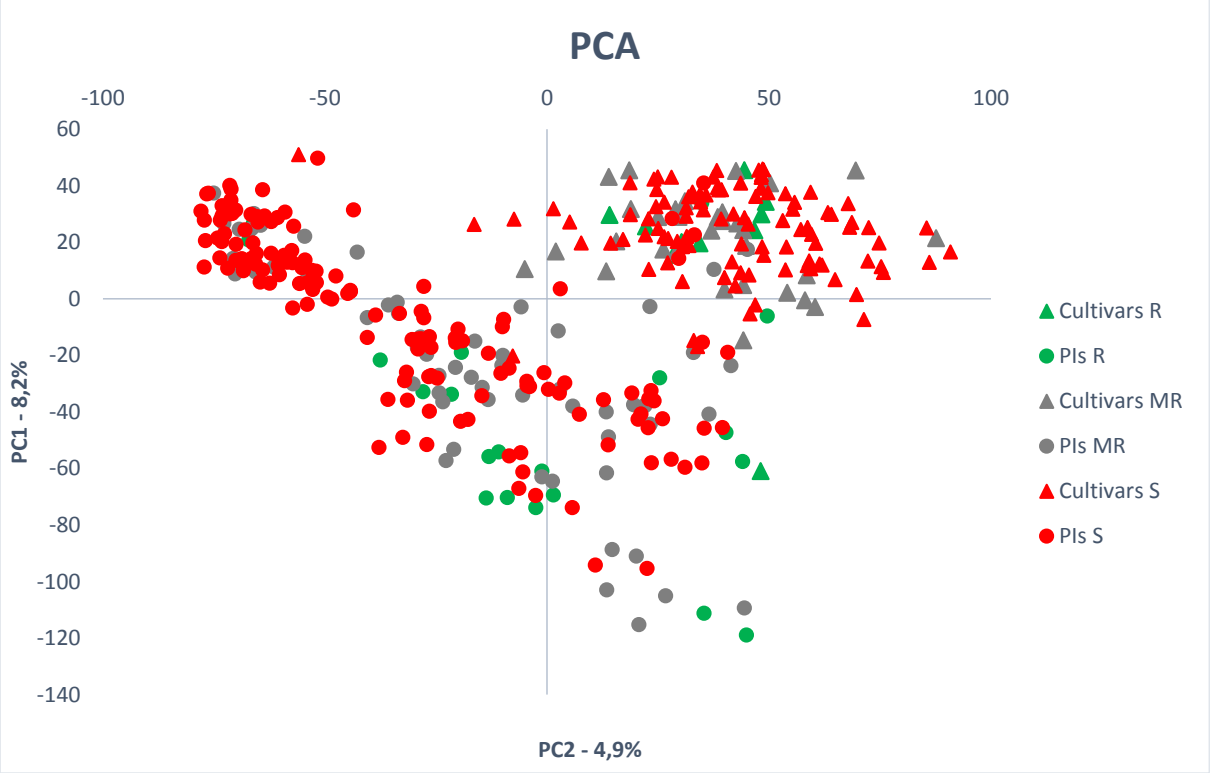


Figure 3 - Genetic structure of accession panels according the first two principal components (PC1 and PC2) with the highest EigenValues. Cultivars and PIs are represented by different dots, been the graphic position determined by genotype.

Association mapping to *M. javanica* resistance

The GWAS was conducted using a cMLM taking into account both genetic relatedness (K matrix) and population structure (P matrix). The plotted p-values obtained by the SNPs through associative mapping confirms that the null association hypothesis does not apply to this *M. javanica* association mapping (Figure 4).

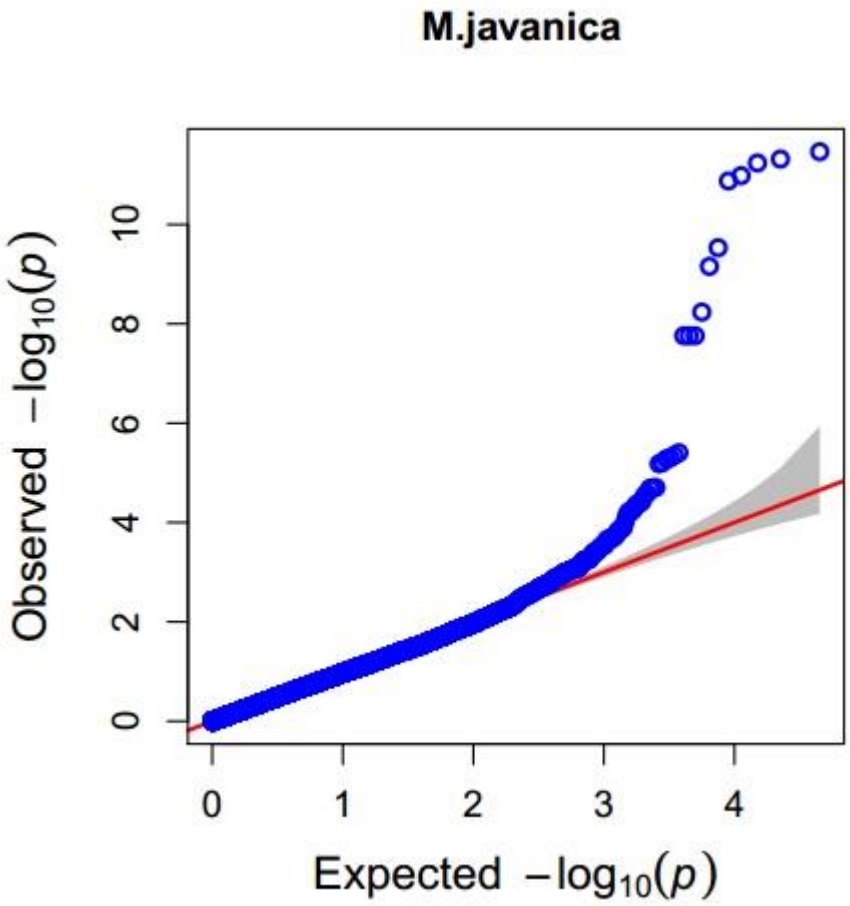


Figure 4 - Quantile-Quantile Plot (QQplot) of P-values. Assuming the p-value follow uniform distribution (>0,1), the X-axis represent expected negative logarithm of P-values, whereas the Y-axis represents observed negative logarithm. The gray line represents the 95% of confidence interval for QQplot under the null hypothesis of no association between the markers and the trait.

Seven SNPs, all located in the same region of chromosome 13, were identified as significantly associated (P=6.97E^{-10} to 3.43E^{-12}) with resistance to *M. javanica* as can be observed in the Manhattan plot (Figure 5). The strongest degree of association (FDR p-value = 8.66E-08) was shared to three SNPs. They are physically located at position 30,804,961; 30.805.500 and 30.805.508. The estimated allelic effect (AE) of SNP1 was -0,46 while SNP2 and SNP3 were -0,45. That values representing both a considerable phenotypic contribution to nematode resistance and a low probability of being a false positive, respectively (Table 1). The MAF obtained was similar among the remaining SNPs ranging from 0.32 to 0.36. Similarly, the R^2 obtained were 0.27 to SNP 1, decreasing to 0.24 on SNP 6 and 7.

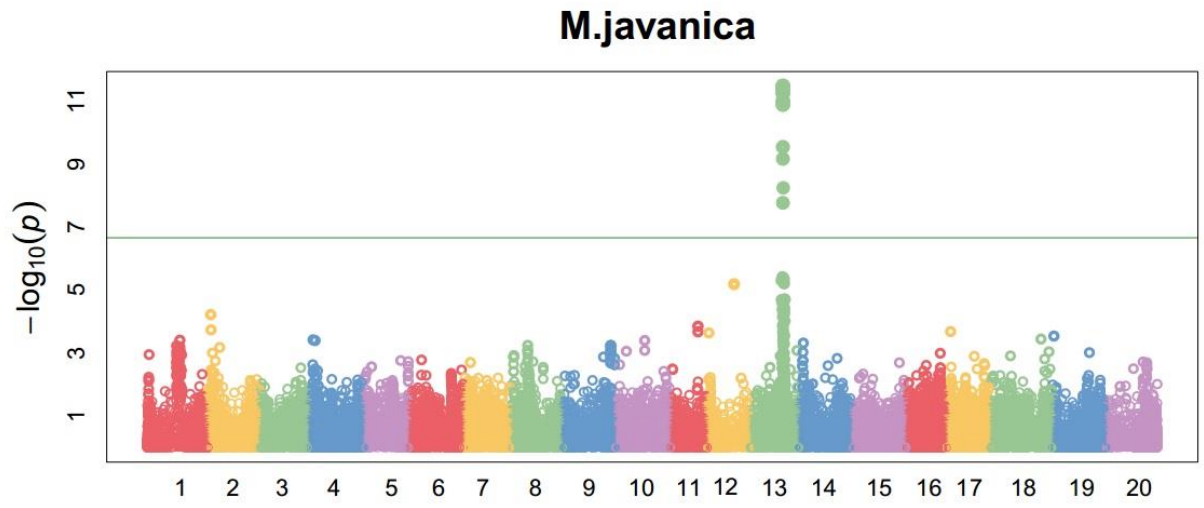


Figure 5 - Graphic distribution of SNPs identified on soybean genome wide association to *M. javanica* resistance. Manhattan plot of GWAS, demonstrating the SNPs distribution along all 20 soybean chromosomes. Significant associations are represented above threshold line.

