

ANA CAROLINA OLIVEIRA MESQUITA

MAPEAMENTO POR ASSOCIAÇÃO GENÔMICA AMPLA PARA  
IDENTIFICAÇÃO DE RESISTÊNCIA AO MOFO BRANCO EM SOJA

Dissertação apresentada à Universidade Federal de  
Uberlândia, como parte das exigências do Programa de  
Pós-graduação em Agronomia – Mestrado, área de  
concentração: Fitopatologia, para obtenção do título de  
“Mestre”.

Orientador

Prof. Dr. Fernando Cezar Juliatti

UBERLÂNDIA  
MINAS GERAIS – BRASIL  
2017

ANA CAROLINA OLIVEIRA MESQUITA

MAPEAMENTO POR ASSOCIAÇÃO GENÔMICA AMPLA PARA  
IDENTIFICAÇÃO DE RESISTÊNCIA AO MOFO BRANCO EM SOJA

Dissertação apresentada à Universidade Federal de  
Uberlândia, como parte das exigências do Programa de  
Pós-graduação em Agronomia – Mestrado, área de  
concentração: Fitopatologia, para obtenção do título de  
“Mestre”.

APROVADA em 10 de fevereiro de 2017.

Prof. Dra. Juliana Araújo Santos Martins

IFTM

Prof. PhD. Steven John Clough

USDA-ARS

Dra. Tâmara Prado de Moraes

UFU

Prof. Dr. Fernando Cezar Juliatti  
ICIAG-UFU  
(Orientador)

UBERLÂNDIA  
MINAS GERAIS – BRASIL  
2017

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema de Bibliotecas da UFU, MG, Brasil.

---

M582m      Mesquita, Ana Carolina Oliveira, 1988  
2017      Mapeamento por associação genômica ampla para identificação de  
resistência ao mofo branco em soja / Ana Carolina Oliveira Mesquita. -  
2017.  
68 f. : il.

Orientador: Fernando Cezar Juliatti.  
Dissertação (mestrado) - Universidade Federal de Uberlândia,  
Programa de Pós-Graduação em Agronomia.  
Inclui bibliografia.

1. Agronomia - Teses. 2. Soja - Resistência a doenças e pragas -  
Teses. 3. Sclerotinia sclerotiorum - Teses. 4. Marcadores moleculares -  
Teses. I. Juliatti, Fernando Cezar. II. Universidade Federal de  
Uberlândia. Programa de Pós-Graduação em Agronomia. III. Título.

---

CDU: 631

## **AGRADECIMENTOS**

A Deus, por todas as oportunidades e bênçãos em minha vida.

À minha mãe Maria Helena Mesquita de Oliveira, meu pai João Batista de Oliveira e minha irmã Fernanda Oliveira de Mesquita pelo exemplo, pelo apoio, compreensão e força em todos os momentos.

Ao professor Fernando Cezar Juliatti que me acolheu novamente como sua orientada e possibilitou que a minha experiência como aluna de mestrado fosse a mais proveitosa possível.

Ao Steven John Clough pela amizade, pela oportunidade de poder fazer parte dessa pesquisa, pelo apoio em minha visita à University of Illinois e pela dedicação do seu tempo, determinantes para o sucesso deste trabalho.

À Adriana de Andrade Figueiró pela amizade, pelos conselhos e pelo apoio, pois sem seu incentivo eu não teria realizado esta pesquisa.

Ao Professor Patrick Brown pela ajuda e esforço para que esta pesquisa pudesse ter um curso proveitoso e produtivo.

À Letícia Marise Lamana, pela amizade, carinho, colaboração sincera e preciosa ajuda que tornaram minha visita a University of Illinois mais fácil.

A todos os alunos de mestrado que ajudaram nas análises de bioinformática, especialmente Xing Wu e Wei Wei, pois sem a ajuda de vocês esta pesquisa não teria sido concretizada.

À professora Maria Imaculada Zuccini, amiga que fiz no laboratório ao longo das análises de bioinformática, agradeço pelo companheirismo.

Aos colegas do LAMIP – Laboratório de Micologia e Proteção de Plantas, pela cordial acolhida, amizade e ajuda.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo fundamental apoio financeiro para o cumprimento desta pesquisa.

A todos os demais que não foram citados, mas que de forma direta ou indireta contribuíram para a realização deste trabalho.

Aos membros da banca, por dedicarem o seu tempo visando a melhoria desse trabalho.

Meus sinceros agradecimentos!



## SUMÁRIO

LISTA DE FIGURAS.....	i
LISTA DE TABELAS.....	ii
RESUMO.....	iii
ABSTRACT.....	iv
1 INTRODUÇÃO.....	1
2 REVISÃO BIBLIOGRÁFICA.....	4
2.1 Aspectos gerais sobre a cultura da soja.....	4
2.2 Mofo branco.....	4
2.3 Métodos de controle do mofo branco.....	7
2.4 Resistência à <i>S. sclerotiorum</i> na soja.....	8
2.5 Mapeamento genético da soja e identificação de QTLs relacionados com a resistência ao mofo branco.....	9
2.6 Mapeamento por associação genômica.....	12
2.7 Softwares utilizados para os estudos de associação ampla do genoma.....	16
3 MATERIAL E MÉTODOS.....	19
3.1 Fenotipagem da resistência à <i>S. sclerotiorum</i> em genótipos de soja.....	19
3.2 Extração de DNA total.....	20
3.3 Construção da biblioteca GBS e sequenciamento.....	21
3.4 Processamento dos dados das leituras das sequências do Illumina e seleção dos SNPs.....	23
3.4.1 GBSSeqToTagDBPlugin.....	24
3.4.2 TagExportToFastqPlugin.....	24
3.4.3 Bowtie 2.....	24
3.4.4 SAMToGBSdbPlugin.....	24
3.4.5 DiscoverySNPCallerPluginV2.....	25
3.4.6 SNPQualityProfilerPlugin.....	25
3.4.7 ProductionSNPCallerPluginV2.....	25
3.4.8 Beagle 4.1.....	25
3.5 Limpeza dos dados genotípicos.....	25
3.6 Análise de associação genômica.....	26
3.7 Análise de desequilíbrio de ligação.....	26
3.8 Identificação de relações entre SNPs e funções biológicas.....	27
4 RESULTADOS.....	28

4.1 Avaliação da resistência à <i>S. sclerotiorum</i> nos genótipos de soja .....	28
4.2 Teste de enzimas de restrição .....	33
4.3 Distribuição e controle de qualidade dos marcadores SNPs.....	35
4.4 Análise da estrutura da população .....	36
4.5 Mapeamento por associação genômica.....	37
4.6 Análise de desequilíbrio de ligação .....	39
4.7 Identificação de relações entre funções biológicas e SNPs .....	41
5 DISCUSSÃO .....	45
6 CONCLUSÕES .....	49
REFERÊNCIAS.....	50

## LISTA DE FIGURAS

<b>Figura 1.</b> Ciclo de vida do fungo <i>Sclerotinia sclerotiorum</i> , agente causal do mofo branco na soja, com destaque no papel das sementes como fonte de inóculo inicial e meio de disseminação.....	7
<b>Figura 2.</b> Posição física dos QTLs previamente identificados associados com resistência à <i>S. sclerotiorum</i> em soja.....	10
<b>Figura 3.</b> Associação dos SNPs com a doença .....	14
<b>Figura 4.</b> Modelo dos fragmentos ao final do término do preparo da biblioteca.....	16
<b>Figura 5.</b> Etapas da inoculação artificial com <i>S. sclerotiorum</i> .....	20
<b>Figura 6.</b> Produto de ligação de um fragmento de DNA genômico (preto) contendo um sítio de restrição HindIII e um sítio de restrição MseI .....	22
<b>Figura 7.</b> Pipeline para filtragem dos SNPs por qualidade.....	23
<b>Figura 8.</b> Distribuição fenotípica do tamanho de lesão causada por <i>S. sclerotiorum</i> em 324 genótipos de soja após inoculação artificial .....	28
<b>Figura 9.</b> Eletroferograma obtido da análise do teste dos pares de enzimas de restrição no Bioanalyzer 2100.....	34
<b>Figura 10.</b> Distribuição dos marcadores SNPs nos 20 cromossomos da soja. ....	35
<b>Figura 11.</b> Gráfico dos valores <i>Eigenvalue</i> e os componentes principais. ....	36
<b>Figura 12.</b> Gráfico de dispersão bidimensional (PC2 x PC3) da análise do componente principal. ....	37
<b>Figura 13.</b> Variância total para resistência à <i>S. sclerotiorum</i> . ....	37
<b>Figura 14.</b> Gráfico quantil-quantil (Q-Q plot), da distribuição dos valores observados e esperados para o estudo de associação genômica para resistência à <i>S. sclerotiorum</i> . ....	38
<b>Figura 15.</b> Gráfico de Manhattan com SNPs significativos que estão associados à resistência a <i>S. sclerotiorum</i> .....	39
<b>Figura 16.</b> A. Representação gráfica dos níveis de Desequilíbrio de ligação (LD) para comparações par a par entre sítios polimórficos que apresentam resultados em todas amostras analisadas, obtido pelo software Haploview para resistência à <i>S. sclerotiorum</i> . B. Detalhe da localização do SNP S01_33164666. C. Detalhe da localização dos SNPs S01_34787519 e S01_35073769. D. Detalhe da localização dos SNPs S01_36048095, S01_36044127, S01_36038838 e S01_36006734. ....	40
<b>Figura 17.</b> Decomposição do desequilíbrio de ligação intracromossômico entre par de marcadores do cromossomo 1 da distância física.....	41

## LISTA DE TABELAS

<b>Tabela 1.</b> Comprimento da lesão da doença nos 324 genótipos resultado da inoculação artificial.....	29
<b>Tabela 2.</b> Tamanho médio de fragmentos do teste de enzimas para o protocolo de preparo da biblioteca. ....	33
<b>Tabela 3.</b> Melhores marcadores significativos associados à resistência à <i>S. sclerotiorum</i> em soja considerando <i>p</i> -values ajustados por FDR <0,05.....	39
<b>Tabela 4.</b> Genes candidatos dentro do bloco de desequilíbrio de ligação ou região genômica vinculada de polimorfismos de nucleotídeo único (SNPs) mais fortemente associados com a resistência à <i>S. sclerotiorum</i> em soja.....	42

## RESUMO

MESQUITA, ANA CAROLINA OLIVEIRA. **Mapeamento por associação genômica ampla para identificação de resistência ao mofo branco em soja**. 2017. 68 p. Dissertação (Mestrado em Agronomia/Fitopatologia) – Universidade Federal de Uberlândia, Uberlândia.<sup>1</sup>

O mofo branco, causado pelo fungo *Sclerotinia sclerotiorum*, é uma doença de grande importância no Brasil. A seleção para sua resistência seria muito facilitada com a utilização de marcadores moleculares e com a descoberta de genes envolvidos com a doença. A soja não tem resistência completa, mas tem diferentes níveis de resistência parciais que podem ser bem-sucedidos em retardar a progressão da doença e permitir que as plantas se recuperem, especialmente se o ambiente favorece mais a planta e menos o patógeno. Marcadores moleculares do tipo SNP foram identificados a partir de genótipos de soja, com a finalidade de associação de marcadores moleculares com resistência ao fungo *S. sclerotiorum* no genoma amplo pela técnica GBS. Foram coletados genótipos de soja reportados como parcialmente resistentes à doença, assim como genótipos de três diferentes programas de melhoramento no Brasil. As plantas foram inoculadas com o isolado Jataí no estágio V2 utilizando o método *Straw test* e avaliou-se o avanço da doença nas hastes das plantas aos quatro dias após a inoculação. Antes da inoculação, extraiu-se e fragmentou-se o DNA das plantas por digestão de restrição utilizando HindIII e MseI, e converteu-se em bibliotecas de genotipagem por sequenciamento utilizando *barcodes* antes do sequenciamento. Após a limpeza das sequências, as leituras foram alinhadas permitindo a identificação de 50.000 polimorfismos de nucleotídeos simples. As amostras foram, ainda, processadas para redução dos genótipos ou SNPs altamente heterozigóticos resultando em 324 genótipos e 49.000 SNPs para usar no software GAPIT para estudo de associação ampla do genoma para SNPs associados com diferentes níveis de resistência ao fungo do mofo branco. Foram identificados oito SNPs com *p*-value ajustados por FDR <0,05. Sete SNPs estavam no cromossomo 1, e quatro destes estavam em desequilíbrio de ligação, significando que estes representavam quatro QTLs. O último SNP estava localizado no cromossomo 12. Os SNPs fornecem possíveis marcadores que poderiam ser utilizados na seleção de soja com resistência ao fungo *S. sclerotiorum*.

Palavras-chave: *screening*, genotipagem por sequenciamento, SNP, marcador, QTL, *Sclerotinia sclerotiorum*

---

<sup>1</sup>Orientador: Fernando Cezar Juliatti – UFU

## ABSTRACT

MESQUITA, ANA CAROLINA OLIVEIRA. **Genome wide association mapping for identification of resistance to white mold in soybean.** 2017. 68 p. Uberlândia: UFU, 2017. 68 p. Dissertation (Master Program Agronomy/Plant Pathology) – Federal University of Uberlândia, Uberlândia.<sup>1</sup>

White mold is a disease of great importance in Brazil. Selection for white mold resistance would be greatly facilitated by the use of molecular markers and the discovery of genes involved with the disease. Soybean do not have complete resistance, but instead have differing levels of partial resistance that could be successful in slowing disease progression enough to allow the plants to recover, especially if the environment favors more the plant and less the pathogen. Molecular markers of the SNP type were identified from soybean genotypes, with the purpose of associating molecular markers with the resistance to fungus *S. sclerotiorum* in the wide genome by the GBS technique. We collected soybean genotypes reported as partially resistant to disease, as well as many from three different breeding programs in Brazil. We inoculated the plants with the isolate Jataí at the V2 stage using the straw test method and scored disease progress down the plant stems at four days post inoculation. Prior to inoculation, the plant DNA was extracted and fragmented by restriction digestion using HindIII and MseI and converted to barcoded genotyping-by-sequencing libraries prior to Illumina sequencing. After sequence cleaning, reads were aligned allowing the identification of 50,000 single nucleotide polymorphisms. Samples were further cleaned to reduce highly heterozygous genotypes or SNPs generating 324 genotypes and 49,000 SNPs to use in GAPIT for a genome wide association study for SNPs associated with varying levels of SSR resistance. We identified eight SNPs at FDR p-value <0.05. Seven of the SNPs were on chromosome 1, and four of these were in linkage disequilibrium, meaning that these seven SNPs represented four QTL. The last SNP was on chromosome 12. The SNPs provide possible markers that could be used in selection of soybean with enhanced resistance to *Sclerotinia sclerotiorum*.

Keywords: screening, genotype-by-sequencing, SNP, marker, QTL, *Sclerotinia sclerotiorum*

---

<sup>1</sup>Major Professor: Fernando Cezar Juliatti– UFU

## 1 INTRODUÇÃO

A soja [*Glycine max* (L.) Merrill] é considerada uma das mais importantes *commodities* produzidas e comercializadas no mundo. Seus grãos apresentam alto teor proteico, tornando-a alimento para a dieta de seres humanos e animais. Além disso, apresenta considerável teor de óleo, que pode ser utilizado na indústria alimentícia e na fabricação de biocombustíveis.

Com o aumento do potencial produtivo da cultura da soja e seu cultivo em diversos ambientes, as doenças assumiram relevante importância dentre os fatores restritivos à produção desta oleaginosa, com destaque para o mofo branco, uma das doenças mais antigas. Devido ao aumento considerável nas lavouras brasileiras a partir de 2008, representa uma das principais doenças da cultura da soja (MEYER et al., 2014).

O mofo branco, causado pelo fungo *Sclerotinia sclerotiorum* (Lib.) de Bary, é uma das doenças que possuiu rápido crescimento de incidência na cultura da soja nas últimas safras, provocando reduções de rendimento de até 70% (MEYER et al., 2013). O progresso da doença é extremamente dependente de condições ambientais favoráveis e seus danos manifestam-se com maior severidade em safras com clima chuvoso e temperatura amena (MEYER; CAMPOS, 2009; MEYER, 2009; JULIATTI; JULIATTI, 2010; MACHADO; CASSETARI NETO, 2010).

Esta doença pode ser manejada por métodos de controle integrado, incluindo a resistência genética. A resistência do hospedeiro, quando disponível, é o método de controle mais confiável e mais econômico para o agricultor e que menos afeta o meio ambiente. Entretanto, sua utilização no controle do mofo branco é limitada pelo desconhecimento do germoplasma brasileiro e por não haver ainda resistência completa da planta frente ao fungo (JULIO, 2016). Por isso, são necessários estudos na busca de genótipos com resistência ao mofo branco.

Para a detecção de resistência genética pode-se mapear o genoma para a identificação de genes candidatos, que conferem resistência à planta, pertencentes a algum caráter quantitativo. O sequenciamento completo do genoma poderia identificar a relação genótipo-fenótipo da planta. Logo, o mapeamento dos genes é imprescindível para elucidar a determinação de diversos fenótipos tanto relacionados às doenças quanto aos traços benéficos.

Os caracteres quantitativos de interesse agrônomo são, em sua maioria, resultantes da ação conjunta de vários genes, o que implica em forte influência ambiental na expressão fenotípica. A denominação QTL (*Quantitative Trait Loci*) é utilizada para nomear as regiões cromossômicas que contêm genes (ou *locus*) que controlam esses caracteres quantitativos. Assim, como ferramenta-suporte aos programas de melhoramento genético, as técnicas baseadas na utilização de marcadores moleculares surgiram para agregar informações à etapa de seleção de genótipos superiores, permitindo estudar com maior facilidade as regiões que influenciam a expressão desses caracteres e respectivos *locus*.

Em uma população, a quantidade de fenótipos dos indivíduos é elevada, então existem fenótipos distintos que são influenciados por vários genes comuns, apesar de alguns fenótipos serem determinados por um ou poucos genes. Assim, buscar genes responsáveis por determinados fenótipos, ou até mesmo variações nos genes que produzam variações em determinada característica, é uma tarefa de grande relevância.

Durante as duas últimas décadas, muitos estudos de mapeamentos de QTLs – mapeamento de ligação - foram conduzidos e muitos destes foram realizados com populações biparentais e alguns com populações de tamanhos limitados (KIM; DIERS, 2000; ARAHANA et al., 2001; HAN et al., 2007, GUO et al., 2008; VUONG; DIERS; HARTMAN, 2008; LI et al., 2009; SEBASTIAN et al., 2010; HUYNH et al., 2010; ZHAO et al., 2015). Com isso, a variação genética explorada é limitada.

Além disso, quando os estudos de mapeamento são realizados, é necessária a validação dos resultados em diferentes populações de melhoramento. O mapeamento por associação, por outro lado, não é um experimento controlado, mas sim um ensaio natural. Os dados genotípicos e fenotípicos são coletados de uma população em que parentesco não é controlado pelo pesquisador, e correlações entre marcadores genéticos e fenótipos são procurados dentro desta população.

Nesse contexto, o estudo de associação do genoma amplo, do inglês *Genome Wide Association Studies* (GWAS), pode ser uma estratégia promissora para a identificação de QTLs para genes de interesse. A estratégia da associação ampla do genoma é suficiente para a cobertura do genoma em que alelos funcionais possam estar em desequilíbrio de ligação (LD) com, pelo menos um, marcador. A cobertura do genoma é obtida pelo uso da técnica de genotipagem por sequenciamento (GBS).

Essa estratégia é uma forma relativamente nova de identificar genes envolvidos em doenças. Os pesquisadores buscam pequenas variações ou polimorfismos no



genoma, que ocorrem mais frequentemente em indivíduos portadores de uma determinada doença dos que não a portam. Habitualmente, neste tipo de análise, são utilizados polimorfismos de nucleotídeo simples (SNPs), que são a forma de variação mais frequente no genoma, que acontece quando um dos nucleotídeos - adenina, citosina, guanina ou timina - difere entre indivíduos em um determinado local cromossômico (*locus*). Os SNPs ajudam a estabelecer localizações, no genoma, de algum fator genético que contribui com a variabilidade. Estes pontos de referência são conhecidos como marcadores genéticos e facilitam a navegação no genoma (OLAZAR, 2013).

A procura por associação entre um certo traço de interesse ou fenótipo e alelos ou genótipos em um *locus* genético com o objetivo de determinar se um alelo particular associa-se com certa doença na população como um todo é o intuito de pesquisadores dessa área. Esta associação pode indicar uma relação causal direta que permitiria investigar os mecanismos e vias metabólicas (*pathways*) na progressão da doença, ou pode indicar uma relação indireta que pode ajudar a localizar a variação causal. Com isso, os pesquisadores podem usar os resultados dos estudos para desenvolver melhores estratégias para detectar, tratar e prevenir doenças.

Os SNPs, associados com a resistência, podem acelerar o processo de melhoramento através de seleção assistida por marcadores ou podem ser incorporados em estratégias de seleção genômica. Além disso, SNPs significativos podem auxiliar no conhecimento sobre a função biológica do polimorfismo e como se relaciona com a resistência.

Neste estudo, marcadores moleculares do tipo SNP foram identificados a partir de genótipos de soja, com finalidade de associação de marcadores moleculares com a resistência ao fungo *Sclerotinia sclerotiorum* no genoma amplo pela técnica GBS.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Aspectos gerais sobre a cultura da soja

A soja [*Glycine max* (L.) Merrill] é uma planta originária da China considerada a mais importante leguminosa cultivada no mundo (GOMES, 1990), e é a principal cultura plantada no Brasil (BRASIL, 2014). É uma planta anual, frequentemente de porte ereto, ramificada esparsamente, possui crescimento arbustivo, com folhas trifolioladas alternas ou compostas e prófilos simples. As flores podem apresentar coloração branca, púrpura diluída ou roxa. As vagens são retas ou ligeiramente curvadas, contendo cada uma, de uma a quatro sementes (GOMES, 1990; ACQUAAH, 2007).

