

WILLIAN HYTALO LUDKE

**ESTUDO DE ASSOCIAÇÃO GENÔMICA E ANÁLISE DE REDES GÊNICAS PARA
RESISTÊNCIA DE SOJA À *Phytophthora sojae***

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

Orientador: Felipe Lopes da Silva

Coorientadores: Ivan Schuser
Lucas Lima Verardo

VIÇOSA – MINAS GERAIS

2019

T

L944e Lüdke, Willian Hytalo, 1991-
2019 Estudo de associação genômica e análise de redes gênicas para
resistência de soja à *Phytophthora sojae* / Willian Hytalo Lüdke. -
Viçosa, MG, 2019.

54 f. : il. (algumas color.) ; 29 cm.

Orientador: Felipe Lopes da Silva.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Soja - Resistência a doenças e pragas. 2. *Glycine max*. 3.
Podridão radicular de fitóftora. I. Universidade Federal de Viçosa.
Departamento de Fitotecnia. Programa de Pós-Graduação em Genética
e Melhoramento. II. Título.

CDD 22. ed. 633.3494

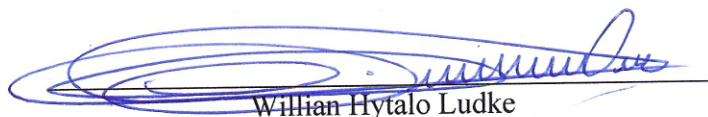
WILLIAN HYTALO LUDKE

**ESTUDO DE ASSOCIAÇÃO GENÔMICA E ANÁLISE DE REDES GÊNICAS PARA
RESISTÊNCIA DE SOJA À *Phytophthora sojae***

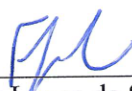
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 26 de julho de 2019.

Assentimento:



Willian Hytalo Ludke
Autor



Felipe Lopes da Silva
Orientador

Aos meus pais, que não mediram esforços para que eu chegasse até aqui.

Dedico

AGRADECIMENTOS

A Deus, por ser a base de tudo em minha vida e por guiar-me pelos melhores caminhos.

Aos meus pais, por sempre estarem ao meu lado, me apoiando, em todas as etapas.

A todos os professores que participaram da minha formação pessoal e profissional, do ensino básico à pós-graduação.

À Universidade Federal de Viçosa (UFV), ao programa de pós-graduação em genética e melhoramento e todo seu corpo docente e quadro de funcionários pelo excelente ensino ofertado, sendo este, público gratuito e de qualidade.

Ao meu orientador e amigo Felipe Lopes da Silva, pelo constante apoio, conselhos, incentivos e pelas horas de conversa, nas quais aprendi a me espelhar em seu exemplo pessoal e profissional.

Aos meus coorientadores Ivan Schuster e Lucas Lima Verardo, por todo apoio na condução deste trabalho.

A Tatiane Dalla Nora Montecelli, Simone Kusano, Elizário Araújo, Dalmaso e demais membros do setor de fitopatologia da Coodetec, pelo apoio na condução das avaliações e pelo inesquecível ambiente amistoso.

Aos meus amigos Fagner Dias, Jeferson Gihel, Michelli Possobom, Ricardo Cabral, Leandro Zuffo, Cleiton Wartha, Leonardo Volpato, Bruno Soares, Diego Santos e Haroldo Rodrigues que são irmãos de coração que ganhei nesta jornada.

A todos os amigos do Programa Soja DFT-UFV, pelo alegre convívio e constante aprendizado.

Ao Grupo de Estudos em Genética e Melhoramento da UFV, que contribuiu de maneira incrível ao meu preparo para o mercado de trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão dos auxílios financeiros.

“Tudo tem o seu tempo determinado, e há tempo para todo o propósito debaixo do céu. Há tempo de nascer, e tempo de morrer; tempo de plantar, e tempo de arrancar o que se plantou”

Eclesiastes 3:1-2

RESUMO

LUDKE, Willian Hytalo, D.Sc., Universidade Federal de Viçosa, julho de 2019. **Estudo de associação genômica e análise de redes gênicas para resistência de soja à *Phytophthora sojae***. Orientador: Felipe Lopes da Silva. Coorientadores: Ivan Schuster e Lucas Lima Verardo.

A soja [*Glycine Max* (L.) Merrill] é considerada uma das cinco mais importantes culturas na produção de alimentos no mundo e no Brasil ocupa do posto de principal *commoditie* agrícola, ocupando mais da metade da área agricultável do país. Um dos fatores mais impactantes na produção de soja é a incidência de doenças, entre estas, a podridão radicular de fitóftora. O objetivo deste estudo foi identificar marcadores SNPs associados à resistência completa de soja à *P. sojae*; identificar genes localizados próximos a marcadores candidatos a resistência parcial e completa a doença e realizar uma análise *post* GWAS através da montagem de uma rede gene – fator de transcrição. Um total de 169 cultivares de soja foram inoculadas com *Phytophthora sojae* através de duas diferentes metodologias e genotipadas com 3.807 marcadores SNPs. As análises realizadas neste estudo foram as de GWAS foi realizada por modelos lineares mistos comprimidos, regressão linear múltipla de *stepwise*, análise de desequilíbrio de ligação, identificação de genes candidatos, análise de sequências regulatórias, análise de enriquecimento de processos biológicos e montagem de rede gene – fator de transcrição. Foram identificados um total de seis marcadores candidatos associados a resistência completa, 31 genes candidatos relacionados a resistência parcial e completa, 279 processos biológicos e centenas de fatores de transcrição. O uso destes marcadores, genes, interações genes – fatores de transcrição permitem uma compreensão mais ampla dos mecanismos relacionados a resistência de soja à *P. sojae* e surgem como ferramenta para aumentar a eficiência dos programas de melhoramento no desenvolvimento de variedades de soja resistentes a *P. sojae*.

Palavras-chave: Soja, *Glycine max* (L.) Merrill. Podridão Radicular de Fitóftora. SNP. GWAS. Desequilíbrio de Ligação. Fatores de Transcrição. Redes Gênicas. Seleção Assistida por Marcadores Moleculares.

ABSTRACT

LUDKE, Willian Hytalo, D.Sc., Universidade Federal de Viçosa, July, 2019. **Genome wide association study and gene network analysis for soybean resistance to *Phytophthora sojae*.** Advisor: Felipe Lopes da Silva. Co-advisors: Ivan Schuster and Lucas Lima Verardo.

Soybeans [*Glycine Max* (L.) Merrill] are considered one of the five most important food crops in the world and in Brazil is the main agricultural commodity, occupying more than half of the country's arable land. One of the most striking factors in soybean production is the incidence of diseases, among them *Phytophthora* root and stem rot. The objective of this study was to identify SNP markers associated with complete resistance to *P. sojae*; localized genes next to marker candidates for partial and complete resistance to *P. sojae* and perform a post GWAS analysis by assembling a gene – transcription factor network. A total of 169 soybean cultivars were inoculated with *Phytophthora sojae* through two different methodologies and genotyped with 3,807 SNP markers. The analyzes carried out in this study were: GWAS performed by compressed linear models; stepwise multiple linear regression; linkage disequilibrium analysis; candidate gene identification; regulatory sequence analysis; biological process enrichment analysis and gene – transcription factor network. A total of six candidate markers associated with complete resistance, 31 candidate genes related to partial and complete resistance, 279 biological processes and hundreds of transcription factor were identified. The use of these SNP markers, genes, gene – transcription factors interactions allow a broader understanding of the mechanisms related to soybean resistance to *P. sojae* and appear as a tool to increase the efficiency of breeding programs in the development of soybean varieties resistant to *P. sojae*.

Keywords: Soybean. *Glycine max* (L.) Merrill. *Phytophthora* root and stem rot. SNP. GWAS. Linkage Disequilibrium. Transcription Factors. Gene Network. Marker Assisted Selection.

LISTA DE ILUSTRAÇÕES

Figura 1. Distribuição, nos 20 cromossomos da soja, dos 3807 SNPs utilizados no presente trabalho	25
Figura 2. Manhattan plot da análise de associação genômica para reação de resistência completa à <i>P. sojae</i>	27
Figura 3. Q-Q plot da distribuição dos valores observados e esperados para da análise de associação genômica para reação de resistência completa à <i>P. sojae</i>	28
Figura 4. Bloco em desequilíbrio de ligação formado por SNPs localizados em uma região com 2,12 Mb no cromossomo 3	29
Figura 5. Média da reação de resistência à <i>P. sojae</i> nos grupos de variedades agrupadas pelos haplótipos dos marcadores Gm04_45613405, Gm03_3225968, Gm03_5341695 e Gm02_46504621	30
Figura 6. Rede gene-fator de transcrição do cromossomo 2	43
Figura 7. Rede gene-fator de transcrição do cromossomo 3	44
Figura 8. Rede gene-fator de transcrição do cromossomo 4	45
Figura 9. Rede gene-fator de transcrição do cromossomo 15	46

LISTA DE TABELAS

Tabela 1. Escala de avaliação de resistência completa à <i>P. sojae</i>	22
Tabela 2. Marcadores significativos associados a resistência completa a <i>P. sojae</i> considerando <i>P-values</i> <0,0001	26
Tabela 3. Marcadores significantes associação a reação de resistencia completa a <i>P. Sojae</i> de acordo com análise de regressão múltipla de stepwise	30
Tabela 4. Marcadores significativos associados a resistência parcial e completa de soja a <i>P. sojae</i>	41
Tabela 5. Genes candidatos relacionados a resistência complete à <i>P. sojae</i> neste estudo	47
Tabela 6. Processos biológicos enriquecidos associados aos fatores de transcrição identificados para resistência à <i>Phytophthora sojae</i>	48
Tabela suplementar S1. Reação de resistência completa à podridão radicular de fitóftora ..	35

SUMÁRIO

1. Introdução geral	12
1.1 Cultura da soja	12
1.2 Podridão radicular de fitóftora	13
1.3 Estudo de associação genômica ampla (GWAS) e redes gênicas	14
1.4 Referências bibliográficas	16
2. ARTIGO 1 – Marcadores SNPs associados a resistência completa de soja à <i>Phytophthora sojae</i>	19
2.1 Resumo	19
2.2 Abstract	19
2.3 Introdução	20
2.4 Material e métodos	21
2.4.1 Material Genético e local do experimento	21
2.4.2 Análise fenotípica	21
2.4.3 Genotipagem com marcadores <i>SNPs</i>	22
2.4.4 Estrutura de população	22
2.4.5 Análise de associação genômica (<i>GWAS</i>)	23
2.4.6 Desequilíbrio de Ligação	24
2.5 Resultados e discussão	24
2.5.1 Análise fenotípica	24
2.5.2 Análise de associação genômica (<i>GWAS</i>)	25
2.6 Conclusão.....	31
2.7 Referências bibliográficas	32

3. ARTIGO 2 – Análise de redes genicas a partir de estudos de associação genômica para resistência de soja à <i>Phytophthora sojae</i>	36
3.1 Resumo	36
3.2 Abstract	36
3.3 Introdução	37
3.4 Material e métodos	38
3.4.1 Análise fenotípica	38
3.4.2 Genotipagem com marcadores <i>SNPs</i> e estrutura de população	39
3.4.3 Análise de associação genômica (<i>GWAS</i>)	39
3.4.4 Análise de redes gênicas	39
3.5 Resultados e discussão	40
3.5.1 Análise fenotípica	40
3.5.2 Análise de associação genômica (<i>GWAS</i>)	41
3.5.3 Análise de redes gênicas	41
3.6 Conclusão	46
3.7 Referências bibliográficas	51

1. Introdução geral

1.1. Cultura da soja

A soja é uma planta anual, herbácea e autógama, com 20 pares de cromossomos. Na classificação botânica, pertence à divisão Magnoliophyta, classe Magnoliopsida, ordem Fabales, família Fabaceae (Leguminosae) e gênero *Glycine* (Sediyama, 2009).

A espécie cultivada *Glycine max* (L.) Merrill foi originada de um evento de domesticação da soja selvagem (*Glycine soja* Seib et Zucc.), ocorrido no centro antigo ou sul da China, aproximadamente há 5000 anos atrás (Gai, 1997; Gai & Guo, 2001). Sua introdução no ocidente deu-se a partir do século XVIII, quando, em 1739, foi cultivada experimentalmente na Europa. No continente americano, atualmente maior produtor mundial de soja, o primeiro relato sobre seu cultivo foi em 1804 e a primeira ocorrência do seu cultivo no Brasil foi em 1882 (Santos, 1988).

O cultivo da soja tem sido altamente concentrado geograficamente em apenas quatro países (EUA, Brasil, Argentina e China), que representam 86% da produção mundial. De acordo com o 1º levantamento da safra mundial de soja 2019/2020, feito pelo Departamento de Agricultura dos estados unidos (USDA) – maio de 2019 – apontou o Brasil como o atual maior produtor de soja do mundo, com 123 milhões de toneladas e os EUA como segundo maior produtor mundial, com 112,95 milhões de toneladas. O mesmo levantamento apontou que atualmente o Brasil é o maior exportador de soja do mundo, exportando cerca de 75,00 milhões de toneladas.

Em condições de campo, diversos fatores afetam a produtividade das culturas agrícolas, sendo estes fatores relacionados a estresses abióticos e bióticos. Em relação aos fatores bióticos, atenção especial tem sido dada aos estresses causados por plantas daninhas, insetos praga, nematoides, oomicetos e doenças bacterianas, fúngicas e virais (Fritsche-neto & Borém, 2012). Especificamente para a soja, estima-se que anualmente 15% a 20% da perda de produção se dá em virtude da ocorrência de doenças (Embrapa, 2013). No Brasil foram identificadas 46 doenças que causam danos à soja, sendo: 32 causadas por fungos, 4 por vírus, 4 por bactérias e 6 por nematóides.

As doenças mais comuns para a cultura da soja e que apresentam maior importância econômica são: ferrugem asiática (*Phakopsora pachyrhizi*), oídio (*Microsphaera diffusa*), mofo branco (*Sclerotinia sclerotiorum*), mancha purpura (*Cercospora kikuchii*), mancha alvo (*Corynespora cassiicola*), mancha olho-de-rã (*Cercospora sojina*), antracnose (*Colletotrichum*

truncatum), podridão de carvão (*Macrophomina phaseolina*), nematoides (*Meloidogyne* spp., *Heterodera glycines*, *Rotylenchulus reniformis*, *Pratylenchus brachyurus*) e podridão radicular de fitóftora (*Phytophthora sojae*) (Embrapa, 2013; Henning, 2009; Matsuo et al., 2012).