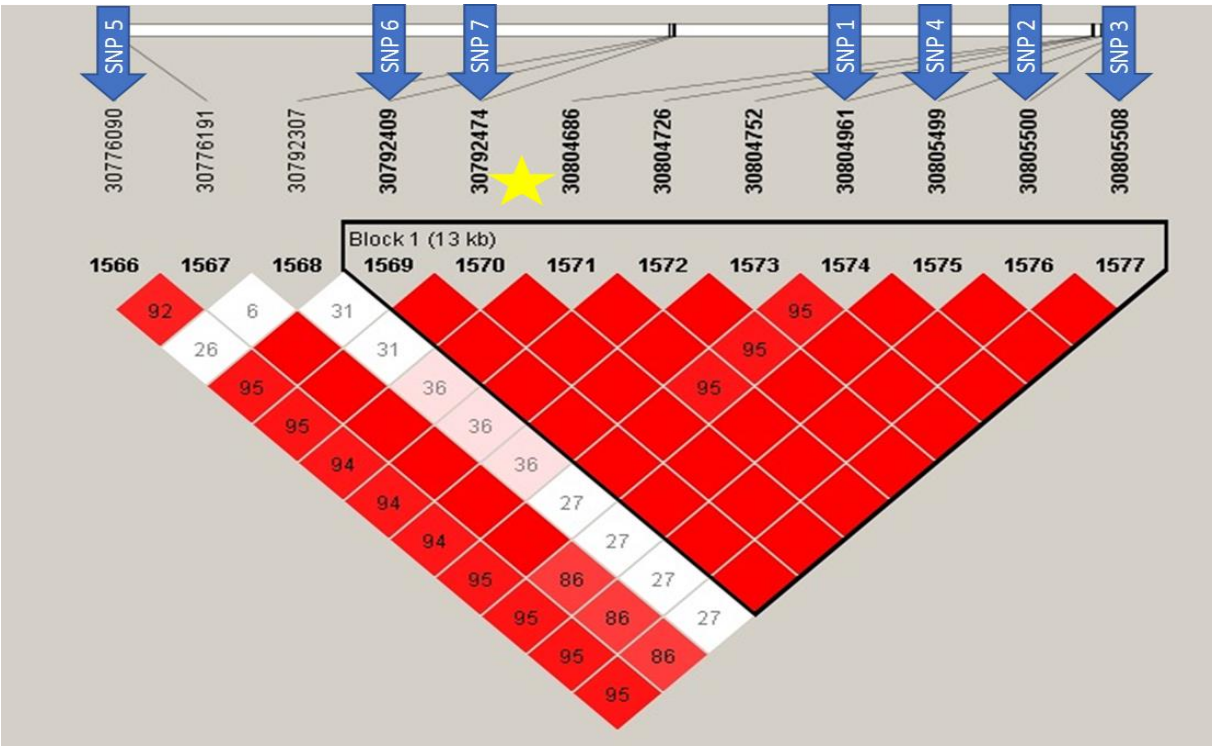
Table 1 – Associated SNPs to *M. javanica* resistance. SNP name are according decreasing p-value (Chr – Chromosome, MAF - Minor Allele Frequence, FDR - False Discovery Rate, AE – Allele Effect – AV – Allelic Variation).

SNP name	SNP Position	Chr	P.value	MAF	FDR	AE	AV	Region	R^2
1	30804961	13	3.43E-12	0.33	8.66E-08	-0.46	C/T	Intergenic	0.27
2	30805500	13	4.82E-12	0.36	8.66E-08	-0.45	A/G	Intergenic	0.26
3	30805508	13	5.77E-12	0.36	8.66E-08	-0.45	T/G	Intergenic	0.26
4	30805499	13	1.04E-11	0.35	1.17E-07	-0.44	C/A	Intergenic	0.26
5	30776090	13	1.33E-11	0.32	1.20E-07	-0.45	C/G	Intergenic	0.26
6	30792409	13	2.94E-10	0.34	2.20E-06	-0.41	T/C	Intergenic	0.24
7	30792474	13	6.97E-10	0.35	4.48E-06	-0.39	A/C	Intergenic	0.24

These seven significant SNPs were distributed along 29.4Kb, ranging between position 30,776,090 and 30,805,508 and were able to define a haplotype that could differentiate resistant from the susceptible accessions.

Among 32 accessions phenotyped as R, twenty-eight (87.5%) contain the haplotypes "CATCCTA", while 96.5% of S accessions presents the haplotype "TGGAGCC" (Supplementary table 5). However, the agreement rate between phenotype and haplotype to S accessions can be increased to 98.5% if only the SNP 1 (30,804,961) was used. On the other hands, four of materials phenotyped as resistant do not share the haplotype of the most significant SNP. These accessions presented the susceptible haplotype. Interestingly, the opposite scenario also was verified, thus, 38 materials present susceptible phenotype and R haplotype and 3 accession presented heterozygosis to SNP1.

According to the LD analysis, six from seven significant SNPs (SNP5 was the exception) were in a unique LD block. This LD block comprised a reduced region of 13.1Kb, which is flanked by SNP 6 and SNP3 (Figure 6). This LD block contain only one gene model (Glyma.13g194700) which coding a protein with LRR domain.



859

860 **Figure 6** - Genomic distribution of mapped SNPs. LD block (delimited by black lines)
861 containing 6 associated SNPs, three non-associated SNPs and one SNP located outside of LD
862 block. (yellow star represents Glyma.13g194700 position).

863

864 The results obtained in the linkage disequilibrium analysis, performed considering full
865 population and 44.992 SNPs, the regression curve fitted to the LD plot falls below r2 = 0.2 at
866 ~0,15Mb (Figure 7a). A similar r2 value was obtained when exclusively chromosome 13 was
867 analyzed (Figure 7b).

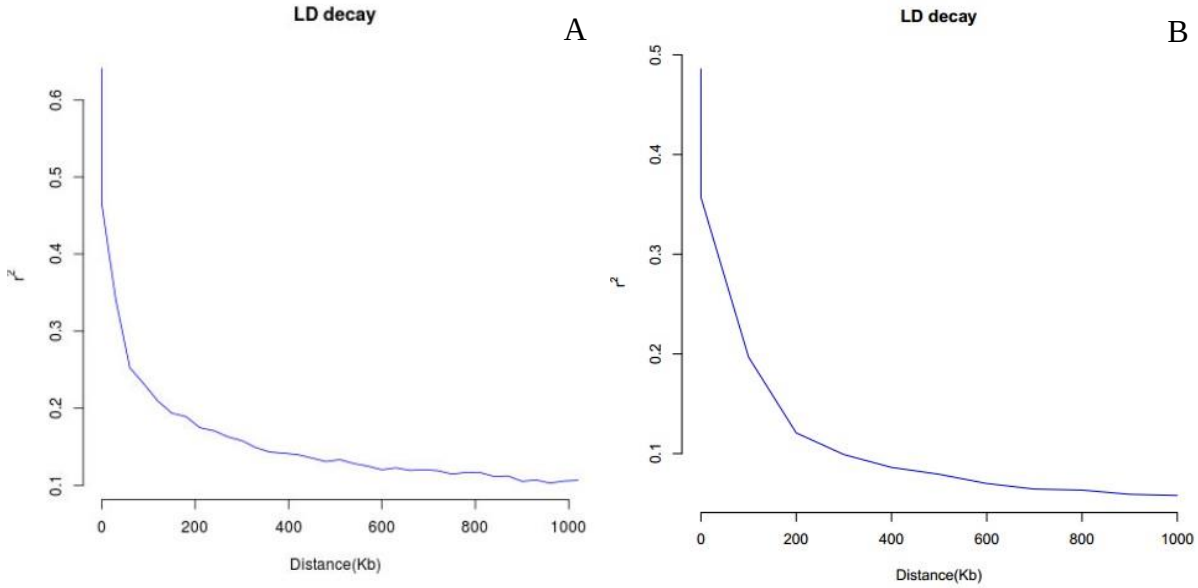


Figure 7 - Linkage disequilibrium decay. a) Whole genome, b) Chromosome 13.

Genotyping assays

The precision of the three different TaqMan genotyping assays developed was checked using fifty-three accession also submitted to GBS. The results obtained on standard reactions demonstrate a coherence between GBS and genotyping assays results higher them 88% (Supplementary table 6).

To identify new materials with strong level of resistance to *M. javanica*, a new panel containing twenty-one accessions were screened. According previous phenotyping literature information were expected fifteen accessions containing high level of resistance to RKN while another six accessions present susceptibility. In fact, that number were obtained, thus, the number of materials with genotype checked to *M. javanica* resistance were increased see (Supplementary table 7). However, two exceptions were observed. The BRSGO8661RR which carrier resistant phenotype, presented susceptible phenotype to all evaluated assays and to the BRS285 the opposite scenario was obtained. The resistance allele was identified even with previous susceptible phenotype (Figure 8)

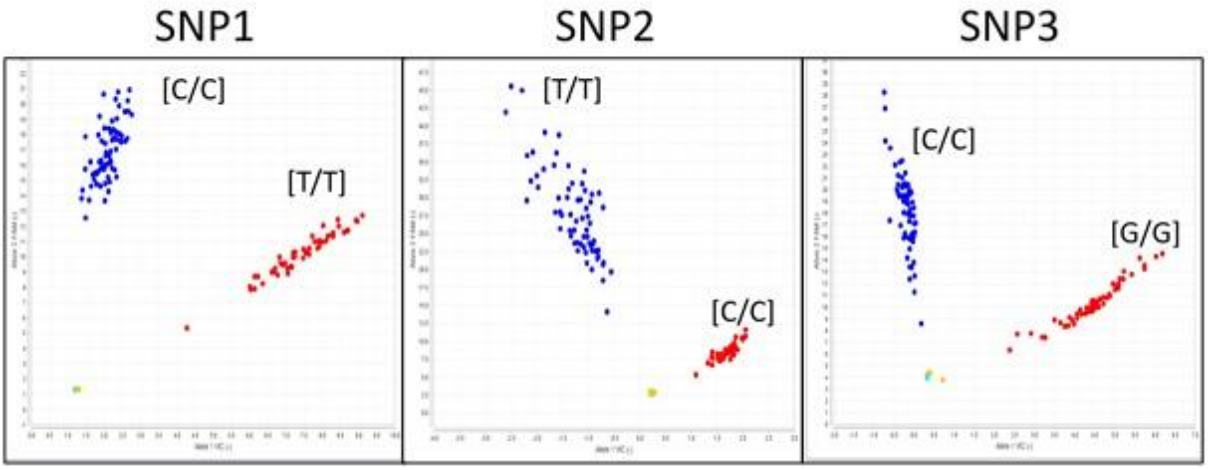


Figure 8 – TaqMan graphs of three SNPs for precision and new set of materials

Discussion

In the present study, an associative panel involving a large population allowed to map a soybean locus associated to *M. javanica* resistance on the chromosome 13. The study allowed the identification of 44,100 SNPs on the population and identified a region of 29.4kb with SNP markers strongly associated with *M. javanica* resistance with significant FDR values. The LD decay value obtained through regression of r^2 indicate that number of SNPs used to association mapping were enough, once the physical distance observed between mapped SNPs and LD block size which contain that, are similar to distance which occur the intersection to threshold.

This study confirms previous information that have mapped this resistance trait on the chromosome thirteen, although in a broader region. In QTL mapping studies, seven RFLP markers distributed approximately between physical positions 17,875,691 and 34,077,986 on chromosome 13, associated to *M. javanica* [12,33]. According to Soybase, this QTL is located in a large interval (16.2Mb) of chromosome (www.soybase.org). Another study [14] also

mapped RKN resistance on this same chromosome, where SSR markers associated to *M. javanica* resistance were identified in a two biparental population [BRS133 (S) x CD201 (R) and BRS133 (S) x PI59099 (R)]. In that study, the markers Satt133 (position 24,949,516..24,949,765) and the SOYHSP176 (27,847,477..27,847,591) were found associated to this trait. Also, using SSR, [15], the authors evaluated between BRS133 and PI595099. The markers Satt114 (28,912,864..28,912,971) and SOYHSP176 were identified explaining phenotypic variation.

