

VITOR SEITI SAGAE

**MAPEAMENTO DE HERDABILIDADE REGIONAL PARA RESPOSTA DE
CULTIVARES DE SOJA (*Glycine max* L.) SUBMETIDAS AO DEFICIT HÍDRICO**

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

Orientador: Moysés Nascimento

Coorientadores: Camila Ferreira Azevedo
Felipe Lopes da Silva

**VIÇOSA – MINAS GERAIS
2022**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

S129e Sagae, Vitor Seiti, 1998-
2022 Mapeamento de herdabilidade regional para resposta de
cultivares de soja (*Glycine max* L.) submetidas ao deficit hídrico
/ Vitor Seiti Sagae. – Viçosa, MG, 2022.
1 dissertação eletrônica (54 f.): il. (algumas color.).

Inclui apêndice.

Orientador: Moysés Nascimento.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Estatística, 2022.

Referências bibliográficas: f. 45-52.

DOI: <https://doi.org/10.47328/ufvbbt.2022.770>

Modo de acesso: World Wide Web.

1. Soja - Melhoramento genético. 2. Soja - Resistência à
seca. 3. Marcadores genéticos. I. Nascimento, Moysés, 1979-
II. Universidade Federal de Viçosa. Departamento de Estatística.
Programa de Pós-Graduação em Genética e Melhoramento.
III. Título.

CDD 22. ed. 631.52


VITOR SEITI SAGAE

**MAPEAMENTO DE HERDABILIDADE REGIONAL PARA RESPOSTA DE
CULTIVARES DE SOJA (*Glycine max* L.) SUBMETIDAS AO DEFICIT HÍDRICO**


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

APROVADA: 15 de setembro de 2022.

Assentimento:

Documento assinado digitalmente
 VITOR SEITI SAGAE
Data: 21/12/2022 19:08:21-0300
Verifique em <https://verificador.iti.br>

Vitor Seiti Sagae
Autor

Documento assinado digitalmente
 MOYSES NASCIMENTO
Data: 12/01/2023 14:55:54-0300
Verifique em <https://verificador.iti.br>

Moisés Nascimento
Orientador

Dedico

Aos meus pais, Vanesia e Fernando

À minha irmã, Andressa

À minha noiva, Sara

AGRADECIMENTOS

Agradeço a Deus pelo dom da vida, por toda proteção, pelos desafios e superações durante este percurso.

À minha família, em especial meus pais Vanesia e Fernando, e minha irmã Andressa pelo amor e apoio durante toda a minha vida.

À minha noiva Sara, por todo carinho, paciência e apoio incondicional para a realização deste sonho.

Aos amigos que mesmo de longe estiveram presentes em todas as conquistas e dificuldades.

Ao meu orientador Prof^o. Moisés, e coorientadores Prof^a. Camila e Prof^o Felipe, por todos os ensinamentos, conselhos e discussões, seja na sala de aula ou fora dela, que contribuíram muito para minha evolução pessoal e profissional.

Aos amigos e colegas do LICAE e Programa Soja da UFV pela convivência, risadas, discussões, pelo trabalho em equipe e companheirismo, que tornaram este período de grande crescimento.

À Universidade Federal de Viçosa, e o Programa de Pós-Graduação em Genética e Melhoramento pela oportunidade de me aperfeiçoar nessa área.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela concessão da bolsa de estudos que possibilitou a realização deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

SAGAE, Vitor Seiti, M.Sc., Universidade Federal de Viçosa, setembro de 2022. **Mapeamento de herdabilidade regional para resposta de cultivares de soja (*Glycine max* L.) submetidas ao déficit hídrico.** Orientador: Moysés Nascimento. Coorientadores: Camila Ferreira Azevedo e Felipe Lopes da Silva.

A soja (*Glycine max* L.) é atualmente o cultivo agrícola de maior importância para o Brasil, ocupando grande parte da área cultivada no país na safra 2021/2022. E apesar da crescente demanda mundial por este grão, a produtividade tem sido afetada frequentemente devido a ocorrência de estiagens durante as fases de desenvolvimento da planta. Uma das alternativas para aumentar a quantidade e qualidade de grãos produzidos é por meio do melhoramento genético de plantas, que vem evoluindo com a melhoria de tecnologias relacionadas ao sequenciamento de DNA. Essas tecnologias possibilitam o uso de informações oriundas de marcadores moleculares em estudos como os de associação genômica ampla (GWAS) para identificação de alelos de interesse. Porém, os métodos tradicionais de associação possuem limitações, como a grande ocorrência de falsos positivos e não detecção de alelos raros, e nesse sentido, novas abordagens como o mapeamento de herdabilidade regional (RHM) foram desenvolvidas. Desta forma, o objetivo do presente trabalho foi avaliar o uso e a eficiência da metodologia de mapeamento de herdabilidade regional na cultura da soja em comparação as abordagens convencionais, para as características período reprodutivo (PR), número de vagens (NV), número de grãos (NG) e porcentagem de vagens imaturas abertas (PVA), que são relacionadas a produção e qualidade de grãos. Os dados analisados correspondem a 99 genótipos sob dois regimes de irrigação (Controle e Estresse), em delineamento de blocos casualizados com 3 repetições. Foram utilizados 4070 marcadores do tipo SNP (*single nucleotide polymorphism*) para as análises de associação genômica via marcas únicas, e o mapeamento de herdabilidade regional com tamanhos de região de 2, 4,75 e 7,5 Mb, ambas com correção para estrutura de população. As metodologias foram comparadas com base no número de associações identificadas e porcentagem da herdabilidade genômica explicada (PE). Foram identificados 1 SNP para PVA e 2 SNPs para NG e NV pela abordagem via marcas únicas. Por outro lado, a análise de RHM identificou 1 região associada a PR, 4

regiões associadas a NV, 4 regiões associadas a NG, e 6 regiões associadas a PVA. O aumento no tamanho das regiões proporcionou maior quantidade de associações para PVA e PR. A análise de RHM explicou maior porção da variância genética aditiva para todas as características avaliadas. Dentro das regiões significativas, foi possível identificar possíveis genes candidatos ao controle das características em resposta as condições de déficit hídrico.

Palavras-chave: Estudo de associação genômica ampla. Efeitos poligênicos. Estresse abiótico.

ABSTRACT

SAGAE, Vitor Seiti, M.Sc., Universidade Federal de Viçosa, September, 2022. **Regional heritability mapping in soybeans (*Glycine max* L.) cultivars response to drought stress.** Adviser: Moysés Nascimento. Co-advisers: Camila Ferreira Azevedo and Felipe Lopes da Silva.

Soybean (*Glycine max* L.) currently is the major crop for Brazil with the most cultivated area in the season 2021/2022, despite that, there is a crescent demand for this grain, which in return had been suffering with losses increasingly often caused by drought stress. Plant breeding allows for raising productivity, and the advances in plant genomics for scan plant DNA allow to use of molecular markers information to perform genome-wide association studies (GWAS) to identify loci of interest, but the traditional methods of GWAS based on single marker associations have some limitations, thus, new approaches such as Regional Heritability Mapping (RHM) were developed to overcome it. This study aimed to compare single marker association and regional heritability mapping approaches. The traits evaluated were reproductive cycle (PR), pod number (NV), seed number (NG), and percentage of premature pod-shattering (PVA) under both control and drought stress conditions. The analyzed data correspond to 99 genotypes under two irrigation regimes (Control and Stress), conducted in a randomized block design with three replications. 4070 SNP were used markers to perform single marker association and RHM approaches, both corrected for population structure, and the window size used for RHM was 2, 4.75 e 7.5 Mb. The methods were compared by the number of identified associations and the percentage of genomic variance explained (PE). For single markers analysis, 1 SNP marker was associated with PVA and 2 SNPs with NG and NV, RHM analysis identified 15 regions associated with the traits evaluated and the increase of window size raised power detection for PVA and PR. Also, RHM analysis showed a high percentage of genomic variance explained compared to single marker method for all traits evaluated. Inside the associated regions was possible to identify some candidate genes that could be related to the traits in response to drought stress conditions.

Keywords: Genome-wide association studies. Polygenic effects. Abiotic stress

LISTA DE ILUSTRAÇÕES

- Figura 1.** Distribuição dos marcadores SNPs entre os cromossomos da Soja (*Glycine max* L.).21
- Figura 2.** Biplot do PC1 x PC2 com os genótipos avaliados sobre os tratamentos de regime de irrigação. As setas em vermelho indicam as características fenotípicas consideradas na análise. PVA: Porcentagem de vagens imaturas abertas. PR: Período reprodutivo. NV: Número de vagens. NG: Número de grãos.27
- Figura 3. (A)** Gráfico de dispersão referente aos diferentes grupos identificados pela análise de ADPC. **(B)** Representação gráfica do agrupamento dos genótipos dentro dos grupos considerados pela análise de ADPC com base nos valores probabilidade de agrupamento. Grupos 1,2 e 3 referem-se às subpopulações identificadas na análise.....28
- Figura 4.** Manhattan plots das análises de associação realizadas para a característica PVA (A), PR (B), NG (C) e NV (D), baseado no valor de $-\log_{10} p^*$ valor no eixo y ao longo dos 20 cromossomos da Soja (*Glycine max* L.), com base na análise via marcas únicas representadas pelos losangos vermelhos. A linha tracejada vermelha indica o nível de significância a 5% considerando os valores de p^* valor corrigidos por FDR. PVA: Porcentagem de vagens imaturas; PR: Período reprodutivo; NG: Número de grãos; NV: Número de vagens.30
- Figura 5.** Desequilíbrio de ligação (r^2) entre os pares de alelos dos marcadores SNPs ao longo do genoma da Soja (*Glycine max* L.) em Mb para todos os genótipos avaliados. A linha azul demonstra o decaimento do DL com os valores ajustados pelo método de regressão loess. A linha tracejada horizontal indica o valor de $r^2 = 0.20$. A linha vertical vermelha indica o valor correspondente a distância física de 2 Mb, a linha azul a distância de 4,75 Mb e a linha verde, a distância de 7,5 Mb.32
- Figura 6.** Distribuição das regiões ao longo dos 20 cromossomos do genoma da Soja (*Glycine max* L.) para a característica PVA. Cada barra lateral indica o tamanho da região utilizada nos casos em que houve significância a 5%. O mapa de calor indica a herdabilidade da região estimada, e as regiões destacadas em verde são aquelas declaradas como significantes pelo teste.35
- Figura 7.** Distribuição das regiões ao longo dos 20 cromossomos do genoma da Soja (*Glycine max* L.) para a característica PR. Cada barra lateral indica o tamanho da região utilizada nos casos em que houve significância a 5%. O mapa de calor indica a herdabilidade da região estimada, e as regiões destacadas em verde são aquelas declaradas como significantes pelo teste.36
- Figura 8.** Distribuição das regiões ao longo dos 20 cromossomos do genoma da Soja (*Glycine max* L.) para a característica NG. Cada barra lateral indica o tamanho da região utilizada nos casos em que houve significância a 5%. O mapa de calor indica a herdabilidade da região estimada, e as regiões destacadas em verde são aquelas declaradas como significantes pelo teste.36

Figura 9. Distribuição das regiões ao longo dos 20 cromossomos do genoma da Soja (*Glycine max* L.) para a característica NV. Cada barra lateral indica o tamanho da região utilizada nos casos em que houve significância a 5%. O mapa de calor indica a herdabilidade da região estimada, e as regiões destacadas em verde são aquelas declaradas como significantes pelo teste.37