Geneticamente, a planta de soja possui um genoma de 1,1-1,15 bilhões de pares de base com predição de 46.430 genes codificadores de proteínas ao longo de 20 cromossomos (SCHMUTZ et al., 2010). O conjunto completo de cromossomos dessa planta apresenta número de cromossomos de  $2n = 40$  (HYMOWTZ et al., 1997).

A soja é uma das principais culturas de verão da produção agrícola brasileira, representando a segunda maior área mundial da oleaginosa, com cerca de 33,24 milhões de hectares cultivados na safra 2015/2016, com produção de 95,4 milhões de toneladas (CONAB, 2016a). Na safra 2016/2017, há uma projeção de crescimento de até 6,7 a 9% na produção, podendo atingir de 101,8 a 104 milhões de toneladas (CONAB, 2016b). Diversos fatores limitam o rendimento da cultura, incluindo diversas doenças de difícil controle. Cerca de 40 doenças causadas por fungos, bactérias, nematoides e vírus afetam a cultura da soja no país (EMBRAPA, 2011).

### 2.2 Mofo branco

O fungo necrotrófico *Sclerotinia sclerotiorum* (Lib.) de Bary é o agente etiológico do mofo branco que acomete plantas de soja e de outras culturas (PURDY, 1979). Algumas outras importantes culturas afetadas são: girassol, colza, amendoim, ervilha, alface, feijão, algodão, tomate, batata e algumas outras hortaliças (BOLAND; HALL, 1994; BOLTON; THOMMA; LEITE, 2005). A doença causada pelo patógeno recebe diferentes denominações em outros hospedeiros, entre elas: podridão branca da haste, mofo branco, podridão da cabeça, podridão aquosa e podridão da haste (PURDY, 1979).

O mofo branco é uma das doenças mais antigas e sem expressão observada em soja no passado. Com a expansão do cultivo da soja no Brasil, tornou-se uma das principais doenças na atualidade, com o aumento da incidência a partir de 2008 (MEYER et al., 2014). Este fungo possui ampla distribuição geográfica, com grande diversidade de hospedeiros, composta por mais de 400 espécies de plantas (PURDY, 1979; BOLAND; HALL, 1994). É considerado um dos patógenos fúngicos mais importantes no mundo e está distribuído em todas as regiões produtoras, sejam elas temperadas, subtropicais ou tropicais (LEITE, 2005). No Brasil, a doença encontra-se disseminada em regiões do Sul e Sudeste que apresentam clima mais ameno e em áreas elevadas do Centro-Oeste e Nordeste, onde ocorrem temperaturas noturnas mais baixas (CASSETARI NETO; MACHADO, SILVA, 2010; JULIATTI; JULIATTI, 2010).

A patologia é designada como um dos principais problemas da cultura da soja e afeta diretamente a economia dos principais países produtores, por acarretar severos prejuízos no rendimento de importantes culturas agrônômicas (BOLAND; HALL, 1994; CUNHA, 2010). A incidência desse patógeno acarreta danos expressivos na produção e qualidade dos grãos, podendo atingir perdas em torno de 70% (MEYER et al., 2013). O monocultivo e a ausência de rotação com espécies não-hospedeiras são as principais causas do aumento da área infestada no Brasil (BALARDIN, 2002; CUNHA, 2010).

Uma medida fundamental para prevenir a ocorrência de doenças causadas por *S. sclerotiorum* é reduzir ao máximo os períodos de alta umidade e baixa temperatura na cultura (LEITE, 2005). As perdas de produtividade de grãos causadas pelo fungo variam de acordo com a incidência do patógeno, região, clima, cultura e variedade cultivada. O fungo encontra condições ótimas de crescimento em temperaturas amenas (15 a 25 °C) e em ambientes com alta umidade. Em temperaturas abaixo de 0° C e acima de 32° C, o fungo apresenta baixa atividade (PURDY, 1979). O patógeno libera ácido oxálico e múltiplos fatores de virulência, aumentando sua capacidade de ser um patógeno eficaz (WEI; CLOUGH, 2016).

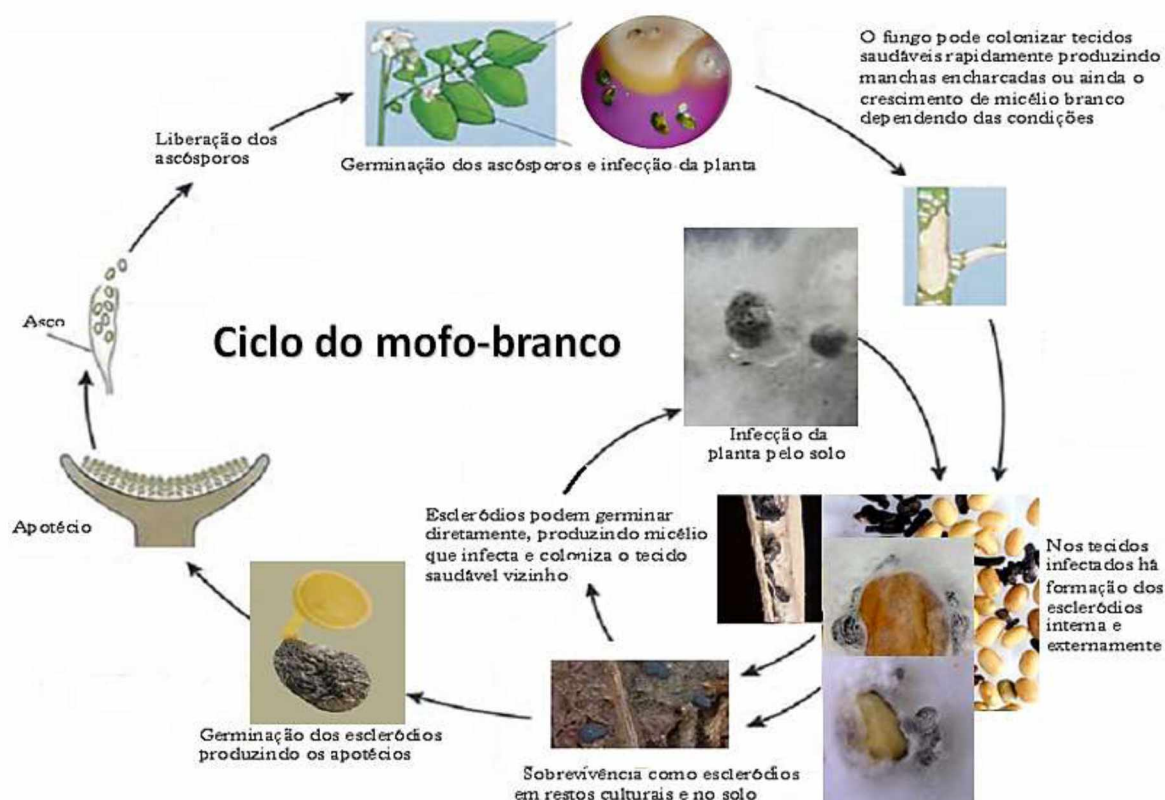
O fungo *S. sclerotiorum* é capaz de infectar qualquer parte da planta, iniciando com maior frequência a partir do surgimento das inflorescências e das axilas das folhas e dos ramos laterais (CAMARGO, 2002). Os sintomas ocorrem, geralmente, no terço médio das plantas, atingindo haste principal, pecíolos, folhas e vagens (LEITE, 2005), que variam de acordo com o hospedeiro, parte afetada do hospedeiro e com as condições ambientais. O sintoma mais comum e característico da doença é a formação de um aglomerado de micélios brancos com aspecto cotonoso sobre a área afetada e o

posterior desenvolvimento dos escleródios, que são estruturas de resistência do fungo a condições adversas do ambiente (CUNHA, 2010).

As plantas doentes podem ser identificadas pelas folhas superiores murchas, de coloração verde-acinzentada, passando à cor marrom, permanecendo presas à planta. Esses sintomas ocorrem entre 7 e 14 dias após o início do desenvolvimento de lesão na haste. As lesões circundam completamente a haste, impedindo o transporte de água, nutrientes e fotoassimilados. O tecido doente na haste, após o desenvolvimento de escleródios, torna-se esbranquiçado e se rompe em tiras com facilidade (BONATO, 2000).

A fase mais vulnerável da soja ao desenvolvimento do patógeno abrange os estádios de floração plena (R2) ao início de formação das vagens (R3/R4). Nesse período, o dossel da planta cobre as entrelinhas de semeadura propiciando condições favoráveis de umidade e temperatura para os escleródios, presentes no solo, germinarem e produzirem os apotécios. Os apotécios, por sua vez, produzem os ascósporos, que são liberados no ar. Os ascósporos, ao encontrarem tecidos senescentes, germinam e iniciam o processo de infecção da planta (CUNHA, 2010).

Para que os escleródios germinem são necessárias condições ambientais favoráveis, como temperatura de 4 a 20° C, alta umidade e estar a uma profundidade do solo de até 5 cm (BARDIN; HUANG, 2001; CLARKSON et al., 2003). Os escleródios podem germinar de duas formas: pela germinação carpogênica, em que ocorre produção de apotécios (órgãos de reprodução sexuada) originando ascósporos. Quando estes estão maduros, serão ejetados para o meio infectando a planta acima da linha do solo; e pela germinação micelogênica, que promove a produção de hifas a partir de microporos, presentes nos escleródios, que atacam os tecidos da planta (BARDIN; HUANG, 2001). A Figura 1 resume de forma esquemática o ciclo de vida do mofo branco na soja, incluindo os processos de infecção via germinação carpogênica e micelogênica.



**Figura 1.** Ciclo de vida do fungo *Sclerotinia sclerotiorum*, agente causal do mofo branco na soja, com destaque no papel das sementes como fonte de inóculo inicial e meio de disseminação. Fonte: JULIATTI et al., 2015.

A germinação carpogênica é considerada, por muitos autores, como a principal responsável pelas epidemias em campo (KARL et al., 1997) produzindo um ou muitos apotécios que liberam milhares de ascósporos que iniciam a doença na parte aérea da planta (BARDIN; HUANG, 2001).

### 2.3 Métodos de controle do mofo branco

O controle da doença está centrado, principalmente, em práticas agronômicas, como o uso de cultivares resistentes ao acamamento, que não propiciam microclima extremamente favorável ao patógeno (LEITE, 2005), maior espaçamento entre linhas, redução da densidade de plantas, plantio direto, rotação com culturas não-hospedeiras e aplicação de fungicidas. Contudo, em condições ambientais favoráveis ao desenvolvimento do patógeno, tais práticas têm-se mostrado pouco eficientes. Desta forma, a alternativa economicamente viável para o controle é o desenvolvimento de genótipos resistentes ao patógeno (CUNHA, 2010).

Quando disponível, a resistência do hospedeiro é o método de controle mais confiável e mais econômico para o agricultor, e o que menos afeta o meio ambiente. Entretanto, sua utilização no controle da podridão branca da haste é limitada pela falta de conhecimento do potencial genético do germoplasma brasileiro (SAGATA, 2010).

#### **2.4 Resistência à *S. sclerotiorum* na soja**

Entre as estratégias de controle de doenças em plantas, o uso da resistência genética é considerado o mais eficiente, ambientalmente seguro (PRIESTLEY; BAYLES, 1988) e mais econômico, por isso, de melhor aceitação pelo agricultor (EMBRAPA, 2013).

A resistência genética a doenças pode ser definida como a habilidade do hospedeiro em paralisar ou reduzir o desenvolvimento do patógeno (PARLEVLIET, 1975). Com essa capacidade, as plantas têm desenvolvido vários mecanismos de defesa para sua proteção contra doenças (DE WIT, 1995). Essa habilidade natural de resistência das plantas pode auxiliar os melhoristas e pesquisadores na seleção ou inserção de mecanismos de defesa (MATIELLO; BARBIERI; CARVALHO, 1997).

O tipo de resistência mais amplamente utilizado pelos melhoristas é a resistência qualitativa (raça-específica), governada por um ou poucos genes de grande efeito sobre o fenótipo, que conferem resistência completa. No entanto, devido a grande variabilidade genética e elevada capacidade de mutação dos patógenos, este tipo de resistência é rapidamente superada (NELSON, 1972).

Alternativas têm sido propostas para diminuir as rápidas superações de resistência e aumentar a durabilidade da resistência nos genótipos modernos, sendo que a resistência parcial configura-se como uma estratégia interessante. Este tipo de resistência é considerada incompleta, não específica as raças do patógeno, de efeito quantitativo e, na maioria dos casos, possui grande número de genes de pequeno efeito atuando conjuntamente na expressão do caráter (ZAMBONATO, 2011).

A resistência parcial restringe o desenvolvimento do patógeno pela redução da infecção, patogênese e esporulação, quantitativamente (PARLEVLIET, 1975). A resistência parcial pode ser resultado da resistência fisiológica ou de características da planta (mecanismos de escape), que propiciam condições menos favoráveis ao desenvolvimento do patógeno (KIM; DIERS, 2000).

A resistência parcial é um caráter quantitativo, ou seja, controlada por múltiplos genes e é uma situação entre a resistência completa e a suscetibilidade completa. Ao

contrário de outros caracteres quantitativos, não só a planta e o ambiente influenciam a variabilidade fenotípica final, mas também a alta variabilidade do patógeno, o que dificulta a seleção de plantas com base no fenótipo (THOMÉ et al., 1999).

Esta resistência é vista como uma opção favorável à exploração da variabilidade genética, através da combinação de genitores de interesse, com o objetivo de gerar e lançar no mercado cultivares com resistências mais duráveis estendendo sua utilização comercial (ZAMBONATO, 2011).

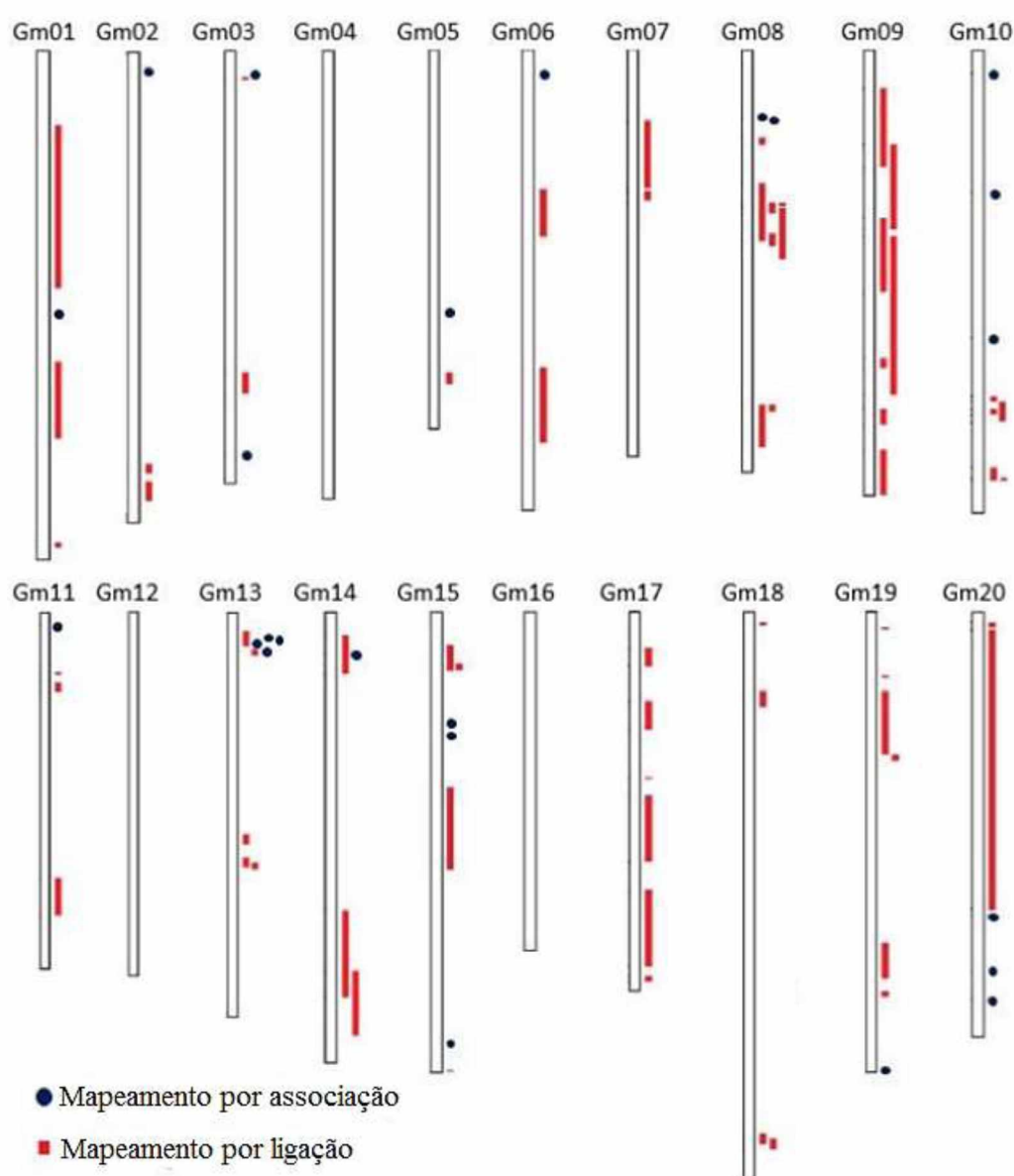
Cultivares de soja geneticamente resistentes ao mofo branco ainda não foram identificados com resistência completa à *S. sclerotiorum* (JULIO, 2016), embora haja diferenças na suscetibilidade de genótipos testados e grandes variações na resistência ao mofo branco (KIM; SNELLER; DIERS, 1999; GARCIA; JULIATTI, 2012; JULIATTI et al., 2013; JULIATTI et al., 2014). Por isso, o desenvolvimento e uso de cultivares que exibem resistência parcial são um dos meios mais eficazes de gerenciar a doença causada pelo fungo *S. sclerotiorum* (KURLE et al., 2001) e é um objetivo importante de programas de melhoramento direcionados para as áreas de cultivo de soja (BASTIEN; SONAH; BELZILE, 2014).

## **2.5 Mapeamento genético da soja e identificação de QTLs relacionados com a resistência ao mofo branco**

Os caracteres quantitativos de interesse agrônomo, como produção de grãos, altura da planta, teor de proteína e resistência a doenças (TANKSLEY, 1993), são, em sua maioria, resultantes da ação conjunta de vários genes, ou seja, pela ação cumulativa de um conjunto de genes, o que implica em forte influência ambiental na expressão fenotípica. A denominação *locus* de caracteres quantitativos, do inglês *Quantitative Trait Loci* (QTL), é utilizada para nomear as regiões cromossômicas que contêm genes (ou *locus*) que controlam esses caracteres quantitativos (FALCONER; MACKAY, 1996).

O melhoramento genético visando caracteres quantitativos, como a resistência parcial a doenças, pode ser dificultado, pois os efeitos de genes individuais que controlam estes caracteres podem não ser facilmente identificados. Para facilitar esta identificação, o uso de marcadores moleculares pode auxiliar e aumentar a eficiência da seleção, identificando segmentos de cromossomos que controlam caracteres quantitativos (DUDLEY, 1993).

A complexidade da resistência ao mofo branco tem levado muitos pesquisadores a utilizarem a análise de QTLs, identificando genes que conferem resistência ao mofo branco para fins de localização de marcadores moleculares eficientes para serem usados em seleção assistida. Diversos estudos têm relatado QTLs, através do mapeamento de ligação, na reação à *S. sclerotiorum* em soja (KIM; DIERS, 2000; ARAHANA et al., 2001; GUO et al., 2008; VUONG; DIERS; HARTMAN, 2008; LI et al., 2009; SEBASTIAN et al., 2010; HUYNH et al., 2010; ZHAO et al., 2015). Estes estudos de ligação revelaram a natureza quantitativa da resistência à *S. sclerotiorum*, identificando em todo o genoma da soja muitos QTLs de efeito menor (Figura 2).



**Figura 2.** Posição física dos QTLs previamente identificados associados com resistência à *S. sclerotiorum* em soja. Gm: cromossomo. Adaptado de MOELLERS (2016).



A distribuição dos marcadores moleculares ao longo do genoma permite a detecção e localização de QTLs (LARA et al., 2014). Estes marcadores podem ser usados para marcar os alelos favoráveis destes genes em populações segregantes de dicotiledóneas e, em seguida, realizar a seleção para a resistência mais eficientemente (JOHNSTON et al., 2015).

A identificação e localização de QTLs que controlam características de importância econômica podem ser realizadas com as ferramentas de mapas genéticos (TANKSLEY, 1993), que são consideradas uma aplicação de grande impacto da tecnologia de desenvolvimento de número ilimitado de marcadores moleculares. O uso dos mapas possibilita a cobertura e análise completa de genomas, a decomposição de características genéticas complexas nos seus componentes Mendelianos, a localização de regiões genômicas que controlam características de importância e a disponibilidade das informações para o uso em programas de melhoramento (FALEIRO, 2011a).

A metodologia proposta para mapeamento de QTLs baseia-se na hipótese de que, se a segregação de um gene marcador pode ser usada para detectar e estimar o efeito de um poligene ligado e, se genes estiverem espalhados ao longo do genoma de uma espécie, será possível mapear e caracterizar todos os poligenes que afetam um caráter (SCHUSTER; CRUZ, 2008). Experimentos de sequenciamento e mapeamento genético mostram que genes de reação a patógenos estão geralmente organizados em agrupamentos muito próximos chamados *clusters* ou blocos gênicos (GEFFROY et al., 1999).