1.2. Podridão radicular de fitóftora

A podridão radicular de fitóftora é uma das doenças com maior potencial destrutivo dentre as já catalogadas para a cultura da soja, com reduções no rendimento de grãos de até 100% em variedades com alto grau de suscetibilidade, causando prejuízos estimados de 1 a 2 bilhões de dólares ao redor do mundo (Dorrance et al., 2018). A doença já foi constatada em países como Argentina, Austrália, Brasil, Canadá, China, EUA, Hungria, Itália, Japão e Nova Zelândia, (Workneh et al., 1999). De acordo com Costamilan et al. (1996) a doença foi diagnosticada no Brasil, especificamente no Rio Grande do Sul na safra agrícola de 1994/95, e, na seguinte safra, foi encontrada no estado de Santa Catarina (Pacheco, 1998).

Os sintomas da doença podem ser identificados em plantas de soja durante todas as fases de desenvolvimento, sendo que as plântulas são mais suscetíveis e morrem mais rapidamente. Em plantas em estádios mais avançados, os sintomas variam de acordo com o seu grau de resistência. As plantas suscetíveis e que sobrevivem a infecção da doença, são menos produtivas que as plantas saudáveis. Em pré-emergência, a doença causa apodrecimento de sementes ou flacidez na radícula. No caso de sementes infectadas, a germinação ocorre de forma lenta e, geralmente, as plântulas morrem e causam falha de estande (Schimithenner et al., 1999). Em plantas adultas, os sintomas ficam restritos ao apodrecimento de raízes secundárias, podendo ocorrer lesões longas, lineares e levemente aprofundadas e de cor marrom e um dos lados da haste principal da planta, com posterior amarelecimento nas folhas, murcha, seca e morte da planta (Costamilan et al., 2010).

O agente causal da podridão radicular de fitóftora é *Phytophthora sojae*, taxonomicamente classificado como pertencente ao filo Oomycota, classe Penosporomycetidae, ordem Pythiales e família Pythiaceae. *P. sojae* é homotático e com micélio cenocítico em colônias novas. Apresenta estruturas denominadas anterídios, que são gametas masculinos e oogônios, que são gametas femininos. Os esporos são formados rapidamente, desde que em condições ótimas (~24°C), podendo germinar diretamente como esporos ou de forma indireta como zoóporos, que são esporos assexuais com flagelos que permitem que este se locomova nos poros saturado por água. De forma geral, a doença é favorecida por ambientes com temperatura igual

ou superior à 25°C e com água disponível no solo, assim como solos com textura argilosa, compactados e com períodos longos de saturação por umidade.

Dentre as medidas de controle da doença, a resistência genética é o principal meio de controle, agindo de forma raça-específica (completa), ou de forma parcial (danos limitam-se aos tecidos radiculares) e o uso de genes resistência à *P. sojae* (genes *Rps*) tem comprovado sua efetividade à resistência ao patógeno (Schmitthenner, 1999). Até agora foram identificados 27 genes *Rps* distribuídos em oito cromossomos da soja, sendo eles, *Rps1a*, *Rps1b*, *Rps1c*, *Rps1d*, *Rps1k*, *Rps2*, *Rps3a*, *Rps3b*, *Rps3c*, *Rps4*, *Rps5*, *Rps6*, *Rps7*, *Rps8*, *Rps9*, *Rps10*, *Rps11*, *RpsYB30*, *RpsYD25*, *RpsSu*, *RpsZS18*, *RpsSN10*, *RpsYD29*, *RpsUN1*, *RpsUN2*, *RpsJS* e o gene *Rps* da cultivar Waseshiroge (Zhu et al. 2007; Yao et al. 2010; Yu et al. 2010; Ortega 2011; Sugimoto et al. 2011; Sun et al. 2011; Wu et al. 2011; Lin et al. 2013; Zhang et al. 2013; Ping et al. 2015).

1.3. Estudo de associação genômica ampla (GWAS) e redes gênicas

De acordo com Oraguzie et al. (2007), o estudo de associação genômica ampla é uma ciência que aborda diferentes áreas, tais como estatística, biologia molecular, genética e bioinformática, sendo a base para identificação de regiões do genoma que são selecionadas, avaliadas e associadas com a variação fenotípica.

Meuwissen et al. (2001) definem associação genômica como análises baseadas em um grande número de marcadores moleculares dispostos ao longo de todo genoma, no estudo de associação genômica ampla (GWAS) utiliza-se destes marcadores para identificar no genoma polimorfismos associados a uma determinada característica fenotípica (Lipka et al., 2015). Nesta abordagem, uma população com variabilidade genética para as características de interesse é utilizada (Zhu et al., 2008) e após a genotipagem de cada indivíduo, um modelo estatístico é utilizado para procuração associações entre os marcadores e as características. De acordo com Lipka et al. (2015) a abordagem estatística mais comumente utilizada é ajustar um modelo em cada marcador, no qual a característica avaliada é a variável resposta e os efeitos dos marcadores são as variáveis explicativas.

Na atualidade, os marcadores genéticos moleculares do tipo SNP (*Single Nucleotide Polymorphism* – Polimorfismo de um único nucleotídeo) tem sido mais utilizado, pois baseiam-se na detecção de polimorfismos resultantes da alteração de uma única base no genoma. O uso desses marcadores apresenta como vantagem a distribuição e frequência no genoma e a automação dos dados, por isso são recomendados em GWAS e seleção genômica ampla.

De acordo com Chakravarti (2014), um importante fator relacionado à eficiência da GWAS é o desequilíbrio de ligação. O desequilíbrio de ligação entre marcadores e QTLs é crucial para a detecção destes últimos visando a seleção auxiliada por marcadores. A seleção assistida por marcadores moleculares proposta por Lande & Thompson (1990) é fundamentada na ocorrência do desequilíbrio de ligação na população estudada e, com isso, os alelos dos marcadores informam sobre a presença e efeitos dos loci que governam os caracteres quantitativos, fornecendo meios para estimação dos efeitos os loci dos QTLs e para o seu eficiente uso na seleção genética (Resende et al., 2013).

Embora muitos trabalhos envolvendo GWAS para diversas características em soja tenham sido realizados, o significado biológico dos genes candidatos localizados próximo os marcadores associados, assim como sua relação funcional com os fatores de transcrição, ainda precisam ser mais bem compreendidos. Com isso, os resultados obtidos via GWAS podem ser utilizados na melhor compreensão dos genes envolvidos nas características fenotípicas observadas, através de uma abordagem de rede para os genes identificados nas regiões genômicas próximas aos SNPs significativos.

Os genes ligados aos SNPs significativos podem ser usados para examinar o compartilhamento de vias e funções biológicas, bem como o enriquecimento de fatores de transcrição significativamente relacionados nos genes selecionados. Com isso, estudando a interação entre um gene, cuja sua função na expressão de uma determinada característica já é conhecida com demais genes-alvo previstos através de análises de sequências regulatórias e contração de redes gene-fatores de transcrição, serve como uma ferramenta para validação de interações entre genes distintos.

Portanto, explorar os processos biológicos dos genes candidatos e seus fatores de transcrição relacionados em análises posteriores à GWAS, pode ser uma ferramenta para apontar quais os genes candidatos de fato estão mais relacionados à expressão de uma dada característica.

1.4. Referências bibliográficas

- Costamilan LM (2010) Podridão radicular de fitóftora (*Phytophthora sojae*). In: Almeida AMR; Seixas CDS. **Soja: Doenças Radiculares e de Hastes e Inter-relações com o Manejo do Solo e da Cultura**. Embrapa soja: 105–126.
- Costamilan LM et al. (1996) Ocorrência de *Phytophthora sojae* no Brasil. **Fitopatologia Brasileira** **21**(3): 395.
- Dorrance, A.E. (2018) Management of *Phytophthora sojae* of soybean: A review and future perspectives. **Can. J. Plant Pathol** (40): 210–219.
- Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA (2013) **Tecnologia de Produção de Soja – Região Central do Brasil 2014**. Embrapa Soja. 266 p.
- Fritsche-neto R.; Borém A. (2012) **Melhoramento de plantas para condições de Estresses Bióticos**. Editora UFV. 240 p.
- Gai J. (1997) Soybean Breeding. In: Gai, J. **Plant Breeding**. Crop Sciences. China: Agricultural Press: 207–251.
- Gai J.; Guo W. History of maodou production on China. In: **Second International Vegetable Soybean Conference**: 41– 47. Washington.
- Henning AA et al. (2009) **Manual de identificação de doenças da soja**. 3 ed. Embrapa Soja. 74 p.
- Lande R; Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics** **124**: 743–756.
- Lin F et al (2013) Molecular mapping of two genes conferring resistance to *Phytophthora sojae* in a soybean landrace PI 567139B. **Theor Appl Genet** **126**: 2177–2185.
- Lipka, AE et al (2015) From association to prediction: statistical methods for the dissection and selection of complex traits in plants. **Curr Opin Plant Biol** **24**: 110–118.
- Matsuo E et al. (2012) Breeding for nematode resistance. In: Borém A; Fritsche-neto R. **Plant breeding for biotic stress resistance**. Springer: 81–102.
- Meuwissen TH et al. (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics** **157** (4): 1819–1829.

- Oraguzie, NC et al. (2007) **Association mapping in plants**. Springer. 278 p.
- Ortega MA, Dorrance AE (2011) Microsporogenesis of *Rps8/rps8* heterozygous soybean lines. **Euphytica** **181**: 77–88.
- Pacheco AC (1998) Registro de *Phytophthora sojae* em Santa Catarina. **Reunião de pesquisa de soja da Região Sul**: 26.
- Ping J et al (2015) Identification and molecular mapping of *Rps11*, a novel gene conferring resistance to *Phytophthora sojae* in soybean. **Theor Appl Genet** 1–7.
- Resende MDV et al (2013) Genética de Associação (GWAS). In: Borém A; Fritsche-neto R. **Biotechnologia Aplicada ao Melhoramento de Plantas**: 119– 150.
- Santos OS (1988) **A cultura da soja – Rio Grande do Sul, Santa Catarina e Paraná**. Editora Globo. 329 p.
- Schimitthenner AF (1999) *Phytophthora* rot. In: Hartman GL; Sinclair JB; Rupe J.C **Compendium of soybean diseases**. 4 ed. Saint Paul: APS Press: 39–42.
- Sediyama T. Tecnologia de produção e usos da soja. Londrina. 2009. 314 p.
- Sugimoto T et al (2011) Genetic analysis and identification of DNA markers linked to a novel *Phytophthora sojae* resistance gene in the Japanese soybean cultivar Waseshiroge. **Euphytica** **182**: 133–145.
- Sun S et al (2011) Characterization and mapping of *RpsYu25*, a novel resistance gene to *Phytophthora sojae*. **Plant Breed** **130**: 139–143.
- Workneh F et a. (1999) Regional assessment of soybean brown rot, *Phytophthora sojae*, and *Heterodera glycines* using área-frame sampling: Prevalence and effects of tillage. **Phytopathology** **89**: 204–211
- Wu X et al (2011) Identification of quantitative trait loci for partial resistance to *Phytophthora sojae* in soybean. **Plant Breed** **130**: 144–149.
- Yao HY et al (2010) Molecular mapping of *Phytophthora* resistance gene in soybean cultivar Zaoshu18. **J Plant Genet Resour** **11**: 213–217.
- Yu AL et al (2010) Genetic analysis and SSR mapping of gene resistance to *Phytophthora sojae* race 1 in soybean cv Suinong 10. **Chin J Oil Crop Sci** **32**: 462–466.

Zhang J et al (2013) Identification and candidate gene analysis of a novel Phytophthora resistance gene *Rps10* in a Chinese soybean cultivar. **PloS One** **8** (7): 1–8.

Zhu ZD et al (2007) Molecular identification of a novel Phytophthora resistance gene in soybean. **Acta Agron Sin** **33**: 154–157.

Zhu, C et al (2008) Status and prospects of association mapping in plants, **The Plant Genome** **1**: 5–20.

2. ARTIGO 1 – Marcadores SNPs associados a resistência completa de soja à

Phytophthora sojae

2.1. Resumo

Podridão radicular de fitóftora é uma das doenças com maior impacto econômico na cultura da soja e, como alternativa de controle, o uso de cultivares resistentes é a medida mais eficiente. Este trabalho buscou identificar marcadores *SNPs* associados a genes que conferem resistência completa de soja à *P. sojae*. Foram utilizadas 156 cultivares de soja inoculada com *Phytophthora sojae* e genotipadas com 3.807 marcadores *SNPs*. A análise associação genômica foi realizada por modelos lineares mistos comprimido, seguido de análise de regressão linear múltipla de *stepwise* e análise de desequilíbrio de ligação. Foram identificados seis marcadores em três cromossomos diferentes (cromossomo 2, 3 e 4). O uso destes marcadores na identificação de genes e a validação destes marcadores podem contribuir no aumento da efetividade dos programas de melhoramento visando a obtenção de cultivares resistentes à *P. sojae*.

Palavras chave: Soja, *Glycine max* (L.) Merrill; Podridão radicular de fitóftora; *SNP*; *GWAS*.

2.2. Abstract

Phytophthora root and stem rot is one of the diseases with the greatest impact on soybean production and, as an alternative control, the use of resistant cultivars is the most efficient measure. This study aimed to identify SNP markers associates with genes that confer complete resistance to *P. sojae*. A total of 156 soybean cultivars inoculated with *Phytophthora sojae* and genotyped with 3,807 SNPs markers were used. A genome wide association study was performed by mixed linear compressed model, followed by stepwise multiple linear regression and linkage disequilibrium analysis. Six markers were identified on three different chromosomes (chromosome 2, 3 and 4). The use of these markers in the identification of genes and the validation of these markers may contribute to the increase of the effectiveness of the breeding programs to obtain cultivars resistant to *P. sojae*.

Key words: Soybean, *Glycine max* (L.) Merrill; Phytophthora root and stem rot; *SNP*; *GWAS*.

2.3. Introdução

A soja [*Glycine Max* (L.) Merrill] destaca-se como a principal *commoditie* agrícola brasileira, colocando o Brasil como o maior produtor e exportador mundial de soja. Para a safra agrícola 2018/2019, estima-se que a área plantada no país seja de 37,20 milhões de hectares, produzindo aproximadamente 123 milhões de toneladas (USDA, 2019). Contudo, em condições de campo, diversos fatores causam a redução da produtividade, especialmente as doenças, responsáveis por uma redução de produção anual na casa de 15% a 20% (Embrapa, 2013).