LISTA DE TABELAS

- Tabela 1.** Localização dos marcadores moleculares nos cromossomos (Crom.) em que foram obtidas significâncias três características pela análise via marcas únicas.31
- Tabela 2.** Distribuição, número de regiões construídas e DL médio por cromossomo dos marcadores moleculares utilizados na análise de mapeamento de herdabilidade regional.33
- Tabela 3.** Identificação da localização (Cromossomo), tamanho e número de SNPs das regiões detectadas como significativas na análise de RHM, a herdabilidade explicada pela região e sua proporção de porcentagem explicada em comparação a herdabilidade genômica.38
- Tabela 4.** Lista de genes reportados na literatura para as características avaliadas identificados dentro das regiões significantes, localizações no genoma, e descrições das proteínas traduzidas.39

LISTA DE SIGLAS E ABREVIATURAS

RHM – *regional heritability mapping*

NV – Número de vagens

PR – Período reprodutivo

NG – Número de grãos

PVA – Porcentagem de vagens imaturas abertas

ANOVA – Análise de variância

MAF – *Minor allele frequency*

SNP – *single nucleotide polymorphism*

QTL – *Quantitative trait locus*

ADCP – Análise de discriminantes por componentes principais

DL – Desequilíbrio de ligação

pb – pares de base

REML/BLUP – *restricted maximum likelihood/Best linear unbiased prediction*

FDR – *false discovery rate*

TRV – Teste de razão de verossimilhança

LISTA DE MATERIAIS SUPLEMENTARES

Tabela complementar 1. Genótipos de Soja utilizados na condução e análises do experimento.....	52
---	----

SUMÁRIO

MAPEAMENTO DE HERDABILIDADE REGIONAL PARA RESPOSTA DE CULTIVARES DE SOJA (*Glycine max* L.) SUBMETIDAS AO DEFICIT HÍDRICO 14

1 Introdução	14
2 Materiais e métodos	18
2.1 Dados fenotípicos.....	18
2.2 Dados moleculares e controle de qualidade	20
2.3 Análise fenotípica e identificação de subpopulações	21
2.4 Análise de associação genômica via marcas únicas.....	22
2.5 Desequilíbrio de ligação	23
2.6 Análise de mapeamento de herdabilidade regional	24
2.7 Comparação das metodologias de associação genômica	25
2.8 Potenciais genes relacionados ao controle das características	26
3 Resultados	27
3.1 Análise fenotípica e de estrutura de população	27
3.2 Análise de associação genômica via marcas únicas.....	29
3.3 Desequilíbrio de ligação	31
3.4 Análise de mapeamento de herdabilidade regional	33
3.5 Genes candidatos ao controle das características.....	38
4 Discussão	40
4.1 Análise fenotípica.....	40
4.2 Comparação entre as metodologias de associação genômica (GWAS)	40
4.3 Genes candidatos	42
5 Considerações finais	44
Referências bibliográficas	45
Material suplementar	52

MAPEAMENTO DE HERDABILIDADE REGIONAL PARA RESPOSTA DE CULTIVARES DE SOJA (*Glycine max* L.) SUBMETIDAS AO DEFICIT HÍDRICO

1 Introdução

Considerada uma importante fonte de proteína e óleo do mundo, a soja (*Glycine max* L.) é também a principal cultura em nível de importância para o Brasil, que teve uma área cultivada de aproximadamente 41 milhões de hectares na safra 2021/2022 (CONAB, 2022). A sua importância está relacionada ao seu uso tanto para a alimentação humana quanto animal, e por isso, há uma crescente demanda global de aumento de produção de soja baseada nas estimativas de aumento populacional nos próximos anos (MASUDA E GOLDSMITH, 2009; RAY et al., 2013).

Uma das formas de se aumentar a produção com o intuito de suprir a demanda de soja, ocorre pela exploração do potencial genético via melhoramento de plantas (ANDERSON, et al., 2019). Estudos desenvolvidos com a cultura da soja demonstram a obtenção de um ganho genético médio anual de 1,1% nos principais países produtores, que é menor do que o ganho necessário para atender a demanda do mundo para o ano de 2050 (DE FELIPE, 2016; RAY et al., 2013). O acréscimo no ganho genético é dificultado muitas vezes pela natureza genética complexa das características de interesse, como a produtividade, em que o valor fenotípico é resultado da ação de vários genes de pequeno efeito, e também, devido à baixa herdabilidade pela ação do efeito aleatório do ambiente sobre a característica (CRUZ, 2012; SILVA et al., 2017).

Considerando os fatores bióticos e abióticos que tem potencial de afetar a produtividade no cultivo da soja no Brasil, o que tem gerado perdas de forma cada vez mais frequente nas últimas safras, é a ocorrência de déficit hídrico em algumas regiões produtoras (HIRAKURI, 2021). Foram registrados períodos de intensa estiagem na região do MATOPIBA em 2015/2016, no Rio Grande do Sul em 2019/2020, e mais recentemente no Oeste do Paraná em 2021/2022, ocasionando severas reduções na produção principalmente pela ocorrência deste fenômeno nas fases reprodutivas da planta (HIRAKURI, 2021; SALVIANO et al., 2021).

Condições de déficit hídrico ocasionam perdas de produtividade por afetar os componentes de produção da planta, como o número de vagens por planta e o número

de grãos por vagem, principalmente quando ocorrem nos períodos de florescimento, formação das vagens e enchimento de grãos (KORTE et al., 1983; NAVARRO JÚNIOR E COSTA, 2002; SUN et al., 2006). Além disso, o tempo reprodutivo da planta também é afetado por condições de déficit hídrico, fazendo com que a quantidade de dias necessários para que ocorra o pleno desenvolvimento das estruturas reprodutivas não seja devidamente atingida (SIONIT E KRAMER, 1977).

A redução de produtividade é o principal problema causado por condições de estresse hídrico, porém, este fenômeno também tem a capacidade de afetar a qualidade dos grãos produzidos (LIU et al., 2022). Há estudos que relacionam a ocorrência de abertura de vagens imaturas à condições de déficit hídrico, que além de causar perdas de produtividade, afeta os componentes de produção, gera queda de grãos, e também diminui a sanidade por aumentar a suscetibilidade das vagens a entrada de patógenos (PALHARINI, 2016).

Tendo em vista o histórico das últimas safras agrícolas, e as previsões acerca das mudanças na disponibilidade e qualidade de água que acompanham as demais alterações climáticas, o melhoramento genético de plantas tem atuado no desenvolvimento de genótipos com alta plasticidade fenotípica, ou seja, plantas com a capacidade de adaptação às mudanças ambientais, como casos de estresses hídricos (SATBHAI et al., 2015; KUNERT et al., 2016). Para o desenvolvimento de genótipos tolerantes a seca, é importante o entendimento acerca da resposta da planta quando exposta a determinadas condições de estresse e quais genes atuam em tais ocasiões, para então identificar possíveis alelos de interesse, e posteriormente associa-los em um novo material superior (BILYEU et al., 2017; DUBEY et al., 2019).

Os estudos realizados com a cultura da soja indicam que a resposta morfológica e fisiológica da planta às condições de estresse hídrico, é condicionada pela ação de vários genes de efeito aditivo, geralmente sendo de baixa herdabilidade, o que dificulta o processo de seleção (KORIR, et al., 2011; MAPHOSA et al., 2012; TEODORO et al., 2019). Dentro do contexto atual de produção agrícola, o desenvolvimento de genótipos com tolerância a seca, além de atender a demanda dos agricultores para evitar prejuízos comerciais, torna-se uma questão de segurança alimentar do mundo moderno, principalmente considerando as mudanças climáticas previstas (DUBEY et al., 2019).

Para detecção de QTLs (*quantitative trait loci*) que participam do controle da produtividade ou de características relacionadas, a utilização de populações geradas a partir de parentais contrastantes foi amplamente utilizada para obtenção de mapas de ligação, possibilitando o estudo da base genética de características complexas (KARIKARI et al., 2019). Porém, esse tipo de abordagem possui suas limitações por explorar somente a variabilidade existente dentro de cruzamentos bi parentais (BHAT, et al., 2022).

A evolução de tecnologias envolvidas nas técnicas de sequenciamento de DNA para obtenção das informações de marcadores moleculares permitiu o avanço de estratégias como a seleção genômica, principalmente devido a maior eficiência e redução no custo para obtenção das informações genotípicas, o que torna o uso da seleção genômica cada vez mais atrativo (MUIR et al., 2016). Um dos tipos de polimorfismo observados nas sequências de genomas de indivíduos são os chamados de SNPs (*single nucleotide polymorphism*), que são considerados os marcadores moleculares mais abundantes e com ampla distribuição por todo o genoma de diferentes espécies, sendo muito utilizados em estudos de associação e predição genômica (AGARWAL et al., 2008).

Um dos principais métodos utilizados para obtenção de informações genômicas é a genotipagem por sequenciamento (*genotyping by sequencing*), que identifica polimorfismos em um único nucleotídeo de forma simultânea a descoberta de novos marcadores. Essa técnica é baseada no uso de enzimas de restrição que cortam o genoma em porções menores, e posteriormente são ligadas nesses sítios de corte a *barcodes* em uma das extremidades e a adaptadores comuns nas outras para posteriormente passar pela amplificação por meio da PCR (*polymerase chain reaction*), depois, esses fragmentos são alinhados com genomas de referência para leitura por meio de técnicas de bioinformática (ELSHIRE et al., 2011).

As plataformas de sequenciamento de nova geração (*next generation sequencing*) permitiram a identificação de variações no genoma como translocações, inversões e polimorfismo considerando locais específicos no genoma por meio de chips de DNA (DE CARVALHO, 2010). Esses chips recebem amostras de RNA com marcações fluorescentes da espécie em estudo, e após a ligação entre as sequências presentes no chip e as sondas oriundas de RNA, as amostras são avaliadas quanto a

emissão da fluorescência, tornando possível identificar o alelo presente nessa amostra. Posteriormente ocorre uma filtragem com o intuito de remover os primeiros nucleotídeos oriundos de adaptadores e diferentes sequências em um único compartimento do chip (GRIFFITHS et al, 2013).

O avanço das técnicas de sequenciamento possibilitou o surgimento de alternativas mais eficazes para detecção de QTLs (*quantitative trait loci*), como por exemplo os estudos de associação genômica ampla (*Genome-Wide Association Studies – GWAS*), que permitem a identificação de associações entre marcadores moleculares e genes que regulam características de interesse, isso devido ao desequilíbrio de ligação existente entre as marcas e os QTLs (ALQUDAH et al., 2020; YU et al., 2019). Pesquisas com diferentes abordagens de GWAS vem sendo amplamente utilizadas para a cultura da soja, e já foram identificados diversos locos responsáveis pelo controle de caracteres relacionados a produtividade (CONTRERAS-SOTO et al., 2017; ZHANG et al., 2019; REN et al., 2020).

O método de GWAS mais utilizado é conhecido como análise via marcas únicas, em que o efeito de cada marcador molecular é estimado como o coeficiente angular de uma análise de regressão, e a hipótese de significância é verificada com uso de testes estatísticos tradicionais, como o teste T (RESENDE et al., 2017). Porém, normalmente neste método há a detecção de uma grande quantidade de falsos-positivos, ou seja, marcadores que apresentam significância pelo teste estatístico, mas que provavelmente não estão em desequilíbrio de ligação com QTLs que controlam a característica em estudo (SUELA et al., 2022).

Dentre os fatores que podem influenciar no poder de detecção dos testes utilizados em estudos de associação, são os efeitos referentes a estrutura de população, que ocorrem quando a significância observada para determinados marcadores, é na realidade relacionada a alelos característicos de determinadas populações, que não são associados ao caractere de interesse (SEHGAL et al., 2016). Além disso, observa-se que mesmo os QTLs associados identificados pela análise via marcas únicas, normalmente são responsáveis por explicar somente uma pequena fração da variância genética (MANOLIO et al., 2009).