Although these QTLs and molecular markers whose were previously described, the closest markers to the region mapped on the present study are Satt114 [15] and B174 [33], at 1.89Mb and 2.57Mb from SNP1, respectively. The mapped SNPs in present study are located inside the region previous mapped, but it represents a greater precision, delimiting a region spanning only 30 Kb.

The present study allowed the identification of the haplotype CATCCTA associated to the resistant accessions, while the haplotype TGGAGCC to susceptible materials. However, due the six from seven mapped SNPs are in a single LD block, and delimits also a single gene (Glyma.13g194700), these haplotypes can be reduced. Using only the first mapped SNP (30,904,961 – C/T), the efficiency on segregation of S accessions are increased up to 2%. It is due full SNPs mapped haplotype composition carrier some aleatory mutation whose were also sampled, reducing the efficiency on segregation of S accessions.

The existence of accessions with R phenotype and S genotype together the R^2 values, gives us evidences of existence of another QTLs (with minor effect) contributing to *M. javanica* resistance whose were not identified by present GWAS, similar results were found in *M. incognita* study [31]. It could be due low sampling of materials carrying this QTL or low frequency on populations.

On the other hand, the opposite scenario was also verified, where the S phenotype and R haplotype was identified to thirty-eight accessions as well as both haplotype were found in MR accessions. It may point the existence of epistasis, once by definition, epistasis is the interaction between genes which modified the segregation ratio [34, 35, 36]

The Glyma.13g194700 is the only gene located inside the small LD block delimited by mapped SNPs. This gene contains high similarity (72.1%) to the gene Medtr8g039910, a gene from *Medicago truncatula,* which encode a protein sharing the same NB-ARC-LRR-TIR domains. The NB-ARC domain is also called as NBS domain [37], while TIR characterize a specific class of NBS-LRR which contain similarity to Interleukin-1 Receptor from human and *Drosophila melanogaster* [38]. In plants the domain NBS is responsible for binding and hydrolysis of ATP and GTP [39]. These domains are present on several genes knowns as to provide resistance to disease resistance (R genes), including to nematodes [40,41,42,43]. The LRR genes play an important function on specificity and recognition in protein-protein interactions [44, 45]. A good example is the Mi-1 gene, from tomato which contain NBS-LRR domain, an important and well characterized gene that confer resistance to *M. incognita* in this species [46,47].

Mapped SNPs were also used as base to design assays based on qPCR methodology for routine use in marker-assisted selection. All markers assays developed here were able to be amplify correctly, forming well separated clusters, the resistant and susceptible. The analysis of a subset of the association panel demonstrated the genotype expected was obtained for more than 88% of samples.

The assays were also tested in a new panel of cultivars to extend the number of material harboring the QTL identified in this study. A total of 16 accessions were genotyped

as *M. javanica* resistant, even the BRSGO8661RR and BRS285 presented phenotype and genotype contrasting.

Therefore, this study was demonstrated relevant in approaching *M. javanica* resistance in an elevated number of individuals of a non-related population through a high accurate SNPs discovery and evaluation technology. Furthermore, it allows us to identification of two accessions (PI507089B and BRS282) previously genotyped as R to *M. incognita* [31] whose were R to *M. javanica* too and development of efficient assays with high potential to be used on marker assisted selection.

Conclusion

This study represents the first GWAS analysis for *M. javanica* resistance. Our data are according previous results where the *M. javanica* resistance was mapped on chromosome 13. Seven SNPs with high association were located in a small range of 29.4kb, which can be reduced to a single LD block 13.09Kb long. It allowed the identification of a haplotype that can be very useful on future marker assisted selection, supporting breeders on RKN resistant plant selection. Furthermore, designed assays showed high efficiency on allelic discrimination between resistant and susceptible plants to *M. javanica.*

SUPPLEMENTARY DATA

Supplementary table 1- List of 369 soybean accessions used for the GWAS panel. (information obtained on GRIN - Germplasm Resources Information Network).

PlantID	Plant name	Maturity Group	Origin
PI628902	BR1	-	Brazil
PI628809	BR16	6.0	Brazil
PI628810	BR27(Cariri)	8.0	Brazil
PI628916	BR36	7.0	Brazil
-	BR40(Itiquira)	-	Brazil
PI 548660	Bragg	-	United States
-	BRS132	-	Brazil
-	BRS133	7.3	Brazil
-	BRS185	6.8	Brazil
-	BRS214	-	Brazil
-	BRS216	8.8	Brazil
-	BRS217[Flora]	7.9	Brazil
-	BRS218[Nina]	-	Brazil
-	BRS233	7.4	Brazil
-	BRS239	6.9	Brazil
-	BRS242RR	6.7	Brazil
-	BRS245RR	7.3	Brazil
-	BRS257	6.7	Brazil
-	BRS263[Diferente]	-	Brazil
-	BRS270RR	-	Brazil
-	BRS271RR	9.3	Brazil
-	BRS278	-	Brazil
-	BRS278RR	-	Brazil
-	BRS282	6.9	Brazil
-	BRS283	6.5	Brazil
-	BRS314	9.2	Brazil
-	BRS320	-	Brazil
-	BRSBalizaRR	-	Brazil
-	BRSCandieiro	-	Brazil
-	BRSCarla	-	Brazil
-	BRSCeleste	-	Brazil
-	BRSCharruaRR	6.8	Brazil
PI675655	BRSCorisco	-	Brazil
-	BRSGiseleRR	8.9	Brazil
-	BRSGO204[Goiania]	-	Brazil
-	BRSGO8060	-	Brazil

-	BRSGO8061	8.6	Brazil
-	BRSGOCaiapônia	-	Brazil
PI675654	BRSGOChapadoes	8.6	Brazil
-	BRSGOEdeia	-	Brazil
-	BRSGOIara	-	Brazil
-	BRSGOMineiros	-	Brazil
-	BRSGralha	8.8	Brazil
-	BRSGuapa	-	Brazil
-	BRSInvernada	6.8	Brazil
-	BRSJuliana	-	Brazil
-	BRSMacota	6.8	Brazil
-	BRSMAPati	-	Brazil
-	BRSMASeridoRCH	-	Brazil
-	BRSMG250[Nobreza]	-	Brazil
-	BRSMG740SRR	7.4	Brazil
-	BRSMG750RR	-	Brazil
-	BRSMG751SRR	-	Brazil
-	BRSMG800A	8.0	Brazil
-	BRSMG850GRR	-	Brazil
-	BRSMGConfianca	-	Brazil
-	BRSMGRenascença	-	Brazil
-	BRSMGVirtuosa	-	Brazil
-	BRSMilena	8.3	Brazil
PI675659	BRSMTPintado	-	Brazil
-	BRSMTUirapuru	-	Brazil
-	BRSNovaSavana	-	Brazil
-	BRSPetala	8.7	Brazil
-	BRSRaimunda	9.0	Brazil
PI675660	BRSSambaiba	9.3	Brazil
-	BRSSinuelo	-	Brazil
-	BRSTorena	-	Brazil
-	BRSValiosaRR	8.1	Brazil
PI628814	CamposGerais	6.0	Brazil
-	CD201	-	Brazil
-	CD202	6.4	Brazil
-	CD219RR	8.2	Brazil
-	CD224	6.9	Brazil
-	CD225RR	5.8	Brazil
-	CD233RR	6.4	Brazil
PI553039	Davis	VI	United States - Arkansas
PI556636	DELTAPINE506	VII	-
-	DM309	-	-
PI628922	Dourados	VIII	Brazil

PI628923	Embrapa1(IAS5RC)	VI	Brazil
PI628925	Embrapa20(DokoRC)	VIII	Brazil
PI628924	Embrapa4(BR4RC)	VI	Brazil
PI663948	Embrapa48	6.8	Brazil
-	Embrapa59	-	Brazil
-	Embrapa60	-	Brazil
-	Embrapa63(Mirador)	-	Brazil
-	EMGOPA301	-	Brazil
-	FC031683	5.0	Unknown
-	FMTMatrinxa	-	Brazil
-	FMTTucunaré	-	Brazil
PI628829	FT10(Princesa)	7.0	Brazil
-	FT2001	-	-
PI628821	FT-Cometa	5.0	Brazil
PI628839	FTEureka	6.0	Brazil
-	FUNDACEP59RR	7.5	Brazil
-	GB881RR	-	Brazil
PI628845	IAC10	7.0	Brazil
PI628850	IAC100	7.0	Brazil
PI628846	IAC11	6.0	Brazil
PI628943	IAC2	9.0	Brazil
PI628844	IAS5	6.0	Brazil
PI628857	Industrial	8.0	Brazil
PI628859	Ivai	8.0	Brazil
PI628812	MG/BR46(Conquista)	8.1	Brazil
PI628813	MG/BR48(GarimpoRCH)	8.0	Brazil
PI628864	Mineira	-	Brazil
-	M-SOY5826	5.6	-
-	M-SOY5942	-	-
-	M-SOY7501	-	-
-	M-SOY7908RR	-	-
-	M-SOY8008	7.6	-
-	M-SOY8199	8.1	-
-	M-SOY8248	8.2	-
-	M-SOY8411	-	-
-	M-SOY8527RR	8.5	-
-	M-SOY8585RR	-	-
-	M-SOY8849RR	8.8	-
-	M-SOY8870	-	-
-	M-SOY8925	-	-
-	Msoy9056RR	9.0	-
PI628811	MT/BR45(Paiaguas)	8.0	Brazil
PI628919	MT/BR50(Parecis)	8.0	Brazil
-	MT/BR53(Tucano)	-	Brazil

PI628873	OCEPAR10	6.0	Brazil
-	OCEPAR12	-	Brazil
PI628867	OCEPAR3(Primavera)	6.0	Brazil
PI628871	OCEPAR8	6.0	Brazil
PI628872	OCEPAR9	6.0	Brazil
PI628879	Parana	5.0	Brazil
-	Pelicano	-	Brazil
PI628882	Perola	8.8	Brazil
-	PI054610	6.0	China
-	PI054610-1	3.0	China
-	PI059845	5.0	Japan
-	PI089772	4.0	China
-	PI090490-2	-	China
-	PI091178	4.0	China
-	PI096118	4.0	North Korea
-	PI096280	4.0	North Korea
-	PI097038	4.0	North Korea
-	PI157428	4.0	North Korea
-	PI158765	4.0	China
-	PI171427	4.0	China
-	PI171431	4.0	China
-	PI171432	4.0	China
-	PI171454	4.0	Japan
-	PI171652	4.0	Turkey
-	PI179826	4.0	China
-	PI196170	4.0	South Korea
-	PI200519	4.0	Japan
-	PI200538	8.0	Japan
-	PI229325	4.0	Japan
-	PI230977	7.0	Japan
-	PI248515	4.0	Japan
-	PI253651D	4.0	China
-	PI253652A	4.0	China
-	PI253654	4.0	China
-	PI253663	4.0	China
-	PI304218	4.0	Taiwan
-	PI323556	4.0	India
-	PI339868B	4.0	South Korea
-	PI371611	4.0	Paskistan
-	PI374189	10.0	India
-	PI398313	4.0	South Korea
-	PI398666	4.0	South Korea
-	PI398848	4.0	South Korea
-	PI398887	4.0	South Korea