Podridão radicular de fitóftora é causada pelo oomiceto *Phytophthora sojae* (Kaufmann & Gerdemann) (Schmitthenner, 1985; Zhu et al, 2003) sendo favorecida em condições de disponibilidade de água no solo e temperaturas iguais ou superiores a 25°C (Costamilan et al. 2007), causando perdas substanciais ou até mesmo totais em produtividade. O uso de cultivares resistentes é o método mais eficiente e econômico no controle da doença e, o uso de genes de resistência à *P. sojae* (genes *Rps*) tem comprovado sua efetividade (Schmitthenner, 1999). 27 genes de resistência a podridão radicular de fitóftora (*Rps*) já foram identificados em soja (Li et al., 2016), sendo estes genes estão localizados em 21 locos e oito cromossomos da soja.

Meuwissen et al. (2001) definem associação genômica como análises baseadas em muitos marcadores moleculares dispostos ao longo de todo genoma. A GWAS determina os efeitos dos genes sobre os valores genéticos dos indivíduos em uma população e, para este fim, usa como meio as associações entre marcadores moleculares e fenótipos (Resende et al., 2012). Esses marcadores, quando usados para fins de seleção e no melhoramento genético, se fundamentam na ligação entre marcadores e locos que governam as características, sejam elas qualitativas ou quantitativas (Resende et al., 2013).

Foi realizado um estudo de associação genômica ampla (GWAS) utilizando 3.807 SNPs de alta qualidade para identificar marcadores e genes associados à resistência completa a podridão radicular de fitóftora. Os resultados encontrados são importantes na identificação de locos que conferem resistência completa à *P. sojae*, permitindo o desenvolvimento de novas variedades de soja com maior diversidade genética e resistência ao patógeno, aumentando assim a durabilidade desta.

2.4. Material e métodos

2.4.1. Material Genético e local do experimento

Na condução dos experimentos, foram utilizadas 156 cultivares de soja, algumas delas históricas e de grande expressão na sojicultura nacional e outras mais recentes e recomendadas para cultivo, sendo estas utilizadas em diversos programas de melhoramento.

O experimento foi realizado em casa de vegetação, no setor de fitopatologia da COODETEC desenvolvimento, produção e comercialização agrícola, na cidade de Cascavel – Paraná.

2.4.2. Análise fenotípica

Para a fenotipagem foi utilizado o inóculo de *P. sojae* com as raças 1d, 2, 5, 6 e 7, obtido da coleção da Embrapa Trigo – Passo Fundo/RS. O meio de cultura contendo o inóculo foi repicado em placas de petri com palitos de madeira esterilizados para posterior inoculação no hipocótilo das plântulas.

O experimento foi instalado em delineamento inteiramente casualizado, com três repetições e a unidade experimental foi constituída por tubete plástico de 280 cm³ com quatro plantas por tubete. Inicialmente o tubete foi preenchido até a metade com solo autoclavado, sendo posteriormente colocada uma camada de areia + solo, autoclavados, na proporção 2:1 até aproximadamente 1 cm abaixo da borda.

A inoculação foi realizada 15 dias após a semeadura, introduzindo-se a ponta do palito colonizado por *P. sojae* no hipocótilo das plântulas, 1 cm abaixo dos cotilédones, conforme teste de hipersensibilidade de hipocótilo (Costamilan et al., 2006). A avaliação do teste foi realizada 10 dias após a inoculação, baseando-se no seguinte critério: número de plantas sadias (PS): plantas que não apresentaram sintomas característicos na doença, com desenvolvimento normal e formação de calo no local de inserção do palito; número de plantas infectadas (PI): plantas que apresentaram necrose no local de inserção do palito, com desenvolvimento do fungo no interior da haste principal e número de plantas mortas (PM): plantas que morreram em virtude da severidade da doença ou apresentou circunferência da haste principal totalmente necrosada.

A classificação quanto a resistência completa à *P. sojae* (tabela 1) baseia-se no percentual de plantas mortas, onde:

$$\%PM = \left(\frac{PM + (PI/2)}{TP} \right) \times 100$$

sendo TP o número total de plantas avaliadas.

Tabela 1. Escala de avaliação de resistência completa à *P. sojae*

Reação	Descrição
Resistente	0 a 15% de plantas mortas
Moderadamente resistente	15 a 25% de plantas mortas
Suscetível	Acima de 25% de plantas mortas

2.4.3. Genotipagem com marcadores SNPs

Amostras de DNA das 156 cultivares de soja foram genotipadas com 6.000 SNPs, utilizando a plataforma BARCSoySOY6K *BeadChip* (Illumina Inc., San Diego, USA), que corresponde a um subconjunto da plataforma SoySNP50K *BeadChip* (Song et al., 2013). A genotipagem foi conduzida pela Deoxi *Biotechnology* Ltda[®] (Araçatuba, São Paulo, Brasil).

Após a eliminação de marcadores monomórficos, ou com mais de 10% de dados perdidos (*Call Rate* igual a 90%) ou *Minor allele frequency* (MAF), menor do que 5%, restaram 3.807 marcadores para as análises de associação genômica. Locus com genótipo heterozigoto para os SNP foram considerados dados perdidos.

2.4.4. Estrutura de população

A estrutura desta população de variedades foi previamente avaliada com este mesmo conjunto de marcadores por Ludke et al. (2019). Em resumo, a análise de estrutura de população foi realizada utilizando-se a metodologia Markov Chain Monte Carlo (MCMC), com algoritmo de estimação Bayesiana, utilizando-se o *software InStruct* (Gao et al., 2007). As probabilidades posteriores foram estimadas utilizando cinco rodadas independentes do algoritmo de amostragem de MCMC para cada número de grupos simulados (*k*), com *k* variando de 2 até 10, sem informações preliminares da população. Foi utilizado *burn in* de 5.000 e um *run length periods* de 50.000 iterações. A melhor estimativa do número *k* de grupos foi definida pelo menor valor da média de verossimilhança (Likelihood) e valores de *Deviance Information Criterion* dentre os nove agrupamentos simulados (Spiegelhalter et al., 2002).

$$DIC = \bar{D} + pD$$

em que $\bar{D}$ é a medida Bayesiana do ajuste do modelo, e é definido como a expectativa *a posteriori* da deviance ($\bar{D} = E_{\theta/y}[-2 \cdot \ln f(y/\theta)]$); pD é o número efetivo de parâmetros, que mede a complexidade do modelo.

2.4.5. Análise de associação genômica (GWAS)

A análise de associação dos dados genotípicos com o fenótipo - reação à resistência completa - foi realizada pelo método de Modelos Lineares Mistos Comprimido (Zhang *et al.*, 2010) para o conjunto de cultivares, utilizando-se o pacote *Genome Association Prediction Tool* (GAPIT) (Lipka *et al.*, 2012) da plataforma analítica R (R core team, 2012). O Modelo Linear Misto (MLM) pode ser descrito utilizando a notação de matriz de Henderson (1982):

$$y = X\beta + Zu + e$$

em que y é o vetor de observações fenotípicas, β é um vetor contendo os efeitos fixos, incluindo os marcadores moleculares e a estrutura da população (Q); u é um vetor de efeitos genéticos aditivos aleatórios dos múltiplos QTLs no *background* das cultivares; X e Z são as matrizes de incidência; e e é o vetor de efeitos residuais aleatórios. Os vetores u e e têm distribuição normal, com média zero e variância:

$$\text{var} \begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}$$

onde $G = \sigma_a^2 K$ em que σ_a^2 é a variância genética aditiva e K a matriz de parentesco. Assume-se variância homogênea para o efeito residual, com média $R = I\sigma_e^2$, em que σ_e^2 é a variância residual.

Marcadores com valores de probabilidade $P < 0.001$ foram considerados significativos na análise de associação SNP/fenótipo e a partir dos dados de *P-value* de significância, foram elaborados os gráficos *Manhattan plot* e Quantile-quantile (*Q-Q plot*).

A proporção da variância total explicada pela variância genética é definida como herdabilidade h^2 :

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

Para a identificação de marcadores não redundantes e, conseqüentemente, número de QTLs, foi realizada análise regressão múltipla com método *stepwise* de seleção do modelo, utilizando apenas os marcadores significativos na análise de modelos lineares mistos. A análise

de regressão múltipla foi realizada com auxílio do programa JMP (SAS, 1990), com 5% de probabilidade de entrada e de saída. Esta análise adiciona um marcador de cada vez ao modelo, discriminando marcadores que sejam iguais ou inferiores, eliminando assim marcadores redundantes (Schuster & Cruz, 2008).

2.4.6. Desequilíbrio de Ligação

A medida de desequilíbrio de ligação (LD) foi calculada com o software Haploview (Barrett et al., 2005), por meio da fórmula $r^2 = (D')^2 / (pAqapBqb)$, em que D' é uma medida de desequilíbrio de ligação e pA , qa , pB e qb são as frequências dos alelos que são comparados. A raiz quadrada de r^2 pode ser definida como o coeficiente de correlação entre os alelos correspondentes.

Os blocos de desequilíbrio de ligação foram obtidos utilizando-se o método Solid Spine, com auxílio do programa Haploview. Foi utilizado ponto de corte de 1%, ou seja, se a adição de um SNP em um bloco resulta em recombinação alélica acima de 1%, então o SNP não é incluído no bloco.

2.5. Resultados e discussão

2.5.1. Análise fenotípica

O conjunto de 156 cultivares foram analisadas para resistência completa a podridão radicular de fitóftora às raças 1d, 2, 5, 6 e 7 de *P. sojae*, sendo que, 47,44% foram resistentes, 5,13% foram moderadamente resistentes e 47,44% foram suscetíveis (tabela suplementar S1). As cultivares CD201 e CD206, consideradas padrões de resistência, não apresentaram plantas mortas devido infecção por *P. sojae* e, com isso, foram classificadas como resistentes.

Os padrões de susceptibilidade BRS 133 e CD 202 apresentaram 50,00 e 62,50% de plantas mortas devido a infecção por *P. sojae*, respectivamente, sendo classificadas como suscetíveis, demonstrando assim a viabilidade e eficiência do inóculo de *P. sojae* utilizado e, sendo o isolado multirraças, pode-se afirmar que no conjunto de variedades avaliadas, existem variedades resistentes ou moderadamente resistentes a mais de uma raça de *P. sojae*.

A herdabilidade da resistência completa à *P. sojae* quando avaliada a reação à resistência completa, foi de 78,80%. Xiao-ling et al. (2011) ao avaliarem duas populações biparentais $F_{2:3}$ inoculadas com isolado virulento aos genes *Rps* 1a, 1b, 1c, 1d, 1k, 2, 3b, 3c, 5, 6 e 7 identificaram valores de herdabilidade de 36,00 e 82,00%. De maneira semelhante, Nguyen et

al. (2012), avaliando populações na geração F_{7:9} inoculadas com isolado virulento aos genes *Rps* 1a, 1b, 1k, 2, 3a, 3b, 3c, 4, 5 e 6 encontraram herdabilidade igual a 78,00%.

2.5.2. Análise de associação genômica (GWAS)

Foram identificados 3.807 SNPs ao longo dos 20 cromossomos da soja, os cromossomos 13 e 18 foram os mais saturados, com 266 e 260 SNPs, respectivamente e, os cromossomos 17 e 1 foram os menos saturados, com 144 e 140 SNPs; de forma geral, os marcadores estão dispostos amplamente em todos os cromossomos do genoma da soja (Figura 1).

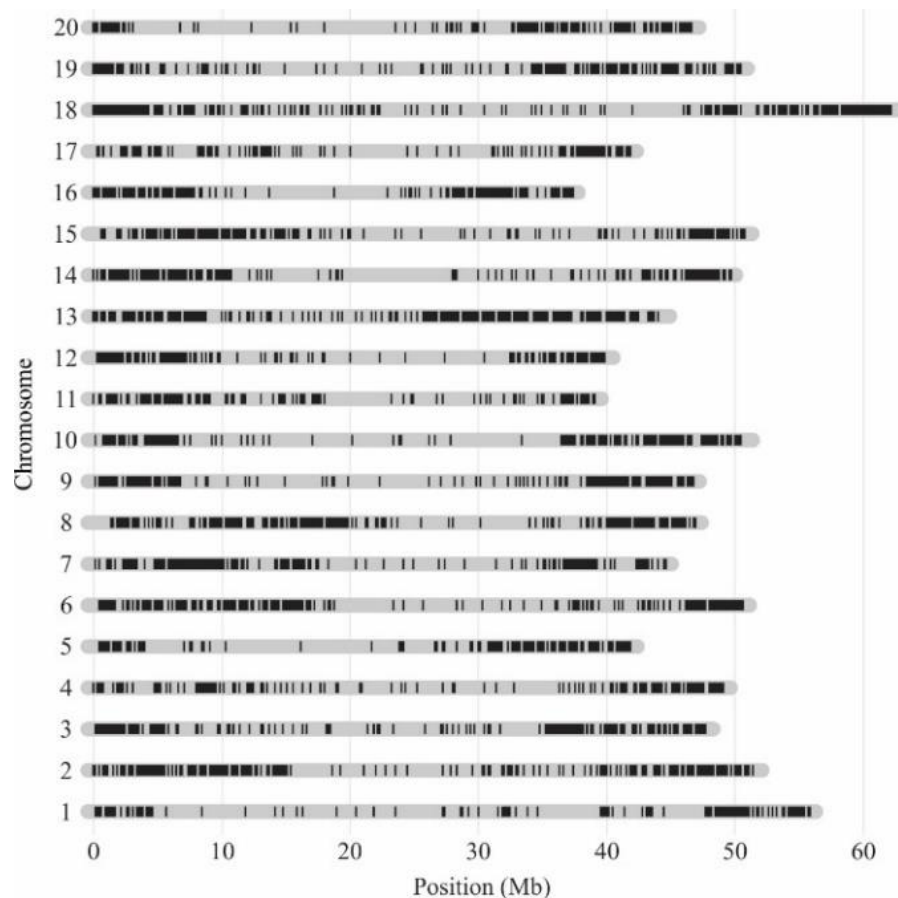


Figura 1. Distribuição, nos 20 cromossomos da soja, dos 3807 SNPs utilizados no presente trabalho

O modelo linear misto inclui efeitos fixos e aleatórios e, para avaliação deste modelo com as matrizes de estrutura populacional (Matriz Q) e de parentesco (K), os resultados encontrados que se basearam no critério de informação bayesiana mostraram ajuste adequado para o modelo (Q + K). A abordagem de modelo linear misto comprimido (CMLM), por sua vez, diminui o tamanho efetivo do conjunto de dados, agrupando os indivíduos em grupos (Zhang *et al.*, 2010).

Atualmente, não existem trabalhos prévios publicados sobre GWAS para resistência completa à *P. sojae*, todavia, alguns trabalhos já foram publicados para resistência parcial e serão abordados nas discussões.

Para o conjunto de dados utilizado, com 156 cultivares de soja e 3.807 SNPs, de acordo com a análise de modelo linear misto comprimido (CMLM), seis marcadores foram significativamente associados à resistência completa de soja à *P. Sojae*, sendo um localizado no cromossomo 2, quatro localizados no cromossomo 3 e um localizado no cromossomo 4 (Tabela 3, Figura 2).