Uma alternativa para aumentar a detecção do efeito completo de um QTL é por meio de estudos genômicos que consideram a informação de vários marcadores

simultaneamente (LEE et al., 2014). Neste sentido, Nagamine et al. (2012) propuseram uma abordagem baseada em componentes de variância chamada *regional heritability mapping* (RHM). Nessa metodologia, marcadores em desequilíbrio de ligação são divididos em regiões, e ocorre a estimação da variância genética aditiva para cada uma delas via o Método da Máxima Verossimilhança Restrita (*REML – Restricted Maximum Likelihood*), utilizando as informações de parentesco gerada a partir dos marcadores que compõem cada região construída.

A abordagem de mapeamento de herdabilidades regionais vem sendo utilizada com sucesso para espécies como *Eucalyptus* (RESENDE et al., 2017), *Phaseolus vulgaris* (RESENDE et al., 2018), *Oriza sativa* (SUELA et al., 2022) e *Manihot esculenta* (OKEKE et al., 2018), principalmente devido ao seu potencial de capturar maiores porções da variância genética em comparação as metodologias tradicionais, e também por apresentar uma menor taxa de falsos positivos (USAI et al., 2014). Todavia, não existem estudos referentes ao uso desse método para a cultura da soja (*Glycine max* L.).

Portanto, o objetivo deste trabalho foi identificar QTLs que atuam no controle da resposta da soja a situações de déficit hídrico para as características número de grãos, número de vagens, período reprodutivo, e porcentagem de vagens imaturas por meio da metodologia de mapeamento de herdabilidades regionais. Para desta forma, auxiliar na localização de novos genes que podem ser responsáveis pelo controle da tolerância ao déficit hídrico, visando a utilização desses em programas de melhoramento para o desenvolvimento de novas cultivares. Também, buscou-se comparar a eficiência da metodologia de mapeamento de herdabilidades regionais com o método de associação genômica via marcas únicas.

2 Materiais e métodos

2.1 Dados fenotípicos

Os dados utilizados foram obtidos por meio da realização de um experimento implantando no ano de 2018. O ensaio foi conduzido em casa de vegetação na

Universidade Federal de Viçosa em Viçosa, Minas Gerais, localizada nas coordenadas de latitude S 20°45'14", longitude W 45°52'54" e altitude de 649 metros.

Foram utilizados 99 genótipos comerciais de soja, originados de várias fontes de germoplasma, possuindo diferentes eventos transgênicos, tipos de crescimento, e grupo de maturidade relativo (Material suplementar). O ensaio foi semeado no delineamento de blocos casualizados com três repetições, sob o esquema fatorial com dois níveis de regimes de irrigação, controle e estresse.

Cada parcela experimental foi constituída por uma planta, com o tratamento controle recebendo, durante todo ciclo da cultura, quantidade suficiente de água para seu desenvolvimento, com o potencial hídrico do substrato próximo a capacidade de campo (CC). No tratamento estresse, a partir do estágio de R3 (início da formação das vagens) foi imposto um déficit hídrico de acordo com a metodologia proposta por Nascimento et al. (2021), mantendo o potencial hídrico do substrato próximo a 50% da CC até o estágio de R6 (vagem cheia), de acordo com a escala proposta por Fehr e Caviness (1977).

Durante a condução do experimento, foram monitoradas a temperatura e a umidade relativa diária, e também, em determinado período foi mensurado o potencial hídrico das plantas para constatação da ocorrência do estresse hídrico, com as avaliações sendo realizadas com o uso de uma bomba de pressão do tipo Scholander (SCHOLANDER et al., 1965). Todos os tratamentos culturais relacionados a adubação, e manejo de pragas e/ou doenças do experimento foram realizados de acordo com as recomendações para a cultura da soja (SILVA et al., 2022).

As características agrônômicas avaliadas ao longo do experimento foram o número de vagens por planta (NV), duração do período reprodutivo (PR), número de grãos por planta (NG), e a porcentagem de vagens abertas por planta (PVA). Os caracteres foram mensurados de acordo com as seguintes metodologias:

- Número de vagens por planta (NV): contagem do número total de vagens da planta no estágio de R8 em quantidade;
- Duração do período reprodutivo (PR): diferença entre o número de dias entre a maturação plena (R8), e o florescimento da planta (R1) em dias;

- Número de grãos por planta (NG): contagem do número total de grãos por planta após a maturação plena (R8) em quantidade;
- Porcentagem de vagens abertas (PVA): razão do número de vagens abertas no estágio de R6 (vagens com grãos cheios), dividido pelo número de vagens totais da planta no mesmo estágio, em porcentagem.

Para fins explanatórios foi realizada a análise multivariada de componentes principais considerando todas as características avaliadas, sendo essas transformadas em variáveis latentes não correlacionadas para condensação e visualização da variação dos dados.

2.2 Dados moleculares e controle de qualidade

A informação molecular dos genótipos utilizados foi gentilmente cedida pela empresa GDM Genética do Brasil S.A. A genotipagem dos materiais ocorreu com o uso de 6000 marcadores do tipo SNP (*single nucleotide polymorphism*), sendo que parte dos materiais utilizados foram sequenciados pelo *Institute for Genome Diversity* (IGD) da Universidade de Cornell localizado em Ithaca/NY, e parte pelo *Biotechnology Center* da Universidade de Wisconsin-Madison localizado em Madison/WI.

A filtragem dos marcadores SNP foi realizada utilizando os critérios de *call rate* igual a 90%, que corresponde aos marcadores com número de informações perdidas inferiores a 10%, e de MAF (*minor allele frequency*) de 5%, que elimina os marcadores monomórficos e/ou aqueles em que o alelo de menor frequência está presente em menos de 5% dos genótipos. Após o controle de qualidade, restaram 4070 marcadores para utilização nas análises genômicas.

Com base nos parâmetros de filtragem estabelecidos para o controle de qualidade, um total de 4070 marcadores SNPs foram selecionados e utilizados em ambas metodologias de associação genômica. Os marcadores e sua distribuição em cada cromossomo estão apresentados na Figura 1.

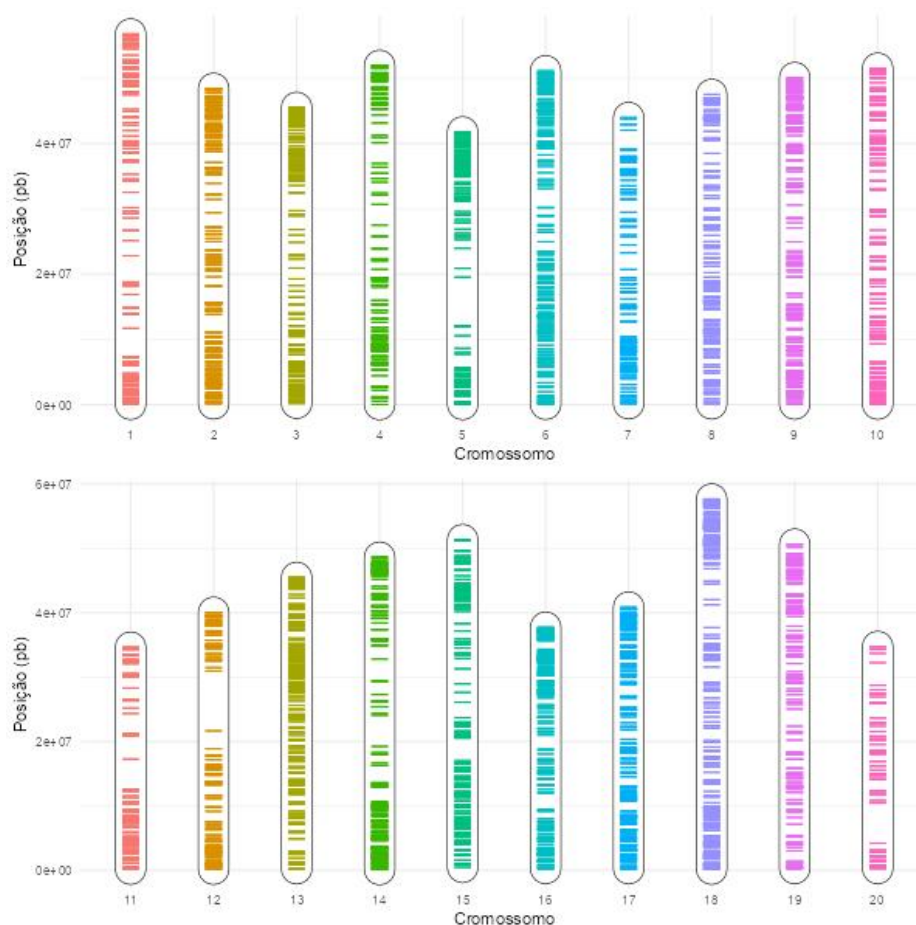


Figura 1. Distribuição dos marcadores SNPs entre os cromossomos da Soja (*Glycine max* L.).

2.3 Análise fenotípica e identificação de subpopulações

Para realização das análises de associação genômica, as características fenotípicas avaliadas foram submetidas à correção para os efeitos de blocos experimentais baseada em modelos lineares mistos (REML/BLUP). O modelo para análise fenotípica utilizado considerou o delineamento de blocos casualizados, os dois tratamentos de irrigação utilizados, a parcela sendo constituída por um indivíduo sem informação de *pedigree*, conforme o seguinte modelo:

$$y = \mu + Xr + W_i + \varepsilon, \quad [1]$$

em que o vetor “y” representa os valores fenotípicos observados, “ μ , r e i” representam os vetores assumidos como fixos para a média geral, blocos e tratamentos de

irrigação, respectivamente, e “ ε ” é o vetor de efeitos residuais com $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. As letras X e W correspondem as matrizes de incidência dos efeitos fixos.

Para identificação do número de subpopulações (K), foi realizada a análise de estrutura de população pelo método multivariado de análise discriminante por componentes principais (ADCP), proposta por Jombart et al. (2010). As análises foram realizadas por meio do software R para ajuste dos modelos, e utilizando o pacote “adegenet” para a análise de ADCP (JOMBART, 2008 e 2010; R CORE TEAM, 2021).

No método de ADCP, inicialmente são geradas variáveis latentes com base em informações moleculares que explicam parte da variabilidade de indivíduos via análise de componentes principais, que neste estudo foi utilizado o número de variáveis latentes responsáveis por uma variação acumulada de 80%. Neste primeiro momento, as variáveis latentes englobam tanto a variabilidade existente entre diferentes grupos (variabilidade de estrutura), e a variabilidade dentro de grupos (variabilidade aleatória).

Com base nas variáveis latentes geradas, a variabilidade é particionada via análise discriminante em variação entre grupos e variação dentro de grupos, produzindo outras variáveis latentes que maximizam a variação entre grupos, considerando diferentes valores de K, além de também minimizar a variação dentro dos subgrupos. Este procedimento identifica o valor de K mais provável, e aloca os indivíduos dentro dos subgrupos, que são as subpopulações, formadas de acordo com o modelo de melhor ajuste. Neste trabalho foi considerado o modelo de melhor ajuste aquele de BIC (*bayesian information criterion*) inferior, sendo este processo repetido 10 vezes para cada valor de K para validação cruzada dos resultados.

2.4 Análise de associação genômica via marcas únicas

As análises de associação genômica foram realizadas pelo software estatístico GenomicLand (AZEVEDO et al., 2019). O estudo de associação pelo método via marcas únicas que estima o efeito do j -ésimo marcador sobre o fenótipo foi baseado no modelo linear misto proposto por Yu et al. (2006):

$$y^* = \mu + Fs + Zg + m_j + \varepsilon, \quad [2]$$

em que “y*” representa o vetor de valores de fenótipos corrigidos (N x 1, com N sendo o número de indivíduos genotipados e fenotipados), “μ” representa a média geral, “s” representa o vetor de efeito fixo de estrutura de população, “g” é o vetor de efeitos poligênicos genômicos aleatório com $g \sim N(0, G\sigma_g^2)$, em que G representa a matriz de parentesco genômica e σ_g^2 a variância poligênica, m_j é o escalar que refere-se ao efeito fixo do j-ésimo marcador, e “ε” é o vetor de efeitos residuais com $\varepsilon \sim N(0, \sigma_e^2)$. A letra F representa a matriz de incidência do efeito fixo e Z a matriz para o efeito aleatório. As análises foram realizadas para cada marcador SNP do conjunto de dados.