-	PI398965	4.0	South Korea
-	PI399020	4.0	South Korea
-	PI404199	4.0	China
-	PI417234	8.0	Japan
-	PI417580	3.0	Japan
-	PI423945	2.0	Japan
-	PI424085A	4.0	South Korea
-	PI424099A	-	-
-	PI424492	4.0	South Korea
-	PI424494	4.0	South Korea
-	PI424495	5.0	South Korea
-	PI424499D	4.0	South Korea
-	PI424504A	4.0	South Korea
-	PI424505	4.0	South Korea
-	PI424506	4.0	South Korea
-	PI424511	4.0	South Korea
-	PI424522	4.0	South Korea
-	PI424523B	4.0	South Korea
-	PI424549A	4.0	South Korea
-	PI424554	4.0	South Korea
-	PI424555B	4.0	South Korea
-	PI424557	4.0	South Korea
-	PI424558A	4.0	South Korea
-	PI424574	4.0	South Korea
-	PI424588	4.0	South Korea
-	PI424597	4.0	South Korea
-	PI424605A	4.0	South Korea
-	PI430460A	1.0	China
-	PI430596	2.0	China
-	PI432359	4.0	Mexico
-	PI437127B	4.0	United States
-	PI437341	2.0	Russia
-	PI437350	1.0	Russia
-	PI437353	1.0	Russia
-	PI437423	1.0	Russia
-	PI437486	1.0	China
-	PI437580	4.0	China
-	PI437636B	1.0	China
-	PI437673	1.0	China
-	PI437679	4.0	China
-	PI437725	4.0	China
-	PI437749	4.0	China
-	PI437773	0.0	China
-	PI437801	4.0	China

-	PI437819	1.0	China
-	PI437829	1.0	China
-	PI437845D	4.0	China
-	PI437912	1.0	China
-	PI438048B	1.0	China
-	PI438187	1.0	China
-	PI438190	1.0	China
-	PI438255	1.0	China
-	PI438302A	4.0	Korea
-	PI438303	4.0	Korea
-	PI438304B	4.0	Korea
-	PI438307	4.0	Korea
-	PI438442A	2.0	Netherlands
-	PI438492	1.0	United States
-	PI442005	1.0	United States
-	PI442010	4.0	South Korea
-	PI442012A	4.0	South Korea
-	PI442018	4.0	South Korea
-	PI442044	0.0	Yugoslavia
-	PI458175C	4.0	South Korea
-	PI458199	4.0	South Korea
-	PI458226	4.0	South Korea
-	PI458234	4.0	South Korea
-	PI458236A	4.0	South Korea
-	PI458249	4.0	South Korea
-	PI458294	4.0	South Korea
-	PI458298	4.0	South Korea
-	PI458306A	4.0	South Korea
-	PI458515	4.0	China
-	PI467316	4.0	China
-	PI470226	4.0	China
-	PI476885	6.0	Vietnan
-	PI483253	9.0	Brazil
-	PI487431	-	-
-	PI495020	4.0	China
-	PI495831	1.0	France
-	PI495832	1.0	France
-	PI506516	4.0	Japan
-	PI506525	4.0	Japan
-	PI506590E	4.0	Japan
-	PI506705	4.0	Japan
-	PI506789	4.0	Japan
-	PI506819	4.0	Japan
-	PI506833	4.0	Japan

-	PI506848	4.0	Japan
-	PI506862	4.0	Japan
-	PI506892	4.0	Japan
-	PI506935	4.0	Japan
-	PI506989	4.0	Japan
-	PI507072	4.0	Japan
-	PI507073	4.0	Japan
-	PI507082A	4.0	Japan
-	PI507089A	4.0	Japan
-	PI507089B	4.0	Japan
-	PI507097	4.0	Japan
-	PI507153	4.0	Japan
-	PI507158	4.0	Japan
-	PI507160	4.0	Japan
-	PI507259	7.0	Japan
-	PI507286C	4.0	Japan
-	PI507316	4.0	Japan
-	PI507317	4.0	Japan
-	PI507407	4.0	Japan
-	PI507408	4.0	Japan
-	PI507430	4.0	Japan
-	PI507432	4.0	Japan
-	PI507443	4.0	Japan
-	PI507449	4.0	Japan
-	PI507480	4.0	Japan
-	PI507492	4.0	Japan
-	PI507501	4.0	Japan
-	PI507571	4.0	Japan
-	PI508296A	4.0	South Korea
-	PI508296G	4.0	South Korea
-	PI509075	4.0	South Korea
-	PI509079	4.0	South Korea
-	PI518719	4.0	China
-	PI518720	4.0	China
-	PI520733	4.0	South Korea
-	PI522189	1.0	Maldova
-	PI532455b	4.0	China
-	PI533655	2.0	United States
-	PI547420	4.0	United States
-	PI547475	4.0	United States
-	PI547672	4.0	United States
-	PI547836	2.0	United States
-	PI547855	3.0	United States
-	PI547856	3.0	United States

-	PI547874	3.0	United States
-	PI548362	3.0	United States
-	PI548493	7.0	Japan
-	PI548542	3.0	United States
-	PI548637	0.0	Canada
-	PI549076A	0.0	China
-	PI561337	1.0	China
-	PI561354	1.0	China
-	PI561379B	6.0	China
-	PI567001B	9.0	Indonesia
-	PI567045	9.0	Indonesia
-	PI567214B	1.0	Russia
-	PI567611	4.0	China
-	PI567648C	4.0	China
-	PI567668	3.0	China
-	PI570668	6.0	Mexico
-	PI573286	6.0	United States
-	PI578506	0.0	China
-	PI587608B	5.0	china
-	PI587618A	6.0	China
-	PI587911A	9.0	China
-	PI587958	9.0	China
-	PI587959	9.0	China
-	PI587991	3.0	China
-	PI588025	4.0	China
-	PI588028	4.0	China
-	PI593956A	1.0	China
-	PI593972	1.0	Japan
-	PI594401B	3.0	China
-	PI594403	4.0	China
-	PI594427C	4.0	China
-	PI594442B	4.0	China
-	PI594470C	8.0	China
-	PI594538A	9.0	China
-	PI594544	9.0	China
-	PI594596	7.0	China
-	PI594775	5.0	China
-	PI595099	8.0	united States
-	PI603200	3.0	United States
-	PI603568	4.0	China
-	PI603694B	4.0	China
-	PI607380	6.0	United States
-	PI608357	5.0	United States
-	PI612758B	3.0	China

-	PI628827	9.0	Brazil
-	PI628889	9.0	Brazil
-	PI632421	4.0	United States
-	PI632428	4.0	United States
-	PI634335	4.0	United States
-	PI634813	0.0	United States
-	PI634903	4.0	United States
-	PI644103	-	Brazil
-	PI646157	-	United States
-	PI657701	-	Nigeria
-	PI89772	4.0	China
-	PI90490-2	4.1	China
-	PI96118	4.2	North Korea
-	PI96280	4.3	North Korea
-	PI97038	4.4	North Korea
PI628883	Planalto	6.0	Brazil
PI628889	SantaRosa	-	Brazil
-	TMG1182RR	8.2	Brazil
-	TMG121RR	8.4	Brazil
PI189968	TOKYO	1.0	France
-	UFV10(Uberaba)	-	Brazil
-	UFV14	-	Brazil
-	UFV18(PatosdeMinas)	-	Brazil
-	UFVS2001	-	Brazil
-	V.MAXRR	5.9	Brazil
PI628899	Vicoja	8.0	Brazil

Supplementary table 2 – List of materials evaluated by TaqMan to increase the number of accession resistant to *M. javanica*

Accession	Phenotype
BRS256RR	R
BRS316RR	R
BRS319RR	R
BRS399RR	R
BRS6970IPRO	R
BRS7380RR	R
BRS7980	R
BRSFavoritaRR	R
BRSGO7963	R
BRSGO8360	R
BRSGO8661RR	R
BRSGOLuziânia	R
BRSMGGarantia	R
CD208	R
BRS211	R
BRS230	S
BRS232	S
BRS285	S
BRS317	S
BRSMG68(Vencedora)	S
BRSMG790A	S

Supplementary table 3 - TaqMan assay

Oligo name	Oligo type/orientation	Sequence (5´- 3´)
Assay 1	Forward primer	GAGTGAATGACAGGTGGA
	Resistant Array	CAATAACCCACGATAGATGGGTGCTT
	Susceptible Array	CAATAACCCACGATAAATGGGTGCTT
	Reverse primer	CACGATTCTCATCCATACAA
Assay 2	Forward primer	CACAGTGCGACTTAGTTC
	Resistant Array	CCATCCCATGACATGATGGTGC
	Susceptible Array	CCATCCCATGCCATGATGGTGC
	Reverse primer	CCAGTAGCAGCATACATC
Assay 3	Forward primer	GTGGGAGAGGTAATGGAG
	Resistant Array	ACTGCTGAACTAGCTGTTCTTCGT
	Susceptible Array	ACTGCTGAACTACCTGTTCTTCGT
	Reverse primer	AAGGGTTTCTCTCATTCC

997 Supplementary table 4 – **List of accessions submitted to GBS and assay analysis**

Genotype
BR16
BR27(Cariri)
BR40(Itiquira)
BRS132
BRS133
BRS185
BRS239
BRSJuliana
BRSMG850GRR
CD219RR
CD224
CD225RR
CD233RR
FC031683
FT-Cometa
Industrial
MG/BR48(GarimpoRCH)
M-SOY5942
PI158765
PI171427
PI230977
PI253651D
PI323556
PI371611
PI398666
PI398887
PI398965
PI399020
PI424588
PI437127B
PI437679
PI438190
PI438307
PI438442A
PI438492
PI442005
PI442010
PI442012A
PI442018
PI442044
PI458175C
PI458199

PI458226
PI506862
PI567214B
PI567648C
PI587911A
PI588025
PI607380
PI608357
PI634813
PI89772
UFV10(Uberaba)