O mapeamento genético dos blocos gênicos (QTLs) detecta as variações genéticas com estratégias diferentes a depender do delineamento do experimento em estudo. Os mapeamentos de ligação e de associação compartilham uma estratégia comum que explora a capacidade de recombinação para fragmentar o genoma em fragmentos que podem ser correlacionados com a variação fenotípica. A principal diferença entre os dois métodos é o controle do pesquisador sobre a recombinação (MYLES et al., 2009).

O mapeamento de ligação é um experimento altamente controlado em que os indivíduos são cruzados para gerar uma população na qual parentesco é conhecido. Nas plantas, estes são geralmente cruzamentos biparentais. Assim, o pesquisador gera um sistema fechado e utiliza pequeno número de marcadores genéticos para inferir as localizações dos pontos de interrupção de recombinação. Com os dados de genótipos de todo o genoma, o pesquisador pode, em seguida, determinar se um fragmento

cromossomal entre os dois pontos de interrupção específicos está associado a um fenótipo (MYLES et al., 2009).

O mapeamento de associação, por outro lado, não é um estudo controlado, mas sim um experimento natural. As informações genóticas e fenóticas são coletadas de uma população em que parentesco não é controlado pelo pesquisador, e correlações entre marcadores genéticos e fenótipos são investigadas dentro desta população. Esta concepção do sistema aberto proporciona maior resolução de mapeamento, quando comparada com o sistema fechado de cruzamentos controlados, mas é mais difícil para inferir quando e onde a recombinação ocorreu. Além disso, o parentesco não controlado entre os indivíduos pode resultar em falsos resultados de associação (MYLES et al., 2009).

Com a geração de um grande número de marcadores dispersos por todo o genoma de um indivíduo, independente da estratégia, alguns destes marcadores, além de estarem muito pertos, estarão em desequilíbrio de ligação (LD) com os QTLs (HASTBACKA et al., 1994) e podem estar relacionados com o fenótipo. O conceito de desequilíbrio de ligação refere-se à associação não aleatória entre alelos de diferentes *loci*, ou seja, entre dois genes ou entre um QTL e um *locus* marcador. É a correlação entre os polimorfismos que são causados pela história compartilhada de mutações e recombinações. As estatísticas utilizadas para a mensuração de LD são as  $r^2$  e  $D'$ , que são baseadas nas frequências alélicas dos alelos dos *loci* em estudo (XU, 2010).

## **2.6 Mapeamento por associação genômica**

Associação genética é um estudo multidisciplinar, que envolve componentes da genética, estatística, biologia molecular e bioinformática, formando a base para a identificação das regiões genômicas selecionadas, avaliadas e associadas por correlação com a variação (ORAGUZIE et al., 2007).

Nos últimos anos, por meio das tecnologias de sequenciamento de nova geração, com redução drástica no custo da genotipagem e, por consequência, aumento da produção de grande número de dados genômicos, considerando as limitações da utilização de mapas de ligação e utilização de QTLs na seleção assistida por marcadores, o mapeamento por associação, uma nova ferramenta de seleção foi evidenciada (BERNARDO; YU, 2007; FALEIRO, 2011a).

Essa nova realidade surgiu como mudança de paradigmas, modificando, sobretudo, a escala das análises. Além disso, surgiu a abordagem que viabiliza a análise

de genomas completos, sem necessidade de identificação ou mapeamento de QTL por ligação e sem restrição no número de QTL considerado na expressão do caráter quantitativo (BERNARDO; YU, 2007).

Nesse sentido, estudos de associação genômica, segundo Meuwissen e colaboradores (2001), são análises baseadas na avaliação de grande número de marcadores amplamente distribuídos ao longo do genoma. Pesquisadores procuram por associação entre um certo traço de interesse ou fenótipo e alelos ou genótipos em um *locus* genético, com o objetivo de determinar se um alelo particular associa-se com certa doença na população como um todo (OLAZAR, 2013).

Os marcadores moleculares detectados, distribuídos pelo genoma, revelam os polimorfismos ou variâncias genéticas. Os marcadores são características de DNA que diferenciam dois ou mais indivíduos e são herdados geneticamente (MILACH, 1998). Os polimorfismos de nucleotídeo em sítio único (SNPs - *Single Nucleotide Polymorphism*) são um tipo de marcador molecular muito utilizado em estudos de mapeamentos genéticos (FALEIRO, 2011b).

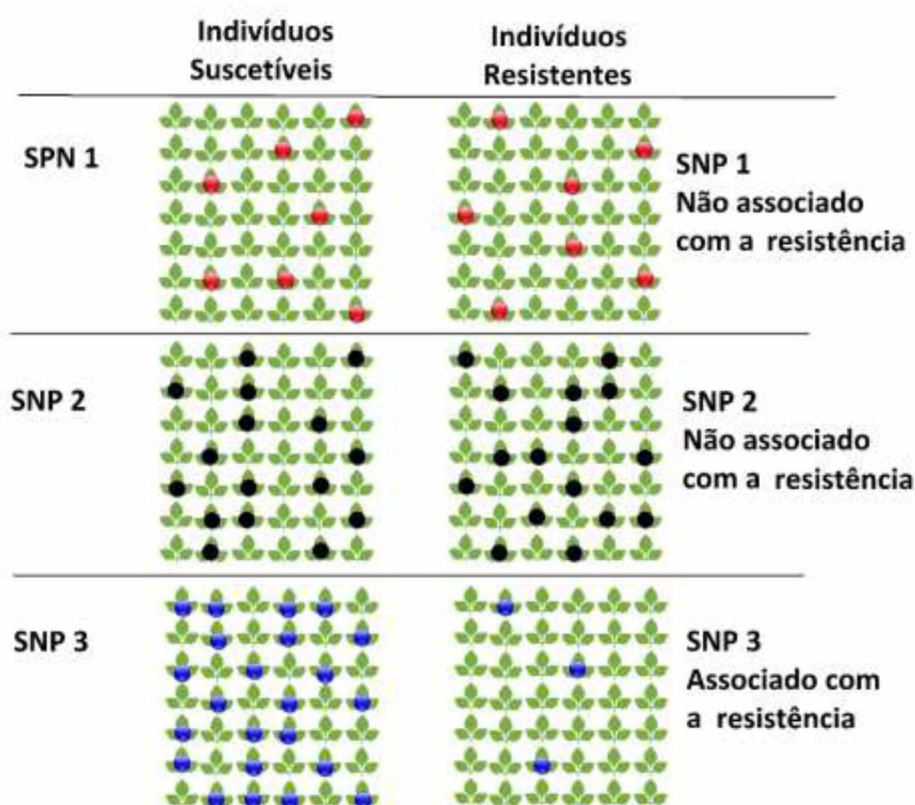
Os SNPs são a forma mais comum de variação no genoma e são, em sua maioria, dialélicos, sendo o alelo menos frequente (MAF) presente em pelo menos 1% da população testada (BROOKES, 1999). Em soja, estima-se a frequência de um SNP a cada 2.038 pares de bases em regiões codificadoras, e um SNP a cada 191 pares de bases em regiões não codificadoras (ZHU et al., 2003), sendo que a presença de um SNP em uma sequência que codifica uma proteína pode induzir mudanças em aminoácidos, resultando em alterações funcionais da proteína (BAN et al., 2010).

Em estudos de Associação Genômica Ampla (GWAS - *Genome Wide Association Study*), o objetivo é a detecção de SNPs com alto desequilíbrio de ligação (atributos altamente correlacionados) com o verdadeiro *locus* da mutação causal (QTLs), podendo o marcador ser integrante do QTL ou ser a própria mutação causadora do fenótipo (OLIVEIRA, 2015). Com o desenvolvimento de marcadores do tipo SNP, com novas tecnologias de sequenciamento e com o aumento da capacidade de análises computacionais, a técnica tornou-se mais atrativa, revelando possibilidades de aplicação prática com importantes benefícios para o melhoramento genético vegetal (BERNARDO; YU, 2007).

Os estudos de GWA, aliados às novas tecnologias de prospecção automatizada de SNP, permitem identificar variações no DNA, que estão correlacionadas com uma dada característica por meio de análises estatísticas, apontando para a localização dos

genes envolvidos com a expressão do fenótipo de interesse e auxiliando na compreensão dos mecanismos biológicos e fisiológicos desses genes (KU; LOY, 2010). A abordagem GWAS representa uma grande evolução em relação à metodologia clássica de análise de genes candidatos, uma vez que não é necessário conhecimento prévio profundo dos mecanismos fisiológicos envolvidos na manifestação da característica investigada (TAYLOR et al., 1998; WANG et al., 2005).

Este tipo de mapeamento de associação envolve a busca de correlações genótipo-fenótipo em indivíduos não aparentados e, muitas vezes, é mais rápido e rentável do que o mapeamento de ligação tradicional (MYLES et al., 2009). Se um tipo de variante (um alelo) é mais frequente em plantas com a resistência, a variante diz estar associada com a doença, ou seja, se determinadas variações genéticas forem significativamente mais frequentes em indivíduos com a resistência, em comparação com plantas suscetíveis, as variações são consideradas associadas à doença, podendo ser observadas na Figura 3 (PEARSON; MANOLIO, 2008).



**Figura 3.** Associação dos SNPs com a doença. Fonte: o autor (2017).

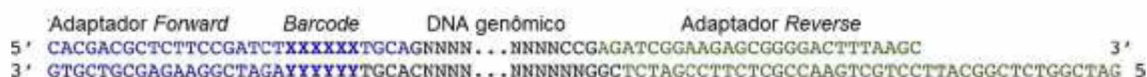
Na cultura da soja, GWAS têm sido utilizados para identificar *loci* associados a características agronômicas, como estresse abiótico e resistência a doenças. Esses estudos de associação do genoma são ferramentas úteis para detectar e identificar características de interesse, incluindo aqueles associados com resistência a doenças em soja (CHANG et al., 2016).

Para aumentar a acurácia das análises, a técnica de genotipagem por sequenciamento (GBS) foi desenvolvida para reduzir a complexidade em grandes genomas, mas com um método simples e robusto (ELSHIRE et al., 2011). A técnica foi desenvolvida como uma ferramenta para estudos de associação genômica em uma variedade de espécies, incluindo aquelas com genomas complexos. A ferramenta GBS utiliza enzimas de restrição para a redução da complexidade do genoma, seguido de sequenciamento multiplex para produzir dados de polimorfismo de alta qualidade a um baixo custo por amostra (POLAND et al., 2012).

O desenvolvimento de bibliotecas de GBS requer pouca quantidade de DNA, evita corte aleatório e seleção de tamanho, e é completada em apenas duas etapas em placas, seguido por amplificação por reação em cadeia da polimerase (PCR) do *pool* da biblioteca. A abordagem original GBS utiliza uma única enzima de restrição para seleção genômica entre locais de restrição (ELSHIRE et al., 2011).

Para garantir melhor qualidade e otimização do protocolo da biblioteca, este foi modificado utilizando um sistema de duas enzimas de restrição, sendo uma enzima com “corte raro” e outra com “corte comum”. Quando combinadas com adaptadores para o sítio de restrição comum, os fragmentos amplificados possuem o adaptador *forward* com o *barcode* e o adaptador *reverse* comum. O *barcode* trata-se de uma sequência de DNA específica que tem por finalidade identificar cada genótipo em ensaios multiplex (POLAND et al., 2012).

A biblioteca produzida é uniforme em que todos os fragmentos possuem: adaptador *forward* - DNA genômico - adaptador *reverse* (Figura 4). Este tipo de construção de biblioteca simplifica muito a quantificação da biblioteca antes do sequenciamento. A abordagem em duas enzimas pode gerar uma redução da complexidade adequada e uniforme (POLAND et al., 2012).



**Figura 4.** Modelo dos fragmentos ao final do término do preparo da biblioteca. Fonte: POLAND et al. (2012).

Com as novas tecnologias de GBS e métodos de bioinformática, tem-se a capacidade de descobrir, genotipar e imputar dados genômicos de populações em qualquer espécie (ANDOLFATTO et al.; 2011; XIE et al.; 2010).

## 2.7 Softwares utilizados para os estudos de associação ampla do genoma

Com as novas tecnologias de genotipagem, juntamente como a abordagem GBS, há aumento do número de SNPs disponibilizados para análise e, por isso, o tempo computacional é cada vez mais um problema na condução desses estudos (AZEVEDO NETO, 2013). Em estudos de GWA, tornou-se essencial o uso de softwares especializados para lidar com a grande dimensionalidade de dados para realização de grande número de testes para investigação de associações (OLLAZAR, 2013).

Grandes avanços têm sido obtidos na área da bioinformática, possibilitando diferentes tipos de análises genéticas cada vez mais acuradas, precisas e rápidas, disponibilizando subsídios para diferentes aplicações práticas de marcadores moleculares em estudos genéticos (FALEIRO, 2011c). Para analisar os dados obtidos do sequenciamento, diversos softwares com programas de bioinformática com *pipelines* para GBS podem ser utilizados com o objetivo de processamento de dados de sequências na pesquisa de SNPs em diferentes genótipos (TORKAMANEH; LAROCHE; BELZILE, 2016). Os softwares que conduzem estudos de associação do genoma completo precisam de metodologias que maximizam a potência estatística, fornecendo exatidão elevada da predição e funcionando de maneira computacionalmente eficiente (LIPKA et al., 2012).

Os programas utilizados para as análises podem analisar os dados com diversos modelos estatísticos, considerando, ou não, efeitos que interferem no resultado das associações. Muitas metodologias podem ser utilizadas para a estimação dos efeitos dos marcadores, sendo que, basicamente, a diferença está na escolha assumida para os efeitos dos marcadores. Independente das pressuposições, todos os métodos baseiam-se na regressão dos fenótipos sobre os marcadores disponíveis, usando um modelo linear de regressão (DE LOS CAMPOS et al., 2013).



Dentre os métodos utilizados, o Modelo Linear Misto (MLM) é um método que realiza o controle de falsos positivos em GWAS. Este modelo incorpora simultaneamente estrutura populacional e o parentesco desigual entre indivíduos (YU et al., 2005). Já o Modelo Linear Misto Comprimido (CMLM) aumenta a potência estatística, ou seja, a capacidade de detectar os marcadores, e reduz drasticamente o tempo computacional em grandes amostras de indivíduos pela clusterização em grupos (ZHANG et al., 2010).

O software TASSEL (*Trait Analysis by Association, Evolution and Linkage*) é um método de Modelo Linear Misto (MLM) que oferece uma variedade de manipulações de dados e resultados obtidos. Geneticistas de plantas, animais e humanos, interessados na realização de análise de associação, encontram diversas utilidades do programa para suas pesquisas (BRADBURY et al., 2007). Uma plataforma modificada foi direcionada para estudos de genotipagem por sequenciamento (GBS), o TASSEL-GBS (GLAUBITZ et al., 2014).

A ferramenta GAPIT (*Genome Association and Prediction Integrated Tool*), um software complementar, foi lançada ao público em 2011 pelo Dr. Zhiwu Zhang. O algoritmo utiliza a plataforma analítica R com a instalação de pacotes específicos implementado para estudos de associação do genoma completo (GWAS) e previsão genômica (ou seleção), incluindo Modelo Linear Misto Comprimido (CMLM) (LIPKA et al., 2012).

O pacote GAPIT implementa estratégias adicionais para lidar com grandes conjuntos de dados genotípicos, podendo estes conter acima de 10.000 indivíduos e 1 milhão de SNPs com tempo computacional mínimo, subdividindo os dados genotípicos em vários arquivos menores. O software disponibiliza resultados detalhados em uma série de tabelas e gráficos concisos para interpretar os resultados (LIPKA et al., 2012).

A qualidade dos resultados obtidos pelos softwares depende do controle de qualidade realizada anterior à análise de GWAS, para evitar que problemas na genotipagem ocasionem associações falso-positivas (PANOUTSOPOULOU; ZEGGINI, 2009). Dentre os critérios utilizados no controle de qualidade das amostras, pode-se verificar a eficiência de genotipagem, a heterozigotidade média, se há amostras que estão em duplicidade, a estratificação da população e a subestrutura da população. Com relação aos SNPs, considera-se a eficiência da genotipagem, os desvios com relação ao equilíbrio de Hardy-Weinberg, frequência do menor alelo (MAF) e SNPs idênticos (TURNER et al., 2011).

A metodologia GWAS, é capaz de identificar regiões cromossômicas em plantas de soja associadas aos fenótipos: diferença nos tamanhos de lesão necrótica causada pelo fungo *S. sclerotiorum*. Estas regiões podem estar relacionadas aos genes e/ ou a vias metabólicas envolvidas na resistência parcial da planta ao fungo.



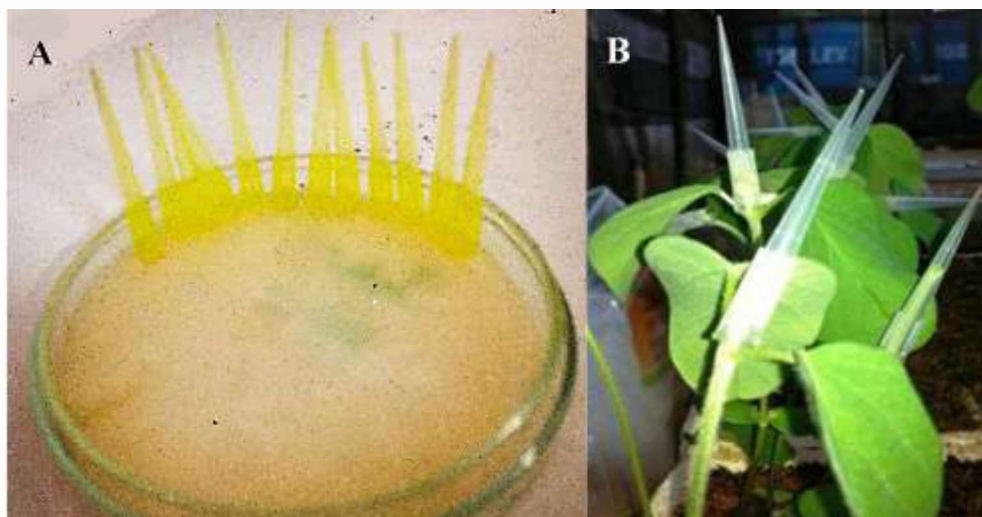
### 3 MATERIAL E MÉTODOS

Os experimentos foram conduzidos no Laboratório de Micologia e Proteção de Plantas (LAMIP), Laboratório de Biologia Molecular do Instituto de Ciências Agrárias, na casa de vegetação da Universidade Federal de Uberlândia (UFU) e nos laboratórios do Instituto Carl R. Woese de Biologia Molecular da Universidade de Illinois Urbana-Champaign, EUA.

#### 3.1 Fenotipagem da resistência à *S. sclerotiorum* em genótipos de soja

Sementes de 352 genótipos de soja das empresas FT Sementes e Tropical Melhoramento & Genética - TMG e do programa de melhoramento de soja da UFU (LAGER-UFU) foram cultivadas em copos plásticos contendo substrato orgânico de crescimento vegetal Bioflora<sup>®</sup> (Bioflora LTDA-ME, Prata, MG, Brasil), baseado em pinho e outras fibras naturais, enriquecido com minerais e nutrientes, em casa de vegetação a aproximadamente 25-30 °C sob ciclo de luz natural em Uberlândia-MG, de dezembro de 2015 a fevereiro de 2016 (fotoperíodo de, aproximadamente, 12 horas). Os experimentos foram conduzidos em delineamento experimental inteiramente casualizado com nove plantas por genótipo.

As brotações imaturas foram removidas no estágio V2 e armazenadas em ultra freezer (- 80° C) para a preservação do tecido até a realização das análises. As plantas foram levadas para o laboratório, dentro de 2-4 dias após a coleta das brotações, para inoculação artificial com *S. sclerotiorum* pelo método *Straw test* (PETZOLDT; DICKSON, 1996), em que as hastes foram seccionadas logo abaixo do segundo nó trifoliado e foram mantidas em contato com discos de micélio oriundos de placas de Petri contendo o fungo (isolado Jataí) (Figura 5A). Os discos de micélio contendo *S. sclerotiorum* foram colocados internamente em ponteiras e estas inseridas sobre o ápice da haste seccionada (Figura 5B).



**Figura 5.** Etapas da inoculação artificial com *S. sclerotiorum*. **A.** Micélio de *S. sclerotiorum*; **B.** Método de inoculação com ponteira (*Straw test*). Fonte: o autor (2017).

Os escleródios de *S. sclerotiorum* isolado Jataí foram inoculados em placas de Petri contendo meio de BDA (Batata Dextrose Ágar). As placas foram mantidas em estufa incubadora BOD a 23° C por 20 dias e fotoperíodo de 12 horas. Após o vigésimo dia na BOD, as placas com crescimento homogêneo foram utilizadas para inoculação das plantas de soja que apresentavam o estágio V2 de crescimento.

As ponteiros foram mantidas nas plantas durante as avaliações, e as plantas foram incubadas em câmara de crescimento à temperatura de 18 a 20° C, com pouca luz, imediatamente após a inoculação. Para a coleta de dados fenotípicos, o comprimento (cm) das lesões necróticas ao longo da haste foi mensurado quatro dias após a inoculação (DAI). As medidas foram coletadas a partir da secção da haste até o final da lesão, com o auxílio de régua graduada em milímetros.

Os dados de comprimento de lesão foram submetidos ao teste de normalidade (Kolmogorov- Smirnov) do programa SPSS Statistics 24.0, com nível de significância de 0,01, para verificação de distribuição normal nos comprimentos de lesão e ao teste de Scott-Knott ( $p < 0,05$ ) através do programa SISVAR 5.3 (FERREIRA, 2011).

### 3.2 Extração de DNA total

O DNA dos brotos das amostras de soja, já previamente coletados, foi extraído utilizando-se o protocolo Doyle; Doyle (1990) baseado no método CTAB e modificado por Mesquita et al. (2015). A integridade do DNA foi verificada em gel de agarose 1%.