A matriz de parentesco genômica “G” foi calculada de acordo com o método proposto por VanRaden (2008), de acordo com a expressão:

$$G = \frac{MM'}{\sum_{j=1}^p 2p_jq_j}, \quad [3]$$

em que M (N x m) é a matriz de incidência dos marcadores, e $p_j q_j$ as frequências alélicas no marcador correspondente.

Após a realização da análise para cada marcador e estimação dos respectivos efeitos, para diminuição da ocorrência de falsos positivos característicos de situações de testes múltiplos, foi realizada uma correção por FDR (*False discovery rate*), que é baseada na proporção do número de resultados falsos positivos (rejeição da hipótese nula quando esta é verdadeira) em relação ao número total de resultados positivos (BENJAMINI E HOCHBERG, 1995). A significância de marcadores foi constatada com a realização de um teste Wald, e por fim, calcularam-se os valores de $-\log_{10}$ com os resultados de p-valores corrigidos (p^*) para plotagem em relação a posição do SNP dentro de cada cromossomo.

2.5 Desequilíbrio de ligação

Para a análise de vários SNPs simultaneamente, a metodologia de RHM subdivide os marcadores de cada cromossomo em regiões, em que os tamanhos

dessas regiões são definidos baseados na informação de desequilíbrio de ligação entre marcadores. Neste trabalho, a medida de desequilíbrio considerada foi a do coeficiente de correlação entre os pares de marcadores (r^2), baseada na informação dos SNPs restantes pós o controle de qualidade.

O cálculo de desequilíbrio foi realizado por meio da função “LD.Decay” implementada em linguagem R no pacote *sommer* (COVARRUBIAS-PAZARAN, 2016), com os valores posteriormente sendo plotados em relação a respectiva distância genética (em pb) entre os pares. Para verificar a tendência de decaimento do DL, foi ajustada uma curva de regressão polinomial local (*loess*), utilizando a função “loess” também em software R (CAMPOY et al., 2016; R CORE TEAM, 2021).

O valor crítico para obtenção da intersecção com a curva de DL considerado foi de “ $r^2 = 0,2$ ”, representado por meio de uma linha reta horizontal, para identificação do respectivo valor de distância a ser utilizado para a construção das regiões genômicas (VOS et al., 2017; OYAMA et al., 2019).

2.6 Análise de mapeamento de herdabilidade regional

Baseando-se nas informações de DL entre os pares de marcadores (r^2), os SNPs de cada cromossomo foram organizados em conjuntos formando regiões genômicas de forma não sobreposta (por exemplo, considerando três regiões compostas por três SNPs: SNP1-SNP3, SNP4-SNP6, SNP7-SNP9).

Para construção das regiões, foram utilizados três tamanhos distintos, sendo um deles estabelecido de acordo com o respectivo valor de distância em que $r^2 \leq 0,2$, de modo semelhante ao utilizado por RESENDE et al. (2018) e SUELA et al. (2022), que correspondeu a distância de 2 Mb. Com o intuito de verificar se a adição de mais marcadores às regiões genômicas as tornariam mais explicativas, ou seja, se capturariam uma maior herdabilidade, foram utilizadas também as distâncias de 4,75 Mb e 7,5 Mb para construção das regiões.

Para a análise de mapeamento de herdabilidade regional, foi utilizado o modelo linear misto de acordo com a metodologia proposta por Nagamine et al. (2012), por meio da seguinte equação:

$$y^* = \mu + Fs + Z_1 g + Z_{2k} r_k + \varepsilon, \quad [4]$$

em que y^* é o vetor de observações corrigidas para os efeitos de blocos e tratamentos de irrigação (conforme Equação 1), “ μ ” representa a média geral, e “ s ” o vetor de efeito fixo de estrutura de população, “ g ” é o vetor de efeitos poligênicos genômicos aleatório com $g \sim N(0, G\sigma_g^2)$, em que G representa a matriz de parentesco genômica, construída sem a presença dos marcadores que formam a região k testada, e σ_g^2 a variância poligênica. A região genômica é representada pelo vetor de efeitos genômicas aditivos r_k , sendo que $r_k \sim N(0, G_{reg_k} \sigma_{r_k}^2)$, com G_{reg_k} referindo-se a matriz de parentesco genômica construída a partir dos marcadores presentes na região k , e $\sigma_{r_k}^2$ corresponde a variância aditiva da k -ésima região. A letra F representa a matriz de incidência para os efeitos fixos, e as matrizes Z_1 e Z_{2k} para os efeitos aleatórios do modelo. Os componentes de variância referentes aos efeitos aleatórios do modelo foram estimados pelo Método da Máxima Verossimilhança Restrita (*Restricted Maximum Likelihood Method – REML*).

Com o intuito de identificar se a variância da região seria significativa ($\sigma_{r_k}^2 > 0$), foi realizado o teste de razão de verossimilhança conforme descrito em Resende et al. (2017), considerando uma hipótese nula (H_0) de que não há variância na região e (H_1) de que há variância significativa na região genômica analisada, fazendo uso do modelo completo apresentado na Equação 4 em comparação a um modelo reduzido sem a presença do efeito da região, de acordo com a seguinte equação:

$$TRV = 2\ln L_1 - 2\ln L_0, \quad [5]$$

em que “ $\ln L_1$ ” é o valor de verossimilhança para o modelo completo, e “ $\ln L_0$ ” o valor de verossimilhança do modelo reduzido, ou seja, sem a presença do efeito para região (r_k), sendo esse teste realizado K vezes de acordo com o número de regiões construídas para cada tamanho.

2.7 Comparação das metodologias de associação genômica

Para comparação de ambas as metodologias de associação, foram estimados os valores de herdabilidade para os marcadores SNPs ou regiões que demonstraram significância ($p^* < 0.05$) a partir das análises. No caso método de associação via

marcas únicas, a herdabilidade foi calculada de acordo com a equação proposta por Resende et al. (2017):

$$h_{M_j}^2 = \frac{2p_jq_jM_j^2}{\sigma_y^2}, \quad [6]$$

Em que p_j e q_j correspondem as frequências alélicas para o j-ésimo marcador, M_j^2 representa o efeito do j-ésimo marcador para a característica considerada, e σ_y^2 a variância fenotípica.

No caso da metodologia de mapeamento de herdabilidade regional, a herdabilidade é obtida por meio da proporção do componente de variância referente a r-ésima região em relação a variância fenotípica, conforme proposto por Resende et al. (2017):

$$h_{r_k}^2 = \frac{\sigma_{r_k}^2}{\sigma_y^2} \quad [7]$$

Com base nos valores obtidos nas equações [6] e [7], a metodologia que se aproximasse do valor de herdabilidade genômico da característica, calculado com base no modelo [4] sem a presença do efeito de região (VANRADEN, 2008), seria aquela que explica maior parte da variação fenotípica do caráter de interesse (RESENDE et al., 2018).

2.8 Potenciais genes relacionados ao controle das características

Após a realização das análises de associação via RHM, baseando-se em informações da literatura acerca das regiões que mostraram significância para as características avaliadas, buscou-se no banco de dados genéticos Phytozome (<https://phytozome-next.jgi.doe.gov>), potenciais genes candidatos que podem estar relacionados ao controle das características fenotípicas avaliadas.

3 Resultados

3.1 Análise fenotípica e de estrutura de população

Os resultados da análise de componentes principais demonstraram a variabilidade causada pela influência dos tratamentos de irrigação sobre os caracteres avaliados, conforme apresentado na Figura 2. De acordo com o PC1 (51,1% da variância dos dados), houve variabilidade para as características NV e NG, de modo que valores maiores de quantidade foram observados no tratamento controle, e com base no PC2 (28% da variância dos dados), é possível observar que houve variabilidade para PVA e PR predominantemente para os tratamentos de estresse, de modo que houve aumento na porcentagem de vagens abertas, e uma redução do período reprodutivo dos genótipos avaliados quando estavam sob estresse.

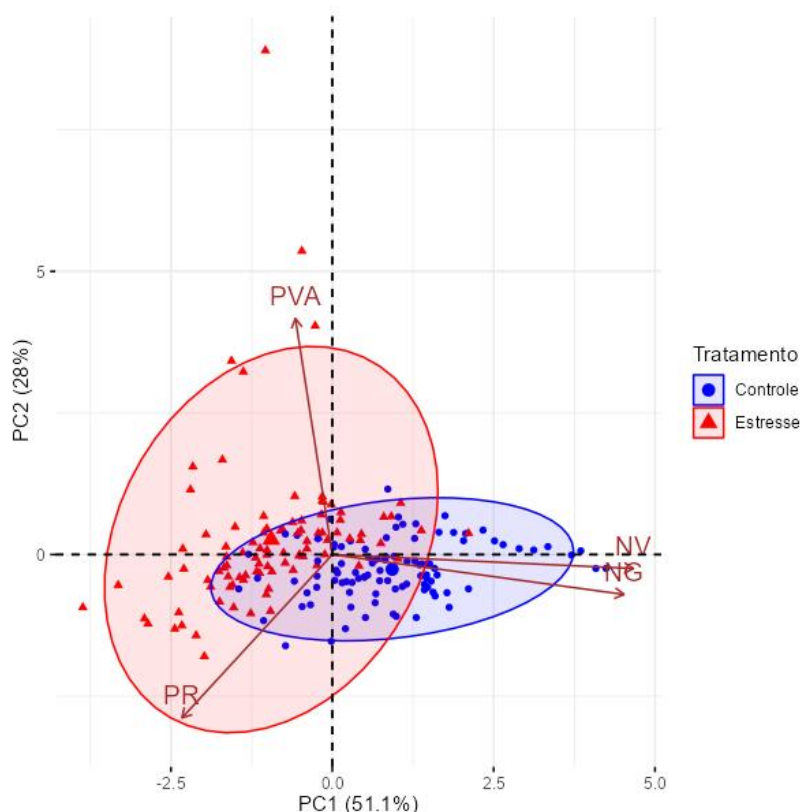


Figura 2. Biplot do PC1 x PC2 com os genótipos avaliados sobre os tratamentos de regime de irrigação. As setas em vermelho indicam as características fenotípicas consideradas na análise. PVA: Porcentagem de vagens imaturas abertas. PR: Período reprodutivo. NV: Número de vagens. NG: Número de grãos.

Com relação a identificação de subpopulações, o modelo de melhor ajuste obtido pelo método de ADPC, foi aquele em que os 99 genótipos foram divididos em 3 diferentes subpopulações, conforme demonstrado na Figura 3. Neste sentido, 24 genótipos foram alocados dentro da subpopulação 1 (Grupo 1), 20 genótipos alocados dentro da subpopulação 2 (Grupo 2), e 55 genótipos na subpopulação 3 (Grupo 3). A Figura 3B mostra a probabilidade de agrupamento dos genótipos dentro dos grupos baseada nas informações de frequências alélicas.