Tabela 2. Marcadores significativos associados a resistência completa a *P. sojae* considerando $P\text{-values} < 0,0001$

SNP	Alelo	Crm ¹	Posição	R ² *	P-valor ²	MAF ³	-Log10(p)
Gm03_5217414	C/T	3	5217414	21,96%	2,58x10 ⁻⁵	0,24	4,59
Gm03_3225968	G/A	3	3225968	19,63%	1,70x10 ⁻⁴	0,39	3,77
Gm03_5341695	A/G	3	5341695	18,33%	5,01x10 ⁻⁴	0,18	3,30
Gm04_45613405	T/C	4	45613405	18,13%	5,92x10 ⁻⁴	0,16	3,23
Gm02_46504621	T/C	2	46504621	17,90%	7,22x10 ⁻⁴	0,10	3,14
Gm03_5106459	T/G	3	5106459	17,64%	8,97x10 ⁻⁴	0,35	3,05

¹ Cromossomo

² Probabilidade de significância

³ Frequência do alelo menos frequente

* Coeficiente de determinação

No cromossomo 2, o marcador Gm02_46504621 apresentou $P\text{-value}$ de 7,22x10⁻⁴, respondendo por 17,90% da variação fenotípica. Este marcador está localizado na posição 4,65 Mb, encontrando-se próximo ao QTL *Phytoph 10-3* para tolerância de soja à podridão radicular de fitóftora (Han et al., 2007).

No cromossomo 3, os marcadores associados foram Gm03_5217414, Gm03_3225968, Gm03_5341695 e Gm03_5106459. Estes marcadores estão localizados em um intervalo de 2,12 Mb, nas posições 5,22, 3,22, 5,34 e 5,11 Mb, respectivamente. Ludke et al. (2019), utilizando o mesmo conjunto de cultivares e marcadores do presente trabalho em estudo de associação genômica para resistência parcial à *P. sojae*, também identificaram os marcadores Gm03_3225968, Gm03_5106459, Gm03_5217414 e Gm03_5341695, reforçando o fato de que estes marcadores estão realmente associados à resistência à *P. sojae*. Demais trabalhos, também com resistência parcial, localizaram marcadores dentro desta mesma região genômica, tais como os marcadores localizados nas posições 3,85, 3,86, 4,28 e 4,31Mb, identificados por Schneider et al. (2016). Qin et al. (2017), identificaram marcadores com alto valor de associação nas posições 3,11, 3,89, 4,48 e 5,49 Mb e Huang et al. (2016) localizaram dois marcadores no

cromossomo 3, nas posições 3,66 e 5,15, corroborando com os resultados observados neste estudo. Todos os marcadores associados a resistência a *P. sojae* localizados no cromossomo 3, exceto Gm03_3225968, encontram-se dentro do QTL *Phytoph 14-5* (Lee et al., 2013). Gm03_5217414 foi o marcador SNP mais significativo da análise, com um *P-value* de $2,58 \times 10^{-5}$, sendo responsável por 21,96% da variação fenotípica observada. Os demais marcadores desta região, foram responsáveis, respectivamente, por 19,63%, 18,33% e 17,64% da variação fenotípica observada.

Por fim, no cromossomo 4, o marcador Gm04_45613405 apresentou *P-value* de $5,92 \times 10^{-4}$, explicando 18,13% da variação fenotípica; este marcador encontra-se localizado dentro do QTL *Phytoph 14-7* que está relacionado à reação de resistência de soja à podridão radicular de fitóftora (Lee et al., 2013).

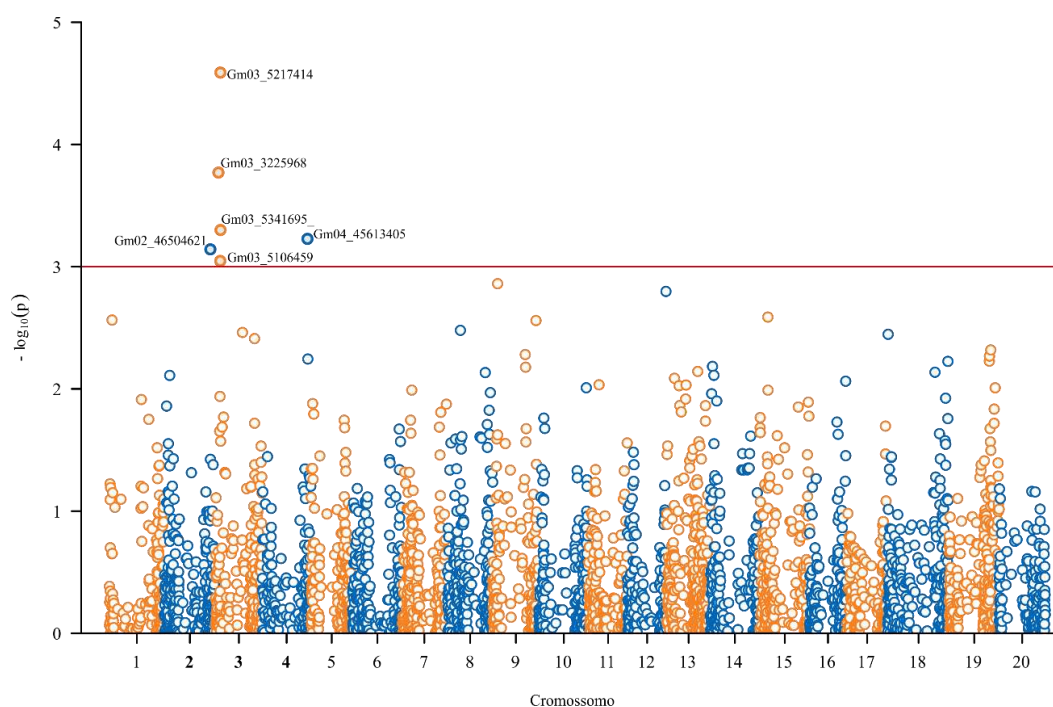


Figura 2. Manhattan plot da análise de associação genômica para reação de resistência completa à *P. sojae*.

O *Q-Q plot* indicou a ocorrência de divergências entre os valores de significância observados e esperados na hipótese nula, pelo teste de χ^2 , com seis pontos deslocados da linha diagonal (Figura 3). De acordo com Patterson et al. (2006), desvios da linha diagonal indicam associações verdadeiras entre marcadores e fenótipo. Como a maioria dos pontos encontram-se posicionados na linha diagonal, comprova-se que o método CMLM foi adequadamente utilizado para a eliminação de associações falso positivas. Os pontos desviados da diagonal,

posicionados acima desta linha, confirmam a associação dos SNPs significativos relacionadas à reação de resistência completa à *P. sojae*.

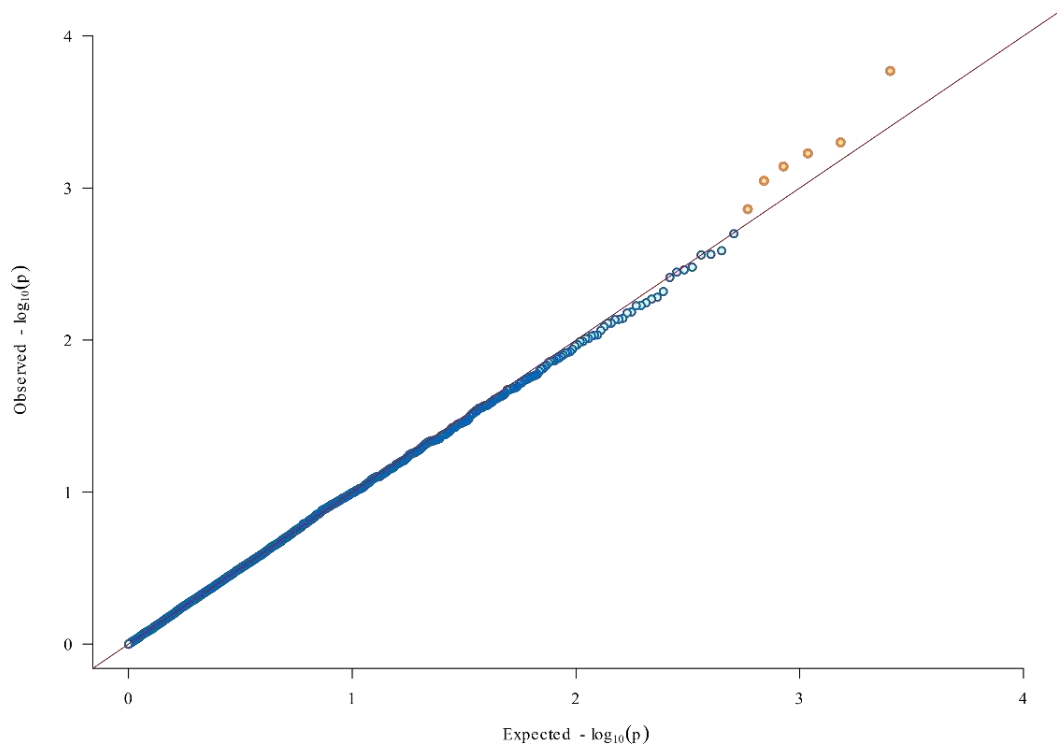


Figura 3. Q-Q plot da distribuição dos valores observados e esperados para da análise de associação genômica para reação de resistência completa à *P. sojae*.

Para avaliar mais detalhadamente a região do cromossomo 3 na qual encontram-se os marcadores significativos identificados no estudo de associação genômica ampla via CMLM, foi realizada análise de desequilíbrio de ligação (LD) de uma região contendo 4,14 Mb, na qual haviam 26 marcadores SNPs, incluindo os significativos.

Esta região apresentou seis blocos em LD (Figura 4). O marcador Gm03_3225968 está em um bloco em DL composto por 247 Kb em que outros dois marcadores estão presentes. O marcador Gm03_5106459 está no bloco 5 em DL, que apresenta 440 Kb e contém mais cinco marcadores e, no sexto bloco em DL, que apresenta 234 Kb, existem três marcadores, dos quais dois foram significativos, Gm03_5217414 e Gm03_5341695, estando estes distantes 124 Kb um do outro.

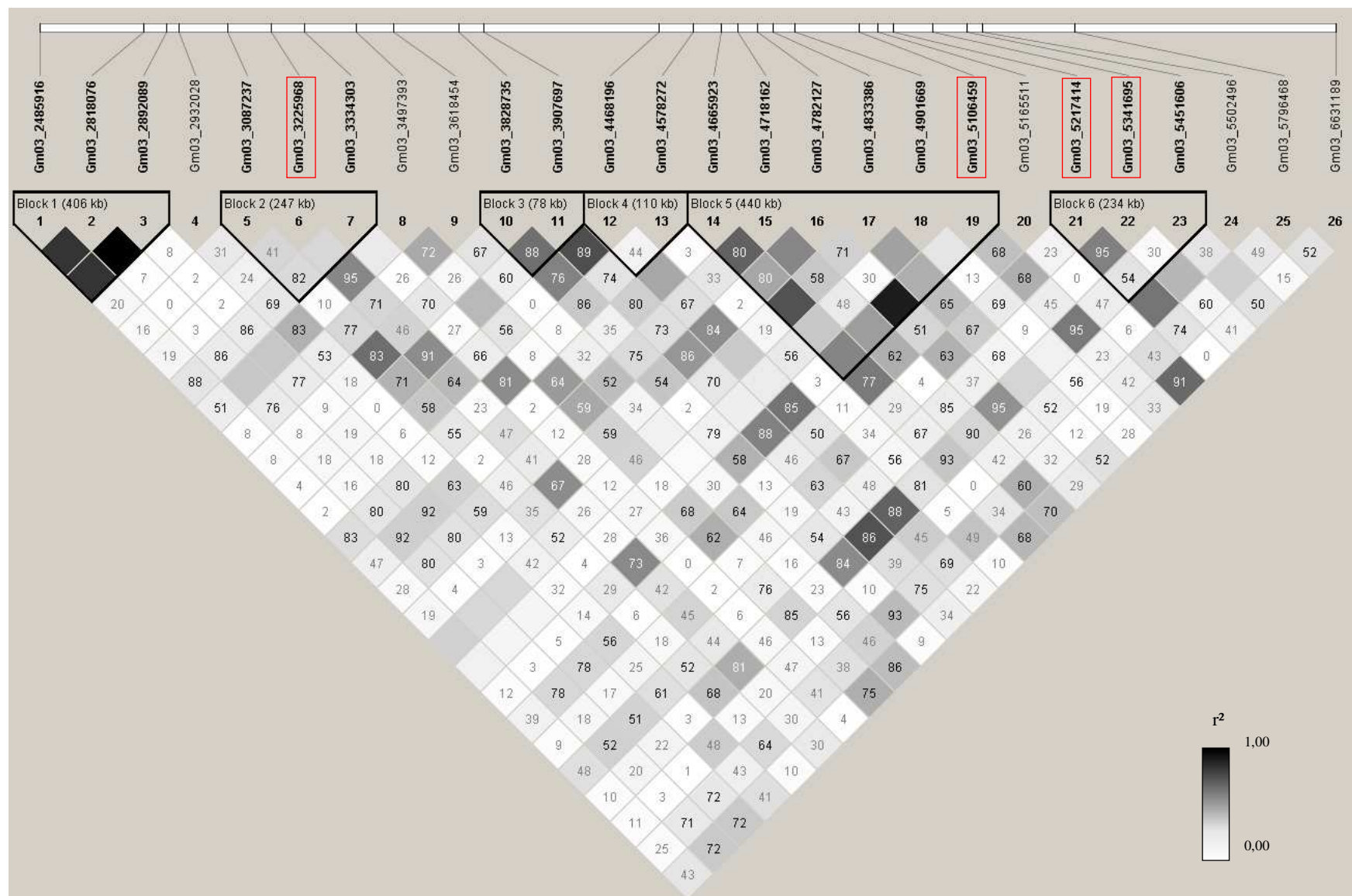


Figura 4. Bloco em desequilíbrio de ligação formado por SNPs localizados em uma região com 2,12 Mb no cromossomo 3

Os seis marcadores associados a reação de resistência completa a *P. Sojae* foram utilizados na análise de regressão múltipla, com método *stepwise* de seleção do modelo (Tabela 4). Os marcadores Gm04_45613405, Gm03_3225968, Gm03_5341695 e Gm02_46504621 foram significativos na análise de regressão múltipla, indicando a existência de quatro QTLs de resistência, dois no cromossomo 3, um no cromossomo 2 e um no cromossomo 4. Estes quatro marcadores explicam em conjunto 40,03% da variação observada para reação de resistência do conjunto de cultivares à *P. sojae*.