998 Supplementary table 5 - **Haplotype obtained on GBS**

Genótipo	**Média**	**Reação**	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
Bragg	1.0	Resistente	C	A	T	C	C	T	A
BRS257	2.2	Resistente	C	A	T	C	C	T	A
BRSGralha	2.5	Resistente	C	A	T	C	C	T	A
BRSJuliana	2.3	Resistente	C	A	T	C	C	T	A
BRSMG850GRR	1.5	Resistente	C	A	T	C	C	T	A
CD219RR	1.8	Resistente	C	A	T	C	C	T	A
CD224	1.3	Resistente	C	A	T	C	C	T	A
CD233RR	2.3	Resistente	C	A	T	C	C	T	A
FC031683	2.5	Resistente	C	A	T	C	C	T	A
PI158765	2.3	Resistente	C	A	T	C	C	T	A
PI171427	2.0	Resistente	C	A	T	C	C	T	A
PI230977	2.5	Resistente	C	A	T	C	C	T	A
PI323556	2.5	Resistente	C	A	T	C	C	T	A
PI437127B	1.0	Resistente	C	A	T	C	C	T	A
PI437679	1.8	Resistente	C	A	T	C	C	T	A
PI506862	1.8	Resistente	C	A	T	C	C	T	A
PI567214B	2.2	Resistente	C	A	T	C	C	T	A
PI567648C	2.3	Resistente	C	A	T	C	C	T	A
PI587911A	2.5	Resistente	C	A	T	C	C	T	A
PI588025	2.3	Resistente	C	A	T	C	C	T	A
PI607380	2.5	Resistente	C	A	T	C	C	T	A
PI608357	2.0	Resistente	C	A	T	C	C	T	A
PI634813	2.5	Resistente	C	A	T	C	C	T	A
PI595099	1.0	Resistente	C	A	T	C	C	T	A

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI89772	2.0	Resistente	C	A	T	C	C	T	A
BRS239	2.3	Resistente	C	A	T	C	C	T	A
CD225RR	2.0	Resistente	C	A	T	C	C	T	A
PI090490-2	2.5	Resistente	C	A	T	C	C	T	A
PI253651D	2.3	Resistente	T	G	G	A	G	C	C
PI424588	2.5	Resistente	T	G	G	A	G	C	C
PI438190	2.5	Resistente	T	G	G	A	G	C	C
UFV10(Uberaba)	2.3	Resistente	T	G	G	A	G	C	C
BRS216	3.5	Mod.Resistent	C	A	T	C	C	T	A
BRS233	3.0	Mod.Resistent	C	A	T	C	C	T	A
BRS278	3.2	Mod.Resistent	C	A	T	C	C	T	A
BRS278RR	3.2	Mod.Resistent	C	A	T	C	C	T	A
BRS282	3.3	Mod.Resistent	C	A	T	C	C	T	A
BRS283	3.3	Mod.Resistent	C	A	T	C	C	T	A
BRSCeleste	3.0	Mod.Resistent	C	A	T	C	C	T	A
BRSCorisco	3.4	Mod.Resistent	C	R	K	M	C	Y	A
BRSGOMineiros	3.2	Mod.Resistent	C	A	T	C	C	T	A
BRSMGVirtuosa	2.7	Mod.Resistent	C	A	T	C	C	T	A
BRSRaimunda	3.0	Mod.Resistent	C	A	T	C	C	T	A
BRSValiosaRR	3.5	Mod.Resistent	C	A	T	C	C	T	A
DELTAPINE506	3.3	Mod.Resistent	C	A	T	C	C	T	A
Embrapa63(Mirador)	3.3	Mod.Resistent	C	A	T	C	C	T	A
IAC2	2.8	Mod.Resistent	C	A	T	C	C	T	A
MG/BR46(Conquista)	2.8	Mod.Resistent	C	A	T	C	C	T	A
M-SOY7908RR	3.2	Mod.Resistent	C	A	T	C	C	T	A
OCEPAR3(Primavera)	3.5	Mod.Resistent	C	A	T	C	C	T	A
PI054610	2.8	Mod.Resistent	C	R	K	M	C	Y	A

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI089772	3.0	Mod.Resistent	Y	A	T	C	C	T	A
PI096118	3.5	Mod.Resistent	C	A	T	C	C	T	A
PI096280	3.3	Mod.Resistent	C	A	T	C	C	T	A
PI179826	3.2	Mod.Resistent	C	A	T	C	S	T	A
PI200538	2.7	Mod.Resistent	C	A	T	C	C	T	A
PI339868B	3.3	Mod.Resistent	C	A	T	C	C	T	A
PI424085A	3.4	Mod.Resistent	C	A	T	C	C	T	A
PI424099A	3.3	Mod.Resistent	C	A	T	C	G	T	A
PI432359	2.8	Mod.Resistent	C	A	T	C	C	T	A
PI437341	3.0	Mod.Resistent	C	A	T	C	C	T	A
PI476885	2.8	Mod.Resistent	C	A	T	C	C	T	A
PI487431	3.2	Mod.Resistent	C	A	T	C	G	T	A
PI495831	3.0	Mod.Resistent	C	A	T	C	C	T	A
PI506705	2.8	Mod.Resistent	C	A	T	C	C	T	A
PI507089A	2.8	Mod.Resistent	Y	A	T	C	S	T	A
PI507089B	2.7	Mod.Resistent	C	A	T	C	C	T	A
PI533655	3.4	Mod.Resistent	C	A	T	C	C	T	A
PI547420	3.2	Mod.Resistent	C	A	T	C	C	T	A
PI547475	3.0	Mod.Resistent	C	A	T	C	C	T	A
PI547672	2.8	Mod.Resistent	C	A	T	C	C	T	A
PI547855	3.3	Mod.Resistent	C	A	T	C	C	T	A
PI548362	3.2	Mod.Resistent	C	A	T	C	C	T	A
PI548542	3.5	Mod.Resistent	C	A	T	C	C	T	A
PI548637	3.2	Mod.Resistent	C	A	T	C	C	T	A
PI567045	3.0	Mod.Resistent	C	A	T	C	C	T	A
PI573286	3.2	Mod.Resistent	C	A	T	C	C	T	A

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI587958	3.2	Mod.Resistent	C	A	T	C	G	T	A
PI588028	3.2	Mod.Resistent	C	A	T	C	C	T	A
PI632421	2.8	Mod.Resistent	C	A	T	C	C	T	A
PI632428	3.0	Mod.Resistent	C	A	T	C	C	T	A
PI657701	3.2	Mod.Resistent	C	A	T	C	C	T	A
PI90490-2	3.5	Mod.Resistent	C	A	T	C	C	T	A
PI96280	3.5	Mod.Resistent	C	A	T	C	C	T	A
SantaRosa	3.5	Mod.Resistent	C	A	T	C	S	T	A
TMG1182RR	2.7	Mod.Resistent	C	A	T	C	C	T	A
UFV18(PatosdeMinas)	3.0	Mod.Resistent	C	A	T	C	C	T	A
V.MAXRR	3.3	Mod.Resistent	C	A	T	C	C	T	A
BR1	3.5	Mod.Resistent	T	G	G	A	G	C	C
BRSGOChapadoes	3.2	Mod.Resistent	T	G	G	A	G	C	C
BRSMG740SRR	2.7	Mod.Resistent	T	G	G	A	G	C	C
BRSMG751SRR	3.2	Mod.Resistent	T	G	G	A	G	C	C
BRSMilena	3.3	Mod.Resistent	T	G	G	A	G	C	C
Embrapa4(BR4RC)	3.4	Mod.Resistent	T	G	G	A	G	C	C
Embrapa59	3.3	Mod.Resistent	T	G	G	A	G	C	C
FMTMatrinxa	3.5	Mod.Resistent	T	G	G	A	G	C	C
FUNDACEP59RR	3.3	Mod.Resistent	T	G	G	A	G	C	C
PI171432	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI253663	3.5	Mod.Resistent	T	G	G	A	G	C	C
PI424494	3.5	Mod.Resistent	T	R	K	M	G	C	C
PI424557	2.7	Mod.Resistent	T	G	G	A	G	C	C
PI424597	3.5	Mod.Resistent	T	G	G	A	G	C	C
PI424605A	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI437725	3.4	Mod.Resistent	T	G	G	A	G	C	C

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI437801	3.4	Mod.Resistent	T	G	G	A	G	C	C
PI437819	3.0	Mod.Resistent	T	G	G	A	G	C	C
PI437912	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI438303	2.8	Mod.Resistent	T	G	G	A	G	C	C
PI458294	3.3	Mod.Resistent	T	G	G	A	G	C	C
PI458515	3.4	Mod.Resistent	T	G	G	A	G	C	A
PI506848	2.8	Mod.Resistent	T	G	G	A	G	C	C
PI507072	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI507073	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI507160	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI507443	3.5	Mod.Resistent	T	G	G	A	G	C	C
PI507571	3.5	Mod.Resistent	T	G	G	A	G	C	C
PI520733	3.0	Mod.Resistent	T	G	G	A	G	C	C
PI549076A	3.5	Mod.Resistent	T	G	G	A	G	C	C
PI561379B	2.7	Mod.Resistent	T	G	G	A	G	C	C
PI567611	2.8	Mod.Resistent	T	R	K	A	G	C	C
PI570668	3.2	Mod.Resistent	T	R	K	M	G	C	C
PI593956A	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI594401B	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI594470C	3.2	Mod.Resistent	T	G	G	A	G	C	C
PI603568	3.0	Mod.Resistent	T	G	G	A	G	C	C
BRS271RR	4.0	Susceptible	C	A	T	C	C	T	A
BRS320	3.8	Susceptible	C	A	T	C	C	T	A
BRSCharruaRR	5.0	Susceptible	C	A	T	C	C	T	A
BRSGiseleRR	3.6	Susceptible	C	A	T	C	C	T	A
BRSGO8060	4.2	Susceptible	C	A	T	C	C	T	A
BRSMacota	3.7	Susceptible	C	A	T	C	C	T	A

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
BRSMAPati	5.0	Susceptible	T	G	G	A	G	C	C
CD201	3.8	Susceptible	C	A	T	C	C	T	A
EMGOPA301	4.8	Susceptible	C	A	T	C	C	T	A
FT2001	4.2	Susceptible	C	A	T	C	C	T	A
FT-Cometa	3.7	Susceptible	C	A	T	C	C	T	A
GB881RR	4.3	Susceptible	C	A	T	C	C	T	A
IAC100	4.7	Susceptible	C	A	T	C	C	T	A
IAC11	4.0	Susceptible	C	A	T	C	C	T	A
Industrial	3.7	Susceptible	C	A	T	C	C	T	A
MG/BR48(GarimpoRCH)	4.5	Susceptible	C	A	T	C	C	T	A
M-SOY5826	4.7	Susceptible	Y	A	T	C	C	T	A
M-SOY5942	4.5	Susceptible	C	A	T	C	C	T	A
M-SOY8008	5.0	Susceptible	Y	A	T	C	C	T	A
M-SOY8248	4.2	Susceptible	C	A	T	C	C	T	A
M-SOY8411	4.8	Susceptible	T	G	G	A	G	C	C
M-SOY8527RR	3.7	Susceptible	C	A	T	C	C	T	A
M-SOY8585RR	3.8	Susceptible	C	A	T	C	C	T	A
M-SOY8849RR	3.8	Susceptible	C	A	T	C	C	T	A
M-SOY8870	5.0	Susceptible	C	A	T	C	C	Y	M
Msoy9056RR	4.4	Susceptible	C	A	T	C	C	T	A
MT/BR50(Parecis)	4.2	Susceptible	Y	A	T	C	S	Y	M
Pelicano	3.8	Susceptible	C	A	T	C	C	T	A
PI097038	4.8	Susceptible	C	A	T	C	C	T	A
PI371611	4.0	Susceptible	C	A	T	C	C	T	A
PI424511	4.2	Susceptible	C	A	T	C	C	T	A
PI438307	4.6	Susceptible	C	A	T	C	C	T	A
PI483253	3.7	Susceptible	C	A	T	C	C	T	A