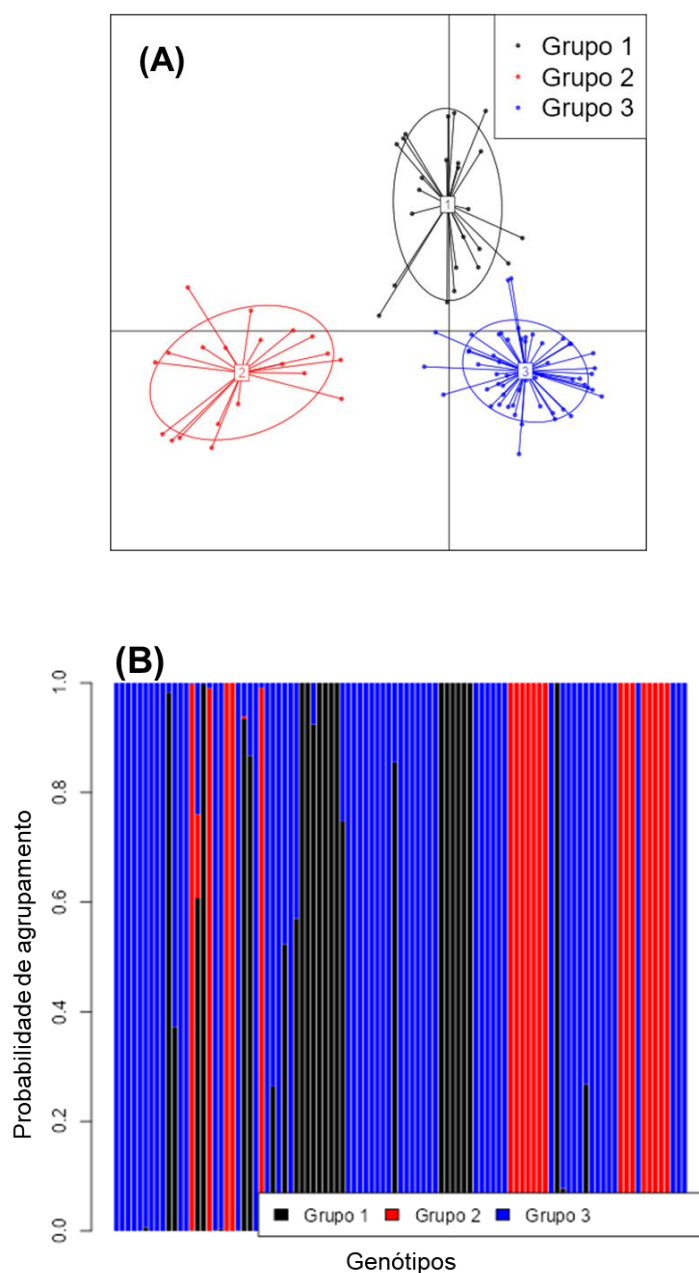
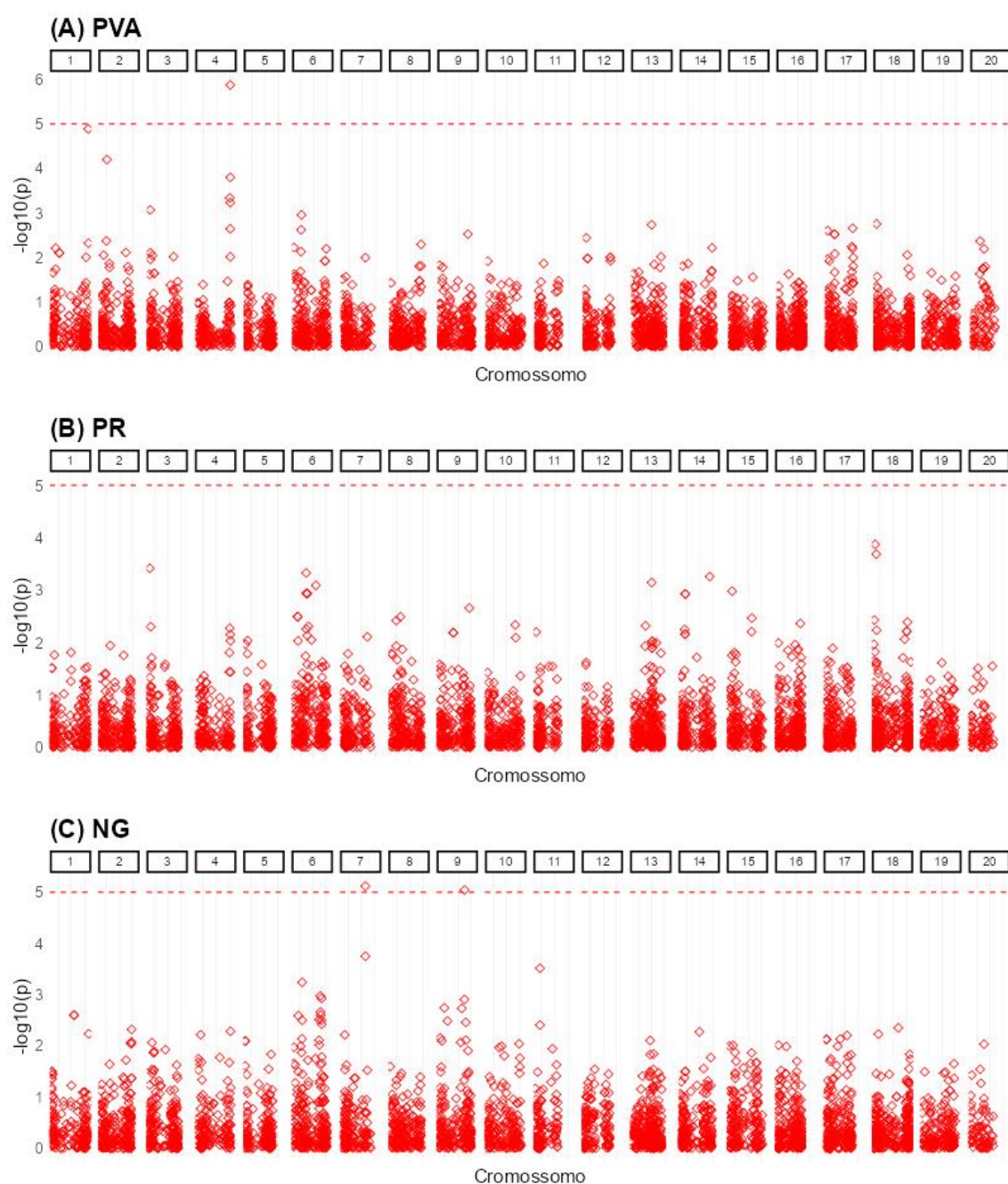


Figura 3. (A) Gráfico de dispersão referente aos diferentes grupos identificados pela análise de ADPC. **(B)** Representação gráfica do agrupamento dos genótipos dentro dos grupos considerados pela análise de ADPC com base nos valores probabilidade de agrupamento. Grupos 1,2 e 3 referem-se às subpopulações identificadas na análise

3.2 Análise de associação genômica via marcas únicas

Foram identificadas ao todo 3 SNPs associados às características PVA, NV e NG. Com relação a característica PVA, uma associação com o marcador SNP (Posição: 50758389 pb) localizado no cromossomo 4 foi significativa (Figura 4A), para PR não foram identificadas associações significativas (Figura 4B), e para NV e NG, 2 marcadores SNPs localizados nos cromossomos 7 e 9 (Posição 36115777; Posição 39954064, respectivamente) demonstraram-se associados, sendo que os mesmos marcadores foram identificados para ambas características (Figura 4C e 4D).



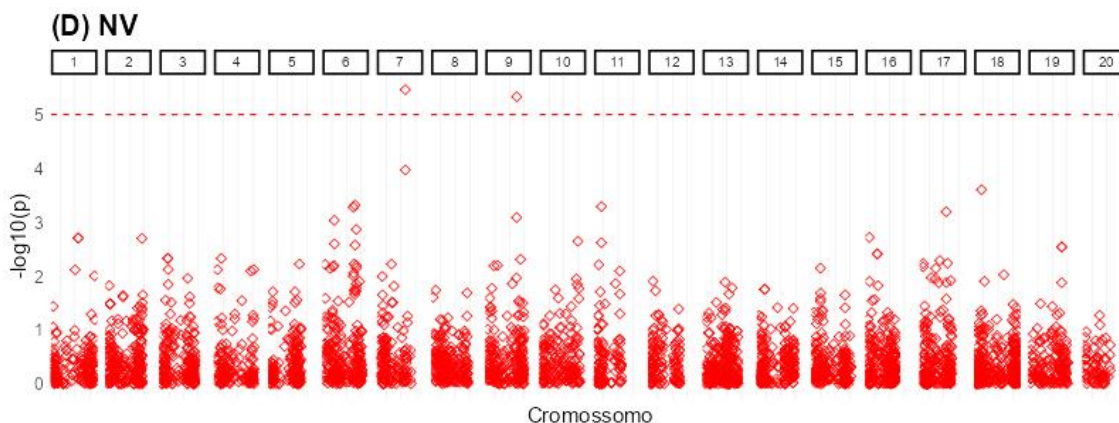


Figura 4. Manhattan plots das análises de associação realizadas para a característica PVA (A), PR (B), NG (C) e NV (D), baseado no valor de $-\log_{10} p^*$ valor no eixo y ao longo dos 20 cromossomos da Soja (*Glycine max* L.), com base na análise via marcas únicas representadas pelos losangos vermelhos. A linha tracejada vermelha indica o nível de significância a 5% considerando os valores de p^* valor corrigidos por FDR. PVA: Porcentagem de vagens imaturas; PR: Período reprodutivo; NG: Número de grãos; NV: Número de vagens.

A herdabilidade para as características avaliadas variou de baixa a moderada (KARASU et al., 2009), conforme apresentado na Tabela 1, sendo o menor valor referente a característica PVA (0,11) e o maior valor relacionado a característica NV (0,52). Um total de 16.280 testes de associação foram realizados pelo método via marcas únicas (4070 para cada característica), e considerando a correção pelo método de FDR ($-\log_{10}(p^*) > 1,5$).

A Tabela 1 apresenta os SNPs associados a QTLs detectados pela análise via marcas únicas, a localização desses SNPs no genoma da planta, e a porcentagem de explicação (PE) da característica baseando-se nos valores de herdabilidade genômica aditiva, que representa o quanto da variação observada foi em decorrência da variação do marcador. Com relação a PVA, a porcentagem de explicação foi de 6,2%, enquanto que para as características relacionadas aos componentes de produtividade da soja variou de 2,23 a 2,91%.

O efeito relacionado ao QTL identificado para a característica PVA apresentou valor negativo, enquanto que aqueles relacionados a NG e NV apresentaram valores positivos. Para a característica PR, não foram identificados marcadores associados por este método. O QTL associado ao marcador SNP para a característica PVA possui efeito (M_j) negativo, já para as características NV e NG, os QTLs associados possuem efeito (M_j) positivo.

Tabela 1. Localização dos marcadores moleculares nos cromossomos (Crom.) em que foram obtidas significâncias três características pela análise via marcas únicas.

Carac.	Crom.	Posição	p* valor	h_M^2	PE
		pb			%
PVA	4	50758389	0,005	0,28	6,2
NG	7	36115777	0,018	0,16	2,23
	9	39954064	0,018	0,21	2,91
NV	7	36115777	0,01	0,18	2,35
	9	39954064	0,01	0,22	2,88

PVA: Porcentagem de vagens abertas; NG: Número de grãos; NV: Número de vagens. h_M^2 : Herdabilidade do marcador. PE: Porcentagem de explicação do marcador.

3.3 Desequilíbrio de ligação

O gráfico de desequilíbrio de ligação (Figura 5) demonstrou um abrupto decaimento, diminuindo até a metade em um valor de r^2 de aproximadamente 0,1. A distância de 2 Mb associou-se com um valor de r^2 aproximado de 0,20, sendo adequada para a construção de regiões para a análise de RHM (RESENDE et al., 2017).

Com base nos tamanhos de regiões considerados, foram construídas 382 regiões para o comprimento de 2 Mb, 186 para 4,75 Mb e 126 para 7,5 Mb, visto que quanto maior o número de marcadores associados em uma mesma região, menor será o número de regiões construídas. Em estudos de associação baseados vários marcadores simultaneamente, o uso de regiões maiores pode estar associado a um maior poder de detecção, porém, é importante encontrar o balanço entre o tamanho de região e quantidade de ruído presente na análise (BENNEWITZ et al., 2017).