### 3.3 Construção da biblioteca GBS e sequenciamento

Para a construção da biblioteca são necessárias duas enzimas para a redução da complexidade do genoma dos genótipos de soja. Por isso, reações de restrição/ ligação foram realizadas em amostras aleatórias de DNA genômico, utilizando seis combinações de duas enzimas de redução da complexidade do genoma: PstI – HinP1I, PstI - MspI, PstI - MseI, HindIII - HinP1I, HindIII - MpsI e HindIII - MseI. As enzimas PstI (CTCGAG) e HindIII (AAGCTT) são cortadores raros, enquanto que MspI (CCGG), HinP1I (GCGC) e MseI (TTAA) são cortadores comuns/ frequentes. Para a seleção de um par de enzimas, foi utilizado o equipamento Bioanalyzer 2100 (Agilent, Santa Clara, CA) com um chip de DNA7500 para estimação da qualidade dos fragmentos obtidos da amplificação por PCR.

As amostras de DNA que foram determinadas como intactas e de alta qualidade foram pipetadas para placas de 96 poços. As alíquotas foram removidas para novas placas para quantificar com PicoGreen™ (Molecular Probes, Eugene, OR) em um leitor de placas. Considerando a concentração de DNA foram selecionadas 352 amostras, juntamente com 32 controles, para a construção da biblioteca de GBS. Para a preparação dos controles, 22 amostras de DNA escolhidas aleatoriamente foram replicadas, e os 10 poços restantes, foram controles negativos (branco) contendo água, distribuídos aleatoriamente entre as quatro placas de 96 poços.

A biblioteca GBS foi construída com o protocolo GBS original modificado utilizando duas enzimas previamente testadas (ELSHIRE et al., 2011; POLAND et al., 2012).

Para a realização do sequenciamento é necessário adaptadores para a leitura dos fragmentos na plataforma do equipamento. Por isso, dois tipos de adaptadores foram utilizados: adaptador A1 (adaptador *barcode*) e adaptador A2 (adaptador comum). Os adaptadores são compostos por dois oligonucleotídeos normais complementares que foram anelados para formação de cadeia dupla, sendo que no adaptador A1 há a presença de uma sequência de 4-9 pb para identificação de cada amostra a ser sequenciada (Figura 6). Os adaptadores foram desenhados com extremidades compatíveis, com fragmentos de DNA que foram cortados com as enzimas HindIII e MseI, que foram utilizadas no protocolo.



**Figura 6.** Produto de ligação de um fragmento de DNA genômico (preto) contendo um sítio de restrição HindIII e um sítio de restrição MseI. O adaptador A1 (azul) liga-se à extremidade gerada por HindIII. O *barcode* e *barcode* complementar de 4-9 pb para este adaptador são identificados pelas letras "X" e "Y", respectivamente. A extremidade gerada pela MseI corresponde ao adaptador A2 reverso (verde).

Os adaptadores comuns foram acondicionados em tubos tipo Eppendorf, e os adaptadores *barcode* foram dispostos em placas de 96 poços. Os adaptadores foram anelados por aquecimento em equipamento BioRad DNA com programação de um ciclo de PCR de fase única a 95 °C durante 1 min e depois diminuindo a temperatura por 1° C cada ciclo durante 65 ciclos.

Os adaptadores foram quantificados após o anelamento para assegurar que a formação de DNA de cadeia dupla foi completa e se estavam na concentração correta (0,1 µM Adaptador A1; 10µM Adaptador A2). A concentração uniforme de adaptadores é crítica para produção de números uniformes de leituras entre amostras quando se sequencia a biblioteca multiplexada.

Reações de restrição foram realizados em placas de 96 poços utilizando 5 µL de cada amostra anteriormente escolhida ajustadas à concentração aproximada de 50 ng/µL de DNA e digestão com HindIII e MseI (New England Biolabs, Ipswich, MA), conforme instruções do fabricante. A digestão foi realizada em equipamento BioRad DNA com programação de 37 °C por 2 horas, seguida por etapa de 20 min a 65 ° C e, ao final, mantiveram-se as placas a 8 °C. Em seguida, nas placas com 0,1 µM de adaptadores *barcode*, reações de ligação foram realizadas com DNA ligase, com o resultado da restrição da etapa anterior e 10 µM de adaptadores A2 por poço, respectivamente, conforme instruções do fabricante.

As bibliotecas foram reunidas em um tubo tipo Eppendorf com 5 µL de cada amostra ligada aos adaptadores e utilizando *beads* Agencourt AMPure XP (Beckman Coulter Inc, Atlanta, GA) foi realizada a purificação, selecionando os fragmentos menores que 100 pb. A biblioteca foi amplificada por PCR utilizando *primers* específicos para os adaptadores (*primers* Illumina *forward e reverse*) durante 15 ciclos utilizando Phusion DNA polimerase (NEB), nas condições de 98 °C por 30 s, 15 ciclos (98 °C por 10 s, 68 °C por 30 s, 72 °C por 30 s), seguida de 72 °C por 5 min, reduzindo a temperatura até 4 °C. Os *primers* Illumina são idênticos aos oligonucleotídeos nas

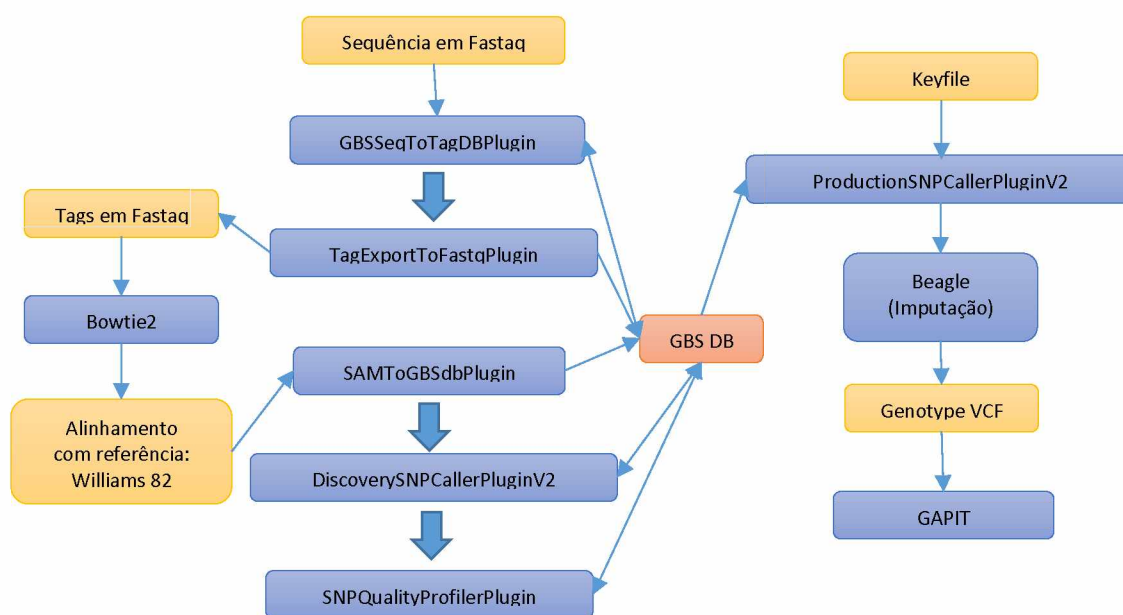


*flowcells* de Illumina com a sequência complementar adicional para o Adaptador 1 (*forward*) e o Adaptador 2 (*reverse*). Utilizou-se novamente o protocolo de purificação com as *beads*, selecionando os fragmentos menores que 100 pb.

O tamanho médio e a concentração dos fragmentos foram estimados no equipamento Bioanalyzer 2100 (Agilent, Santa Clara, CA) utilizando um chip de DNA7500. As bibliotecas foram ajustadas para 10 nmol e sequenciadas com 100 pb, no modo *single-read* no sequenciador HiSeq4000 (Illumina, San Diego, CA) usando um kit de sequenciamento HiSeq SBS versão 4 no Centro de Biotecnologia Roy J. Carver da Universidade de Illinois (Urbana, IL).

### 3.4 Processamento dos dados das leituras das sequências do Illumina e seleção dos SNPs

Utilizando sequências resultantes do sequenciamento em formato FASTQ, foram utilizadas para o processamento as etapas da pipeline Tassel 5 GBS v2 para espécies com genoma de referência (Figura 7).



**Figura 7.** Pipeline para filtragem dos SNPs por qualidade. SAM: formato de arquivo; DB: banco de dados; VCF: formato de arquivo.

### 3.4.1 GBSSeqToTagDBPlugin

O programa GBSSeqToTagDBPlugin utilizou os arquivos fastQ como entrada, identificou tags e a amostra em que elas aparecem, e armazenou esses dados em um banco de dados local. Foram mantidas apenas boas leituras com um *barcode* e um local de corte e nenhum N na parte útil da sequência. As sequências foram desanexadas dos *barcodes* e foram truncadas sequências que têm um segundo local de corte ou leitura no adaptador comum. Esse software é executado para identificar as tags distintas da base de dados e exportá-las de modo que elas podem ser alinhadas com o genoma de referência com vários alinhadores (GLAUBITZ et al., 2014).

### 3.4.2 TagExportToFastqPlugin

O programa TagExportToFastqPlugin recuperou tags distintas armazenadas na base de dados e as reformatou para um ficheiro FASTQ que pode ser lido pelo programa de alinhamento Bowtie2 (GLAUBITZ et al., 2014).

### 3.4.3 Bowtie 2

Bowtie 2 é uma ferramenta ultrarrápida e eficiente para alinhar sequências obtidas pelo sequenciamento com sequências de referência de longo comprimento (LANGMEAD; SALZBERG, 2012). Os dados processados foram alinhados com o Williams 82 Gmax v9.0 do portal Phytozome como o genoma de referência. O arquivo de saída é um arquivo formato SAM, necessário para chamar SNPs em etapas posteriores da análise GBS (SCHMUTZ et al., 2010).

### 3.4.4 SAMToGBSdbPlugin

SAMToGBSdbPlugin realizou a leitura dos arquivos SAM para determinar as potenciais posições de tags contra o genoma de referência. A ferramenta atualizou o banco de dados atual com informações sobre as posições das tags. Foi verificado se haviam tags encontradas no arquivo SAM que não podiam ser combinadas com tags no banco de dados. Este plugin realizou o armazenamento das informações de posição para cada tag alinhada (GLAUBITZ et al., 2014).

### **3.4.5 DiscoverySNPCallerPluginV2**

DiscoverySNPCallerPlugin foi utilizado para identificar SNPs a partir das tags alinhadas utilizando o banco de dados GBS. Tags posicionadas no mesmo local físico foram alinhadas uma contra a outra; os SNPs foram chamados a partir das tags alinhadas, e a posição dos SNPs e os dados de alelos foram enviados para o banco de dados (GLAUBITZ et al., 2014).

### **3.4.6 SNPQualityProfilerPlugin**

SNPQualityProfilerPlugin foi utilizado para marcar todos os SNPs descobertos em diversas coberturas, profundidade e estatísticas genóticas para um determinado conjunto de amostras. O plugin cria um arquivo CSV (arquivo separado por vírgula) contendo a informação de qualidades armazenadas para cada posição. Estes dados foram utilizados para criar uma pontuação de qualidade para ser armazenado com as posições (GLAUBITZ et al., 2014).

### **3.4.7 ProductionSNPCallerPluginV2**

Essa ferramenta realizou a conversão dos dados de keyfile (tabela com dados de Flowcell, Lane, *Barcode* e FullSampleName) dos genótipos, em seguida, executou a adição deste para um arquivo formato VCF (GLAUBITZ et al., 2014). Leituras de sequências de baixa qualidade que foram alinhadas com a sequência do genoma de referência foram removidas a partir dos conjuntos de dados (LANGMEAD; SALZBERG, 2012).

### **3.4.8 Beagle 4.1**

O software Beagle realizou a chamada de genótipos, a ordenação dos marcadores ao longo dos cromossomos, a imputação de marcadores não relacionados com o genótipo e detectou segmentos identificados por descendência, com rapidez e acurácia, eliminando os SNPs monomórficos (BROWNING; BROWNING, 2016).

## **3.5 Limpeza dos dados genotípicos**

Para a seleção dos melhores genótipos, utilizando o programa Excel, foram identificados e removidos aqueles que apresentaram desvio padrão maior que 2,0 e média fenotípica maior que 5,0 cm. Ao final foram utilizados para as análises 324 genótipos.

Com o pipeline TASSEL foram identificados os possíveis SNPs e estes foram filtrados, sendo removidos os que não estavam presentes em, no mínimo, 75% dos genótipos e apresentavam altos níveis de heterozigosidade entre os genótipos, filtrando a 10%. Portanto, se um SNP era heterozigoto em mais de 10% das amostras, ele foi removido, e se uma amostra tivesse mais de 10% de SNP heterozigotos, esta foi removida. Após a remoção dos SNPs, estes foram submetidos a outras etapas do processamento de dados.

### 3.6 Análise de associação genômica

Para a análise de associação foi utilizado o programa GAPIT na plataforma analítica R, pelo Modelo Linear Misto Comprimido (CMLM) com os parâmetros de componentes principais (PCA) no valor de 3, frequência alélica média (MAF) maior que 0,05 e coeficiente de variação com 2 *batches*, separando os genótipos em lotes de acordo com diferentes datas de avaliação. O algoritmo utiliza a plataforma com a instalação de pacotes específicos da ferramenta implementado para estudos de associação do genoma completo (GWAS).

Para melhor confiabilidade da análise, foram retirados os alelos raros filtrando os dados de SNPs imputados com uma frequência mínima alélica (MAF) menor que 0,05. Para confirmar a significância das associações entre SNPs, foi utilizado o método FDR (*False Discovery Rate*), que tem por finalidade reduzir o número de falsos positivos. Os marcadores foram definidos como estando significativamente associados com base no *p*-value ajustado pelo FDR de 0,05, ou seja, todos SNPs associados com um valor menor que 0,05 foram considerados significativos durante a análise (STOREY, 2002).

A partir dos dados de *p*-value de significância ajustados, foram construídos gráficos de Manhattan para se observar a dispersão das associações entre marcadores SNP e o caráter de interesse. Com a finalidade de se verificar a qualidade dos SNPs que foram associados, gráficos do tipo Q-Q foram construídos, objetivando relacionar as associações encontradas em relação às associações esperadas.

### 3.7 Análise de desequilíbrio de ligação

Uma função embutida no software TASSEL foi usada para determinar os valores de  $r^2$  em pares, isto é, a probabilidade de dois SNPs ocorrerem conjuntamente, para todos os 251 SNPs da região de 10 Mb contendo sete SNPs significativos, abrangendo a posição 30 Mb a 40 Mb no cromossomo 1. O diagrama de decaimento LD foi



construído com base nos valores  $r^2$  e distâncias entre cada par de SNPs nesta região. Para calcular o decaimento de LD, foi utilizado um modelo não-linear para estimar os valores de E esperados ( $r^2$ ) (HILL; WEIR, 1988; REMINGTON et al., 2001). O gráfico de LD foi construído usando Haploview 4.2 para uma região de 4 Mb onde os sete SNPs significativos no cromossomo 1 foram localizados (BARRETT et al., 2005).

### **3.8 Identificação de relações entre SNPs e funções biológicas**

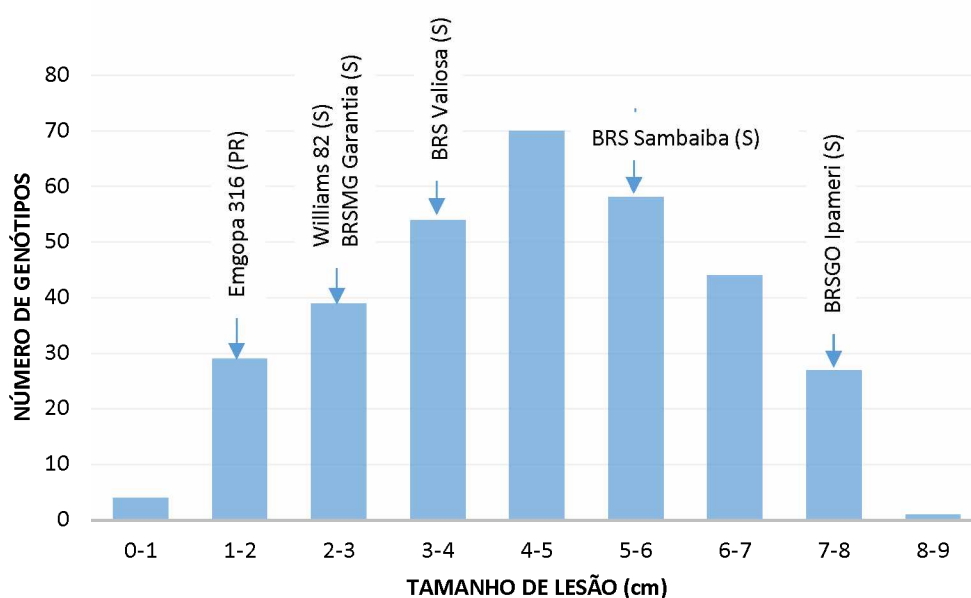
Para a identificação da atividade biológica dos genes, foram utilizadas anotações do servidor SGED, que usou BLASTx de sequências de DNA para o banco de dados de nucleotídeos de proteína do NCBI e foram selecionados os 10 melhores hits. O BLASTx identifica potenciais produtos proteicos codificados por uma consulta nucleotídica (ALTSCHUL et al., 1997).

## 4 RESULTADOS

### 4.1 Avaliação da resistência à *S. sclerotiorum* nos genótipos de soja

As plantas que germinaram e cresceram saudáveis foram utilizadas para extração de DNA e fenotipagem. A maioria (108 dos 352 genótipos utilizados) foram representados por nove plantas, 86 por oito, 52 por sete, 48 por seis, 26 por cinco, 20 por quatro e apenas oito genótipos baseados em apenas três plantas. Os outros quatro genótipos: Emgopa 316, PI503336, Williams 82 e FMT05-60 foram plantados e inoculados duas vezes, com 16, 13, 10 e 7 plantas combinadas para análises, respectivamente.

Após a inoculação artificial com *S. sclerotiorum* nos genótipos de soja, a distribuição do tamanho das lesões necróticas entre os mesmos está apresentada na Figura 8. Os comprimentos de lesão cobriram uma gama muito ampla (0,67 a 8,43 cm, média de 4,46 cm) e a distribuição foi normal ( $p=0,046$ ), com poucos altamente resistentes ou altamente suscetíveis e a grande maioria dos genótipos exibiram reações intermediárias. É válido ressaltar que todos os genótipos desenvolveram uma lesão; mesmo o genótipo mais resistente desenvolveu uma lesão curta, indicativa de que estava infectado, mas foi capaz de parar o desenvolvimento do fungo.



**Figura 8.** Distribuição fenotípica do tamanho de lesão causada por *S. sclerotiorum* em 324 genótipos de soja após inoculação artificial. PR: parcialmente resistente; S: suscetível.

Segundo estudos de Juliatti e colaboradores (2013; 2014), o genótipo Emgopa 316 (1,20 cm) é considerado parcialmente resistente, e os genótipos Williams 82, BRSMG Garantia, BRS Valiosa RR, BRS Sambaiba e BRSGO Ipameri são denominados suscetíveis (3,00, 3,00, 3,60, 4,58 e 7,58 cm, respectivamente) conforme Figura 8 e Tabela 1. Com o teste de comparações de médias Scott-Knott foram detectados 11 grupos em relação a resistência ao mofo branco. Dezesete genótipos apresentaram tamanho de lesão média menor que 1,20 cm, limiar dos tamanhos de lesão para genótipos mais resistentes e 34 genótipos apresentaram tamanho de lesão média de 6,83 cm ou superior (Tabela 1).

**Tabela 1.** Comprimento da lesão da doença nos 324 genótipos resultado da inoculação artificial.

Genótipo	Comprimento médio de lesão (cm)*	Genótipo	Comprimento médio de lesão (cm)*
P1C142-158.243	0,43 a	FT-1	0,70 a
BRS Charrua RR	0,58 a	P1C142-128.686	0,75 a
FMT05-40.907/1	0,58 a	PI194639	0,85 a
IAC 100	0,60 a	P1C142-112.208	0,88 a
P1C142-116.940	1,03 b	BRSGO Araçu	1,28 b
V-Max RR	1,08 b	P1C142-170.130	1,28 b
P1C142-169.621	1,18 b	TMG 1188 RR (GENO_TKG_35)	1,33 b
P1C142-156.243	1,20 b	NK S19-90	1,43 b
EMGOPA 316	1,20 b		
P1C142-169.669	1,50 c	V-Top RR	1,75 c
FT-3	1,50 c	P98Y30	1,78 c
IAS 5	1,53 c	P1C142-168.243	1,80 c
P1C142-107.220	1,53 c	P1C142-120.884	1,83 c
L79-1404	1,55 c	P1C142-123.850	1,85 c
P1C142-110.417	1,58 c	P1C142-154.809	1,95 c
P1C142-127.559	1,60 c	FMT04-48.182	1,95 c
P1C142-102.143	1,60 c	P1C142-151.674	1,98 c
FMT03-40.673	1,70 c	P1C142-126.986	1,98 c
P1C142-151.788	1,73 c	P1C142-168.971	1,98 c
P1C142-116.766	1,73 c	FT-37	2,00 c
P1C142-169.073	1,75 c	P1C142-120.656	2,00 c
BRSGO Indira	1,75 c		
BMX Turbo RR	2,08 d	A2506	2,38 d
L284	2,10 d	P1C142-184.849	2,40 d
P1C142-126.380	2,10 d	P1C142-119.350	2,43 d
FT-4	2,13 d	P1C142-168.153	2,43 d
L91-8052	2,20 d	P1C142-124.193	2,48 d

Continua...

Continuação Tabela 1.