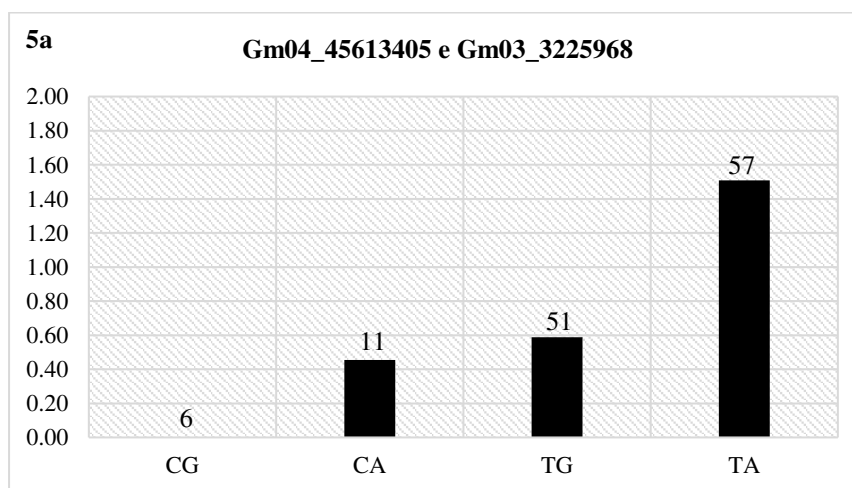
Tabela 3. Marcadores significantes associação a reação de resistência completa a *P. Sojae* de acordo com análise de regressão múltipla de *stepwise*.

SNP	Alelos	Cr ^{m1}	Posição	r ^{2*}	P-valor
Gm04_45613405	T/C	4	45613405	0,0651	0,0013
Gm03_3225968	G/A	3	3225968	0,2540	0,0000
Gm03_5341695	A/G	3	5341695	0,3823	0,0000
Gm02_46504621	T/C	2	46504621	0,4003	0,0348

¹ Cromossomo

* Coeficiente de determinação do modelo de regressão, com a adição dos marcadores.

Ao se realizar o agrupamento das cultivares de acordo com os haplótipos identificados, utilizando os marcadores significativos na regressão múltipla, observou-se o efeito da associação dos marcadores com a resistência à *P. sojae* (Figura 5). Observando-se diferença na média da reação de resistência das variedades contendo os haplótipos com os alelos de resistência para os marcadores Gm04_45613405, Gm03_3225968 (alelos C e G), e as variedades contendo o haplótipos com os alelos de suscetibilidade nestes marcadores (T e A) se nota que na presença dos alelos de resistência a média é zero e aumenta gradativamente a medida que os alelos de resistência vão sendo substituídos pelos alelos de suscetibilidade, como pode ser visto na figura 5a.



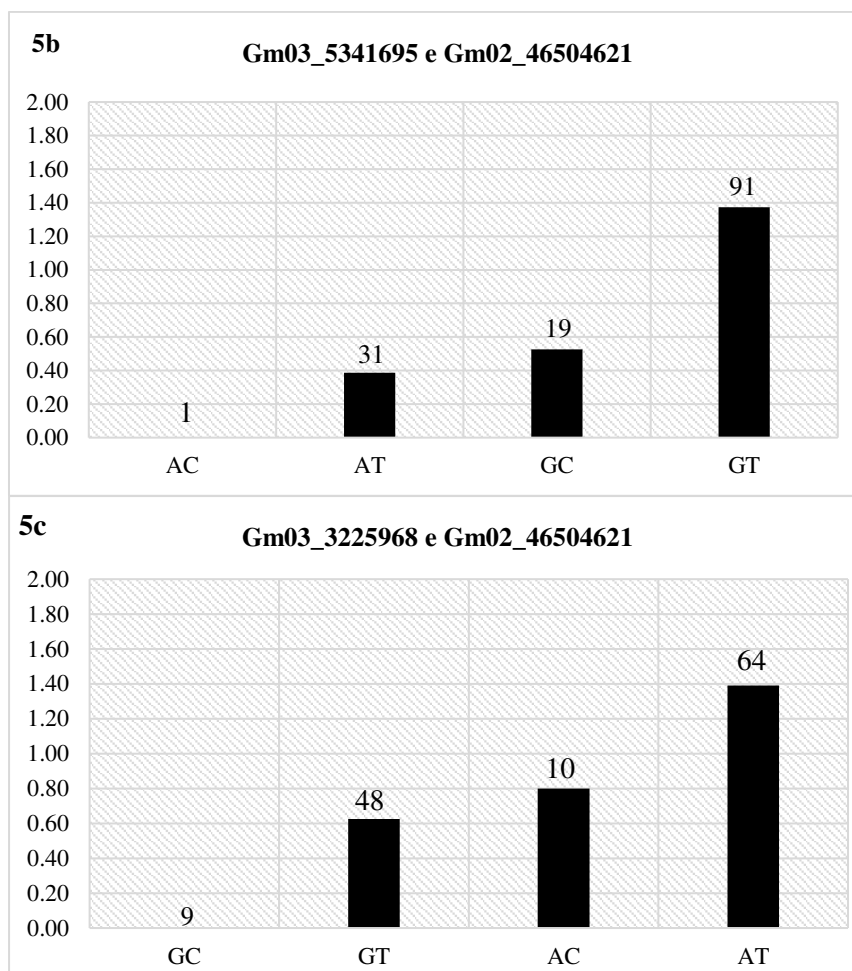


Figura 5. Média da reação de resistência à *P. sojae* nos grupos de variedades agrupadas pelos haplótipos dos marcadores Gm04_45613405, Gm03_3225968, Gm03_5341695 e Gm02_46504621

O mesmo acontece com a presença dos marcadores Gm03_5341695 e Gm02_46504621 (figura 5b) e Gm03_3225968 e Gm02_46504621 (figura 5c).

De maneira geral, observou-se que a presença de dois marcadores com alelos de resistência resulta em imunidade, a presença de um único alelo de resistência as médias de reação ficam bem menores que 1, ou seja, moderadamente resistente e, na ausência dos alelos de resistência a suscetibilidade é consideravelmente maior.

2.6. Conclusão

Neste estudo foram identificados 6 marcadores que podem ser utilizados nos programas de melhoramento de soja visando aumentar a frequência de linhagens com resistência completa de soja a *P. sojae*, além disso, a seleção de plantas com a presença de haplótipos de resistência contendo os alelos T, G, A e T dos marcadores Gm04_45613405, Gm03_3225968, Gm_5341695 e Gm02_4654621 reduz o percentual de plantas mortas.

2.7. Referências bibliográficas

- Barrett JC et al (2005) Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**
- Contreras-soto RI et al (2017) A Genome-Wide Association Study for Agronomic Traits in Soybean Using SNP Markers and SNP-Based Haplotype Analysis. **PLoS ONE** **12**(2): 1 – 22.
- Costamilan LM (2006) Reação de cultivares de soja à *Phytophthora sojae*. In: Reunião de pesquisa de soja da região sul. **Ata e resumos**. Pelotas: Embrapa Clima Temperado.
- Costamilan LM et al (2007) **Podridão radicular de fitóftora em soja – Documentos Online 79**. Passo Fundo, RS: Embrapa Trigo. 23 p.
- Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA (2013) **Tecnologia de Produção de Soja** – Região Central do Brasil 2014. Embrapa Soja. 266 p.
- Gao H et al (2007) A Markov Chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. **Genetics** **176** (3): 1635–1651.
- Han Y et al (2007) Mapping QTL tolerance to *Phytophthora* root rot in soybean using microsatellite and RAPD/SCAR derived markers. **Euphytica** **162**: 231–239.
- Henderson Jr CR (1982). Analysis of Covariance in the Mixed Model: Higher-Level, Nonhomogeneous, and Random Regressions. **Biometrics** **38** (3): 623–640.
- Huang J et al (2016) Phenotypic evaluation and genetic dissection of resistance to *Phytophthora sojae* in the Chinese soybean mini core collection. **BMC genetics** **17** (85): 1 – 14.
- Lee S et al (2013) Novel quantitative trait loci for partial resistance to *Phytophthora sojae* in soybean PI398841. **Theor. Appl. Genet** **126** (4): 1121-1132.
- Li L et al (2016) Loci and candidate gene identification for resistance to *Phytophthora sojae* via association analysis in soybean [*Glycine max* (L.) Merr.] **Mol Genet Genomics** **1** – 7.
- Lipka AE et al (2012) GAPIT: genome association and prediction integrated tool. **Bioinformatics** **28**: 2397–9.
- Ludke et al (2019) SNP markers associated with soybean partial resistance to *Phytophthora sojae*. **Crop Breeding and Applied Biotechnology** **19**: 31–39.

- Nguyen VT (2012) Mapping of Quantitative Trait Loci Associated with Resistance to *Phytophthora sojae* and Flooding Tolerance in Soybean. **Crop Science** **52**: 2481–2493.
- Patterson N et al (2006) Population structure and eigenanalysis. **PLoS Genetics** **2** (12): 1 – 21.
- Qin J et al (2017) Genome-wide association mapping of resistance to *Phytophthora sojae* in a soybean [*Glycine max* (L.) Merr.] germplasm panel from maturity groups IV and V. **PLoS ONE** **12**(9): 1–11.
- R Core Team (2012) R. A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- SAS Institute (1990) JMP: Software for statistical visualization, version 2. Cary, N.C.
- Schmitthenner AF (1985) Problems and Progress in Control of Phytophthora Root-Rot of Soybean. **Plant Dis.** **69**: 362–368.
- Schmitthenner AF (1999) Phytophthora rot of soybean. **Compendium of soybean diseases, 4th edition**. The American Phytopathological Society Press, Sr. Paul: 39–42.
- Schneider R et al (2016) Genome-wide association mapping of partial resistance to *Phytophthora sojae* in soybean plant introductions from the Republic of Korea. **BMC Genomics** **17**(607): 1–14.
- Schuster I, Cruz CD **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. ed. 2, Viçosa, MG. 2008. 568 p.
- Song Q et al (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. **PLoS One** **8**(1): 1–12.
- Spiegelhalter DJ et al (2002) Bayesian measures of model complexity and fit. **J. R. Statist. Soc. Series B** **64**(4): 583–639.
- Xiao ling W. et al (2011) Identification, Genetic Analysis and Mapping of Resistance to *Phytophthora sojae* of Pm28 in Soybean. **Agricultural Sciences in China** **10** (10): 1506–1511.
- Zhang Z et al (2010) Mixed linear model approach adapted for genome-wide association studies. **Nature genetics** **42**: 355–360.
- Zhu ZD et al (2007) Molecular identification of a novel *Phytophthora* resistance gene in soybean. **Acta Agron Sin** **33**: 154–157.

Zhu ZD et al (2003) Distribution and virulence diversity of *Phytophthora sojae* in China. **Agri Sci China** **3**: 116–123.

Tabela suplementar s1. Reação de resistência completa à podridão radicular de fitóftora.

Genótipo	Reação	Genótipo	Reação	Genótipo	Reação	Genótipo	Reação
M6009 RR	S	CD206	R	A6411 RR	R	A6001	S
FUNDACEP59 RR	S	CD208	R	P98Y70	R	R7	S
CD2800	R	CD201	R	CD248 RR	S	FT-GUAIRA	S
FUNDACEP63 RR	S	Bragg	S	CD206 RR	R	OCEPAR14	R
FUNDACEP61 RR	S	CAC-1	S	BRS282	S	MSOY2001	R
BRS284	S	CD2840	R	M7211 RR	S	MSOY7201	S
M7578 RR	S	TMG1067 RR	R	P98Y51	R	CD231RR	R
BRS259	S	TMG7161 RR	R	CD5969	R	CD247RR	S
A4725 RG	S	TMG4001 RR	R	CD2860	R	CD244RR	R
CD238 RR	S	CD232	MR	BRS246 RR	S	CD202	S
5D660 RR	R	TMG1161 RR	R	RA516 RR	S	A6001RR	R
CD250 RRSTS	S	Fundacep33	R	BRS213	S	ANTA 82	MR
5D688 RR	R	CD242 RR	R	M7908 RR	S	EMBRAPA59	S
CD2792 RR	S	CD237 RR	S	BRS258	S	BRSMG68 (VENCEDORA)	S
CD254 RR	S	USP1	R	BRS262	R	BRSMS LAMBARI	R
CD251 RR	R	CD236 RR	S	BRS268	S	BRSMT CRIXAS	S
CD245 RR	MR	CD239 RR	R	RA628 RR	S	FMT MATRINXÃ	S
5G770 RR	S	5G830 RR	S	M9144 RR	R	BRSMG LIDERANÇA	S
NA5909 RR	R	CD252	R	RA518 RR	MR	FMT TABARANA	R
TMG1066 RR	S	CD235 RR	R	BRS184	MR	BRSMTPintado	R
BRS CELESTE	S	CD226 RR	R	BRS230	S	OCEPAR3- PRIMAVERA	S
SPRING53	R	BRS232	S	BRS133	S	EMGOPA304 (CAMPEIRA)	S
FMT TUCUNARE	R	BRS185	R	TMG123 RR	R	UFV16 (CAPINOPOLIS)	R
MERCEDES70A	R	EMBRAPA48	S	A7321 RG	S	DOMMARIO 70I RR	R
CD5807	R	BRS257	R	CD221	R	BMX POTENCIA RR	R
M-SOY6101	S	BRS283	MR	CD228	S	BMX ENERGIA RR	R
FUNDACEP39	S	TMG103 RR	R	CD229 RR	R	BMX TURBO RR	R
CD2585 RR	S	5D690 RR	R	M8527 RR	MR	BRSMG Renascenca	R
CD2737 RR	R	CD219 RR	R	CD246	S	FUNDACEP57 RR	S
BMX FORÇA RR	R	5D711 RR	R	CD217	R	FUNDACEP53 RR	R
NS4823	R	CD234 RR	S	CD233 RR	R	FUNDACEP58 RR	S
FT-ESTRELA	R	NA4990 RR	S	CD204	S	BMX TITAN RR	R
EMGOPA 302	S	CD215	S	CD225 RR	S	BMX MAGNA RR	R
IAS5	R	CD2630 RR	S	CD224	R	BMX ATIVA RR	MR
FUNDACEP 38	S	CD240 RR	S	CD205	R	CD249 RRSTS	S
M-SOY7901	S	CD253	R	CD230 RR	R	MGBR46 (CONQUISTA)	S
CDFAPA 220	S	P98Y11	R	CD216	S	FUNDACEP55RR	S
CD2721 RR	S	BRS243 RR	R	CD214 RR	R	FUNDACEP56 RR	R

R: Resistente

MR: Moderadamente resistente

S: Suscetível

3. ARTIGO 2 – Análise de redes genicas a partir de estudos de associação genômica para resistência de soja à *Phytophthora sojae*

3.1. Resumo

Phytophthora sojae é um oomiceto patogênico da cultura da soja que causa a podridão radicular de fitóftora, uma das doenças que mais afetam a cultura ao redor do mundo, causando prejuízos na casa de bilhões de dólares. Atualmente a medida de controle mais eficiente e ambientalmente correta é o uso de cultivares resistentes. Este estudo buscou identificar genes localizados próximos a marcadores candidatos a resistência à *P. sojae* e posteriormente elaborar uma rede gene-fator de transcrição para melhor compreensão da resistência. Foram utilizados estudos de associação genômica ampla previamente realizados para resistência parcial e completa para a busca de genes candidatos localizados próximos aos marcadores SNPs significativos e posteriormente foi realizado um estudo *post* GWAS via rede gene-fator de transcrição. Foram identificados 31 genes candidatos e 279 processos biológicos ligados a centenas de fatores de transcrição. O uso destes genes e sua interação com diferentes fatores de transcrição podem contribuir na compreensão do mecanismo de resistência de soja à *P. sojae* e descoberta de genótipos resistentes para serem utilizados em programas de melhoramento.