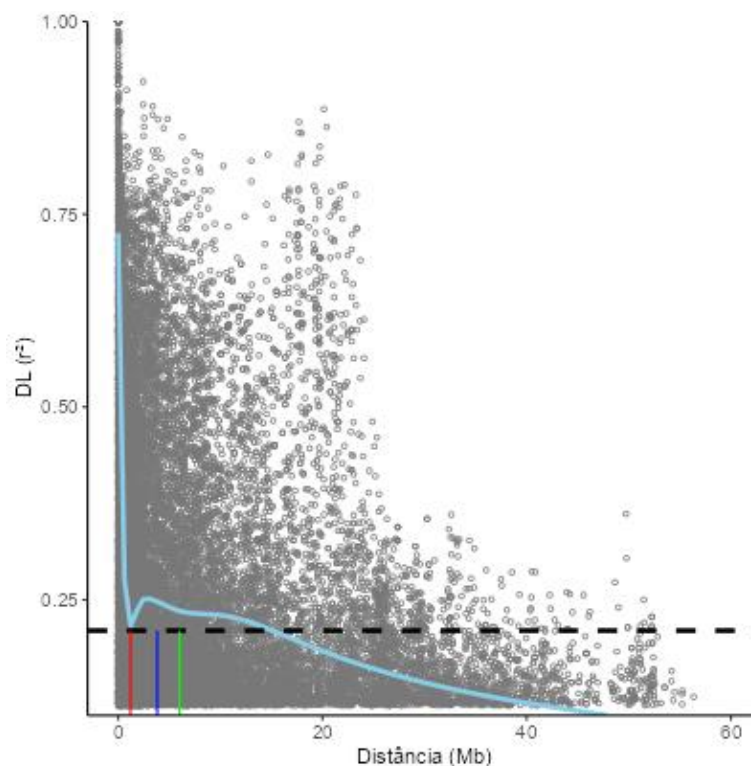


Figura 5. Desequilíbrio de ligação (r^2) entre os pares de alelos dos marcadores SNPs ao longo do genoma da Soja (*Glycine max* L.) em Mb para todos os genótipos avaliados. A linha azul demonstra o decaimento do DL com os valores ajustados pelo método de regressão loess. A linha tracejada horizontal indica o valor de $r^2 = 0,20$. A linha vertical vermelha indica o valor correspondente a distância física de 2 Mb, a linha azul a distância de 4,75 Mb e a linha verde, a distância de 7,5 Mb.

Os marcadores foram organizados em regiões para cada cromossomo, e a Tabela 2 apresenta as informações referentes à distribuição dos marcadores entre os cromossomos e às regiões construídas. Para o tamanho de 2 Mb, foram construídas em média 19 regiões por cromossomo, para o tamanho de 4,75 foram em média 9 regiões, e para o tamanho de 7,5 Mb foram 6 regiões. Os valores médios de DL obtidos variaram de 0,03 até 0,15 para regiões de 2 Mb, 0,01 até 0,14 para regiões de 4,75 Mb e 0,01 até 0,11 para regiões de 7,5 Mb, ou seja, de acordo com o esperado, as estimativas de DL foram maiores quanto menor o tamanho da região construída, com exceção para o cromossomo 13, em que o DL aumentou conforme o tamanho da região foi maior, o que pode estar associado à uma maior densidade de marcadores com ampla distribuição neste cromossomo.

Tabela 2. Distribuição, número de regiões construídas e DL médio por cromossomo dos marcadores moleculares utilizados na análise de mapeamento de herdabilidade regional.

Cromos.	Tamanho (Mb)	nº de Regiões			DL médio (r^2)		
		2 Mb	4.75 Mb	7.5 Mb	2 Mb	4.75 Mb	7.5 Mb
1	56,74	21	11	7	0,14	0,03	0,01
2	48,35	21	10	7	0,11	0,06	0,03
3	45,31	20	9	6	0,05	0,02	0,02
4	51,96	21	11	7	0,11	0,12	0,11
5	41,75	15	8	5	0,03	0,01	0,04
6	51,10	23	11	7	0,04	0,01	0,01
7	43,84	17	9	6	0,07	0,08	0,01
8	47,42	21	10	6	0,15	0,02	0,09
9	50,02	21	10	7	0,06	0,11	0,04
10	51,47	20	10	7	0,07	0,06	0,02
11	34,49	13	6	5	0,05	0,01	0,01
12	39,90	15	7	5	0,08	0,01	0,01
13	45,25	21	10	6	0,04	0,08	0,09
14	48,52	19	9	7	0,15	0,14	0,09
15	50,90	21	10	7	0,12	0,08	0,11
16	37,48	18	8	5	0,08	0,06	0,01
17	40,63	18	8	6	0,07	0,08	0,02
18	57,50	24	12	8	0,12	0,06	0,05
19	50,39	21	11	7	0,11	0,13	0,06
20	34,49	12	6	5	0,09	0,01	0,01
Total	927,51	382	186	126	-	-	-

3.4 Análise de mapeamento de herdabilidade regional

Foram avaliadas pela análise de mapeamento de herdabilidade regional um total de 2776 regiões ($382 \times 4 / 186 \times 4 / 126 \times 4$) com um número variável de SNPs para cada região, sendo que nestas, foram identificadas 15 regiões associadas as características avaliadas baseadas na correção por FDR ($-\log_{10}(p^*) > 1,5$).

Os gráficos de calor construídos para cada característica (Figura 6, 7, 8 e 9) mostram a herdabilidade explicada ao longo de todas as regiões analisadas para cada tamanho em que foram identificadas associações significativas. Os quadrados destacados em verde indicam aquelas regiões que atingiram o nível de significância corrigida pelo método de FDR. Por meio dessas figuras, é possível observar que houve regiões que demonstraram potencial para estarem associadas as características, apesar de não terem sido significativas pelo teste realizado. A característica em que se observou regiões com maior herdabilidade regional foi NG, enquanto que as menores foram PVA. Porém, esta característica foi a que teve maior número de regiões associadas.

O aumento no tamanho da região contribuiu com o poder de detecção do teste somente para PVA, já para as outras características essa relação não pôde ser observada. É interessante observar que, apesar das associações terem sido identificadas apenas em alguns cromossomos, para todas as características pode-se observar indícios de associações espalhados ao longo de todo o genoma da soja.

As 3 regiões associadas com a característica PVA (Figura 6) foram encontradas em 3 diferentes cromossomos (4, 2 e 1), com a região do cromossomo 4 sendo identificada em todos os tamanhos de regiões, já para o cromossomo 2, foram identificadas 2 regiões somente de tamanho 4,75 e 7,5 Mb, e para o cromossomo 1, apenas no tamanho de 7,5 Mb essa associação foi identificada. A região associada com a característica PR (Figura 7) foi identificada somente no tamanho de 7,5 Mb estando localizada no cromossomo 18. Para a características NG (Figura 8), foram identificadas 2 regiões de tamanho 4,75 Mb e 2 regiões no tamanho de 7,5 Mb no mesmo cromossomo (9 e 6), e considerando o NV (Figura 9), 2 regiões foram significativas para tamanho de 2 Mb e 2 regiões para o tamanho de 4,75 Mb localizadas nos cromossomos 9 e 6.

A análise de RHM conseguiu identificar regiões associadas para todas as características avaliadas ($p < 0,05$, pelo TRV $\sim \chi^2$ com p graus de liberdade), porém, o tamanho da região em que essa relação foi observada variou conforme a característica. Para o caso de PVA e NG, quanto maior o tamanho da região utilizada, mais QTLs associados foram identificados. Para PR, apenas na região de 7,5 Mb foi encontrada associação, porém, para a característica NV, as associações foram identificadas nas regiões de menor tamanho. As associações identificadas pela análise de RHM englobaram os SNPs que demonstraram significância na análise de associação via marcas únicas, porém, por meio da análise de mapeamento de herdabilidade regional foram identificadas novas associações que antes não haviam sido reconhecidas pelo método convencional de análise.

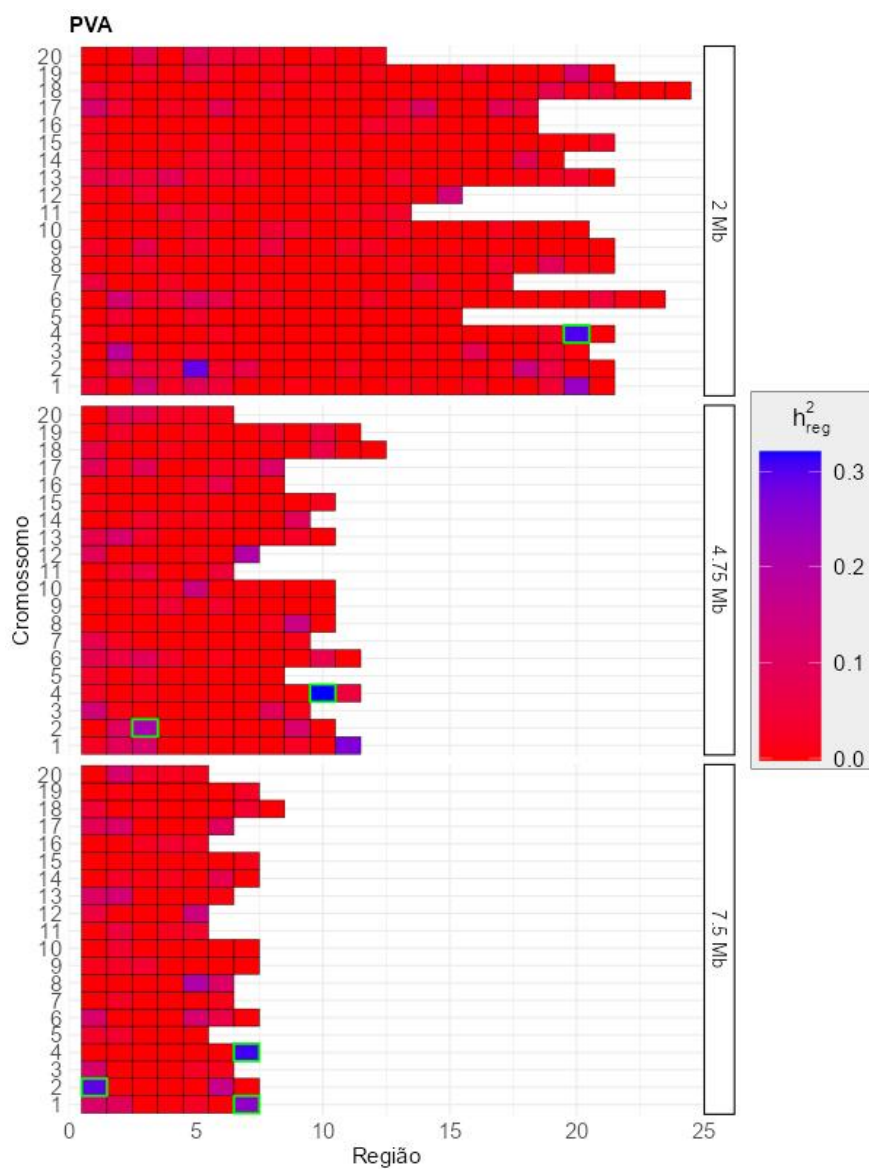


Figura 6. Distribuição das regiões ao longo dos 20 cromossomos do genoma da Soja (*Glycine max* L.) para a característica PVA. Cada barra lateral indica o tamanho da região utilizada nos casos em que houve significância a 5%. O mapa de calor indica a herdabilidade da região estimada, e as regiões destacadas em verde são aquelas declaradas como significantes pelo teste.