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI547836	3.8	Susceptible	C	A	T	C	C	T	A
PI547856	4.2	Susceptible	C	A	T	C	C	T	A
PI547874	4.2	Susceptible	Y	R	K	M	S	Y	M
PI567001B	4.2	Susceptible	C	A	T	C	C	T	A
PI587959	4.8	Susceptible	C	A	T	C	G	T	A
PI594427C	3.7	Susceptible	C	A	T	C	G	T	A
PI594442B	4.3	Susceptible	C	A	T	C	C	T	A
PI603200	3.8	Susceptible	C	A	T	C	C	T	A
PI628889	3.7	Susceptible	C	A	T	C	C	T	A
PI634335	4.2	Susceptible	C	A	T	C	C	T	A
PI634903	3.8	Susceptible	C	A	T	C	C	T	A
PI646157	3.7	Susceptible	C	A	T	C	C	T	A
PI96118	4.2	Susceptible	C	A	T	C	C	T	A
PI97038	4.2	Susceptible	C	A	T	C	C	T	A
TMG121RR	4.7	Susceptible	C	A	T	C	C	T	A
UFV14	4.5	Susceptible	T	G	G	A	G	C	C
UFVS2001	3.8	Susceptible	C	A	T	C	C	T	A
BR16	3.8	Susceptible	T	G	G	A	G	C	C
BR27(Cariri)	4.2	Susceptible	T	G	G	A	G	C	C
BR36	4.5	Susceptible	T	G	G	A	G	C	C
BR40(Itiquira)	4.7	Susceptible	T	G	G	A	G	C	C
BRS132	4.2	Susceptible	T	G	G	A	G	C	C
BRS133	4.5	Susceptible	T	G	G	A	G	C	C
BRS185	5.0	Susceptible	T	G	G	A	G	C	C
BRS214	5.0	Susceptible	T	G	G	A	G	C	C
BRS217[Flora]	4.7	Susceptible	T	G	G	A	G	C	C
BRS218[Nina]	4.8	Susceptible	T	G	G	A	G	C	C

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
BRS242RR	4.8	Susceptible	T	G	G	A	G	C	C
BRS245RR	4.8	Susceptible	T	G	G	A	G	C	C
BRS263[Diferente]	4.7	Susceptible	T	G	G	A	G	C	C
BRS270RR	4.0	Susceptible	T	G	G	A	G	C	C
BRS314	4.7	Susceptible	T	G	G	A	G	C	C
BRSBalizaRR	5.0	Susceptible	T	G	G	A	G	C	C
BRSCandieiro	4.5	Susceptible	T	G	G	A	G	C	C
BRSCarla	4.2	Susceptible	T	G	G	A	G	C	C
BRSGO204[Goiania]	4.8	Susceptible	T	G	G	A	G	C	C
BRSGO8061	4.5	Susceptible	T	G	G	A	G	C	C
BRSGOCaiapônia	4.5	Susceptible	T	G	G	A	G	C	C
BRSGOEdeia	4.0	Susceptible	T	G	G	A	G	C	C
BRSGOIara	5.0	Susceptible	T	G	G	A	G	C	C
BRSGuapa	4.5	Susceptible	T	G	G	A	G	C	C
BRSInvernada	4.8	Susceptible	T	G	G	A	G	C	C
BRSMASeridoRCH	4.8	Susceptible	T	G	G	A	G	C	C
BRSMG250[Nobreza]	4.2	Susceptible	T	G	G	A	G	C	C
BRSMG750RR	3.8	Susceptible	T	G	G	A	G	C	C
BRSMG800A	4.3	Susceptible	T	G	G	A	G	C	C
BRSMGConfianca	5.0	Susceptible	T	G	G	A	G	C	C
BRSMGRenascença	4.5	Susceptible	T	G	G	A	G	C	C
BRSMTPintado	4.8	Susceptible	T	G	G	A	G	C	C
BRSMTUirapuru	5.0	Susceptible	T	G	G	A	G	C	C
BRSNovaSavana	4.5	Susceptible	T	G	G	A	G	C	C
BRSPetala	5.0	Susceptible	T	G	G	A	G	C	C
BRSSambaiba	4.8	Susceptible	T	G	G	A	G	C	C
BRSSinuelo	5.0	Susceptible	T	G	G	A	G	C	C

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
BRSTorena	4.2	Susceptible	T	G	G	A	G	C	C
CamposGerais	4.7	Susceptible	T	G	G	A	G	C	C
CD202	4.5	Susceptible	T	G	G	A	G	C	C
Davis	4.8	Susceptible	T	G	G	A	G	C	C
DM309	3.8	Susceptible	T	G	G	A	G	C	C
Dourados	3.8	Susceptible	T	G	G	A	G	C	C
Embrapa1(IAS5RC)	4.3	Susceptible	T	G	G	A	G	C	C
Embrapa20(DokoRC)	4.3	Susceptible	T	G	G	A	G	C	C
Embrapa48	4.3	Susceptible	T	G	G	A	G	C	C
Embrapa60	4.2	Susceptible	T	G	G	A	G	C	C
FMTTucunaré	5.0	Susceptible	T	G	G	A	G	C	C
FT10(Princesa)	4.5	Susceptible	T	G	G	A	G	C	C
FTEureka	4.2	Susceptible	T	G	G	A	G	C	C
IAC10	4.5	Susceptible	T	G	G	A	G	C	C
IAS5	4.5	Susceptible	T	G	G	A	G	C	C
Ivai	4.0	Susceptible	T	G	G	A	G	C	C
Mineira	5.0	Susceptible	T	G	G	A	G	C	C
M-SOY7501	4.7	Susceptible	T	G	G	A	G	C	C
M-SOY8199	4.3	Susceptible	T	G	G	A	G	C	C
M-SOY8925	4.2	Susceptible	T	G	G	A	G	C	C
MT/BR45(Paiaguas)	4.7	Susceptible	T	G	G	A	G	C	C
MT/BR53(Tucano)	5.0	Susceptible	T	G	G	A	G	C	C
OCEPAR10	3.8	Susceptible	T	G	G	A	G	C	C
OCEPAR12	5.0	Susceptible	T	G	G	A	G	C	C
OCEPAR8	5.0	Susceptible	T	G	G	A	G	C	C
OCEPAR9	4.7	Susceptible	T	G	G	A	G	C	C
Parana	4.8	Susceptible	T	G	G	A	G	C	C

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
Perola	4.7	Susceptible	T	G	G	A	G	C	C
PI054610-1	4.5	Susceptible	T	G	G	A	G	C	C
PI059845	4.5	Susceptible	T	G	G	A	G	C	C
PI091178	5.0	Susceptible	T	G	G	A	G	C	C
PI157428	4.4	Susceptible	T	G	G	A	G	C	C
PI171431	3.6	Susceptible	T	G	G	A	G	C	C
PI171454	5.0	Susceptible	T	G	G	A	G	C	C
PI171652	3.7	Susceptible	T	G	G	A	G	C	C
PI196170	4.5	Susceptible	T	G	G	A	G	C	C
PI200519	4.0	Susceptible	T	G	G	A	G	C	C
PI229325	4.8	Susceptible	T	G	G	A	G	C	C
PI248515	4.2	Susceptible	T	G	G	A	G	C	C
PI253652A	4.7	Susceptible	T	G	G	A	G	C	C
PI253654	3.7	Susceptible	T	G	G	A	G	C	C
PI304218	4.6	Susceptible	T	G	G	A	G	C	C
PI374189	5.0	Susceptible	T	G	G	A	G	C	C
PI398313	4.3	Susceptible	T	G	G	A	G	C	C
PI398666	5.0	Susceptible	T	G	G	A	G	C	C
PI398887	5.0	Susceptible	T	G	G	A	G	C	C
PI398965	4.7	Susceptible	T	G	G	A	G	C	C
PI399020	5.0	Susceptible	T	G	G	A	G	C	C
PI404199	5.0	Susceptible	T	R	K	M	G	C	C
PI417234	4.3	Susceptible	T	G	G	A	G	C	C
PI417580	4.7	Susceptible	T	G	G	A	G	C	C
PI423945	4.8	Susceptible	T	G	G	A	G	C	C
PI424492	4.0	Susceptible	T	G	G	A	G	C	C
PI424495	4.0	Susceptible	T	G	G	A	G	C	C

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI424499D	4.0	Susceptible	T	G	G	A	G	C	C
PI424504A	4.2	Susceptible	T	G	G	A	G	C	C
PI424505	4.8	Susceptible	T	G	G	A	G	C	C
PI424506	3.8	Susceptible	T	G	G	A	G	C	C
PI424522	3.8	Susceptible	T	G	G	A	G	C	C
PI424523B	3.8	Susceptible	T	G	G	A	G	C	C
PI424549A	4.5	Susceptible	T	G	G	A	G	C	C
PI424554	3.8	Susceptible	T	G	G	A	G	C	C
PI424555B	4.7	Susceptible	T	G	G	A	G	C	C
PI424558A	3.8	Susceptible	T	G	G	A	G	C	C
PI424574	3.6	Susceptible	T	R	G	A	G	C	C
PI430460A	4.8	Susceptible	T	G	G	A	G	C	C
PI430596	4.5	Susceptible	T	G	G	A	G	C	C
PI437350	4.3	Susceptible	T	G	G	A	G	C	C
PI437353	3.8	Susceptible	T	G	G	A	G	C	A
PI437423	3.7	Susceptible	T	G	G	A	G	C	C
PI437486	5.0	Susceptible	T	G	G	A	G	C	C
PI437580	4.0	Susceptible	T	G	G	A	G	C	C
PI437636B	3.7	Susceptible	T	G	G	A	G	C	C
PI437673	4.0	Susceptible	T	G	G	A	G	C	C
PI437749	4.6	Susceptible	T	G	G	A	G	C	C
PI437773	4.8	Susceptible	T	G	G	A	G	C	C
PI437829	3.7	Susceptible	T	G	G	A	G	C	C
PI437845D	4.0	Susceptible	T	G	G	A	G	C	C
PI438048B	5.0	Susceptible	T	G	G	A	G	C	C
PI438187	3.7	Susceptible	T	G	G	A	G	C	C
PI438255	4.5	Susceptible	T	G	G	A	G	C	C