TC132-10.924	2,23 d	AXN-1-55	2,48 d
P1C142-124.500	2,28 d	P1C142-117.050	2,53 d
P1C142-122.870	2,30 d	VINTON 81	2,58 d
FT-9	2,63 e	P1C142-107.552	3,13 e
CB13-4260	2,65 e	FT-59	3,15 e
P1C142-115.054	2,75 e	L120	3,15 e
FT-7	2,78 e	P1C142-123.484	3,23 e
Syn90-74 RR	2,78 e	FMT05-40.701/1	3,25 e
CORSOY 79	2,88 e	WESTAG 97	3,25 e
FT-46	2,88 e	TC130-00.089	3,28 e
P1C142-116.753	2,95 e	HODGSON	3,28 e
WILLIAMS 82	3,00 e	TMG 132 RR (GENO_TKG_34)	3,35 e
BRSMG Garantia	3,00 e	L536 (T24)	3,35 e
P1C142-100.250	3,03 e	P1C142-121.241	3,35 e
TC132-11.755	3,03 e	HARDIN	3,38 e
L82-1187	3,08 e	FT-10	3,38 e
Anta 82 RR (GENO_TKG_41)	3,08 e	BMX Potencia RR	3,43 e
AC-CALIBRI	3,08 e	BMX Desafio RR	3,43 e
FT-2	3,08 e	L471 (T35)	3,43 e
HODGSON 78	3,08 e	L474 (T34)	3,48 e
P1C142-166.448	3,08 e	BRSMG 725S	3,48 e
FT-42	3,10 e	FT-23	3,48 e
FT-11	3,10 e		
BSR 101	3,53 f	L547	3,88 f
BRS Raimunda	3,55 f	KENWOOD 94	3,88 f
L514 CT	3,58 f	L510 PL ≠1	3,88 f
BRB 16003200233060	3,58 f	BRB 8607561 RR	3,90 f
FT-8	3,60 f	BRS 257	3,90 f
BRS Valiosa RR	3,60 f	L548	3,93 f
TC130-00.079	3,63 f	FMT05-60.497/1	3,93 f
BMX Ativa RR	3,65 f	FT-45	4,00 f
BRSGO Mineiros RR	3,65 f	L61-5047	4,03 f
DASSEL	3,68 f	L347	4,05 f
FT-60	3,73 f	PI595099	4,08 f
BRSGO Jataí	3,73 f	BRSG0 6159	4,08 f
FT-33	3,73 f	TC11-2-37.987/G012	4,10 f
FT-26	3,75 f	BRSMG 790A	4,13 f
FT-24	3,78 f	TMG 1067 RR	4,18 f
L530	3,80 f	L632	4,18 f
L510 PL ≠2	3,80 f	NA 5909 RG	4,18 f
L587	3,80 f	NS 11631 RR	4,20 f
FT-5	3,83 f	L577	4,20 f
BRB16063 2001 - 33000	3,85 f	L635 (T14)	4,23 f

Continua...

Continuação Tabela 1.

BMX Veloz RR	3,85 f		
L98	4,25 g	FT-21	4,70 g
L669 (T5)	4,28 g	CHAPMAN	4,73 g
BRS GO Raissa	4,28 g	BRS Silvânia RR	4,73 g
P1C142-182.964	4,30 g	TC132-11.235	4,75 g
CD202	4,30 g	FT-27	4,75 g
L470	4,33 g	FT-14	4,78 g
TMG 7262 RR (GENO_TKG_15)	4,35 g	L611	4,78 g
P1C142-183.810	4,38 g	BRS/GO Chapadões	4,80 g
L526	4,40 g	L258	4,80 g
TC132-11.450	4,40 g	FT-61	4,80 g
FMT Tucunaré	4,45 g	TMG 7161 RR (GENO_TKG_13)	4,80 g
L663 (T8)	4,48 g	FT-63	4,83 g
P1C142-101.828	4,48 g	FT-12	4,83 g
TMG 4182 (GENO_TKG_37)	4,50 g	BRASRO50611 RR	4,83 g
L457	4,50 g	TC130-00.056	4,83 g
DOKO	4,55 g	TMG 4190	4,88 g
L608	4,55 g	L166 (T91)	4,88 g
TC132-10.251	4,58 g	L296 (T57)	4,90 g
FT-30	4,58 g	TC132-12.191/A	4,90 g
P1C142-171.127	4,58 g	L633	4,90 g
BRS Sambaiba	4,58 g	L630 CT	4,93 g
TMGCB14-22920	4,60 g	P1C142-124.344	4,93 g
P1C142-172.624	4,60 g	FT-68	4,93 g
TMGCB14-22919	4,60 g	L77-1585	4,93 g
L578	4,63 g	FT-58	4,95 g
L380 (T44)	4,63 g	L404 (T47)	4,95 g
BRS Tracajá	4,65 g	L615 (T15)	4,95 g
BMX Apolo RR	4,68 g	FT-67	4,98 g
L343 CT	4,68 g	P1C142-184.652	5,00 g
P98Y11	4,70 g		
L241 (T79)	5,10 h	L644	5,38 h
TC130-00.093	5,10 h	BRS 133	5,43 h
L572 (512)	5,10 h	PI90763	5,43 h
L463 (T37)	5,15 h	TMG 4185	5,43 h
P1C142-180.177	5,18 h	L22 CT	5,43 h
L268	5,18 h	Lee 74	5,45 h
L588	5,20 h	L430 (T42)	5,48 h
RR12-739 (GENO_TKG_433)	5,20 h	L15	5,48 h
L260	5,20 h	TMG RTRNCS (GENO_TKG_62)	5,50 h
L10	5,23 h	P1C142-176.307	5,50 h
FT-65	5,25 h	Bedford	5,50 h

Continua...

Continuação Tabela 1.

L532	5,25 h	L242 (T78)	5,50 h
BRS 185	5,28 h	Hartwig	5,53 h
TC132-10.041/A	5,28 h	L338 CT	5,58 h
PI503336	5,28 h	TC11-2-37.987/G026	5,60 h
L585	5,30 h	L579	5,63 h
PI391589B	5,30 h	L638	5,68 h
TC130-00.021	5,30 h	Pickett	5,68 h
L434 (T41)	5,33 h	TC12-2-59.830	5,70 h
BR0581114 RR	5,35 h	M7739 IPRO	5,73 h
L291 (T60)	5,38 h	PI398637	5,75 h
L641 (T13)	5,38 h	P1C142-186.022	5,80 h
AxN-1-55-HP	5,38 h	L35U SEL (T51)	5,83 h
L531	5,38 h	TC11-2-37.987/G025	5,85 h
L336 (T50)	5,38 h	L591 (T19) PL ≠1	5,85 h
L423 (T44)	5,88 i	TC12-0-47.107	6,28 i
BR6I 600320023306	5,90 i	RR12-715 (GENO_TKG_418)	6,30 i
L576	5,90 i	TC12-0-52.294/G026	6,30 i
L456 (T39)	5,93 i	L664 (T7)	6,30 i
L429 (T43)	5,93 i	P1C142-170.329	6,33 i
L25U (T72)	5,95 i	TC12-0-52.213	6,38 i
FT-62	6,00 i	P1C142-194.871	6,38 i
BR05736154	6,00 i	L88 (T104)	6,38 i
P1C142-170.382	6,05 i	L165 (T92)	6,40 i
BRAR 050611G	6,08 i	FT-71	6,53 i
P1C142-170.215	6,08 i	TC12-0-48.687	6,53 i
P1C142-186.380	6,10 i	L525 (T28)	6,60 i
BRA 54651611 RR	6,10 i	P1C142-188.840	6,60 i
PI88788	6,13 i	TC12-2-60.290	6,60 i
L248 (T76)	6,13 i	TC12-0-46.100	6,63 i
PI399073	6,15 i	L671 (T3)	6,63 i
L351	6,18 i	PI189861	6,65 i
L140 (T94)	6,18 i	TC12-1-47.590	6,68 i
AxN-1-55-HC	6,20 i	BR0+X742:CG7425736 156	6,68 i
P1C142-185.974	6,20 i	TC12-0-52.024	6,70 i
P1C142-195.052	6,23 i	TC12-0-51.331	6,75 i
L412 L417 (T45)	6,23 i		
FT-49	6,83 j	L263/264 CT	7,08 j
PI358318A	6,83 j	FT-28	7,13 j
G004501413260 - A	6,88 j	TC12-0-51.186	7,15 j
FT-29	6,90 j	ANTA 82 RR	7,20 j
BR058114 RR	6,93 j	TMGER15 31787	7,23 j
TC12-0-52.294	6,95 j	L19	7,30 j
TC12-0-47.648	6,98 j	FT-64	7,33 j

Continua...

Continuação Tabela 1.

TC12-0-50.332/G027	6,98 j	TC12-0-49.220	7,33 j
TMGER15 31786	7,00 j	BRSGO Ipameri	7,35 j
TC130-00.011	7,03 j	L379 (T50)	7,38 j
L199	7,03 j	TC12-2-56.681	7,38 j
TC12-0-50.332	7,05 j	FT-22	7,40 j
MSOY 7908 RR	7,05 j	L32	7,40 j
TC12-1-50.655	7,48 k	L25	7,78 k
TC12-0-52.243/G022	7,53 k	L196	7,88 k
BRS RAISSA RES 1/3	7,60 k	L136	8,00 k
NCS	7,63 k	L90 (CT)	8,43 k
PI196157			

\*Médias seguidas de mesma letra minúscula na coluna não diferem entre si. Teste Scott-Knott ( $p < 0,05$ ).

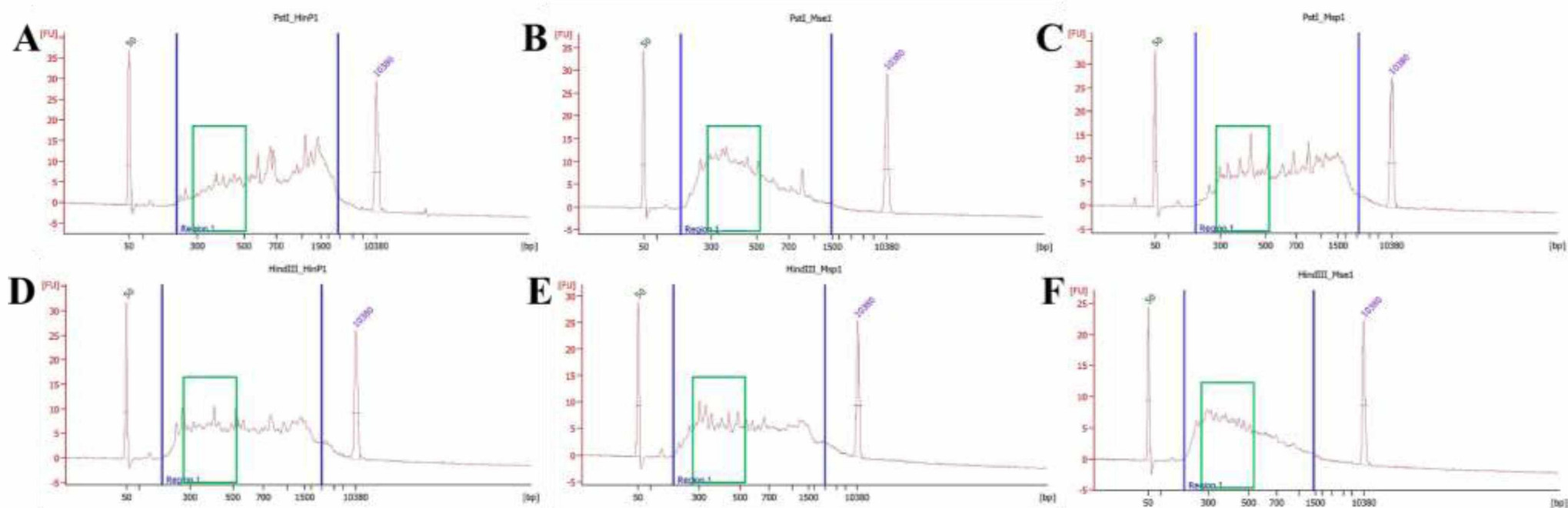
## 4.2 Teste de enzimas de restrição

O teste para a escolha do melhor par de enzimas para o preparo da biblioteca GBS preconiza homogeneidade dos picos do eletroferograma para fragmentos de 300 a 500 pb. Considerando o tamanho médio de fragmentos, os pares de enzimas PstI / MseI e HindIII / MseI foram as melhores opções e o par PstI / HinPII seguido pelo par PstI / MspI obtiveram a pior qualidade (Tabela 2).

**Tabela 2.** Tamanho médio de fragmentos do teste de enzimas para o protocolo de preparo da biblioteca.

Par de enzimas	Tamanho médio dos fragmentos (pb)
PstI / HinPII	914
PstI / MseI	490
PstI / MspI	837
HindIII / HinPII	760
HindIII / MspI	741
HindIII / MseI	508

Com isso, foi possível identificar que as enzimas HindIII e MseI obtiveram melhor performance na homogeneidade dos picos concentrados na região pretendida, não sendo observada no par PstI / MseI, pois picos significativos estavam presentes fora da faixa de 300 a 500 pb, como pode-se detectar na Figura 9.



**Figura 9.** Eletroferograma obtido da análise do teste dos pares de enzimas de restrição no Bioanalyzer 2100. **A.** PstI / HinP1I; **B.** PstI / MseI; **C.** PstI / MspI; **D.** HindIII / HinP1I; **E.** HindIII / Msp1I; **F.** HindIII / MseI. O retângulo verde indica a região favorável (300 a 500 bp) do tamanho dos fragmentos para o sequenciamento. FU: unidade de fluorescência.



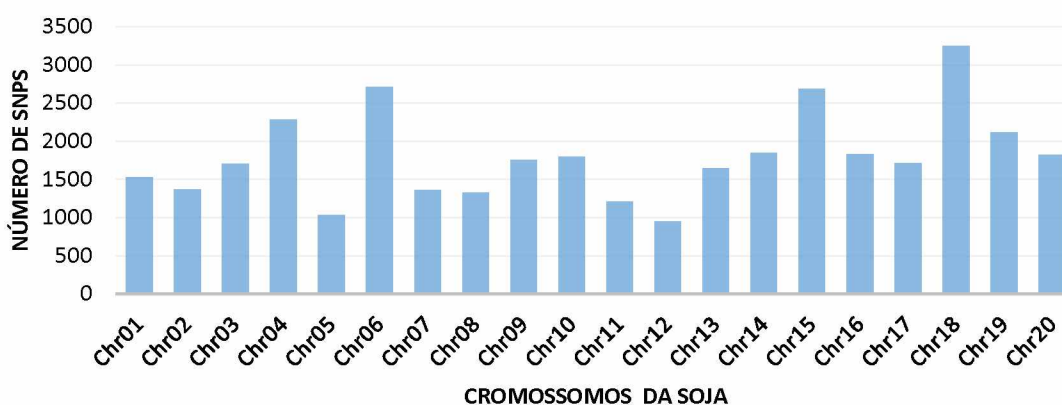
Verificou-se que o par de enzimas de restrição HindIII e MseI apresentou maior qualidade e quantidade de fragmentos de tamanho ideal (entre 300-500 pb) para a construção da biblioteca de GBS.

#### 4.3 Distribuição e controle de qualidade dos marcadores SNPs

As bibliotecas GBS codificadas com *barcodes* foram sequenciadas em uma lâmina (*flow cell*) em uma canaleta (*lane*) produzindo 373 milhões de leituras de cadeia simples de 100 nucleotídeos de comprimento, com resultados de alta qualidade, para os 352 genótipos.

Com o programa TASSEL, foram identificados 50.000 SNPs e estes foram filtrados, sendo removidos os que não estavam presentes em no mínimo 75% dos genótipos e apresentavam heterozigosidade entre os genótipos. Além disso, foram removidos 19 genótipos que apresentaram desvio padrão de seus valores fenotípicos maiores que 2,0, cinco genótipos em que o intervalo de maiores e menores valores fenotípicos foi superior a 5,0 cm, uma amostra pareceu ser infectada por vírus e uma que foi mal identificada. Após a remoção de 26 genótipos, 324 genótipos e 35.955 SNPs foram submetidos ao programa GAPIT para estudo de associação ampla do genoma para SNPs associados com diferentes níveis de resistência SSR.

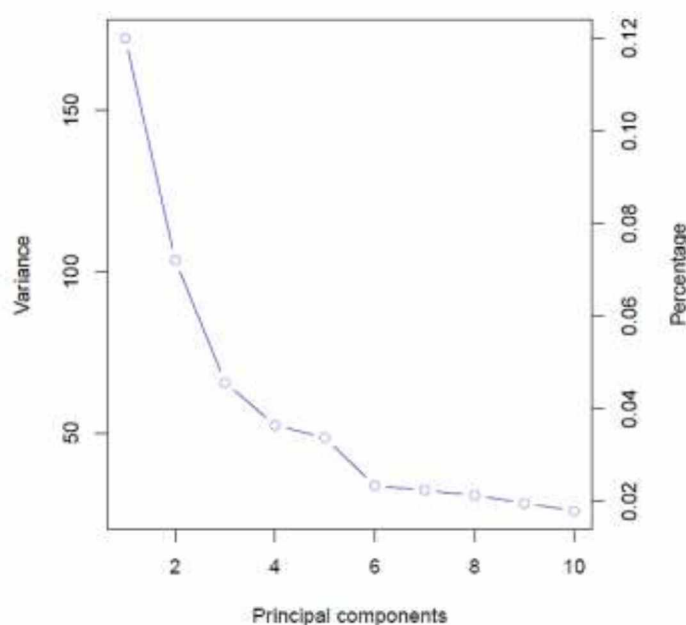
Após as análises realizadas pelo software GAPIT, 35.955 SNPs foram identificados ao longo dos 20 cromossomos da soja em que houve média de 1.798 SNPs por cromossomo, sendo que o maior número de marcadores foi identificado no cromossomo 18 (3.247 SNPs), seguido pelo cromossomo 6 (2.712 SNPs) e o menor número de SNPs foi observado no cromossomo 12 (954 SNPs) e no 5 (1.035 SNPs) (Figura 10).



**Figura 10.** Distribuição dos marcadores SNPs nos 20 cromossomos da soja.

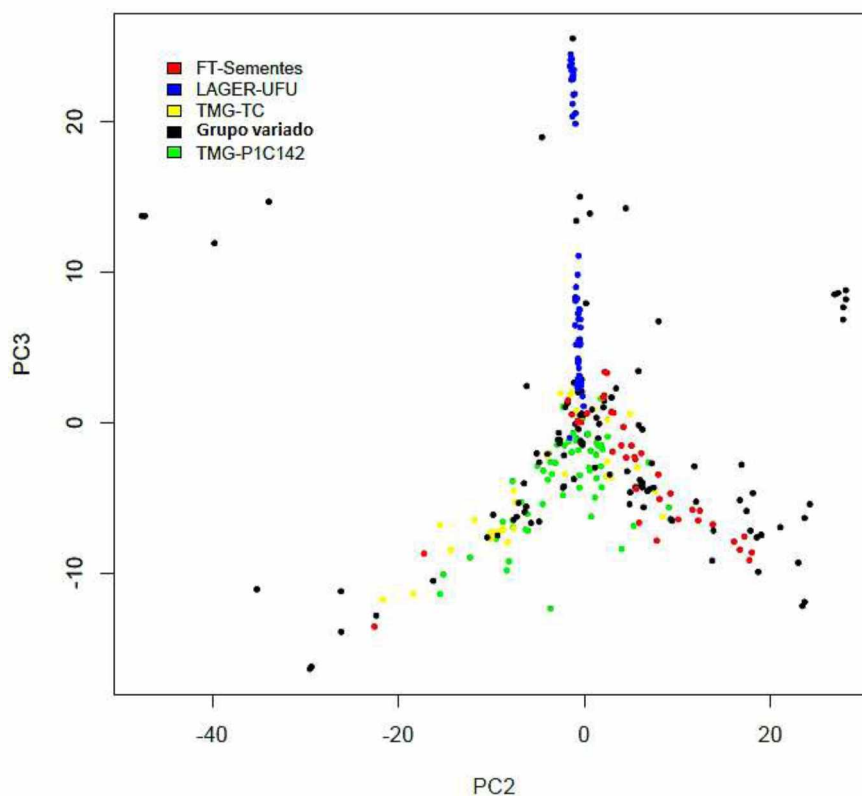
#### 4.4 Análise da estrutura da população

Antes da condução do mapeamento por associação, a estrutura genética dos 324 genótipos foi explorada por PCA usando o conjunto de 35.955 SNPs para estimar a influência da estrutura da população nos resultados do mapeamento. Observando o gráfico de *Eigenvalue* para o efeito de cada princípio, o componente mostra que PC1 explicou aproximadamente 12% da variação fenotípica, PC2 cerca de 7% e PC 3 cerca de 4,5% (Figura 11).



**Figura 11.** Gráfico dos valores *Eigenvalue* e os componentes principais.

O gráfico de dispersão bidimensional (PC2 e PC3) envolvendo todos os genótipos apresentou cinco subpopulações principais (FT sementes, LAGER-UFU, TMG-TC, TMG-P1C142 e grupo variado) conforme Figura 12. Apesar da formação de grupos, houve genótipos com agrupamento menos evidente, indicando baixa estrutura da população dentro do mapeamento. Assim, os três primeiros PCs foram adicionados ao Modelo Linear Misto Comprimido (CMLM) utilizado para as análises de associação.