Palavras chave: Soja, *Glycine max* (L.) Merrill; Podridão radicular de fitóftora; SNP; GWAS; Rede gênica.

3.2. Abstract

Phytophthora soja is a pathogenic oomycete of soybean crop that causes phytophthora root and stem rot, one of the diseases that most affect the soybean production around the world, causing losses in the billions of dollars. Currently the most efficient and environmental friendly to reduce disease damages is the use of resistant cultivars. This study aimed to identify genes located close to candidate markers for resistance to *P. sojae* and later to elaborate a gene – transcription factor network for a better understanding of the resistance. Was used a genome wide association study previously performed for partial and complete resistance to the search of candidate genes located near the significant SNP markers and later a post GWAS study was performed via the gene – transcription factor. Was identified 31 candidate genes and 279 biological processes linked to hundreds of transcription factors. The use of these genes and their interactions with different transcription factors may contribute to the understanding of the resistance mechanism of soybean to *P. sojae* and the discovery of genotypes to be used in breeding programs

Key words: Soja, *Glycine max* (L.) Merrill; phytophthora root and stem rot; *SNP*; *GWAS*; transcription factors; gene network.

3.3. Introdução

Considerada uma das cinco mais importantes culturas na produção de alimentos a nível global (Savary et al., 2019), a soja é produzida em aproximadamente 6% da área cultivável do mundo (Bandara et al, 2019) e a rápida expansão desta área tem sido acompanhada por um aumento na incidência de doenças na cultura, impactando diretamente na produtividade de grãos (Hartman et al., 2015).

Dentre as doenças que afetam a cultura da soja, uma das mais importantes é a podridão radicular de fitóftora, causada pelo patógeno *Phytophthora sojae* (Kaufmann & Gerdemann) (Schmitthenner, 1985; Zhu et al, 2003), especialmente em condições de solos com alta umidade e temperaturas ambientais altas (iguais ou superiores a 25°C) (Costamilan et al. 2007), possuindo alto potencial destrutivo. O uso de materiais que apresentam resistência genética é o mais eficiente método de controle, podendo reduzir de forma significativa a incidência da doença. De acordo com Sugimoto et al. (2012), existem cultivares de soja que apresentam resistência completa e resistência parcial ao patógeno. A resistência parcial é controlada por múltiplos genes e possui amplo espectro (Kou & Wang, 2010), todavia, esta resistência pode ser perdida sob alta pressão de inóculo (Dorrance, 2003). Por outro lado, um dos métodos mais eficientes no controle da doença é o uso de cultivares com resistência completa, que é controlada por um pequeno grupo de genes (Niu et al., 2017).

Com o rápido avanço no conhecimento do genoma vegetal e do papel fisiológico e molecular dos genes, houve uma revolução na genética molecular e sua eficiência no melhoramento de plantas (Nadeem et al., 2017). Neste contexto, a aplicação de estudos de associação genômica tem se mostrado uma importante ferramenta, uma vez que pode ser aplicada na explicação da variância genética de uma dada característica através da sua

associação entre fenótipo-genótipo (Boyle et al., 2017), contribuindo na identificação de genes e QTLs (Fang et al., 2017).

Genes candidatos relacionados a resistência identificado próximo a marcadores significativos são uma importante ferramenta na compreensão dos mecanismos de resistência na soja (Huang et al., 2016), neste âmbito, realizando-se análise de redes gênicas, é possível explorar os processos biológicos destes genes e seus fatores de transcrição e, com isso, apontar os genes candidatos mais prováveis (Verardo et al., 2016).

Foi realizado um estudo de associação genômica ampla (GWAS) utilizando 3.807 SNPs de alta qualidade para identificar marcadores e genes associados à resistência completa a podridão radicular de fitóftora e posteriormente realizada uma análise *post GWAS* via rede genes-fatores de transcrição. Os resultados encontrados são importantes na identificação de locos que conferem resistência completa à *P. sojae*, permitindo o desenvolvimento de novas variedades de soja com maior diversidade genética e resistência ao patógeno, aumentando assim a durabilidade desta.

3.4. Material e métodos

3.4.1. Análise fenotípica

O conjunto de genótipos utilizado neste trabalho foi composto por 169 cultivares comerciais de soja de diferentes épocas e empresas, obtendo-se assim grande representatividade da variabilidade genética dos materiais empregados na sojicultura nacional. Para ambos os experimentos, o inóculo de *P. sojae* utilizado foi o mesmo, este possui fórmula de virulência 1d, 2, 5, 6 e 7, sendo esta a mais representativa da população do patógeno no Brasil (Costamilan et al., 2013).

Para avaliação de resistência parcial, todas as cultivares foram fenotipadas, o experimento foi conduzido em casa de vegetação, na Universidade Federal de Viçosa, Viçosa – Minas Gerais. A inoculação ocorreu no momento da semeadura, utilizando-se o método da camada de micélio (Schmitthenner & Bhat, 1994). A fenotipagem foi realizada 21 dias após a semeadura, avaliando-se as raízes de acordo com escala de notas e posterior classificação de reação à podridão radicular de fitóftora propostas por Dorrance et al. (2003).

Para avaliação de resistência completa, foram utilizadas 156 cultivares deste grupo, algumas foram descartadas por terem apresentado problemas relacionados ao vigor de semente e baixa germinação. O experimento foi realizado em casa de vegetação, no setor de fitopatologia da COODETEC Desenvolvimento, Produção e Comercialização agrícola, Cascavel – Paraná. O método de fenotipagem utilizado foi o teste de hipersensibilidade de hipocótilo e a avaliação final foi realizada 10 dias após a inoculação (25 dias após a semeadura) (Costamilan et al., 2006).

3.4.2. Genotipagem com marcadores SNPs e estrutura de população

Conforme descrito por Ludke et al. (2019), amostras das 169 cultivares foram genotipadas com 6.000 marcadores SNPs, posteriormente, foram eliminados marcadores monomórficos, ou com mais de 10% de dados perdidos (*Call Rate* igual a 90%) ou *Minor allele frequency* (MAF), menor do que 5%, restando 3.807 marcadores. Além disso, genótipos heterozigotos para os SNPs foram considerados dados perdidos.

A análise de estrutura de população foi realizada utilizando-se a metodologia Markov Chain Monte Carlo (MCMC), com algoritmo de estimação Bayesiana, utilizando-se o software InStruct (Gao et al., 2007).

3.4.3. Análise de associação genômica (GWAS)

A análise de associação dos dados genotípicos com o fenótipo foi realizada pelo método de Modelos Lineares Mistos Comprimido (Zhang et al., 2010), utilizando-se o pacote Genome Association Prediction Tool (GAPIT) (Lipka et al., 2012) da plataforma analítica R. Marcadores com valores de probabilidade $P < 0,001$ foram considerados significativos na análise de associação SNP/fenótipo.

3.4.4. Análise de redes gênicas

Inicialmente foi realizada a identificação de genes candidatos através do banco de dados de soja da plataforma NCBI (RefSeq assembly accession: GCF_000004515.5) (O’Leary et al., 2016). Foram selecionados os genes mais próximos dos marcadores associados à característica (entre 16.373 pb *downstream* a 12.833 pb *upstream*) e, posteriormente, identificaram-se as anotações existentes para estes genes. As distâncias utilizadas nesse estudo para definir a proximidade gene-SNP estão de acordo com Brodie et al. (2016), onde observaram que um gene responsável pela expressão de uma determinada característica pode estar localizado até 200.000 pb do SNP identificado como associado para esta característica.

A partir da identificação dos genes candidatos obtidos em cada GWAS, informações obtidas sobre suas interações com fatores de transcrição (FT) e alvos identificados através de análises de sequências regulatórias podem ser usados como uma validação *in silico* para as interações FT- alvo na rede gênica (Fortes et al., 2010). Assim, foi usado o *software* PlantTFDB 4.0 (Jin et al., 2017) para a análise de regiões regulatórias identificados (2.000 pb *upstream* e 200 pb *downstream* do sítio de início de transcrição) do conjunto de genes candidatos. Os FT foram identificados ($p < 0,00001$) e usados para uma análise de enriquecimento de processos biológicos via aplicativo Bingo (Biological Networks Gene Ontology Tool) (Maere et al., 2005) do *software* Cytoscape (Shannon et al., 2003). Os FT que apresentaram os processos biológicos mais relacionados com a característica em estudo (ex. Defesa em resposta à bactérias e fungos) foram selecionados para obtenção de uma rede com suas interações (FT-alvo).

A partir dos FT selecionados e os genes candidatos de cada análise de GWAS, foi possível obter uma rede gene-FT. Assim, de acordo com o número de sítios de ligação para os fatores de transcrição (SLFT), cada elo (genes e fatores de transcrição) da rede mais enriquecidos foram identificados. Ou seja, genes com um maior número de SLFT possuem maior número de ligações na rede, possibilitando maior enriquecimento dos mesmos. Desta forma, a rede gene-FT obtida permitiu uma visão global dos genes identificados nas análises de GWAS, bem como, destacar os genes candidatos mais prováveis para a característica em estudo.

3.5. Resultados e discussão

3.5.1. Análise fenotípica

Os resultados da análise fenotípica para resistência parcial podem ser encontrados com maior nível de detalhamento em Ludke et al. (2016), onde 24,85% dos genótipos testados foram altamente suscetíveis, 18,34% dos genótipos foram moderadamente suscetíveis, 7,69% foram moderadamente resistentes e 49,11% foram altamente resistentes.

Em relação a análise fenotípica para resistência completa, os dados encontram-se no artigo 1 deste presente trabalho, onde 47,44% foram resistentes, 5,13% foram moderadamente resistentes e 47,44% foram suscetíveis.

Comparando-se as duas análises, chegou-se a uma coincidência de 87% quanto aos resultados obtidos em relação a reação dos genótipos aos dois métodos de inoculação.

3.5.2. Análise de associação genômica (GWAS)

Como pode ser observado na tabela 4, a análise de associação genômica para resistência parcial à *P. sojae* para as 169 cultivares identificou a presença de cinco marcadores significativos no cromossomo 3 e três marcadores significativos no cromossomo 15 (Ludke et al, 2019).

Conforme observado no artigo 1, para a reação de resistência completa à *P. sojae* à 156 cultivares do conjunto de genótipos utilizados, foram identificados um marcador com associação significativa no cromossomo 2, quatro marcadores no cromossomo 3 e um marcador no cromossomo 4 (tabela 5).

Os marcadores Gm03_3225968, Gm03_5106459, Gm03_5217414 e Gm03_5341695 foram significativos tanto na análise de resistência parcial quanto na análise de resistência completa.

Tabela 4. Marcadores significativos associados a resistência parcial e completa de soja a *P. sojae*

Avaliação	SNP	Alelos	Cro ¹	Posição
Resistencia parcial	Gm03_3225968	G/A	3	3225968
	Gm03_4782127	T/C	3	4782127
	Gm03_5106459	T/G	3	5106459
	Gm03_5217414	C/T	3	5217414
	Gm03_5341695	A/G	3	5341695
	Gm15_11496274	T/C	15	11496274
	Gm15_18422604	T/C	15	18422604
	Gm15_19326210	T/C	15	19326210
Resistencia completa	Gm03_5217414	C/T	3	5217414
	Gm03_3225968	G/A	3	3225968
	Gm03_5341695	A/G	3	5341695
	Gm04_45613405	T/C	4	45613405
	Gm02_46504621	T/C	2	46504621
	Gm03_5106459	T/G	3	5106459

¹Cromossomo

3.5.3. Análise de redes gênicas

A partir dos 10 diferentes marcadores SNPs identificados nas GWAS para resistência parcial e completa, foram localizados 28 genes próximos a estes marcadores, sendo que, 10 destes genes ainda não foram caracterizados quanto a sua função biológica (Tabela 5). Após a identificação destes genes, os mesmos foram utilizados na busca por fatores de transcrição relacionados a diversos processos biológicos.

Na busca por fatores de transcrição baseado no *set* de genes candidatos, foram identificados 279 processos biológicos relacionados a centenas de fatores de transcrição distintos. Objetivando a elaboração de uma rede com maior especificidade, foram selecionados processos biológicos ligados de forma mais direta a interação planta-patógeno e mecanismos de defesas da planta contra fatores bióticos e, baseados neste critério, foram selecionados 31 processos biológicos distintos e 48 fatores de transcrição (tabela 7). Os processos biológicos e os fatores de transcrição podem repetir-se nos diferentes cromossomos e visando avaliar a rede de maneira mais específica, a mesma foi montada individualmente para cada cromossomo.

Na rede gene-FT gerada para cromossomo 2 (figura 6), foram identificados quatro genes, entretanto, o marcador Gm02_46504621 não se encontra inserido dentro de nenhum destes. Dentre este *set* de genes, o gene LOC100816582 foi o que apresentou maior enriquecimento na rede, este possui como anotação *mRNA-regulatory protein NPR3, transcript variant X1* e está ligado aos fatores de transcrição AT2G01570, AT2G25000, AT3G20840, AT4G18960, AT5G01900 e AT5G10510, que estão relacionados a processos biológicos envolvidos no estímulo síntese de ácido salicílico, desenvolvimento de sistema radicular e parte aérea e resposta a bactérias. Diversos membros da família das proteínas NPR (*non-expressor of pathogenesis-related protein*) foram identificados como receptores de ácido salicílico (AS) em plantas (Fu et al., 2012; Wu et al., 2012), Moreu et al. (2012) também demonstram que NPR3 se liga ao AS e resulta em mecanismo de defesa de plantas, da mesma maneira, diversos outros trabalhos mostram a correlação entre a presença desta família e a expressão de AS (Malamy et al., 1990; An & Mou, 2011; Yan & Dong, 2014), o que é relevante pelo fato de o AS ser um hormônio vegetal que está envolvido em diversos mecanismos nas plantas, dentre eles, respostas a estresses abióticos e bióticos (Klessig et al., 2016).