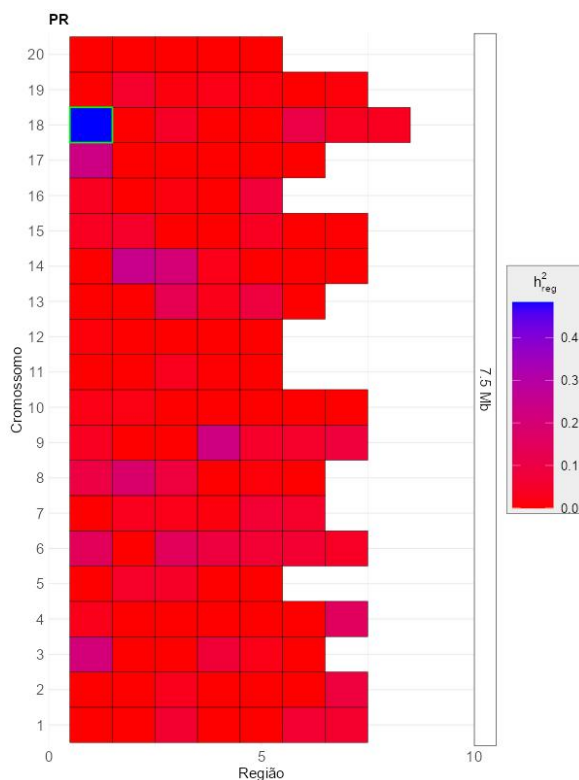


Figura 7. Distribuição das regiões ao longo dos 20 cromossomos do genoma da Soja (*Glycine max* L.) para a característica PR. Cada barra lateral indica o tamanho da região utilizada nos casos em que houve significância a 5%. O mapa de calor indica a herdabilidade da região estimada, e as regiões destacadas em verde são aquelas declaradas como significantes pelo teste.

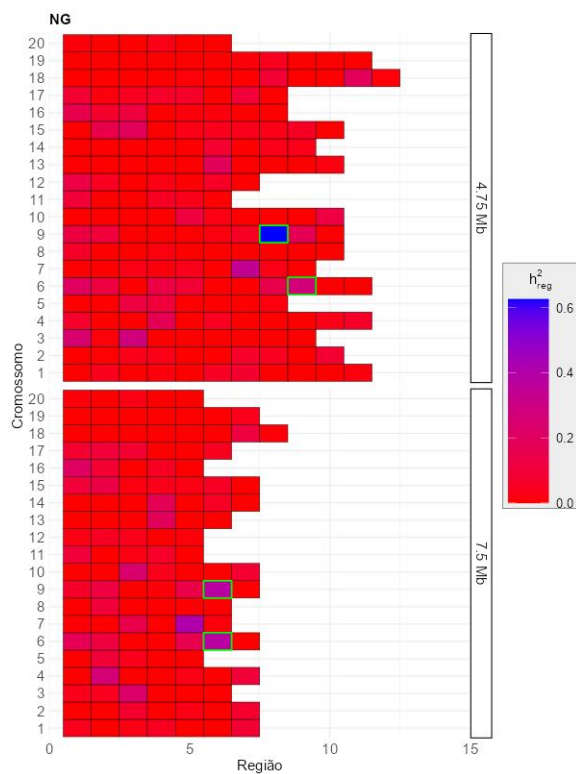


Figura 8. Distribuição das regiões ao longo dos 20 cromossomos do genoma da Soja (*Glycine max* L.) para a característica NG. Cada barra lateral indica o tamanho da região utilizada nos casos em que

houve significância a 5%. O mapa de calor indica a herdabilidade da região estimada, e as regiões destacadas em verde são aquelas declaradas como significantes pelo teste.

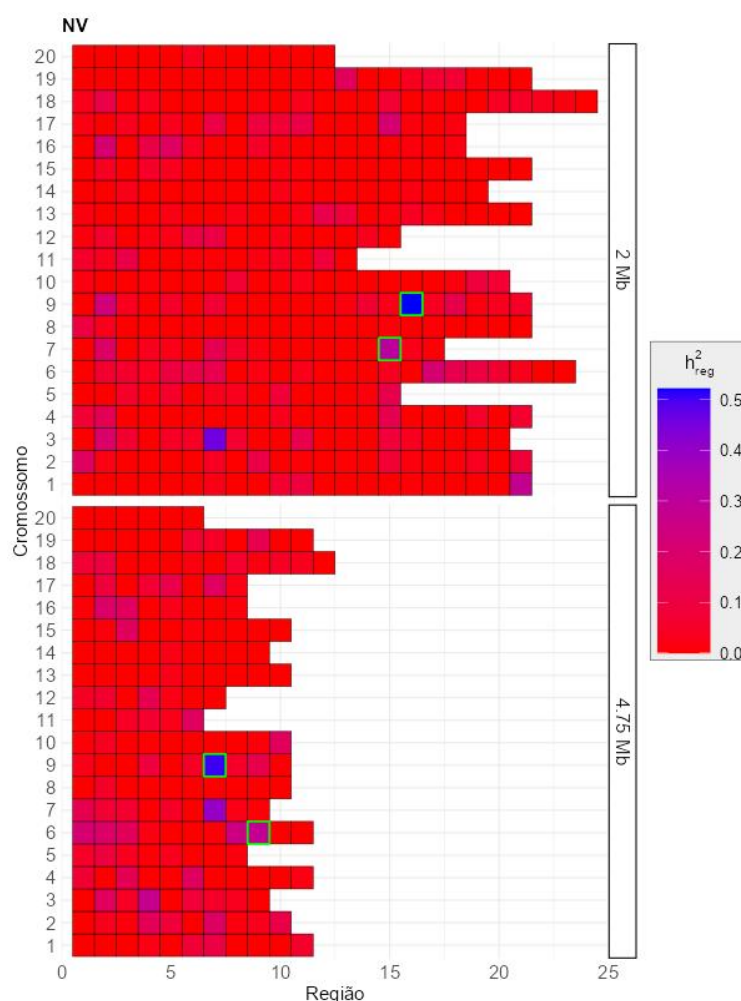


Figura 9. Distribuição das regiões ao longo dos 20 cromossomos do genoma da Soja (*Glycine max* L.) para a característica NV. Cada barra lateral indica o tamanho da região utilizada nos casos em que houve significância a 5%. O mapa de calor indica a herdabilidade da região estimada, e as regiões destacadas em verde são aquelas declaradas como significantes pelo teste.

Observando as regiões em que foram detectadas significância estatística na Tabela 3 pode-se constatar que, além de novas associações, essas regiões englobaram também todos aqueles marcadores identificados pela análise via marcas únicas. O número de SNPs nessas regiões variou de 11 (NV e PVA) a 98 (PR), a média de marcadores utilizados em cada tamanho de região foi de 14,3 para 2 Mb, 23,16 para 4,75 Mb e 47,16 para 7,5 Mb, já os valores de herdabilidade das regiões obtidos variaram de 0,2114 (PVA) até 0,5215 (NV). Foram encontradas associações significativas em 7 diferentes cromossomos da planta (1, 2, 4, 6, 7, 9, e 18). De modo

geral, a região de 2 Mb foi a que apresentou menor detecção de QTLs para as características avaliadas. Para as regiões de 4,75 Mb e 7,5 Mb, o poder de detecção foi semelhante, porém, as associações identificadas não foram as mesmas nessas regiões.

A característica com maior porcentagem de explicação em uma única região foi PR com 10,89%, porém, foi a região que englobou o maior número de marcadores SNPs (98), enquanto isso, o menor valor de PE observado foi de 3,77% para NG. porcentagem de explicação das regiões identificadas foi superior àquelas observadas na análise via marcas únicas, e também, houve casos em que foram identificadas mais de uma região associada a uma característica, fazendo com que a PE como um todo fosse maior.

Tabela 3. Identificação da localização (Cromossomo), tamanho e número de SNPs das regiões detectadas como significativas na análise de RHM, a herdabilidade explicada pela região e sua proporção de porcentagem explicada em comparação a herdabilidade genômica.

Tam. Região	Carac.	Crom.	Posição inicial	Posição final	Nº de SNPs	p* valor	h_{RHM}^2	PE
Mb			pb					%
2	PVA	4	49545548	51490484	19	< 0,001	0,3029	6,43
	NV	9	38479478	39954064	13	< 0,005	0,5215	8,89
		7	35712247	37686216	11	0,043	0,3082	5,25
4.75	PVA	4	46043483	50761700	31	< 0,001	0,3206	7,10
		2	9685144	14418314	12	0,037	0,2114	4,68
	NG	9	35682957	39954064	21	< 0,001	0,6255	8,73
		6	40323692	44876991	27	0,038	0,2701	3,77
	NV	6	40323692	44876991	27	0,048	0,2821	4,81
		9	35682957	39954064	21	0,049	0,5044	8,60
7.5	PVA	4	47797028	51965135	31	<0,001	0,3080	6,83
		1	55251501	56782521	11	0,009	0,2514	5,57
	PR	2	8024996	15523496	34	0,027	0,2914	6,46
		18	166456	7285289	98	0,039	0,4827	10,89
	NG	6	41108191	48422424	57	0,047	0,3755	5,24
		9	38479478	45969769	52	0,049	0,3780	5,28

PVA: Porcentagem de vagens imaturas; NV: Número de vagens; NG: Número de grãos; PR: Período reprodutivo; h_{RHM}^2 : Herdabilidade da região. PE: Porcentagem de explicação da região.

3.5 Genes candidatos ao controle das características

Com base nos bancos de dados e literatura existente, foi possível identificar genes responsáveis pelo controle das características de interesse (Tabela 4). Os genes apresentados são aqueles que se encontram dentro das regiões em que foram identificadas as associações significantes, porém, em alguns casos o gene listado é responsável pela produção de uma proteína homóloga àquela em que a associação é reportada na literatura. Pode-se observar que existem genes relacionados a todos os

caracteres avaliados, porém, para PVA no cromossomo 1 e NV no cromossomo 9, não foram encontrados relatos prévios de associação de genes que controlam essas características nas regiões significantes neste estudo.

Tabela 4. Lista de genes reportados na literatura para as características avaliadas identificados dentro das regiões significantes, localizações no genoma, e descrições das proteínas traduzidas.

Carac.	Crom.	Gene candidato	Descrição	Localização (pb)	Referência
				pb	
		Glyma.04G214100	PTHR12565:SF125 - TRANSCRIPTION FACTOR HEC3-RELATED	48576688-48577420	
		Glyma.04G199000	PF02365 - No apical meristem (NAM) protein (NAM)	47121566-47123284	
		Glyma.04G212000	PF02365 - No apical meristem (NAM) protein (NAM)	48385895-48387157	Dong et al. (2013)
		Glyma.04G213300	PTHR31989:SF14 - NAC DOMAIN CONTAINING PROTEIN 103	48507580-48512003	
		Glyma.04G226700	PTHR31989:SF5 - F11O4.3-RELATED	49623093-49626512	
		Glyma.04G208300	PTHR31719:SF3 - GRAB1-LIKE PROTEIN	48062885-48064803	
		Glyma.02G109000	PTHR11945:SF169 - MADS-BOX FAMILY PROTEIN	10440414-10444977	
		Glyma.02G121500	PTHR11945:SF194 - AGAMOUS-LIKE MADS-BOX PROTEIN AGL3-RELATED	12065639-12082937	
PVA	1, 2 e 4	Glyma.02G121600	PTHR11945:SF155 - AGAMOUS-LIKE MADS-BOX PROTEIN AGL8-RELATED	12084713-12089043	Liu et al. (2022)
		Glyma.04G245500	PTHR11945:SF97 - FLORAL HOMEOTIC PROTEIN PISTILLATA/Exhibits flower specific expression	51318457-51321641	
		Glyma.02G109300	PTHR10108:SF827 - METHYLTRANSFERASE-LIKE PROTEIN-RELATED/ Coexpressed with genes in leaf specific coexpression subnetwork	10489230-10493389	
		Glyma.04G238900	PF06017 - Unconventional myosin tail, actin- and lipid-binding (Myosin_TH1) PF00560//PF07714//PF08263//PF13855 - Leucine Rich Repeat (LRR_1) //	50756828-50759467	Moura (2021)
		Glyma.04G239000	Protein tyrosine kinase (Pkinase_Tyr) // Leucine rich repeat N-terminal domain (LRRNT_2) // Leucine rich repeat (LRR_8)	50759936-50764122	
		Glyma.04G239100	-	50762901-50763075	
PR	18	Glyma.18G062200	PTHR14493:SF54 - ZINC FINGER CCCH DOMAIN-CONTAINING PROTEIN 66	5675032-5678138	Hu E Zuo (2021)
		Glyma.06G257400	PTHR27002:SF9 - CYSTEINE-RICH RECEPTOR-LIKE PROTEIN KINASE 27-RELATED	44043937-44047058	Chen et al. (2007)
NV	6, 7 e 9	Glyma.06G257500	PTHR27002:SF9 - CYSTEINE-RICH RECEPTOR-LIKE PROTEIN KINASE 27-RELATED	44050090-44053834	
		Glyma.07G201800	PTHR31079:SF5 - NAC DOMAIN CONTAINING PROTEIN 10	37074545-37078552	Kuroda et al. (2013)
		Glyma.06G248200	PTHR10641//PTHR10641:SF574 - MYB-LIKE DNA-BINDING PROTEIN MYB	41892791-41893544	Funatsuki et al. (2005)
NG	6 e 9	Glyma.09G150300	PTHR10795:SF322 - SUBTILISIN SERINE PROTEASE-RELATED	37097964-37100295	
		Glyma.09G195100	PTHR10795//PTHR10795:SF382 - PROPROTEIN CONVERTASE SUBTILISIN/KEXIN	41972423-41978298	Liang et al. (2010)