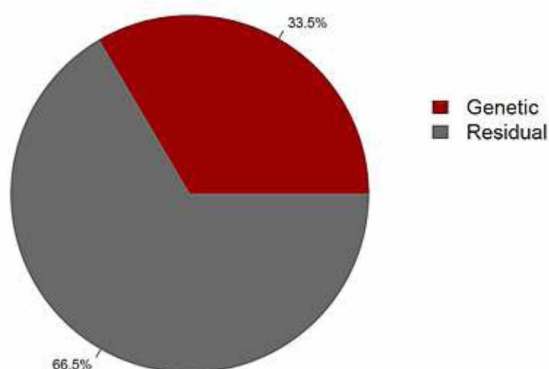


**Figura 12.** Gráfico de dispersão bidimensional (PC2 x PC3) da análise do componente principal.

Segundo os estudos de Zhao e colaboradores (2015), o método CMLM é mais adequado para o mapeamento GWA para resistência ao mofo branco, indicando que o efeito de ambos, estrutura da população e parentesco, deve ser considerado.

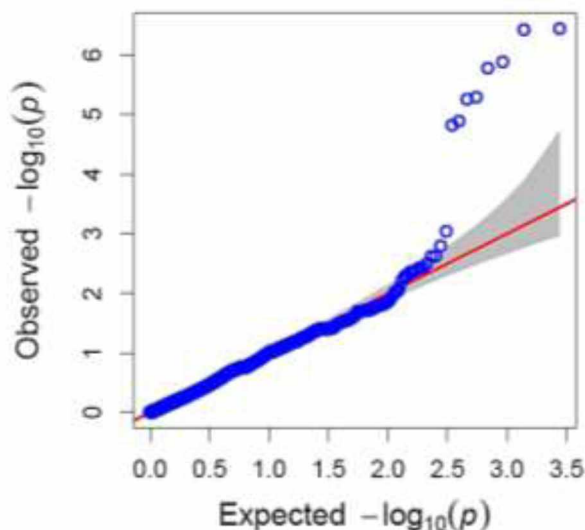
#### 4.5 Mapeamento por associação genômica

Foram realizados testes para detectar associações entre marcadores SNP e resistência à *S. sclerotiorum* em soja. A característica exibiu herdabilidade de 33,5%, como mostrado como o componente genético na Figura 13.



**Figura 13.** Variância total para resistência à *S. sclerotiorum*.

O gráfico quantil-quantil (Q-Q plot) indicou a ocorrência de divergência entre os valores de significância observados e esperados na hipótese nula (linha vermelha) do teste estatístico  $\chi^2$ , com 8 pontos deslocados da linha diagonal (Figura 14).



**Figura 14.** Gráfico quantil-quantil (Q-Q plot), da distribuição dos valores observados e esperados para o estudo de associação genômica para resistência à *S. sclerotiorum*.

Para administrar o grande número de marcadores testados e para estabelecer uma taxa de descoberta falsa razoável, com finalidade de reduzir o número de falsos positivos, estipulou-se valores menores que 0,05 para  $p$ -values ajustados por FDR. Com isso, foram identificados oito SNPs com valores de MAF maiores que 0,05 e obtiveram  $p$ -values ajustados por FDR menores que 5%. Os valores de  $p$ -values ajustados por FDR dos marcadores significativos variaram de 0,0528% a 0,5327%, com valores bem menores que o limite de 5%.

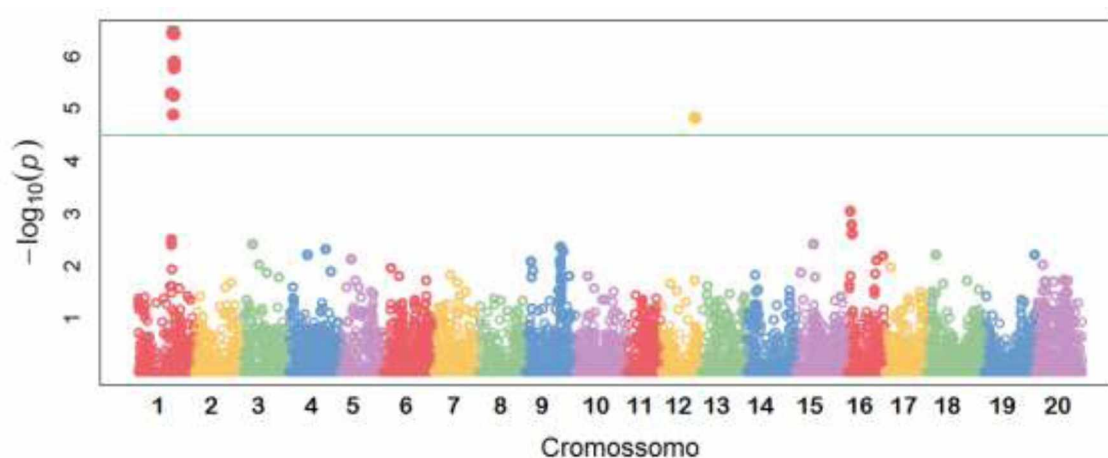
A associação mais significativa foi encontrada com marcador SNP no cromossomo 1 ( $p$ -value=  $3,69 \times 10^{-7}$ ;  $p$ -values ajustados por FDR = 0,000528). A estimativa da proporção de variância para este marcador determinou 6,27% da variância fenotípica nesta população que transportam o alelo de resistência, sendo que os outros marcadores significativos foram enquadrados em uma variância entre 6,26 a 4,48% (Tabela 3).

**Tabela 3.** Melhores marcadores significativos associados à resistência à *S. sclerotiorum* em soja considerando  $p$ -values ajustados por FDR <0,05.

Cromossomo	Posição (pb)	$p$ -value	R <sup>2</sup> (%)	MAF	$p$ -values ajustados por FDR
1	34787519	$3,69 \times 10^{-7}$	6,27	0,248	0,000528
1	36048095	$3,81 \times 10^{-7}$	6,26	0,242	0,000528
1	36044127	$1,31 \times 10^{-6}$	5,66	0,248	0,001160
1	36038838	$1,67 \times 10^{-6}$	5,54	0,236	0,001160
1	36006734	$5,65 \times 10^{-6}$	5,00	0,251	0,002611
1	33164666	$5,16 \times 10^{-6}$	4,96	0,374	0,002611
1	35073769	$1,28 \times 10^{-5}$	4,57	0,245	0,005070
12	33968718	$1,54 \times 10^{-5}$	4,48	0,113	0,005327

R<sup>2</sup>: coeficiente de determinação; MAF: *Minor Allele Frequency* (alelo de menor frequência); FDR: *False Discovery Rate* (taxa de falsa descoberta).

Considerando os limites de significância, sete dos SNPs estavam no cromossomo 1 e um SNP estava no cromossomo 12, como pode ser observado no gráfico Manhattan (Figura 15) e na Tabela 3.



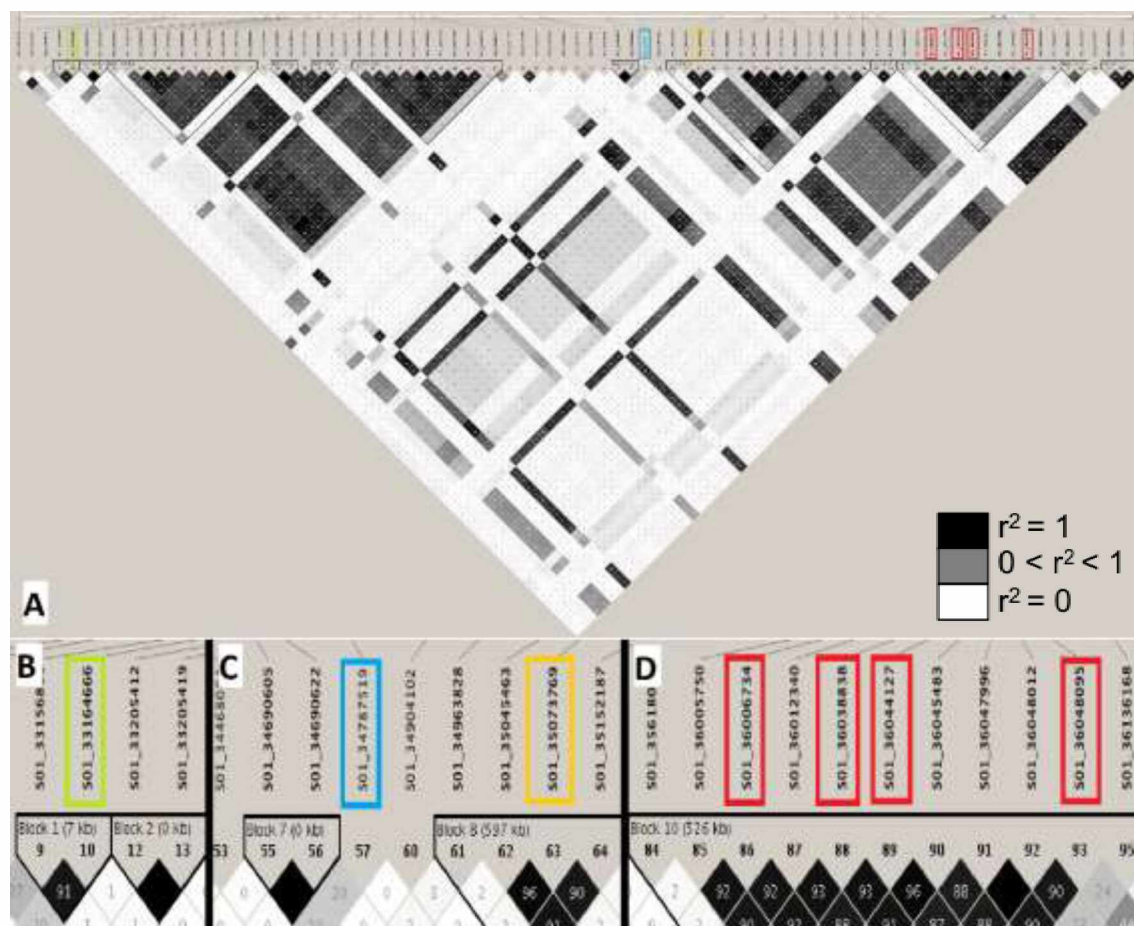
**Figura 15.** Gráfico de Manhattan com SNPs significativos que estão associados à resistência a *S. sclerotiorum*. Os cromossomos têm cores diferentes e a linha horizontal indica o limite com  $p$ -values correspondentes ao valor de  $p$ -value ajustado por FDR < 0,05.

#### 4.6 Análise de desequilíbrio de ligação

Para obter uma visão abrangente da localização genômica do *locus* principal no cromossomo 1, a análise de desequilíbrio de ligação (LD) do Cr01 foi realizada em parte do cromossomo (4 Mb) em que haviam 104 SNPs, onde os sete SNPs foram

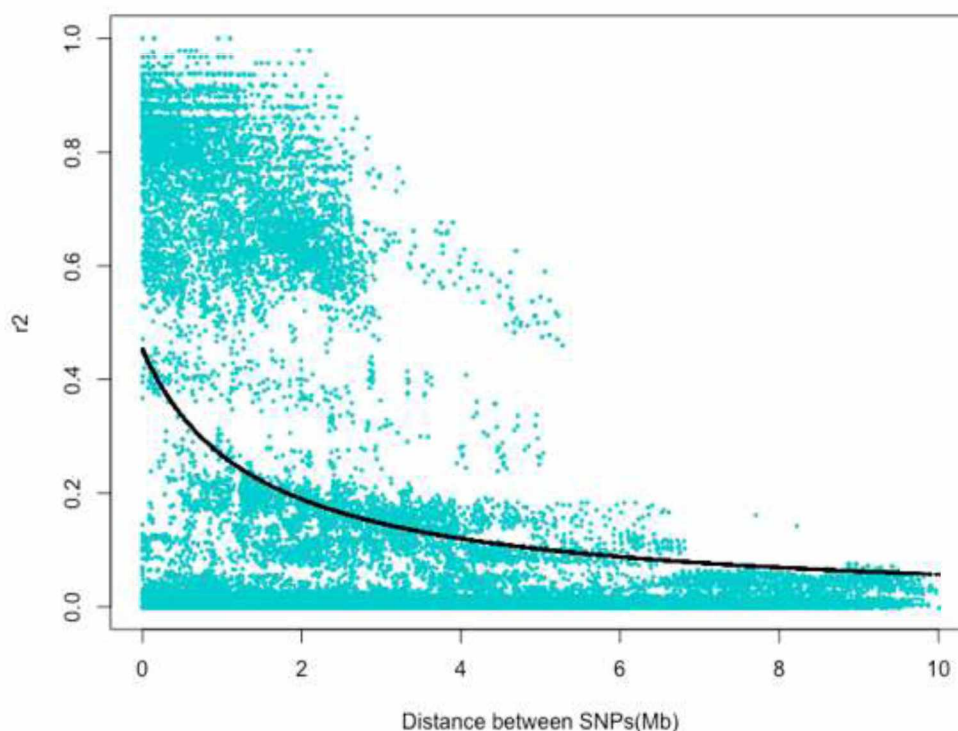


identificados com associação com a resistência de soja ao mofo branco em um  $p$ -value ajustado por FDR menor que 0,05. O gráfico LD (Figura 16) mostra que seis (azul, amarelo, vermelho) dos sete SNPs foram em alto LD ( $r^2 = 0,81-0,96$ ), enquanto que um (verde) foi de LD baixo-moderado ( $r^2 = 0,40 - 0,47$ ).



**Figura 16.** A. Representação gráfica dos níveis de Desequilíbrio de ligação (LD) para comparações par a par entre sítios polimórficos que apresentam resultados em todas amostras analisadas, obtido pelo software Haploview para resistência à *S. sclerotiorum*. B. Detalhe da localização do SNP S01\_33164666. C. Detalhe da localização dos SNPs S01\_34787519 e S01\_35073769. D. Detalhe da localização dos SNPs S01\_36048095, S01\_36044127, S01\_36038838 e S01\_36006734.

A curva de regressão de LD de uma região de 10 Mb contendo os sete SNPs significativos no cromossomo 1 apresenta alto decaimento de LD, mostrando decaimento em  $r^2$  a 0,2 a ser cerca de 1,8 Mb (Figura 17).



**Figura 17.** Decomposição do desequilíbrio de ligação intracromossômico entre par de marcadores do cromossomo 1 da distância física.

#### 4.7 Identificação de relações entre funções biológicas e SNPs

A resistência ou suscetibilidade à infecção de um genótipo vegetal particular dentro de uma espécie é determinada por linhas de defesa entrelaçadas, incluindo barreiras constitutivas e reações induzidas (THORDAL-CHRISTENSEN, 2003). O controle das reações induzidas estabelece importância na compreensão dos mecanismos pelos quais as plantas percebem os sinais ambientais e os transmite através da maquinaria celular para ativar respostas adaptativas (AFZAL; WOOD; LIGHTFOOT, 2008).

Para a identificação dos genes com função biológica relacionada com a defesa da planta em que os SNPs significativos estavam localizados, foi estipulada uma distância de até 15 Mb para a análise. Após a filtragem dos domínios de genes de proteínas não caracterizadas, 10 genes que continham os SNPs do cromossomo 1 possuíam proteínas conhecidas e 7 genes localizados no cromossomo 12 também apresentavam proteínas caracterizadas (Tabela 4).

**Tabela 4.** Genes candidatos dentro do bloco de desequilíbrio de ligação ou região genômica vinculada de polimorfismos de nucleotídeo único (SNPs) mais fortemente associados com a resistência à *S. sclerotiorum* em soja.

Gene	Cromossomo	Começo	Final	Distância do SNP (pb)	Anotação
Glyma.01G099600	1	33115843	33125726	48823	Serina/treonina quinase envolvida ao complexo ATG1 [ <i>Glycine max</i> ]
Glyma.01G099700	1	33208869	33217763	44203	Uracil fosforibosiltransferase [ <i>Glycine soja</i> ]
Glyma.01G099800	1	33261108	33274284	96442	Delta-1-pirrolina-5-carboxilato sintetase (isoforma X2) [ <i>Glycine max</i> ]
Glyma.01G104100	1	35076719	35085254	2950	Isocorismato sintase 2, tipo cloroplástica (isoforma X1) [ <i>Glycine max</i> ]
Glyma.01G104200	1	35104237	35111049	30468	BEL1 de interação ao homeodomínio da proteína 9 [ <i>Glycine max</i> ]
Glyma.01G106100	1	35881007	35883036	125727	Glutaciona S-transferase [ <i>Glycine soja</i> ]
Glyma.01G106200	1	35918034	35919778	88700	Proteína universal do estresse tipo A [ <i>Glycine soja</i> ]
Glyma.01G106300	1	35961504	35964818	45230	Xiloglucano endotransglicosilase/hidrolase (proteína 6) [ <i>Glycine max</i> ]
Glyma.01G106400	1	36051588	36052756	3493	Digalactosildiacilglicerol sintase 1, tipo cloroplástica [ <i>Glycine soja</i> ]
Glyma.01G106500	1	36103010	36105425	54915	Serina/treonina quinase tipo pakD [ <i>Glycine max</i> ]
Glyma.12G179200	12	33903699	33910581	65019	Regulador autoimune [ <i>Glycine soja</i> ]
Glyma.12G179300	12	33928417	33933941	40301	E3 ubiquitina ligase (NEURL1B) [ <i>Glycine soja</i> ]
Glyma.12G179400	12	33972888	33973824	4170	Serina carboxipeptidase tipo 51 (isoforma X1) [ <i>Glycine max</i> ]
Glyma.12G179500	12	33977725	33980896	9007	Alfa-farneseno sintase [ <i>Glycine max</i> ]
Glyma.12G179600	12	33986377	33992239	17659	Carboxipeptidase lisossomal Pro-X [ <i>Glycine max</i> ]
Glyma.12G179700	12	34002160	34006655	33442	lysosomal Pro-X carboxypeptidase-like [ <i>Glycine max</i> ]
Glyma.12G179800	12	34009158	34014157	40440	Protease aspártica (proteína 2) [ <i>Glycine max</i> ]

Fonte: Blastx.



Três proteínas dos domínios dos genes identificados não estão envolvidas na resposta das plantas à doença, sendo digalactosildiacilglicerol sintase 1, tipo cloroplástica, proteína encontrada na membrana dos tilacóides (GAUDE et al., 2004), uracil fosforibosiltransferase, envolvida na recuperação de uracil para a síntese de nucleotídeos de pirimidina (MAINGUET et al., 2009), e BEL1 de interação ao homeodomínio da proteína 9, fator de transcrição envolvido na regulação da função meristemática (KANRAR; ONGUKA; SMITH, 2006). Não existem informações funcionais sobre o papel, nas plantas, da classe de proteínas carboxipeptidase lisossomal Pro-X (Glyma.12G179600 e Glyma.12G179700) (TRIPATHI; SOWDHAMINI, 2006) e da proteína regulador autoimune (Glyma.12G179200). As demais proteínas estão envolvidas em alguma etapa da dinâmica da resistência a estresses bióticos.

A sinalização de defesa da planta é ativada pelos receptores de detecção de patógenos (DODDS; RATHJEN, 2010). Os receptores ativados frequentemente provocam sinais de transdução e tipicamente levam à modificação de proteínas que controlam a transcrição do gene e a ativação da expressão do gene associado com a defesa (LI et al., 2016).

No cromossomo 1, os genes Glyma.01G099600 e Glyma.01G106500 possuem anotação para as proteínas serina/treonina quinase envolvida ao complexo ATG1 e serina/treonina quinase tipo pakD, respectivamente, que participam do processamento do reconhecimento de padrões moleculares associados a patógenos (PAMPs) (AFZAL; WOOD; LIGHTFOOT, 2008), participando dos eventos primários da dinâmica molecular referente aos estágios de interação com o fungo *S. sclerotiorum* (WEI; CLOUGH, 2016).

A proteína delta-1-pirrolina-5-carboxilato sintetase (isoforma X2), codificada pelo gene Glyma.01G099800, está envolvida com processo de biossíntese de prolina. As funções da prolina na planta, relacionadas com a defesa primária, são *scavenger* de radicais livres (SARADHI et al., 1995) e relacionadas com defesas induzidas pós-infecção, por ser constituinte principal de proteínas da parede celular (NANJO et al., 1999), sugerindo reforço da parede celular. A proteína xiloglucano endotransglicosilase/hidrolase (proteína 6) (Glyma.01G106300) também é um componente essencial da parede celular primária e, portanto, participa na construção da parede celular de tecidos em crescimento (HARA et al., 2014).

A enzima isocorismato sintase 2, tipo cloroplástica (isoforma X1) (Glyma.01G104100) está envolvida na síntese de ácido salicílico, necessário para a

resistência adquirida local e sistêmica (GARCION et al., 2008), que está localizada a 2950 pb do SNP S01\_35073769.

No decorrer do estresse biótico, as proteínas glutathione S-transferase (Glyma.01G106100) são induzidas por infecção ou por respostas que invocam reações de defesa de plantas, bem como por estresse osmótico e temperaturas extremas (MARRS, 1996).

Tem sido demonstrada a função da proteína universal do estresse tipo A (Glyma.01G106200) na resposta a diversos estresses ambientais, tais como salinidade, seca, frio, calor e estresse oxidativo, exibindo também forte função de chaperona em condições de estresse (JUNG et al., 2015).

No Cromossomo 12, todas as proteínas identificadas estão relacionadas com à resistência a doença. A proteína E3 ubiquitina ligase (NEURL1B) (Glyma.12G179300) desempenha papel central na sinalização celular que regula os processos, incluindo a degradação da proteína e resposta imunológica. A ubiquitinação é um componente de várias respostas imunes. Sem o bom funcionamento da ubiquitina, a invasão de patógenos e outras moléculas nocivas aumentaria dramaticamente devido ao enfraquecimento das defesas imunitárias (TRUJILLO; SHIRASU, 2010).

A enzima protease aspártica (proteína 2) (Glyma.12G179800) está relacionada com respostas ao estresse (SIMÕES; FARO, 2004) e seu acúmulo é resposta a ataques de patógenos, sugerindo seu potencial envolvimento na ativação de mecanismos de resistência induzidos pela geração de um peptídeo sinal (XIA et al., 2004).