Outro gene identificado no cromossomo dois foi o LOC102661690 possui como anotação *mRNA-putative Myb family transcription factor At1g14600*, estando ligado aos fatores de transcrição AT1G13260, AT2G01570 e AT5G47230, que por sua vez se relacionam com a estrutura de desenvolvimento de raízes e parte aérea e também nas respostas celulares ao estímulo por etileno. A família de proteínas MYB é grande, funcionalmente diversa, encontrada em todos os eucariotos e estão envolvidas no controle de vários processos, tais como respostas a estresses abióticos e bióticos (Ambawat et al., 2013). Seo & Park (2010) submeteram plantas de *Arabidopsis*, que expressavam as proteínas MYB, a infecção por patógenos e identificaram que estas plantas ativavam a ação de genes relacionados a defesa da planta.

No cromossomo 3 foram identificados doze genes, dos quais, quatro ainda não foram caracterizados. Como pode ser observado na figura 7, o gene que apresentou maior enriquecimento de rede foi o LOC100801499 o qual possui como anotação *mRNA-phosphoglucomutase, cytoplasmic* e está ligado aos fatores de transcrição AT2G01570, AT4G18960, AT4G27950, AT5G03790 e AT5G47230, que se relacionam com os processos biológicos de resposta ao estímulo ao etileno e desenvolvimento do sistema radicular. Liao et al. (2012) identificou a proteína *Phosphoglucomutase, cytoplasmic* em estudo de interação solo-raiz, juntamente com proteínas induzidas pelo ataque de patógenos, todavia, não existem mais relatos desta proteína em estudos de relação planta-patógeno e mecanismos de defesa de plantas.

O segundo gene mais enriquecido na rede foi o LOC106798029 e dentro dele está localizado o marcador Gm03_3225968. Este gene ainda não foi caracterizado, entretanto, o mesmo encontra-se ligado a fatores de transcrição diretamente relacionados com processos biológicos diretamente ligados a interação planta-patógeno e aos mecanismos de defesa, tais como o de resposta celular ao estímulo de ácido jasmônico, que é um hormônio relacionado ao mecanismo de defesa das plantas (Berendsen et al., 2012), mudando a composição dos exudatos das raízes, dificultando assim a entrada de fitopatógenos pelo sistema radicular (Badri et al., 2008). Da mesma maneira, o gene LOC100799892 também não foi caracterizado, porém, encontra-se ligado ao fator de transcrição AT3G50410, que está presente em todos os processos biológicos identificados neste trabalho, relacionados ao ácido jasmônico.

O gene LOC100794060 não apresentou grande enriquecimento na rede, estando ligado aos fatores de transcrição AT2G01570, entretanto, sua caracterização é *mRNA-putative disease resistance RPP13-like protein 1, transcript variant X1*. O gene RPP13 e as proteínas derivadas da ação deste gene estão diretamente relacionadas a controle da resistência de plantas à patógenos, como oomycetos (Rose et al., 2004). Em soja, Zhang et al ao realizarem um estudo de associação genômica para arquitetura genética dos sintomas apresentados em plantas infestadas por *Fusarium virguliforme*, também identificaram genes candidatos com a mesma caracterização do gene LOC100794060. Zhong et al. (2017), em um estudo realizado com *Phytophthora sojae* identificou no cromossomo 3 genes candidatos com mesma anotação aos identificados neste trabalho.

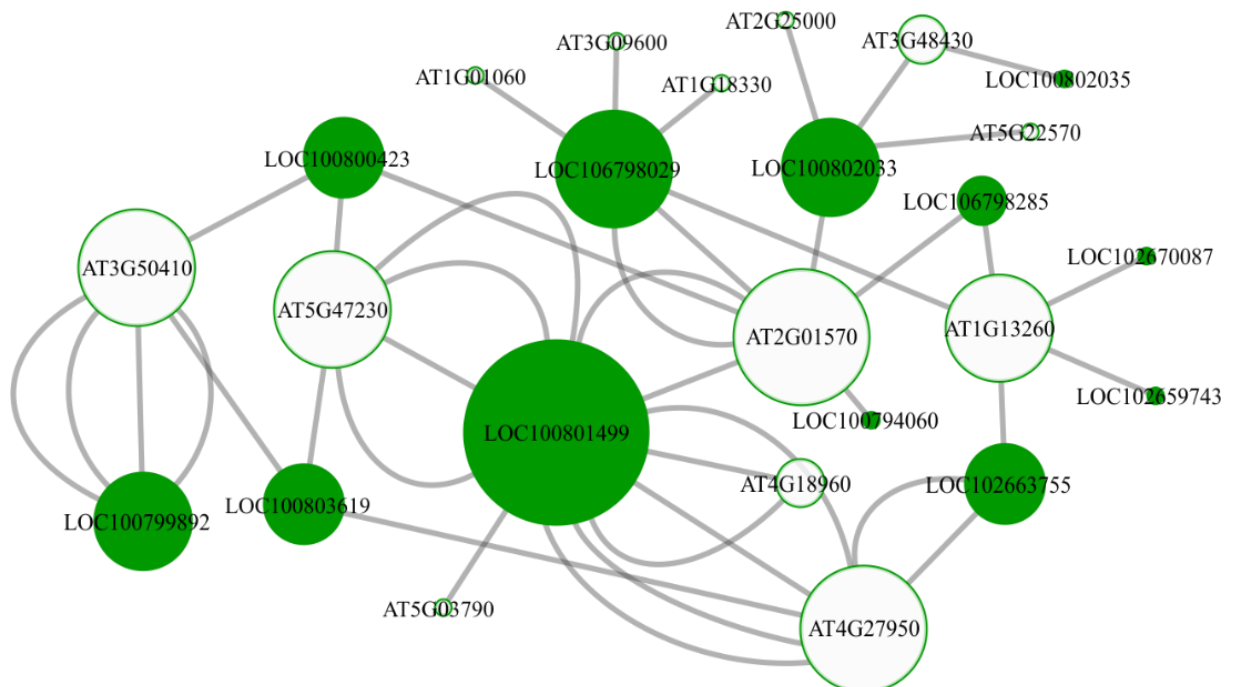


Figura 7. Rede gene-fator de transcrição do cromossomo 3

O gene mais enriquecido no cromossomo 4 (figura 8) foi o LOC102663858, estando ligado com onze fatores de transcrição distintos. Este gene não foi caracterizado, entretanto, está ligado a fatores de transcrição chave no mecanismo de resistência de plantas, como o AT5G47220, presente nos processos biológicos de resposta celular ao estímulo de ácido jasmônico, resposta celular a estresse abiótico, mecanismo de defesa contra bactérias, indução do sistema imune e regulação da ativação do sistema de defesa da planta. Outros FT importantes que estão ligados a este gene são o AT2G01570 e AT3G50410, que relacionados a resposta ao estímulo da produção de ácido salicílico.

A rede gene-ft no cromossomo 15 (figura 9) possui sete genes ligados a 38 fatores de transcrição distintos, o gene que apresentou maior enriquecimento foi o LOC100795745, que se ligou a 21 fatores de transcrição e tem como caracterização *mRNA-ATP-citrate synthase beta chain protein 1*. Na literatura não constam relatos de que exista correlação entre a síntese de citrato e o mecanismo de defesa das plantas, entretanto, ao se analisar os fatores de transcrição ligados a este gene, observa-se que estes encontram-se presentes em distintos processos biológicos relacionados a defesa de plantas, como resposta celular ao ácido jasmônico, resposta a fungos, resposta a bactérias, regulação de sistema imune, indução sistemática de resistência e resposta de defesa ao ataque de fungos.

O gene LOC106796110 está ligado a cinco fatores de transcrição distintos e tem como caracterização *mRNA-putative disease resistance protein RGA3*. Como demonstram alguns estudos, o uso de análogos de genes de resistência a doenças (RGAs) é extremamente importante na identificação de novos genes, uma vez que são parciais ou totalmente ligados a estes (Collins et al., 2001; Madsen et al., 2003) e estudos realizados com a cultura da soja que avaliaram a relação planta-patogeno e os mecanismos de defesa comprovam isso (Priolli et al., 2010; Tucker et al., 2010; Bhupender et al., 2014; Vieira et al., 2016).

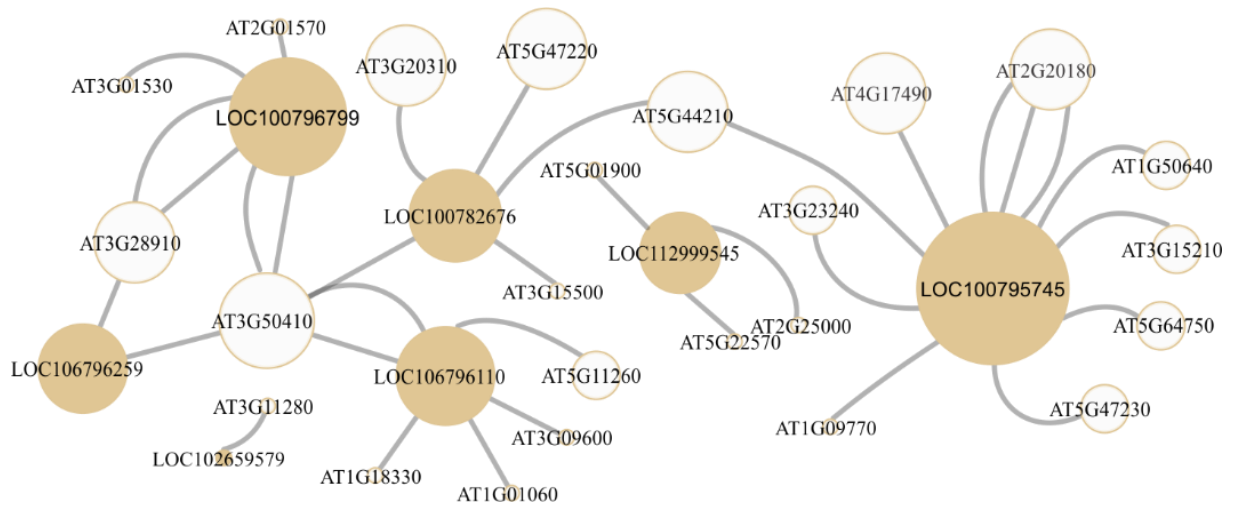


Figura 9. Rede gene-fator de transcrição do cromossomo 15

3.6 Conclusão

Neste estudo, após a identificação de 31 genes dispostos em quatro cromossomos, foram encontrados 279 processos biológicos relacionados a distintos fatores de transcrição e com estas informações foram montadas diferentes redes gene-FT. A montagem destas redes gene-FT permitiu uma visualização mais ampla e melhor compreensão dos mecanismos envolvidos, direta e indiretamente, na resistência de soja à *P. sojae*.

Tabela 5. Genes candidatos relacionados a resistência completa à *P. sojae* neste estudo.

Marcador	Crm ¹	Posição	Gene	Distância do SNP ²		Caracterização do gene
Gm02_46504621	2	46504621	LOC100815525	2.940	Upstream	mRNA-transmembrane protein 56-B, transcript variant X2
			LOC102661690	11.382	Downstream	mRNA-putative Myb family transcription factor At1g14600
			LOC100816582	7.858	Downstream	mRNA-regulatory protein NPR3, transcript variant X1
			LOC100817119	16.373	Downstream	mRNA-phytoene synthase 2, chloroplastic
Gm03_3225968	3	3225968	LOC106798285	2.677	Upstream	uncharacterized LOC106798285
			LOC106798029		Dentro do gene	mRNA-uncharacterized LOC106798029
			LOC102659743	2.617	Downstream	mRNA-probable pectinesterase/pectinesterase inhibitor 17
Gm03_4782127	3	4782127	LOC100803619		Dentro do gene	mRNA-peroxidase P7
			LOC100794060	8.442	Downstream	mRNA-putative disease resistance RPP13-like protein 1, transcript variant X1
			LOC102663755	5.098	Downstream	uncharacterized LOC102663755
			LOC100802035	9.408	Upstream	mRNA-peroxidase P7
Gm03_5217414	3	5217414	LOC100799892	5.155	Upstream	mRNA-uncharacterized LOC100799892
			LOC100800423	3.394	Upstream	mRNA-senescence associated gene 20
Gm03_5341695	3	5341695	LOC100801499	1.799	Upstream	mRNA-phosphoglucomutase, cytoplasmic
			LOC102670087	3.052	Downstream	mRNA-NAP1-related protein 2
			LOC100802033	9.043	Downstream	mRNA-7-methylxanthosine synthase 1
Gm04_45613405	4	45613405	LOC102665657	1.800	Upstream	mRNA-uncharacterized LOC102665657
			LOC100798680	1.877	Downstream	mRNA-cell division cycle protein 48 homolog
			LOC100804876	9.125	Downstream	mRNA-kinetochore protein SPC25 homolog, transcript variant X1
			LOC102663858	7.269	Downstream	uncharacterized LOC102663858
			LOC100803814	9.178	Upstream	mRNA-jmjC domain-containing protein 4, transcript variant X2
Gm15_11496274	15	11496274	LOC102659579		Dentro do gene	mRNA-uncharacterized LOC102659579
			LOC100795745	12.833	Upstream	mRNA-ATP-citrate synthase beta chain protein 1
			LOC100796799	14.445	Downstream	mRNA-TPD1 protein homolog 1, transcript variant X1

Continua...