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI438302A	5.0	Susceptible	T	G	G	A	G	C	C
PI438304B	5.0	Susceptible	T	G	G	A	G	C	C
PI438442A	4.5	Susceptible	T	G	G	A	G	C	C
PI438492	4.2	Susceptible	T	G	G	A	G	C	C
PI442005	5.0	Susceptible	T	G	G	A	G	C	C
PI442010	4.3	Susceptible	T	G	G	A	G	C	C
PI442012A	4.3	Susceptible	T	G	G	A	G	C	C
PI442018	4.5	Susceptible	T	G	G	A	G	C	C
PI442044	3.7	Susceptible	T	G	G	A	G	C	C
PI458175C	3.8	Susceptible	T	G	G	A	G	C	C
PI458199	3.7	Susceptible	T	G	G	A	G	C	C
PI458226	4.7	Susceptible	T	G	G	A	G	C	C
PI458234	3.8	Susceptible	T	G	G	A	G	C	C
PI458236A	4.2	Susceptible	T	G	G	A	G	C	C
PI458249	4.0	Susceptible	T	G	G	A	G	C	C
PI458298	5.0	Susceptible	T	G	G	A	G	C	C
PI458306A	4.5	Susceptible	T	G	G	A	G	C	C
PI467316	4.6	Susceptible	T	G	G	A	G	C	C
PI470226	3.8	Susceptible	T	G	G	A	G	C	C
PI495020	3.8	Susceptible	T	R	K	M	G	C	C
PI495832	3.7	Susceptible	T	G	G	A	G	C	C
PI506516	4.4	Susceptible	T	G	G	A	G	C	C
PI506525	4.2	Susceptible	T	G	G	A	G	C	C
PI506590E	3.8	Susceptible	T	G	G	A	G	C	C
PI506789	4.5	Susceptible	T	G	G	A	G	C	C
PI506819	5.0	Susceptible	T	G	G	A	G	C	C
PI506833	4.7	Susceptible	T	G	G	A	G	C	C

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI506892	4.2	Susceptible	T	G	G	A	G	Y	M
PI506935	4.0	Susceptible	T	G	G	A	G	C	C
PI506989	4.5	Susceptible	T	G	G	A	G	C	C
PI507082A	3.8	Susceptible	T	G	G	A	G	C	C
PI507097	4.8	Susceptible	T	G	G	A	G	C	C
PI507153	4.5	Susceptible	T	G	G	A	G	C	C
PI507158	4.0	Susceptible	T	G	G	A	G	C	C
PI507259	4.0	Susceptible	T	G	G	A	C	C	C
PI507286C	3.7	Susceptible	T	G	G	A	G	C	C
PI507316	4.3	Susceptible	T	G	G	A	G	C	C
PI507317	3.8	Susceptible	T	G	G	A	G	C	C
PI507407	3.8	Susceptible	T	G	G	A	G	C	C
PI507408	4.8	Susceptible	T	G	G	A	G	C	C
PI507430	4.0	Susceptible	T	G	G	A	G	C	C
PI507432	4.3	Susceptible	T	G	G	A	G	C	C
PI507449	4.0	Susceptible	T	G	G	A	G	C	C
PI507480	5.0	Susceptible	T	G	G	A	G	C	C
PI507492	4.0	Susceptible	T	G	G	A	G	C	C
PI507501	4.2	Susceptible	T	G	G	A	G	Y	M
PI508296A	4.3	Susceptible	T	G	G	A	G	C	C
PI508296G	4.5	Susceptible	T	G	G	A	G	C	C
PI509075	5.0	Susceptible	T	G	G	A	G	C	C
PI509079	4.2	Susceptible	T	G	G	A	G	C	C
PI518719	4.8	Susceptible	T	G	G	A	G	C	C
PI518720	4.8	Susceptible	T	G	G	A	G	C	C
PI522189	4.7	Susceptible	T	G	G	A	G	C	C
PI532455b	4.2	Susceptible	T	G	G	A	G	C	C

Genótipo	Média	Reação	SNP 1 30804961	SNP 2 30805500	SNP 3 30805508	SNP 4 30805499	SNP 5 30776090	SNP 6 30792409	SNP 7 30792474
PI548493	3.7	Susceptible	T	G	G	A	G	C	C
PI561337	4.3	Susceptible	T	G	G	A	G	C	C
PI561354	4.8	Susceptible	T	G	G	A	G	C	C
PI567668	4.0	Susceptible	T	G	G	A	G	C	C
PI578506	3.7	Susceptible	T	G	G	A	G	C	C
PI587608B	4.2	Susceptible	T	G	G	A	G	C	C
PI587618A	3.8	Susceptible	T	G	G	A	G	C	C
PI587991	4.4	Susceptible	T	G	G	A	G	C	C
PI593972	4.2	Susceptible	T	G	G	A	G	C	C
PI594403	4.0	Susceptible	T	G	G	A	G	C	C
PI594538A	4.2	Susceptible	T	G	G	A	G	C	C
PI594544	3.7	Susceptible	T	G	G	A	G	C	C
PI594596	4.2	Susceptible	T	G	G	A	G	C	C
PI594775	5.0	Susceptible	T	G	G	A	G	C	C
PI603694B	4.5	Susceptible	T	G	G	A	G	C	C
PI612758B	4.0	Susceptible	T	G	G	A	G	C	C
PI628827	4.7	Susceptible	T	G	G	A	G	C	C
PI644103	4.3	Susceptible	T	G	G	A	G	C	C
Planalto	4.3	Susceptible	T	G	G	A	G	C	C
TOKYO	4.0	Susceptible	T	G	G	A	G	C	C
Viçoja	4.5	Susceptible	T	G	G	A	G	C	C
PI398848	5.0	Susceptible	T	G	G	A	G	C	C

1001 Supplementary table 6 – **TaqMan analysis of accessions also containing GBS analysis**

Genotype	SNP1	SNP5	SNP7	5.13E-12 30804961	1.39E-11 30805500	1.47E-11 30805508	2.24E-11 30805499	3.37E-11 30792409	3.91E-11 30792474	8.72E-11 30776090
BR16	S	S	S	T	G	G	A	C	C	G
BR27(Cariri)	R	R	R	T	G	G	A	C	C	G
BR40(Itiquira)	R	R	R	T	G	G	A	C	C	G
BRS132	R	R	R	T	G	G	A	C	C	G
BRS133	S	S	S	T	G	G	A	C	C	G
BRS185	S	S	S	T	G	G	A	C	C	G
BRS239	R	R	R	T	G	G	A	C	C	G
BRSJuliana	R	R	R	C	A	T	C	T	A	C
BRSMG850GRR	R	R	R	C	A	T	C	T	A	C
CD219RR	R	R	R	C	A	T	C	T	A	C
CD224	R	R	R	C	A	T	C	T	A	C
CD225RR	R	R	R	T	G	G	A	C	C	G
CD233RR	R	R	R	C	A	T	C	T	A	C
FC031683	R	R	R	C	A	T	C	T	A	C
FT-Cometa	R	R	R	Y	A	T	C	T	A	C
Industrial	R	R	R	C	A	T	C	T	A	C
MG/BR48(GarimpoRCH)	R	R	R	C	A	T	C	T	A	C
M-SOY5942	R	R	R	C	A	T	C	T	A	C
PI158765	R	R	R	C	A	T	C	T	A	C
PI171427	R	R	R	Y	A	T	C	Y	M	S
PI230977	R	R	R	C	A	T	C	T	A	C
PI253651D	S	S	S	T	G	G	A	C	C	G
PI323556	R	R	R	C	A	T	C	T	A	C
PI371611	R	R	R	C	A	T	C	T	A	C

PI398666	S	S	S	T	G	G	A	C	C	G
PI398887	R	R	R	T	G	G	A	C	C	G
PI398965	S	S	S	T	G	G	A	C	C	G
PI399020	S	S	S	T	G	G	A	C	C	G
PI424588	S	S	S	T	G	G	A	C	C	G
PI437127B	R	R	R	C	A	T	C	T	A	C
PI437679	R	NA	R	C	A	T	C	T	A	C
PI438190	S	S	S	T	G	G	A	C	C	G
PI438307	R	R	R	C	A	T	C	T	A	C
PI438442A	S	S	S	T	G	G	A	C	C	G
PI438492	S	S	S	T	G	G	A	C	C	G
PI442005	S	S	S	T	G	G	A	C	C	G
PI442010	S	S	S	T	G	G	A	C	C	G
PI442012A	S	S	S	T	G	G	A	C	C	G
PI442018	S	S	S	T	G	G	A	C	C	G
PI442044	S	S	S	T	G	G	A	C	C	G
PI458175C	S	S	S	T	G	G	A	C	C	G
PI458199	S	S	S	T	G	G	A	C	C	G
PI458226	S	S	S	T	G	G	A	C	C	G
PI506862	R	R	R	C	A	T	C	T	A	C
PI567214B	R	R	R	C	A	T	C	T	A	C
PI567648C	R	R	R	C	A	T	C	Y	M	C
PI587911A	R	NA	R	C	A	T	C	T	A	G
PI588025	R	NA	R	C	A	T	C	T	A	C
PI607380	R	NA	R	C	A	T	C	T	A	C
PI608357	R	NA	R	C	A	T	C	T	A	C
PI634813	R	NA	R	C	A	T	C	T	A	C
PI89772	R	NA	R	C	A	T	C	T	A	C

UFV10(Uberaba)	S	S	S	T	G	G	A	C	C	G

Supplementary table 7 – **List of accessions submitted to GBS and assay analysis**

Accession	Phenotype	SNP1	SNP5	SNP7
BRS256RR	R	R	R	R
BRS316RR	R	R	R	R
BRS319RR	R	R	R	R
BRS399RR	R	R	R	R
BRS6970IPRO	R	R	R	R
BRS7380RR	R	R	R	R
BRS7980	R	R	R	R
BRSFavoritaRR	R	R	R	R
BRSGO7963	R	R	R	R
BRSGO8360	R	R	R	R
BRSGO8661RR	R	S	S	S
BRSGOLuziânia	R	R	R	R
BRSMGGarantia	R	R	R	R
CD208	R	R	R	R
BRS211	R	R	R	R
BRS230	S	S	S	S
BRS232	S	S	S	S
BRS285	S	R	R	R
BRS317	S	S	S	S
BRSMG68(Vencedora)	S	S	S	S
BRSMG790A	S	S	NA	S

DECLARATIONS

Acknowledgements

We would like to thanks the financial support of the Coordination for the Improvement of Higher Level for Education program (CAPES) and National Counsel of Technological and Scientific Development (CNPq) for all the support of this study and Dra. Ivani O. N. Lopes for invaluable assistance with statistical analyses and Claudemir O. de Lima assistance with root-knot nematode phenotyping.

Authors' contribution

J.C.A. – Performed plant sampling, DNA extraction, data analysis and writing manuscript. A.L.L.P. – DNA extraction, data generation and analysis and manuscript writing. W.P.D. – Plant seedling and phenotyping, F.B. – GBS pipeline support, manuscript review, F.C.M.G. and R.V.A. – conceived and planned the study and review the manuscript. All authors read and approved the final manuscript.