A classe de proteínas serina carboxipeptidase tipo 51 (isoforma X1) (Glyma.12G179400) possui função de degradação e processamento de proteínas para incluir atividades tais como a produção de metabólitos secundários (FRASER; RIDER; CHAPPLE, 2005), que está localizada a 4.170 pb do único SNP do cromossomo 12.

A proteína alfa-farneseno sintase (Glyma.12G179500) está relacionada à produção de etileno (WATKINS; BARDEN; BRAMLAGE, 1993) e está localizada a 9.007 pb do SNP do cromossomo 12. Os mecanismos de fatores de transcrição e hormônios, atividade do sistema imunológico da planta, são regulados em parte por hormônios de sinalização, tais como o etileno (MOORE; LOAKE; SPOEL, 2011).

## 5 DISCUSSÃO

A resistência da soja ao fungo *S. sclerotiorum* é complexa em seu mecanismo patogênico e via de resistência, e é um caráter quantitativo (KIM; DIERS, 2000; VUONG; DIERS; HARTMAN, 2008). O comprimento da necrose no caule, após a inoculação artificial do fungo, segue uma distribuição normal (ARAHANA et al., 2001; BASTIEN et al., 2014). Com isso, pode-se inferir que a resistência da soja ao fungo pode ser uma característica quantitativa. Não há resistência completa à *S. sclerotiorum* na soja porque a resistência é controlada quantitativamente por numerosos genes ou *loci* de traços quantitativos (QTLs) (IQUIRA; SONAH; BELZILE, 2015).

A análise de componentes principais indicou que o conjunto de genótipos estudado tinha quatro subpopulações distinguíveis. As subpopulações FT sementes, LAGER-UFU, TMG-TC e TMG-P1C142 refletem a diferença dos programas de melhoramento, com diferentes estratégias de seleção, refletindo em divergências nos genótipos de soja. As subpopulações, com grande parte da variação genética compartilhadas, geralmente variam geneticamente porque estão sujeitas a pressões seletivas diferentes em ambientes distintos (SADAVA et al., 2009).

A avaliação da decomposição de LD fornece uma estimativa do número de marcadores necessários para detectar QTLs. O método mais utilizado para descrever o decaimento de LD consiste em realizar uma regressão de  $r^2$  contra a distância física ou genética e encontrar a intersecção da regressão com um limiar estabelecido, que no caso deste estudo foi de  $r^2 = 0,2$ , por ser o valor mais utilizado em pesquisas anteriores. Em nossa coleção de genótipos de soja, LD decaiu em 1,8 Mb, ao longo do cromossomo 1.

É relatado que populações que foram submetidas à intensa seleção, como o melhoramento genético, podem sofrer extensão do decaimento de LD pela perda de diversidade genética (HYTEN et al., 2007; LAM et al., 2010). Essa redução pode ser detectada nos maiores valores de LD em  $r^2 = 0,2$  deste estudo com valor de 1,8 Mb e nas pesquisas de Bastien e colaboradores (2014) com valor de 0,85 Mb, que utilizaram populações oriundas de programas de melhoramento e os menores valores dos estudos de Zhao e colaboradores (2015) com 0,133 Mb e Iquira e colaboradores (2015) com 0,5 Mb, com análise de populações primitivas e elite.

Os estudos de mapeamento por associação já realizados de resistência à *S. sclerotiorum* utilizaram diferentes quantidades de indivíduos. Para o estudo genômico, quanto maior a população, maior a probabilidade de detecção de QTLs e marcadores

relacionados com os mesmos (SCHUSTER; CRUZ, 2013), e isso pode ser comprovado no estudo de Zhao e colaboradores (2015) com a identificação de 25.179 SNPs em 458 genótipos e, neste estudo com 324 genótipos e 35.955 SNPs. Já com número reduzido de genótipos, o número de marcadores diminui, como pode ser detectado nas pesquisas de Bastien e colaboradores e Iquirá e colaboradores, com 7.864 SNPs e 130 linhagens e 8.339 SNPs e 101 genótipos, respectivamente.

Na análise de mapeamento por associação, um dos testes estatísticos utilizados é o teste  $\chi^2$ , que é demonstrado no gráfico quantil-quantil (Q-Q plot). O gráfico mostra a ocorrência de divergência entre os valores de significância observados e aqueles esperados na hipótese nula do teste. Segundo Paterson e colaboradores (2006), desvios da diagonal do gráfico (hipótese nula) geralmente sugerem associação verdadeira dos SNPs com o fenótipo, exceto em caso que o desvio ocorre de forma generalizada, indicando inflação do teste.

Como a maioria dos pontos estão situados na linha diagonal, sugere-se que o Modelo Linear Misto Comprimido está adequado para a contabilização de falsos positivos decorrentes da estrutura da população e parentesco. Os pontos desviados da diagonal no canto superior direito do gráfico são provavelmente verdadeiros positivos. Nota-se que há oito pontos mais desviados, confirmando a associação dos oito SNPs com a característica estudada (Figura 14).

No total, a análise GWAS dos dados de GBS possibilitou identificar oito SNPs que foram significativamente associados com a resistência à *S. sclerotiorum*, e que, considerando as análises de desequilíbrio de ligação, estes oito SNPs representam provavelmente cinco QTLs, quatro no cromossomo 1 e um no cromossomo 12.

Como observado no resultado do software Haploview, os SNPs S01\_36048095, S01\_36044127, S01\_36038838, S01\_36006734 são estatisticamente inseparáveis, por isso podem estar localizados no mesmo QTL, assim, o cromossomo 1 possui quatro QTLs e o cromossomo 12 contém um QTL.

Os quatro QTLs mapeados no cromossomo 1 estão localizados nas mesmas regiões do estudo dos pesquisadores Bastien, Sonah e Belzile (2014) que também identificaram SNPs para resistência à *S. sclerotiorum*, mas não os mesmos, apoiando os dados apresentados, acrescentando maior credibilidade de que esta região no cromossomo 1 está associada a aumentar a resistência à doença do mofo branco.

O SNP significativo no cromossomo 12 parece ser uma nova descoberta, pois não foram encontrados relatos semelhantes em mapeamentos relacionados com a

resistência a este patógeno neste cromossomo, como pode ser observado em estudos publicados sobre estudos de associação amplos (GWAS) com soja a respostas a mofo branco. Em estudos prévios que identificaram QTLs relacionados com a resistência, foram relatados nos cromossomos 1, 15, 19 e 20 (BASTIEN; SONAH; BELZILE, 2014); 13 (ZHAO et al., 2015); 3, 8 e 20 (IQUIRA; SONAH; BELZILE, 2015). O uso de um conjunto diferente de genótipos e método de inoculação em relação a trabalhos anteriores pode explicar essa falta de sobreposição das regiões cromossômicas.

Neste estudo, em testes para detecção de associações entre marcadores SNP e resistência à *S. sclerotiorum*, a característica exibiu herdabilidade de 33,5% e, no estudo de Iquira e colaboradores (2015) relacionado com resistência ao mesmo fungo foi de 67% (IQUIRA; SONAH; BELZILE, 2015), sendo que os valores possuíram médio valor de herdabilidade, ou seja, há um nível de correspondência entre o fenótipo e o valor genético, com média correlação entre o valor fenotípico e genotípico (CRUZ, 2005). O valor mediano pode ser atribuído ao comportamento puramente quantitativo desse caráter em função do grande número de genes que o controlam, permitindo maior influência ambiental e, conseqüentemente, diminuição da relação entre a variância genética e fenotípica (ROSSMANN, 2001).

Pode-se detectar que 11 genes estavam relacionados com diferentes vias metabólicas envolvidas na defesa da planta ao fungo *S. sclerotiorum*, sendo sete no cromossomo 1 e quatro no cromossomo 12. No cromossomo 1, proteínas de defesa codificadas por genes da planta estão envolvidas no reconhecimento do fungo na membrana celular da planta (PAMPs), biossíntese de prolina, suposto reforço da parede celular, via do ácido salicílico e envolvimento na resposta ao estresse biótico. Já no cromossomo 12, há relação com sinalização celular, resposta a ataque de patógenos, formação de metabólitos secundários e na via do etileno. Essas relações evidenciam forte associação desses marcadores genéticos com a resistência ao fungo causador do mofo branco.

Os genótipos mais resistentes identificados em nosso estudo (P1C142-158.243, BRS Charrua RR, FMT05-40.907/1, IAC 100, FT-1, P1C142-128.686, PI194639, P1C142-112.208, P1C142-116.940, V-Max RR, P1C142-169.621, P1C142-156.243, EMGOPA 316, BRSGO Araçu, P1C142-170.130, TMG 1188 RR (GENO\_TKG\_35) e NK S19-90) podem ser introduzidos em programas de melhoramento de soja, pois podem possuir alelos de resistência para os marcadores SNP mais significativos. Assim, os melhoristas poderiam selecionar um ou mais genótipos para introduzir essa

resistência em seu programa. No entanto, a resistência global dos genótipos parece ser controlada por um número relativamente elevado de *loci*. Devido a este elevado número, o melhoramento para a resistência à *S. sclerotiorum* provavelmente requererá estratégias capazes de explorar múltiplos QTLs, como a seleção genômica, e os SNPs identificados neste estudo podem fornecer possíveis marcadores que poderiam ser utilizados na seleção de soja com maior resistência ao mofo branco.

Os SNPs associados com a resistência poderão ser utilizados pelos programas de melhoramento de soja para identificar linhagens/cultivares com genes de interesse para a resistência ao fungo do mofo branco, acelerando o processo de melhoramento através de seleção assistida por marcadores ou podem ser incorporados em estratégias de seleção genômica. Além disso, SNPs significativos podem auxiliar no conhecimento sobre a função biológica do polimorfismo e na relação com a resistência.

## 6 CONCLUSÕES

Em uma coleção de genótipos de soja foi realizado um estudo de mapeamento de associação, auxiliado pela abordagem de genotipagem por sequenciamento, para identificar *loci* associados à resistência a *S. sclerotiorum* em soja. A descoberta de marcadores do tipo SNP mapeados perto de QTLs de resistência à doença previamente descobertos comprovam que esta abordagem é um método experimental valioso com aplicações potencialmente amplas para genética de soja. Foi confirmado um *locus* de resistência no cromossomo 1 previamente identificado, com sete SNPs não identificados anteriormente neste cromossomo, e foi identificado um aparente novo QTL no cromossomo 12.

Estudos posteriores serão necessários para confirmação da verdadeira ligação dos marcadores SNP aos QTLs previamente detectados para resistência à *S. sclerotiorum*, principalmente para o marcador do cromossomo 12.

## REFERÊNCIAS

- ACQUAHH, G. **Principles of plant genetics and breeding**. Malden: Blackwell Publishing, 2007. 569 p.
- AFZAL, A. J.; WOOD, A. J.; LIGHTFOOT, D. A. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. **Molecular Plant-Microbe Interactions**, Saint Paul, v. 21, n. 5, p. 507-517, 2008.
- ALTSCHUL, S. F.; MADDEN, T. L.; SCHÄFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, Oxford, v. 25, v. 17, p. 3389-3402, 1997.
- ANDOLFATTO, P.; DAVISON, D.; EREZYILMAZ, D.; HU, T. T.; MAST, J.; SUNAYAMA-MORITA, T.; STERN, D. L. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. **Genome Research**, Cold Spring Harbor, v. 21, n. 4, p. 610-617, 2011.
- ARAHANA, V. S.; GRAEFA, G. L.; SPECHTA, J. E.; STEADMAND, J. R.; ESKRIDGE, K. M. Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in soybean. **Crop science**, Madison, v. 41, n. 1, p. 180-188, 2001.
- AZEVEDO NETO, J. O. O. **Troost**: busca de interações entre trios de SNPs em estudos de associação de genoma inteiro. 2013. 154 f. Tese (Doutorado) – Instituto de Biociências, Universidade de São Paulo, São Paulo, 2013.
- BALARDIN, R. S. **Doenças da soja**. Santa Maria: UFMS, 2002. 107 p.
- BAN, H. J. HEO, J. Y.; OH, K. S.; PARK, K. J. Identification of type 2 diabetes-associated combination of SNPs using support vector machine. **BMC genetics**, London, v. 11, n. 1, p. 11-26, 2010.
- BARDIN, S. D.; HUANG, H. C. Research on biology and control of *Sclerotinia* diseases in Canada. **Canadian Journal of Plant Pathology**, Ottawa, v. 23, n. 1, p. 88-98, 2001.
- BARRETT, J. C.; FRY, B.; MALLER, J.; DALY, M. J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, Oxford, v. 21, n. 2, p. 263-265, 2005.
- BASTIEN, M.; SONAH, H.; BELZILE, F. J. Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. **The Plant genome**, Madison, v. 7, n. 1, 2014.
- BERNARDO, R.; YU, J. Prospects for genome wide selection for quantitative traits in maize. **Crop Science**, Madison, v. 47, n. 3, p. 1082–1090, 2007.



BOLAND, G. J.; HALL, R. Index of plant hosts of *Sclerotinia sclerotiorum*. **Canadian Journal of Plant Pathology**, Ottawa, v. 16, n. 2, p. 93-108, 1994.

BONATO, E. R. **Estresses em soja**. Passo fundo: Embrapa Trigo, 2000. 254 p.

BOLTON, M. D.; THOMMA, P. H. J. B.; NELSON, B.D. *Sclerotinia sclerotiorum* (Lib.) de Bary: Biology and molecular traits of a cosmopolitan pathogen. **Molecular Plant Pathology**, Oxford, v. 7, n. 1, p. 1-16, 2006.

BRADBURY, P. J.; ZHANG, Z.; KROON, D. E.; CASSTEVEN, T. M.; RAMDOSS, Y.; BUCKLER, E. S. TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics**, Oxford, v. 23, n. 19, p. 2633-2635, 2007.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Projeções do agronegócio**: Brasil 2013/2014 a 2023/2024. Brasília, 2014. 100 p.

BROOKES, A. J. The essence of SNP. **Gene**, Netherlands, v. 234, n. 2, p. 177-186, 1999.

BROWNING, B. L.; BROWNING, S. R. Genotype imputation with millions of reference samples. **The American Journal of Human Genetics**, Baltimore, v. 98, n. 1, p. 116-126, 2016.

CASSETARI NETO, D.; MACHADO, A. Q.; SILVA, R. A. **Manual de doenças de soja**. São Paulo: Cheminova Brasil, 2010. 57 p.

CAMARGO, T. V. Boletim de Pesquisa de Soja. **Podridão Branca da Haste ou Podridão de Sclerotinia – *Sclerotinia sclerotiorum***. Rondonópolis: Fundação Mato Grosso, 2002. 147 p.

CHANG, H. X.; BROWN, P. J.; LIPKA, A. E.; DOMIER, L. L.; HARTMAN, G. L. Genome-wide association and genomic prediction identifies associated loci and predicts the sensitivity of Tobacco ringspot virus in soybean plant introductions. **BMC Genomics**, London, v. 17, n. 153, 2016.

CLARKSON, J. P.; STAVELEY, J.; PHELPS, K.; YOUNG, C. S.; WHIPPS, J. M. Ascospore release and survival in *Sclerotinia sclerotiorum*. **Mycological Research**, Cambridge, v. 107, n. 2, p. 213-222, 2003.

COMPANHIA NACIONAL DE ABASTECIMENTO (Brasil). **Acompanhamento safra brasileira grãos**, Brasília, DF, 2016a. 182 p. (Safr 2015/16, v. 12, n. 6).

COMPANHIA NACIONAL DE ABASTECIMENTO (Brasil). **Acompanhamento safra brasileira grãos**, Brasília, DF, 2016b. 164 p. (Safr 2016/17 v. 4, n.1).

CRUZ, C. D. **Princípios de genética quantitativa**. 1. ed. Viçosa, MG: Ed. UFV, 2005. 391 p.

CUNHA, W. G. **Resistência a *Sclerotinia sclerotiorum* em plantas de soja geneticamente modificadas para expressar o gene da oxalato descarboxilase de**

***Flammulina velutipes***. 2010. 90 f. Tese (Doutorado) - Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF, 2010.

DE LOS CAMPOS, G.; HICKEY, J. M.; PONG-WONG, R.; DAETWYLER, H. D.; CALUS, M. P. L. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, Austin, v. 193, p. 327–345, 2013.

DE WIT, P. L. G. M. Fungal avirulence genes and plant resistance genes: unraveling the molecular basis of gene-for-gene interactions. **Advances in Botanical Research**, New York, v. 21, n. 1, p. 147-185, 1995.

DODDS, P. N.; RATHJEN, J. P. Plant immunity: Towards an integrated view of plant–pathogen interactions. **Nature Reviews Genetics**, London, v. 11, n. 8, p. 539–548, 2010.

DOYLE, J. J.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, Rockville, v. 12, n. 1, p. 13-15, 1990.

DUDLEY, J. W. Molecular markers in plant improvement manipulation of genes affecting quantitative traits. **Crop Science**, Madison, v. 33, n. 4, p. 660-668, 1993.

ELSHIRE, R. J.; GLAUBITZ, J. C.; SUN, Q.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E. S.; MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS ONE**, San Francisco, v. 6, n. 5, 2011.

EMBRAPA. **Tecnologias de produção de soja**: Região Central do Brasil 2012 e 2013. Londrina: Embrapa Soja, 2011. 261 p. (Sistemas de Produção/ Embrapa Soja, n. 15).

EMBRAPA. **Tecnologias de produção de soja**: Região Central do Brasil 2014. Londrina: Embrapa Soja, 2013. 265 p. (Sistema de Produção/ Embrapa Soja, n. 16).

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4. ed. Edinburgh: Longman Group Limited, 1996. 464 p.

FALEIRO, F. G. Aplicações de marcadores moleculares como ferramenta auxiliar em programas de conservação, caracterização e uso de germoplasma e melhoramento genético vegetal. In: FALEIRO, F. G.; ANDRADE, S. R. M.; REIS JUNIOR, F. B. R. **Biotechnologia estado da arte e aplicações na agricultura**. Planaltina, DF: Embrapa Cerrados, 2011a. p. 55-118

FALEIRO, F. G. **Marcadores genéticos-moleculares aplicados a programas de conservação e uso de recursos genéticos**. Planaltina, DF: Embrapa Cerrados, 2011b. 102 p.

FALEIRO, F. G. Princípio científico e análises genéticas utilizando marcadores moleculares. In: FALEIRO, F. G.; ANDRADE, S. R. M.; REIS JUNIOR, F. B. R. **Biotechnologia estado da arte e aplicações na agricultura**. Planaltina, DF: Embrapa Cerrados, 2011c. p. 31-52

- FERREIRA, D. F. Sisvar: a computer statistical analysis system. **Ciência e Agrotecnologia**, Lavras, v. 35, n. 6, p. 1039-1042, 2011.
- FRASER, C. M.; RIDER, L. W.; CHAPPLE, C. An expression and bioinformatics analysis of the Arabidopsis serine carboxypeptidase-like gene family. **Plant Physiology**, Bethesda, v. 138, n. 2, p. 1136–1148, 2005.
- GAUDE, N.; TIPPMANN, H.; FLEMETAKIS, E.; KATINAKIS, P.; UDVARDI, M.; DOERMANN, P. The galactolipid digalactosyldiacylglycerol accumulates in the peribacteroid membrane of nitrogen-fixing nodules of Soybean and Lotus. **The Journal of Biological Chemistry**, Bethesda, v. 279, n. 33, p. 34624-34630, 2004.
- GARCIA, R. A.; JULIATTI, F. C. Avaliação da resistência da soja a *Sclerotinia sclerotiorum* em diferentes estádios fenológicos e períodos de exposição ao inóculo. **Tropical Plant Pathology**, Brasília, v. 37, n. 3, p. 196-203, 2012.
- GARCION, C.; LOHMANN, A.; LAMODIERE, E.; CATINOT, J.; BUCHALA, A.; DOERMANN, P.; METRAUX, J. P. Characterization and biological function of the ISOCHORISMATE SYNTHASE2 gene of Arabidopsis. **Plant Physiology**, Bethesda, v. 147, n. 3, p. 1279-1287, 2008.
- GEFFROY, V.; SICARD, D.; OLIVEIRA, J. C. F.; SÉVIGNAC, M.; COHEN, S.; GEPTS, P.; NEEMA, C.; LANGIN, T.; DRON, M. Identification of an ancestral resistance gene cluster involved in the coevolution process between *Phaseolus vulgaris* and its fungal pathogen *Colletotrichum lindemuthianum*. **Molecular Plant-Microbe Interactions**, Saint Paul, v. 12, n. 9, p. 774-782, 1999.
- GLAUBITZ, J. C.; CASSTEVEN, T. M.; LU, F.; HARRIMAN, J.; ELSHIRE, R. J.; SUN, Q.; BUCKLER, E. S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. **PLoS ONE**, San Francisco, v. 9, n. 2, 2014.
- GOMES, P. **A soja**. 5. ed. São Paulo: Nobel, 1990. 152 p.
- GUO, X.; WANG, D.; GORDON, S. G.; HELLIWELL, E.; SMITH, T.; BERRY, S. A.; ST. MARTIN, S. K.; DORRANCE, A. E. Genetic mapping of QTLs underlying partial resistance to in soybean PI 391589A and PI 391589B. **Crop Science**, Madison, v. 48, n. 3, p. 1129–1139, 2008.
- HAN, F.; KATT, M.; SCHUH, W.; WEBB, D. M. **QTL controlling Sclerotinia stem rot resistance in soybean**. US 7250552 B2. 7 jun. 2002, 31 jul. 2007. United States patent application publication.
- HARA, Y.; YOKOYAMA, R.; OSAKABE, K.; TOKI, S.; NISHITANI, K. Function of xyloglucan endotransglucosylase/hydrolases in rice. **Annals of Botany**, London, v. 114, n. 6, p. 1309-1318, 2014.
- HASTBACKA, J.; CHAPPELLE, A.; MAHTANI, M. M.; CLINES, G.; REEVE-DALY, M. P.; DALY, M.; HAMILTON, B. A.; KUSUMI, K.; TRIVEDI, B.; WEAVER, A. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning

by fine-structure linkage disequilibrium mapping. **Cell**, Cambridge, v. 78, n. 1, p. 1073–1087, 1994.