Tabela 5. Continuação

Gm15_19326210	15	19326210	LOC106796110	10.608	Upstream	mRNA-putative disease resistance protein RGA3
			LOC100782676	540	Downstream	uncharacterized LOC100782676
			LOC106796259	11.741	Downstream	uncharacterized LOC106796259
			LOC112999545	8.054	Downstream	uncharacterized LOC112999545

¹ Cromossomo² Distancia do SNP em pares de base**Tabela 6.** Processos biológicos enriquecidos associados aos fatores de transcrição identificados para resistência à *Phytophthora sojae*

Crom ¹	GO-ID ²	<i>P.valeu</i> ³	Descrição do processo biológico	Fatores de transcrição
2	9863	7,53x10 ⁻³	salicylic acid mediated signaling pathway	AT5G01900, AT5G22570
	71446	7,53x10 ⁻³	cellular response to salicylic acid stimulus	AT5G01900, AT5G22570
	9617	6,40x10 ⁻³	response to bacterium	AT5G01900, AT5G22570, AT1G20980, AT3G20770, AT2G25000
	22622	5,27x10 ⁻³	root system development	AT4G10350, AT1G13260, AT1G79580, AT5G10510, AT3G20840
	48364	5,27x10 ⁻³	root development	AT4G10350, AT1G13260, AT1G79580, AT5G10510, AT3G20840
	22621	3,66x10 ⁻³	shoot system development	AT4G27950, AT1G13260, AT5G03790, AT3G48430, AT1G30210, AT4G18960
	48367	3,43x10 ⁻³	shoot development	AT4G27950, AT1G13260, AT5G03790, AT3G48430, AT1G30210, AT4G18960
	42742	2,50x10 ⁻³	defense response to bacterium	AT5G01900, AT5G22570, AT1G20980, AT3G20770, AT2G25000
	71369	1,53x10 ⁻³	cellular response to ethylene stimulus	AT1G73730, AT3G20770, AT5G47230
	9873	1,44x10 ⁻³	ethylene mediated signaling pathway	AT1G73730, AT3G20770, AT5G47230
	9751	5,08x10 ⁻⁵	response to salicylic acid stimulus	AT3G09600, AT5G01900, AT5G22570, AT1G01060, AT3G50410, AT2G25000
	9723	2,56x10 ⁻⁶	response to ethylene stimulus	AT3G09600, AT1G73730, AT1G01060, AT1G18330, AT2G01570, AT3G20770, AT5G47230
3	9863	7,53x10 ⁻³	salicylic acid mediated signaling pathway	AT5G01900, AT5G22570
	9617	6,40x10 ⁻³	response to bacterium	AT5G01900, AT5G22570, AT1G20980, AT3G20770, AT2G25000
	22622	5,27x10 ⁻³	root system development	AT4G10350, AT1G13260, AT1G79580, AT5G10510, AT3G20840

Continuação...

Tabela 6. Continuação

3	48364	5,27x10 ⁻³	root development	AT4G10350, AT1G13260, AT1G79580, AT5G10510, AT3G20840
	22621	3,66x10 ⁻³	shoot system development	AT4G27950, AT1G13260, AT5G03790, AT3G48430, AT1G30210, AT4G18960
	48367	3,43x10 ⁻³	shoot development	AT4G27950, AT1G13260, AT5G03790, AT3G48430, AT1G30210, AT4G18960
	42742	2,50x10 ⁻³	defense response to bacterium	AT5G01900, AT5G22570, AT1G20980, AT3G20770, AT2G25000
	71369	1,53x10 ⁻³	cellular response to ethylene stimulus	AT1G73730, AT3G20770, AT5G47230
	9873	1,44x10 ⁻³	ethylene mediated signaling pathway	AT1G73730, AT3G20770, AT5G47230
	9751	5,08x10 ⁻⁵	response to salicylic acid stimulus	AT3G09600, AT5G01900, AT5G22570, AT1G01060, AT3G50410, AT2G25000
4	9723	2,56x10 ⁻⁶	response to ethylene stimulus	AT3G09600, AT1G73730, AT1G01060, AT1G18330, AT2G01570, AT3G20770, AT5G47230
	22621	1,75x10 ⁻²	shoot system development	AT4G27950, AT2G02820, AT4G01500, AT4G31920, AT3G48430
	48367	1,66x10 ⁻²	shoot development	AT4G27950, AT2G02820, AT4G01500, AT4G31920, AT3G48430
	9867	1,59x10 ⁻²	jasmonic acid mediated signaling pathway	AT5G47220, AT2G01570
	71395	1,59x10 ⁻²	cellular response to jasmonic acid stimulus	AT5G47220, AT2G01570
	9753	6,24x10 ⁻³	response to jasmonic acid stimulus	AT1G49010, AT5G47220, AT1G01060, AT2G01570
	9751	4,52x10 ⁻³	response to salicylic acid stimulus	AT1G49010, AT1G01060, AT3G50410, AT2G04880
	71214	3,10x10 ⁻³	cellular response to abiotic stimulus	AT3G22170, AT4G15090, AT2G20180
	71369	3,13x10 ⁻⁹	cellular response to ethylene stimulus	AT3G20310, AT5G47220, AT4G17490, AT1G73730, AT5G44210, AT5G47230, AT4G16110
	9873	2,67x10 ⁻⁹	ethylene mediated signaling pathway	AT3G20310, AT5G47220, AT4G17490, AT1G73730, AT5G44210, AT5G47230, AT4G16110
15	9723	6,20x10 ⁻¹⁰	response to ethylene stimulus	AT3G20310, AT5G47220, AT4G17490, AT1G73730, AT5G44210, AT1G01060, AT1G18330, AT2G01570, AT5G47230, AT4G16110
	9620	1,71x10 ⁻²	response to fungus	AT5G61600, AT1G09770, AT5G64750, AT4G34410, AT2G25000
	42742	1,08x10 ⁻²	defense response to bacterium	AT3G15210, AT5G47220, AT5G01900, AT5G22570, AT1G09770, AT2G25000

Continuação...

Tabela 6. Continuação

	9617	8,63x10 ⁻³	response to bacterium	AT3G15210, AT5G47220, AT5G01900, AT5G22570, AT1G09770, AT3G28910, AT2G25000
	50776	7,01x10 ⁻³	regulation of immune response	AT3G15210, AT5G47220, AT1G09770
	2682	7,01x10 ⁻³	regulation of immune system process	AT3G15210, AT5G47220, AT1G09770
	10104	6,61x10 ⁻³	regulation of ethylene mediated signaling pathway	AT3G15210, AT1G50640
	45088	5,66x10 ⁻³	regulation of innate immune response	AT3G15210, AT5G47220, AT1G09770
	50832	5,35x10 ⁻³	defense response to fungus	AT5G61600, AT1G09770, AT5G64750, AT4G34410, AT2G25000
	9682	3,88x10 ⁻³	induced systemic resistance	AT3G15210, AT5G47220
	31347	3,47x10 ⁻³	regulation of defense response	AT3G15210, AT5G47220, AT1G09770, AT2G25000
	31349	3,17x10 ⁻³	positive regulation of defense response	AT3G15210, AT5G47220, AT1G09770
15	71214	2,44x10 ⁻³	cellular response to abiotic stimulus	AT3G59060, AT5G11260, AT2G20180, AT2G43010
	45089	2,12x10 ⁻³	positive regulation of innate immune response	AT3G15210, AT5G47220, AT1G09770
	50778	2,12x10 ⁻³	positive regulation of immune response	AT3G15210, AT5G47220, AT1G09770
	2684	2,12x10 ⁻³	positive regulation of immune system process	AT3G15210, AT5G47220, AT1G09770
	2218	1,69x10 ⁻³	activation of innate immune response	AT3G15210, AT5G47220, AT1G09770
	2253	1,69x10 ⁻³	activation of immune response	AT3G15210, AT5G47220, AT1G09770
	9864	1,33x10 ⁻³	induced systemic resistance, jasmonic acid mediated signaling pathway	AT3G15210, AT5G47220
	9867	1,77x10 ⁻⁶	jasmonic acid mediated signaling pathway	AT3G15210, AT5G47220, AT3G23240, AT2G01570, AT3G01530, AT3G15500
	71395	1,77x10 ⁻⁶	cellular response to jasmonic acid stimulus	AT3G15210, AT5G47220, AT3G23240, AT2G01570, AT3G01530, AT3G15500

¹ Cromossomo² Identificação da ontologia gênica³ Probabilidade de significância

3.7 Referências bibliográficas

- Ambawat, S et al (2013) MYB transcription factor genes as regulators for plant responses: an overview. **Physiol. Mol. Biol. Plants** **19** (3): 307–321.
- An, C; Mou, Z (2011) Salicylic acid and its function in plant immunity. **J. Integr. Plant Biol.** **53**: 412–428.
- Badri, DV et al (2008) Transcriptome analysis of Arabidopsis roots treated with signaling compounds: a focus on signal transduction, metabolic regulation and secretion. **New Phytol.** **179**: 209–223.
- Berendsen, RL et al (2012) The rhizosphere microbiome and plant health. **Trends in Plant Science** **17**(8): 478–486.
- Bhupender, K et al (2014) Mapping of yellow mosaic virus (YMV) resistance in soybean (*Glycine max* L. Merr.) through association mapping approach. **Genetica** **143** (1): 1–10.
- Bhupender, K et al (2014) Screening of soybean [*Glycine max* (L.) Merr.] genotypes for yellow mosaic virus (YMC) disease resistance and their molecular characterization using RGA and SSRs markers. **Australian J. of Crop Sci.** **8** (1): 27–34.
- Boyle, EA et al (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. **Cell.** 169 (7): 1177–1186.
- Brodie A et al (2016) How far from the SNP may the causative genes be? **Nucleic Acids Research** **44**: 6046 – 6054.
- Collins, N et al (2001) Resistance gene analogs in barley and their relationship to rust resistance genes. **Genome** **44**: 375–381.
- Costamilan LM (2006) Reação de cultivares de soja à *Phytophthora sojae*. In: Reunião de pesquisa de soja da região sul. **Ata e resumos**. Pelotas: Embrapa Clima Temperado.
- Costamilan LM et al (2007) **Podridão radicular de fitóftora em soja – Documentos Online 79**. Passo Fundo, RS: Embrapa Trigo. 23 p.
- Costamilan LM et al (2013) Pathogenic diversity of *Phytophthora sojae* pathotypes from Brazil. **European Journal of Plant Pathology** **135**(4): 845–853.
- Dorrance AE et al (2003) Effect of partial resistance on *Phytophthora* stem rot incidence and yield of soybean in Ohio. **Plant disease** **81**: 308-312.

- Dorrance, AE (2003) Effect of partial resistance on Phytophthora stem rot incidence and yield of soybean in Ohio. **Plant Dis.** **87**: 308–312.
- Fang, C et al (2017) Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. **Genome Biology** **18** (1): 161.
- Fortes MR et al (2010) Association weight matrix for the genetic dissection of puberty in beef cattle. **Proc. Natl. Acad. Sci. USA.** **107**: 13642–13647.
- Fu, ZQ et al (2012) NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. **Nature** **486**: 228–232.
- Gao H et al (2007) A Markov Chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. **Genetics** **176** (3): 1635–1651.
- Hartman, GL et al (2015). **Compendium of soybean diseases and pests**. 5 ed. American Phytopathological Society Press, Saint Paul.
- Huang J et al (2016) Phenotypic evaluation and genetic dissection of resistance to Phytophthora sojae in the Chinese soybean mini core collection. **BMC genetics** **17** (85): 1 – 14.
- Jin J et al (2017) PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interaction in plants. **Nucleic Acids Res.** **45** (1): 1040–1045.
- Klessig, DF et al (2016) Multiple targets of salicylic acid and its derivatives in plants and animals. **Front. Immunol.** **7**: 206.
- Kou, Y; Wang, S (2010) Broad-spectrum and durability: understanding of quantitative disease resistance. **Curr. Opin. Plant Biol.** **13**: 181–185.
- Liao C et al (2012) Comparative analyses of three legume species reveals conserved and unique root extracellular proteins. **Proteomics** **12**: 3219–3226.
- Lipka AE et al (2012) GAPIT: genome association and prediction integrated tool. **Bioinformatics** **28**: 2397–9.
- Madsen, LH et al (2003) Barley disease resistance gene analogs of the NBS-LRR class: identification and mapping. **Mol. Gen. and Genom.** **269**: 150–161.
- Maere S et al (2005) BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. **Bioinformatics** **21**: 3448–3449.

- Malamy, J et al (1990) Salicylic acid: a likely endogenous signal in the resistance response of tobacco to viral infection. **Science** **250**: 1002–1004.
- Moreu, M et al (2012) Salicylic acid binds NPR3 and NPR4 to regulate NPR1-dependent defense responses. **Cell Research** **22**: 1631–1633.
- Nadeem, MA et al (2017) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. **Biol. and Biotech. equip.** **32**: 1–25.
- Niu, J et al (2017) Fine Mapping of a Resistance Gene RpsHN that controls *Phytophthora sojae* using recombinant inbred lines and secondary populations. **Frot. Plant Sci.** **8**: 538.
- O’Leary, NA et al (2016) Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion and functional annotation. **Nucleic Acids Research** **44** (1): 733–745.
- Priolli, RHG et al (2010) Genetic diversity among Brazilian soybean cultivars based on SSR loci and pedigree data. **Braz. Arch. Biol. Technol.** **53** (3): 519–531.
- Rose, EL et al (2004) The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in *Arabidopsis thaliana*. **Genetics** **166**: 1517–1527.
- Savary S et al (2019) The global burden of pathogens and pests on major food crops. **Nat. Ecol. Evol.** **3**: 430–439.
- Schmitthenner AF (1985) Problems and Progress in Control of Phytophthora Root-Rot of Soybean. **Plant Dis.** **69**: 362–368.
- Schmitthenner AF, Bhat RG (1994) Useful methods for studying *Phytophthora* in the laboratory. **Res. Dev. Cent. Spec. Circ.** **143** 10 p.
- Seo, PJ; Park, CM (2010) MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in *Arabidopsis*. **New Phytologist** **186**: 471–483.
- Shannon P et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome Research** **13**(11): 2498–2504.
- Sugimoto, T et al (2012) Pathogenic diversity of *Phytophthora sojae* and breeding strategies to develop Phytophthora-resistant soybeans. **Bree. Sci.** **61**: 511–522.
- Tucker, DM et al (2010) Mapping quantitative trait loci for partial resistance to *Phytophthora sojae* in a soybean interspecific cross. **Crop Sci.** **50** (2): 628–635.

- Verardo LL et al (2016) After genome-wide association studies: Gene networks elucidating candidate genes divergences for number of teats across two pig populations. **J. Anim. Sci.** **94**: 1446–1458.
- Vieira, PMH et al (2016) Comparative analysis of soybean genotype resistance to *Heterodera glycines* and *Meloidogyne* species via resistance gene analogs. **Genet. Mol. Res.** **15** (3): gmr.15038562
- Wu, Y et al (2012) The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. **Cell Rep.** **1**: 639–647.
- Yan, S; Dong, X (2014) Perception of the plant immune signal salicylic acid. **Curr. Opin. In Plant Biol.** **20**: 64–68.
- Zhang Z et al (2010) Mixed linear model approach adapted for genome-wide association studies. **Nature genetics** **42**: 355–360.
- Zhong, C et al (2017) Next-generation sequencing to identify candidate genes and develop diagnostic markers for a novel Phytophthora resistance gene, *RpsHC18*, in soybean. **Theoretical and Applied Genetics** **131** (3): 525–538.
- Zhu ZD et al (2003) Distribution and virulence diversity of Phytophthora sojae in China. **Agri Sci China** **3**: 116–123.