Conflicts of interest

"The authors declare that they have no conflicts of interests"

REFERENCE

1. United States Department of Agriculture – USDA. World Agricultural Production. November. 2017
2. Lima SOF, Correa VR, Nogueira SR, Santos PRR. Nematodes affecting soybean and sustainable practices for their management. In Soybean – The basis of yield, biomass and productivity. Intechopen. Available at https://www.intechopen.com/books/soybean-

1029 the-basis-of-yield-biomass-and-productivity/nematodes-affecting-soybean-and-
1030 sustainable-practices-for-their-management. 2017.
1031 3. Jones JT, Haegeman A, Danchin EGJ, Gaur HS, Helder J, Jones MGK, et al. Top 10
1032 plant-parasitic nematodes in molecular plant pathology. Molecular Plant Pathology.
1033 2013; doi: 10.1111/mpp.12057
1034 4. Doucet ME, Pinochet J. Occurrence of *Meloidogyne* spp. In Argentina. Journal of
1035 Nematology. 1992; 4S:765-770.
1036 5. Fourie H, McDonald AH, Loots GC. Plant-parasitic nematodes in field crops in South
1037 Africa. Nematology. 2001; doi: 10.1163/156854101753250773
1038 6. Garcia RM, Rich JR. Root-knot nematodes in north-central Florida soybean fields.
1039 Nematropica. 1985; 15:1.
1040 7. SILVA, J.F.V. Problemas fitossanitários da soja no Brasil, com ênfase em nematoides.
1041 In: Congresso brasileiro de nematologia, XXI, Maringá (PR). Resumos, p. 16-20. 1998.
1042 8. Silva JFV, Ferraz LCCB, Arias CA. Herança da resistência a *Meloidogyne javanica* em
1043 soja. Nematropica. 2001a; 31:2.
1044 9. Ferraz, L.C.C.B. As meloidoginoses da soja: passado, presente e futuro. In: Ferraz,
1045 L.C.C.B.; Asmus, G.L.; Carneiro, R.G.; Mazaffera, P.; Silva, J.F.V. Relações parasito-
1046 hospedeiro nas meloidoginoses da soja. Londrina: Embrapa Soja, 127p. 2001
1047 10. Silva JFV, Carneiro GES, Yorinori JT, Almeida AMR, Arias CAA, Kiihl RAS, et al.
1048 Desenvolvimento de linhagens de soja com resistência a patógenos. 2002; Boletim de
1049 pesquisa e desenvolvimento/Embrapa Soja.
1050 11. Dias WP, Garcia A, Silva JFV, Carneiro GES. Nematóides em soja: Identificação e
1051 controle. In: Circular Técnica76. 2010.
1052 12. Tamulonis JP, Luzzi BM, Hussey RS, Parrott WA, Boerma HR. DNA marker associated
1053 with resistance to javanses root-knot nematode in soybean. 1997a. 37:783-788.

13. Mienie CMS, Fourie H, Smit MA, Staden V, Botha FC. Identification of AFLP markers in soybean linked to resistance to *Meloidogyne javanica* and conversion to sequence characterized amplified regions (SCARS). Plant Growth Regulation. 2002; doi: 10.1023/A:1020585023976
14. Silva JFV, Ferraz LCBC, Arias CAA, Abdelnoor RV. Identificação de marcadores moleculares de microssatélites associados à resistência de genótipos de soja a *Meloidogyne javanica*. Nematologia Brasileira, 2001b; 25:1 79-83.
15. Fuganti R, Beneveti MA, Silva JFV, Arrias CAA, Marin SRR, Binneck E, Nepomuceno AL. Identificação de marcadores moleculares de microssatélites para seleção de genótipos de soja resistentes a *Meloidogyne javanica*. Nematologia Brasileira. 2004; 28:2, 125-130p.
16. 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; doi.org/10.1371/journal.pone.0054603
17. Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL. et al. Fingerpirnitng soybean germplasm and its utility in genomic research. G3: Genes | Genomes | Genetics. 2015; doi: 10.1534/g3.115.019000.
18. 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. Nature Biotechnology. 2015; doi:10.1038/nbt.3096
19. Lee Y, Jeong N, Kim JH, Lee K, Kim KH, Pirani A, Ha B. et al. Development, validation and genetic analysis of a large soybean SNP genotyping array. The plant Journal. 2015; doi: 10.1111/tpj.12755

20. Hwang E, Song Q, Jia G, Specht JE, Hyten DL, Costa J. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics. 2014; doi: 10.1186/1471-2164-15-1
21. Korte A, Farlow A. The advantages and limitarions of trait analysis with GWAS: a review. Plant Methods. 2013; doi: 10.1186/1746-4811-9-29.
22. Jannink J. Dynamics of long-term genomic selection. Genetics Selectino Evolution. 2010; https://doi.org/10.1186/1297-9686-42-35
23. Voung TD, Sonah H, Meinhardt CG, Deshmukh R, Kadam S, Nelson RL, et al. Genetic architecture of cyst nematode resistance reealed by genome-wide association study in soybean. BMC Genomics. 2015; doi: 10.1186/s12864-015-1811-y
24. Autshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science, 2008. doi:10.1126/science.115640
25. Brachi B, Morris GP, Borevitz OJ. Genome-wide association studies in plants: the missing heritability is in in the field. Genome Biology, 2011. doi.org/10.1186/gb-2011-12-10-232
26. Bastien M, Sonah H, Belzile F. Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. 2014. doi: 10.3835/plantgenome2013.10.0030
27. Fang C, Ma Y, Wu S, Liu Z, Wang Z, et al. Genome-wide association studies dissect the geneti networks underlying agronomical traits in soybean. 2017. DOI 10.1186/s13059-017-1289-9.
28. Zhang Z, Ersoz E, Lai C, Todhunter R, Tiwari, HK. Mixed linear model approach adapterd for genome-wide association studies. 2010. doi:10.1038/ng.546

29. Sonah H, O'Donoughue L, Cober E, Rajcan I, Belzile F. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnology Journal. 2014; doi: 10.1111/pbi.12249
30. Mamidi S, Lee RK, Goos JR, McClean PE. Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). Plos One. 2014; doi: 10.1371/journal.pone.0107469
31. Passianoto ALL, Sonah H, Dias WP, Marcelino-Guimaraes FC, Belzile F, Abdelnoor RV. Genome wide association for resistance to the southern root-knot nematode (*Meloidogyne incognita*) in soybean. Molecular Breeding. 2017; https://doi.org/10.1007/s11032-017-0744-3
32. Kang, Y.J.; Kim, K.H.; Shim, S.; Yoon, M.Y.; Sun, S.; Kim, M.Y.; Van, K.; Lee, S. Genome wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biology. 2012; doi: 10.1186/1471-2229-12-139
33. Tamulonis JP, Luzzi BM, Hussey RS, Parrott WA, Boerma HR. RFLP mapping of resistance to southern root-knot nematode in soybean. Crop Sci. 1997b. 37:6, 1903-1909p.
34. Kearsey MJ and Pooni HS (1998) The Genetical Analysis of Quantitative Traits. Chapman and Hall, London, 381 pp.
35. Eshed Y and Zamir D (1996) Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics 143:1807- 1817.
36. Falconer DS (1989) Introduction to Quantitative Genetics. John Wiley and Sons, New York, 340 pp.
37. McHale L, Tan X, Koehl P, Michilmore RW. Plant NBS-LRR proteins: adaptable guards. Genome Biology. 2006; doi: 10.1186/gb-2006-7-4-212

38. Ooijen Gernen Van, Van den Burg Harrold, Cornelissen BJC, Takken FLW. Structure and function of resistance proteins in solanaceous plants. Annual Review Phytopathology. 2007, doi: 10.1146/annurev.phyto.45.062806.094430

39. Tameling WI, Elzinga SD, Darmin PS, Vossen JH, Takken FL, Haring MA, Cornelissen BJ (2002) The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. Plant Cell 14 2929–2939

40. Ithal N, Recknor J, Nettleton D, Hearne L, Maier T, Baum TJ. Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. The American Phytopathological Society. 2007a; doi: 10.1094/MPMI-20-3-0293

41. Ithal, N.; Recknor, J.; Nettleton, D.; Hearne, L.; Maier, T.; Baum, T.J.; Mitchum, M.G. Developmental transcript profiling of cyst nematode feeding cells in soybean roots. 2007b; doi: 10.1094/MPMI-20-5-0510

42. Kandoth PK, Ithal N, Recknor J, Maier T, Nettleton D, Baum TJ, Mitchum MG. The soybean Rhg1 locus for resistance to the soybean cyst nematode Heterodera glycines regulates the expression of a large number of stress and defense related genes in degeneration feeding cells. Plant physiology. 2011; doi: 10.1104/pp.110.167536

43. Beneventi MA, da Silva Jr OB, Sá MEL, Firmino AAP, de Amorim RMS, et al. Transcription profile of soybean-root-knot nematode interaction reveals a key role of phythormones in the resistance reaction. BMC Genomics. 2013. doi:10.1186/1471-2164-14-322.

44. Kobe B., Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature. 1995; 374:183–186.

45. Leister R.T., Katagiri F. A resistance gene product of the nucleotide binding site – leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo. Plant J. 2000;22:345–354.
46. Vos P, Simons G, Jesse T, Wijbrandi J, Heinen L, Hoger R. The tomato Mi-1 gene confers resistance to both root-knot nematodes and potato aphids. Nature Biotechnology. 1998; doi: 10.1038/4350
47. Tameling, W.I.; Elzinga, S.D.; Darmin, P.S.; Vossen, J.H.; Takken, F.L.; Haring, M.A.; Cornelissen, B.J. The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. Plant Cell. 2002; doi: 10.1105/tpc.005793
48. Abney TS, Crochet WD. The uniform soybean tests: Northern region 2004. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1065&context=ars. Accessed 06 Apr 2018.
49. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE6, 2011, https://doi.org/10.1371/journal.pone.0019379.
50. Scheet P.; Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. The American Journal of Human Genetics. 2006; doi: 10.1086/502802
51. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, 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. Fly. 2012. doi: 10.4161/fly.19695
52. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ. et al GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012; doi: 10.1093/bioinformatics/bts444

53. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; doi:10.1093/bioinformatics/bth457

54. Gabriel, S.B.; Schaffner, S.F.; Nguyen, H.; Moore, J.H.; Roy, J.; Blumenstile, B.; Higgins, J.; DeFelice, M.; Lochner, A.; Faggart, M.; Liu-Cordero, S.N.; Rotimi, C.; Adeyemo, A.; Cooper, R.; Ward, R.; Lander, E.R.; Daly, M.J.; Altshuler, D. The structure of haplotype blocks in the human genome. Science. 2002, doi: 10.1126/science.1069424