HILL, W. G.; WEIR, B. S. Variances and covariances of squared linkage disequilibrium in finite populations. **Theoretical Population Biology**, New York, v. 33, n. 1, p. 54–78, 1988.

HUYNH, T. T.; BASTIEN, M.; IQUIRA, E.; TURCOTTE, P.; BELZILE, F. Identification of QTLs associated with partial resistance to white mold in soybean using field-based inoculation. **Crop Science**, Madison, v. 50, n. 3, p. 969–979, 2010.

HYMOWITZ, T.; SINGH, R. J.; KOLLIPARA, K. P. Biosystematics of the genus *Glycine*. **Soybean Genetics Newsletter**, Baltimore, v. 24, p. 119-120, 1997.

HYTEN, D. L.; CHOI, I. Y.; SONG, Q.; SHOEMAKER, R. C.; NELSON, R. L.; COSTA, J. M.; SPECHT, J. E.; CREGAN, P. B. Highly variable patterns of linkage disequilibrium in multiple soybean populations. **Genetics**, Austin, v. 175, n. 4, p. 1937-1944, 2007.

IQUIRA, E.; SONAH, H.; BELZILE, F. J. Association mapping of QTLs for *Sclerotinia* stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. **BMC Plant Biology**, London, v. 15, n. 5, 2015.

JOHNSTON, F. H.; LENEXA, M. K.; WEST DES MOINES, W. S.; ZIONSVILLE, D. M. W. **QTL controlling *Sclerotinia* stem rot resistance in soybean**. US 8956872, 17 fev. 2015.

JULIATTI, F. C.; CAIRES, A. M.; JULIATTI, B. C. M.; BORIN, M. R.; SOUZA FILHO, D. J. Reação de genótipos de soja transgênicos e convencionais à podridão-branca da haste. **Bioscience Journal**, Uberlândia, v. 29, n. 3, p. 921-931, 2013.

JULIATTI, F. C.; FIGUEIRÓ, A. A.; GARCIA, R. A.; SANTOS, J. B.; MACHADO, J. C.; POZZA, E. A.; LÔBO JÚNIOR, M.; CIVARDI, E. A.; PAULA JÚNIOR, T. J.; VIEIRA, R. F.; LEHNER, M. S.; LIMA, R. C.; JACCOUD FILHO, D. S. *Sclerotinia sclerotiorum* e mofo branco: estudos básicos e aplicados. **Revisão Anual de Patologia de Plantas**, Passo Fundo, v. 23, p. 159-194, 2015.

JULIATTI, F. C.; JULIATTI, F. C. **Podridão branca da haste da soja: Manejo e uso de fungicidas em busca da sustentabilidade nos sistemas de produção**. Uberlândia: Composer, 2010. 53 p.

JULIATTI, F. C.; SAGATA, E.; JACCOUD FILHO, D. S.; JULIATTI, B. C. M. Métodos de inoculação e avaliação da resistência de genótipos de soja à *Sclerotinia sclerotiorum*. **Bioscience Journal**, Uberlândia, v. 30, n. 4, p. 958-968, 2014.

JULIO, V. A. **Manejo de mofo branco na cultura da soja**. 2016. Disponível em: <<http://www.pioneersementes.com.br/blog/119/manejo-de-mofo-branco-na-cultura-da-soja>>. Acesso em: 21 out. 2016.

JUNG, Y. J.; MELENCION, S. M. B.; LEE, E. S.; PARK, J. H.; ALINAPON, C. V.; OH, H. T.; YUN, D. J.; CHI, Y. H.; LEE, S. Y. Universal stress protein exhibits a redox-dependent chaperone function in *Arabidopsis* and enhances plant tolerance to heat shock and oxidative stress. **Frontiers in Plant Science**, Lausanne, v. 6, n. 1141, 2015.

KANRAR, S.; ONGUKA, O.; SMITH H.M.S. *Arabidopsis* inflorescence architecture requires the activities of KNOX-BELL homeodomain heterodimers. **Planta**, Berlin, v. 224, n. 5, p. 1163-1173, 2006.

KARL, A. C.; NASSER, L. C. B.; CAFÉ FILHO, A. C. Mofo branco do feijoeiro, *Sclerotinia sclerotiorum* (Lib.) de Bary, em áreas irrigadas nos cerrados. In: ENCONTRO DE FITOPATOLOGIA, 2., Viçosa, MG, 1997. **Palestras...** Viçosa, MG: UFV, 1997. p. 18-23.

KIM, H. S.; SNELLER, C. H.; DIERS, B. W. Evaluation of soybean cultivars for resistance to *Sclerotinia* stem rot in field environments. **Crop Science**, Madison, v. 39, n. 1, p. 64-68, 1999.

KIM, H. S.; DIERS, B. W. Inheritance of partial resistance to *Sclerotinia* stem rot in soybean. **Crop Science**, Madison, v. 40, n. 1, p. 55-61, 2000.

KU, C. S.; LOY, E. Y. The pursuit of genome-wide association studies: where are we now? **Journal of Human Genetics**, New York, v. 55, n. 4, p. 195-206, 2010.

KURLE, J. E.; GRAU, C. R.; OPLINGER, E. S.; MENGISTU, A. Tillage, crop sequence, and cultivar effects on *Sclerotinia* stem rot incidence and yield in soybean. **Agronomy Journal**, Madison, v. 93, n. 5, p. 973-982, 2001.

LAM, H. M.; XU, X.; LIU, X.; CHEN, W.; YANG, G.; WONG, F. L.; LI, M. W.; HE, W.; QIN, N.; WANG, B.; LI, J.; JIAN, M.; WANG, J.; SHAO, G.; WANG, J.; SUN, S. S.; ZHANG, G. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. **Nature Genetics**, New York, v. 42, n. 12, p. 1053-1059, 2010.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, New York, v. 4, n. 9, p. 357-359, 2012.

LARA, L. A. C.; SANTOS, J. B.; VELOSO, J. S.; BALESTRE, M.; ALVES, F. C.; LEITE, M. E. Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in carioca common bean by the moving away method. **ISRN Molecular Biology**, Cairo, v. 2014, 2014.

LEITE, R. M. V. B. C. **Ocorrência de doenças causadas por *Sclerotinia sclerotiorum* em girassol e soja**. Londrina: Embrapa Soja, 2005. 3 p. (Comunicado técnico n. 76).

LI, B.; MENG, X.; SHAN, L.; HE, P. Transcriptional regulation of pattern-triggered immunity in plants. **Cell Host & Microbe**, Cambridge, v. 19, n. 5, p. 641-650, 2016.

- LI, D.; SUN, M.; HAN, Y.; TENG, W.; LI, W. Identification of QTL underlying soluble pigment content in soybean stems related to resistance to soybean white mold (*Sclerotinia sclerotiorum*). **Euphytica**, Wageningen, v. 172, n. 1, p. 49–57, 2009.
- LIPKA, A. E.; TIAN, F.; WANG, Q.; PEIFFER, J.; LI, M.; BRADBURY, P. J.; GORE, M. A.; BUCKLER, E. S.; ZHANG, Z. GAPIT: genome association and prediction integrated tool. **Bioinformatics**, Oxford, v. 28, n. 18, p. 2397–2399, 2012.
- MACHADO, A. Q.; CASSETARI NETO, D. Epidemia branca. **Cultivar Grandes Culturas**, Pelotas, v. 12, n. 130, p. 20-23, 2010.
- MAINGUET, S. E.; GAKIÈRE, B.; MAJIRA, A.; PELLETIER, S.; BRINGEL, F.; GUÉRARD, F.; CABOCHE, M.; BERTHOMÉ, R.; RENOU, J. P. Uracil salvage is necessary for early Arabidopsis development. **Plant Journal**, Oxford, v. 60, n. 2, p. 280–291, 2009.
- MARRS, K. A. The functions and regulation of glutathione S-transferases in plants. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 47, p. 127–158, 1996.
- MATIELLO, R. R.; BARBIERI, R. L.; CARVALHO, F. I. F. Resistência das plantas a moléstias fúngicas. **Ciência Rural**, Santa Maria, v. 27, n. 1, p. 161-168, 1997.
- MESQUITA, A. C. O.; FIGUEIRÓ, A. A.; COUTO, K. R.; OLIVEIRA, M. F.; JULIATTI, F. C. Optimization of soybean DNA extraction under different storage and development periods. **Bioscience Journal**, Uberlândia, v. 31, n. 4, p. 1102-1106, 2015.
- MEUWISSEN, T. H.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Austin, v. 157, n. 4, p. 1819-1829, 2001.
- MEYER, M. C.; CAMPOS, H. D. Guerra ao mofo. **Cultivar Grandes Culturas**, Pelotas, v. 11, n. 120, p. 16-18, 2009.
- MEYER, M. C. Manejo de mofo branco em soja. **Boletim Passarela da Soja**, Luís Eduardo Magalhães, v. 1, n. 1, p-16, 2009.
- MEYER, M. C.; CAMPOS, H. D.; GODOY, C. V.; UTIAMADA, C. M. **Ensaio cooperativos de controle químico de mofo branco na cultura da soja: safras 2009 a 2012**. Londrina: Embrapa soja, 2014. 100 p. (Documentos/ Embrapa soja, n. 345).
- MEYER, M. C.; CAMPOS, H. D.; NUNES JUNIOR, J.; VENANCIO, W. S.; GODOY, C. V. Chemical control of white mold (*Sclerotinia sclerotiorum*) on soybean in Brazil. **Acta Phytopathologica Sinica**, Budapest, v. 43 p. 137. 2013.
- MILACH, S. C. K. **Marcadores moleculares em plantas**. Porto Alegre: UFRGS, 1998.
- MOELLERS, T. C. **Genome-wide association and epistasis studies of *Sclerotinia sclerotiorum* resistance in soybean**. 2016. Disponível em:

<<http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=6046&context=etd>>. Acesso em: 10 dez. 2016.

MOORE, J. W.; LOAKE, G. J.; SPOEL, S. H. Transcription dynamics in plant immunity. **The Plant Cell**, Rockville, v. 23, n. 8, p. 2809–2820, 2011.

MYLES, S.; PEIFFER, J.; BROWN, P. J.; ERSOZ, E. S.; ZHANG, Z.; COSTICH, D. E.; BUCKLER, E. S. Association mapping: Critical considerations shift from genotyping to experimental design. **The Plant Cell**, Rockville, v. 21 n. 8 p. 2194-2202, 2009.

NANJO, T.; KOBAYASHI, M.; YOSHIBA, Y. SANADA, Y.; WADA, K.; TSUKAYA, H.; KAKUBARI, Y.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Biological functions of proline in morphogenesis and osmotolerance revealed in antisense transgenic *Arabidopsis thaliana*. **The Plant Journal**, Oxford, v. 18, n. 2, p. 185-193, 1999.

NELSON, R. R. Stabilizing racial populations of plant pathogens by use of resistance genes. **Journal of Environmental Quality**, Madison, v. 1, n. 3, p. 220-227, 1972.

OLAZAR, M. R. **Uma metodologia para a descoberta de marcadores genéticos**. 2013. 133 f. Tese (Doutorado) – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013.

OLIVEIRA, F. C. **Um método para seleção de atributos em dados genômicos**. 2015. 273 f. Tese (Doutorado) - Instituto de Ciências Exatas, Universidade de Juiz de Fora, Juiz de Fora, 2015.

ORAGUZIE, N. C.; RIKKERINK, E. H. A.; GARDINER, S. E.; SILVA, H. N. (eds). **Association mapping in plants**. Berlim: Springer, 2007. 278 p.

PANOUTSOPOULOU, K.; ZEGGINI, E. Finding common susceptibility variants for complex disease: past, present and future. **Briefings in functional genomics and proteomics**, Oxford, v. 8, n. 5, p. 345-352, 2009.

PARLEVLIET, J. E. Partial resistance of barley to leaf rust, *Puccinia Hordei* I. effect of cultivar and development stage on latent period. **Euphytica**, Wageningen, v. 24, n. 1, p. 21-27, 1975.

PATTERSON, N.; PRICE, A. L.; REICH, D. Population structure and eigenanalysis. **PLoS Genetics**, San Francisco, v. 2, n. 12, 2006.

PEARSON, T. A.; MANOLIO, T. A. How to interpret a genome-wide association study. **JAMA**, Chicago, v. 299, n. 11, p. 1335-1344, 2008.

PETZOLDT, R.; DICKSON, M. Straw test for resistance to white mold in beans. **Annual Report of the Bean Improvement Cooperative**, Fort Collins, v. 39, p. 142-143, 1996.

POLAND, J. A.; BROWN, P. J.; SORRELLS, M. E.; JANNINK, J. L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. **PLoS ONE**, San Francisco, v. 7, n. 2, 2012.

PRIESTLEY, R. H.; BAYLES, R. A. The contribution and value of resistant cultivars to disease control in cereals. In: CLIFFORD, B. C.; LESTER, E. (Ed.). **Control of plant diseases: Costs and benefits**. Oxford: Blackwell Scientific Publications, 1988. p. 53-65.

PURDY, L. H. *Sclerotinia sclerotiorum*: history, diseases and symptomatology, host range, geographic distribution, and impact. **Phytopathology**, Saint Paul, v. 69, n. 8, p. 875-880, 1979.

REMLINGTON, D. L.; THORNSBERRY, J. M.; MATSUOKA, Y.; WILSON, L. M.; WHITT, S. R.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M. M.; BUCKLER, E. S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 98, n. 20, p. 11479–11484, 2001.

ROSSMANN, H. **Estimativas de parâmetros genéticos e fenotípicos de uma população de soja avaliada em quatro anos**. 2001. 80 f. Dissertação (Mestrado) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2001.

SADAVA, D.; HELLER, H. C.; ORIAN, G. H.; PURVES, W. K.; HILLIS, D. M. **Vida: a ciência da biologia**. 8. ed. Porto Alegre: Artmed, 2009. v. 2.

SAGATA, E. **Métodos de inoculação e avaliação da resistência de genótipos de soja à *Sclerotinia sclerotiorum***. 2010. 58 f. Dissertação (Mestrado) – Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia, 2010.

SARADHI, P. P.; ALIA, A.; ARORA, S.; PRASAD, K. V. S. K. Proline accumulates in plants exposed to UV radiation and protects them against UV induced peroxidation. **Biochemical and Biophysical Research Communications**, Orlando, v. 209, n. 1, p. 1-5, 1995.

SCHUSTER, I.; CRUZ, C. D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. 2. ed. Viçosa, MG: UFV, 2008. 568 p.

SCHMUTZ, J.; CANNON, S. B.; SCHLUETER, J.; MA, J.; MITROS, T.; NELSON, W.; HYTEN, D. L.; SONG, Q.; THELEN, J. J.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, G. D.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYYA, M. K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.; LIBAULT, M.; SETHURAMAN, A.; ZHANG, X-C.; SHINOZAKI, K.; NGUYEN, H. T.; WING, R. A.; CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R. C.; JACKSON, S. A. Genome sequence of the palaeopolyploid soybean. **Nature**, London, v. 463, n. 7278, p. 178–183, 2010.



SEBASTIAN, S.; LU, H.; HAN, F.; KYLE, D.; HEDGES, B.; ROACH, M.; STEPHENS, P.; LIU, J. **Genetic loci associated with Sclerotinia tolerance in soybean**. US7790949 B2. 21 jul. 2009, 7 set. 2010. United States patent application publication.

SIMÕES, I.; FARO, C. Structure and function of plant aspartic proteinases. **European Journal of Biochemistry**, Berlin, v. 271, n. 11, p. 2067–2075, 2004.

STOREY, J. D. A Direct approach to false discovery rates. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Oxford, v. 64, n. 3, p. 479–498, 2002.

TANKSLEY, S. D. Mapping polygenes. **Annual Review of Genetics**, Palo Alto, v. 27, n. 1, p. 205–233, 1993.

TAYLOR, J. F.; COUTINHO, L. L.; HERRING, K. L.; GALLAGHER JUNIOR, D. S. BRENNEMAN, R. A.; BURNEY, M.; SANDERS, J. O.; TURNER, J. W.; SMITH, S. B.; MILLER, R. K.; SAVELL, J. W.; DAVIS, S. K. Candidate gene analysis of GH1 for effects on growth and carcass composition of cattle. **Animal Genetics**, Oxford, v. 29, n. 3, p. 194–201, 1998.

THOMÉ, G. C. H.; MILACH, S. C. K.; CRUZ, R. P.; FEDERIZZI, L. C. Melhoramento para resistência parcial a moléstias fúngicas em cereais. **Ciência Rural**, Santa Maria, v. 29, n. 2, p. 365–371, 1999.

THORDAL-CHRISTENSEN, H. Fresh insights into processes of nonhost resistance. **Current Opinion in Plant Biology**, London, v. 6, n. 4, p. 351–357, 2003.

TORKAMANEH, D.; LAROCHE, J.; BELZILE, F. Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. **PLoS ONE**, San Francisco, v. 11, n. 8, 2016.

TRIPATHI, L. P.; SOWDHAMINI, R. Cross genome comparisons of serine proteases in Arabidopsis and rice. **BMC Genomics**, London, v. 7, n. 200, 2006.

TRUJILLO, M.; SHIRASU, K. Ubiquitination in plant immunity. **Current Opinion in Plant Biology**, London, v. 13, n. 4, p. 402–408, 2010.

TURNER, S.; ARMSTRONG, L. L.; BRADFORD, Y.; CARLSON, C. S.; CRAWFORD, B. C.; CREINSHAW, A. T.; ANDRADE, N.; DOHENY, K. F.; HAINES, J. L.; HAYES, G.; JARVIK, G.; JIANG, L.; KULLO, I. J.; LI, R.; LING, H.; MANOLIO, T. A.; MATSUMOTO, M.; MCCARTY, C. A.; MCDAVID, A. N.; MIREL, D. B.; PASCHALL, J. E.; PUGH, E. W.; RASMUSEN, L. V.; WILKE, R. A.; ZUVICH, R. L.; RITCHIE, M. D. Quality control procedures for genome wide association studies. **Current protocols human genetics**, New York, v. 1. n. 19, 2011.

VUONG, T. D.; DIERS, B. W.; HARTMAN, G. L. Identification of QTL for resistance to *Sclerotinia* stem rot in soybean plant introduction 194639. **Crop science**, Madison, v. 48, n. 6, p. 2209–2214, 2008.

WANG, W. Y. S.; BARRATT, B. J.; CLAYTON, D. G.; TODD, J. A. Genome wide association studies: theoretical and practical concerns. **Nature Reviews Genetics**, London, v. 6, n. 2, p. 109–118, 2005.

WATKINS, C. B.; BARDEN, C. L.; BRAMLAGE, W. J. Relationships among  $\alpha$ -farnesene, conjugated trienes and ethylene production with superficial scald development of apples. **Acta Horticulturae**, Holanda, v. 343, p. 155-160, 1993.

WEI, W.; CLOUGH, S. J. *Sclerotinia sclerotiorum*: Molecular aspects in plant-pathogenic interactions. **Revisão Anual de Patologia de Plantas**, Passo Fundo, v. 24, p. 174-189, 2016.

XIA, Y.; SUZUKI, H.; BOREVITZ, J.; BLOUNT, J.; GUO, Z.; PATEL, K.; DIXON, R. A.; LAMB, C. An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. **EMBO Journal**, Oxford, v. 23, n. 4, p. 980–988, 2004.

XIE, W.; FENG, Q.; YU, H.; HUANG, X.; ZHAO, Q.; XING, Y.; YU, S.; HAN, B.; MEYER, Q. Parentindependent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. **Proceedings of the National Academy of Sciences of United States of America**, Washington, v. 107, n. 23, p. 10578-10583, 2010.

XU, Y. **Molecular plant breeding**. London: CAB International, 2010. 734 p.

YU, J.; PRESSOIR, G.; BRIGGS, W. H.; BI, I. V.; YAMASAKI, M.; DOEBLEY, J. F.; MCMULLEN, M. D.; GAUT, B. S.; NIELSEN, D. M.; HOLLAND, J. B.; KRESOVICH, S.; BUCKLER, E. S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, New York, v. 38, n. 2, p. 203-208, 2005.

ZAMBONATO, F. **Caracterização fenotípica e genética da resistência parcial à ferrugem da folha em aveia (*Avena sativa* L.)**. 2011. 89 f. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

ZHAO, X.; HAN, Y.; LI, Y.; LIU, D.; SUN, M.; ZHAO, Y.; LV, C.; LI, D.; YANG, Z.; HUANG, L.; TENG, W.; QIU, L.; ZHENG, H.; LI, W. Loci and candidate gene identification for resistance to *Sclerotinia sclerotiorum* in soybean (*Glycine max* L. Merr.) via association and linkage maps. **Plant Journal**, Oxford, v. 82, n. 2, p. 245-255, 2015.

ZHANG, Z.; ERSOZ, E.; LAI, C. Q.; TODHUNTER, R. J.; TIWARI, H. K.; GORE, M. A.; BRADBURY, P. J.; YU, J.; ARNETT, D. K.; ORDOVAS, J. M.; BUCKLER, E. S. Mixed linear model approach adapted for genome-wide association studies. **Nature Genetics**, New York, v. 42, n. 4, p. 355–360, 2010.

ZHU, Y. L.; SONG, Q. J.; HYTEN, D. L.; VAN TASSELL, C. P.; MATUKUMALLI, L. K.; GRIMM, D. R.; HYATT, S. M.; FICKUS, E. W.; YOUNG, N. D.; CREGAN, P. B. Single-nucleotide polymorphisms in soybean. **Genetics**, Austin, v. 163, n. 3, p. 1123-1134, 2003.