4 Discussão

4.1 Análise fenotípica

As análises fenotípicas iniciais indicaram que os regimes de irrigação impostos sobre os genótipos influenciaram em todas as características avaliadas, com exceção de PR. Este resultado se assemelha ao encontrado no trabalho realizado por Nascimento et al. (2021), que verificou grande redução do número de vagens e grãos sob condições de estresse hídrico, porém, em determinado ano obteve uma pequena variação para a característica relacionada ao período reprodutivo.

Para a característica PVA, apesar de ainda ser pouco estudada, existem relatos com soja que associam o aumento da abertura de vagens imaturas com períodos de estresse durante a formação das vagens, seguidos de excesso de chuvas (VANN et al., 2021). Também com relação a PVA, resposta diferencial dos genótipos foram observadas, corroborando com os relatos que demonstram que este caráter tem uma arquitetura genética complexa, e que há variabilidade entre os genótipos existentes (LIU et al., 2022).

4.2 Comparação entre as metodologias de associação genômica (GWAS)

Os 4070 marcadores do tipo SNP utilizados nas análises de associação genômica foram eficientes para detectar associações entre marcadores e QTLs responsáveis pelo controle das características avaliadas. Soares (2018) utilizando um número menor de marcadores, teve êxito para identificação de associações significativas avaliando caracteres quantitativos relacionados a defesa contra efeitos adversos bióticos na soja. E com relação aos estudos de estresse hídrico, Dhungana et al. (2020) identificaram QTLs relacionados a resposta da planta em condições de estresse hídrico utilizando 2648 marcadores polimórficos de base única.

A metodologia de análise discriminante por componentes principais foi eficiente para determinar a estrutura de subpopulações com base nas informações dos marcadores, visto que conseguiu identificar 3 subgrupos de genótipos. Além de corrigir vieses relacionados a estrutura de população, que é um dos maiores

causadores da ocorrência de falsos positivos nas análises de associação (KALER et al., 2020), o fato da análise de RHM incluir ainda os efeitos relacionados ao parentesco genético entre os indivíduos, faz com que o efeito do SNP seja ajustado considerando o efeito conjunto de todos os marcadores simultaneamente (RESENDE et al., 2017).

A obtenção de associação para os mesmos marcadores SNPs para as características NV e NG pelo método via marcas únicas pode ter ocorrido devido ao fato de ambas as características serem correlacionadas, conforme relatado por Dalchiavon et al. (2012).

O desequilíbrio de ligação nas espécies em que há predominância do sistema de cruzamento de autofecundação tende a cair lentamente (FLINT-GARCIA et al., 2003), e apesar de haver uma grande variação no padrão de DL conforme o germoplasma em estudo, os resultados obtidos neste trabalho corroboraram com os padrões identificados por Contreras-Soto et al. (2017).

Para as abordagens de associação que avaliam vários locos simultaneamente, a definição do tamanho das regiões genômicas podem ser feita com base na informação de DL (ZHAO et al., 2007). O tamanho da região estabelecida pelo pesquisador pode causar alterações no poder de detecção do teste de associação (GUO et al., 2016), por isso, é importante efetuar os testes com diferentes tamanhos de região, conforme realizado neste estudo. Considerar a informação de DL para definir o tamanho das regiões também garante que essa informação seja distribuída ao longo das regiões geradas (RESENDE et al., 2018).

Com relação as características PR e PVA, com relação ao tamanho das regiões utilizadas nas análises de RHM, os resultados foram semelhantes aos obtidos por Suela et al. (2022) e Bennewitz et al. (2017), que verificaram um aumento no poder de detecção no teste conforme o aumento na região genômica analisada. Já para as características relacionadas aos componentes de produtividade NV e NG, não houve grande variação no poder de detecção com a mudança no tamanho das regiões testadas, da mesma forma, resultados consistentes entre diferentes tamanhos de regiões utilizando a análise de RHM foram relatados por Riggio E Pong-Wong (2012) utilizando dados de simulação de características de animais.

De forma semelhante aos resultados obtidos por outros autores (RESENDE et al., 2018; SHIRALI et al., 2016; SUELA et al., 2022), a análise de mapeamento de herdabilidade regional teve a capacidade de explicar uma maior porção da variância genética aditiva em comparação à metodologia de associação via marcas únicas. Isso ocorre devido ao efeito combinado de vários alelos próximos que formam a região. Porém, o oposto pode ocorrer, em que um alelo de efeito maior não é localizado devido ao efeito de diluição dentro da região que está inserido, sendo uma provável causa das associações identificadas pela análise via marcas únicas, mas que não foram localizadas pelo método de RHM (RESENDE et al., 2017).

4.3 Genes candidatos

Os genes relacionados a porcentagem de vagens imaturas abertas (PVA) identificados pelos SNPs associados via marcas únicas localizados no cromossomo 4 (*Glyma.04g238900*, *Glyma.04g239000*), são relatados como genes que traduzem proteínas que participam dos processos relacionados ao crescimento, desenvolvimento da planta e resposta a estresses, como por exemplo, os receptores quinase de repetição ricos em leucina, que auxiliam na regulação do fechamento dos estômatos da planta (ZHOU et al., 2016). Com relação as características NV e NG, em que foram identificados os mesmos marcadores associados para ambas as características, o gene identificado no cromossomo 9 (*Glyma.09g175100*) tem relação com proteínas PPR (*Pentatricopeptide-repeat*), que também são reportadas como proteínas que atuam no crescimento, desenvolvimento e resposta da planta a estresses abióticos (SU et al., 2019).

Dentre os genes identificados pela análise de mapeamento de herdabilidade regional para a característica PVA, destacam-se os genes *Glyma.02G109000*, *Glyma.02G121500*, *Glyma.02G121600*, *Glyma.04G239000*, *Glyma.04G238900* e *Glyma.04G213300*, em que os genes mencionados *Glyma.02* são responsáveis pela síntese de fatores de transcrição (*MADS box*) que regulam a produção de isoflavonas do tipo das agliconas (CHU et al., 2017; CHEN et al., 2021), que são proteínas homólogas àquelas oriundas do gene *GmAGL8* relatado por Liu et al. (2022), e comprovado via transgenia por Qin et al. (2017) como um dos genes responsáveis

pela característica de abertura de vagem na soja. De modo semelhante aos anteriores, os genes *Glyma.04G239000* e *Glyma.04G238900* também participam da biossíntese de aglicona (CHEN et al., 2021). Já o gene *Glyma.04G213300*, é responsável pela produção de uma proteína homóloga àquela produzida pelo gene *SHAT11-5*, que participa da regulação das células FCC (*fibre cap cells*) relacionadas à resistência a deiscência das vagens.

Para a característica PR, o gene identificado *Glyma.18G062200* é reportado como responsável pela produção de proteínas do tipo *CCCH zinc finger*, que comumente são observadas durante o período reprodutivo da soja, e estão envolvidas nos processos de crescimento e adaptação da planta (HU E ZUO, 2021).

Os genes *Glyma.06G257400* e *Glyma.06G257500* identificados para NV estão de acordo com os localizados por Chen et al. (2007), e também, há relatos em *Arabidopsis thaliana* sobre as proteínas de transmembrana do tipo CRK (*cysteine-rich receptor-like protein kinases*) oriundas desses genes estarem associadas a densidade de estômatos e resposta da planta ao estresse hídrico (ARELLANO-VILLAGÓMEZ et al., 2021). O gene *Glyma.07G201800* é pertencente à família *NAC*, que tem por característica estarem presentes no aumento da tolerância de estresse hídrico pelas plantas para características relacionadas a produtividade, que é diretamente influenciada pelo componente de produção NV (HUSSAIN et al., 2017).

Com relação ao NG, apesar de existirem trabalhos que demonstram a associação dos genes indicados com esta característica, como aqueles realizados por Funatsuki et al. (2005), Li et al. (2010) e Liang et al. (2010), ainda não há estudos aprofundados de como esses genes afetam esse caractere. Porém, os fatores de transcrição do tipo *MYB* como aqueles produzidos pelo gene *Glyma.06G248200*, podem estar associados ao desenvolvimento da planta e resposta contra estresses, de modo semelhante ao gene *GmMYB84* reportado por WANG et al. (2017), e além disso, também atuam na biossíntese de isoflavonas de modo semelhante àquelas descritas para PVA (CHU et al., 2017). Já o gene *Glyma.09G150300*, é responsável pela biossíntese de serinas proteases do tipo subtilisina, em que estudos em *Arabidopsis thaliana* sugerem a associação dessas proteases com a densidade e distribuição de estômatos na planta (BERGER E ALTMANN, 2000).

5 Considerações finais

A ocorrência do déficit hídrico na soja pode afetar características de interesse, tanto aquelas relacionadas a componentes de produtividade como o número de vagens, grãos, e ciclo, quanto a qualidade da semente por meio do aumento da abertura de vagens imaturas.

Com o uso da metodologia de mapeamento de herdabilidade regional, foi possível identificar QTLs que atuam no controle da resposta da planta a situações de déficit hídrico, como o que foi imposto neste trabalho. Foi possível identificar associações para as características número de grãos, número de vagens, período reprodutivo, e porcentagem de vagens imaturas.

A análise de mapeamento de herdabilidade regional em soja permitiu a explicação de uma maior porção da herdabilidade genômica, e consequentemente a identificação de um maior número de genes quando comparada ao método convencional de associação, para as características período reprodutivo, número de vagens, número de grãos e abertura de vagens imaturas.

Há a necessidade de novas pesquisas com a cultura da soja para verificação de possíveis genes responsáveis pelo controle das características número de vagens, no cromossomo 9, e abertura de vagens imaturas no cromossomo 1.

Dentre os genes identificados a partir da análise de RHM para este estudo, houve potenciais genes candidatos relacionados tanto às características que foram avaliadas, quanto genes de resposta da planta as condições de estresse hídrico.

A metodologia de RHM foi eficiente em localizar as mesmas associações identificadas pelo método convencional, porém, o oposto não ocorreu, indicando que houve uma superioridade do método em comparação a metodologia de associação via marcas únicas.

Referências bibliográficas

AGARWAL, M., SHRIVASTAVA, N., & PADH, H. Advances in molecular marker techniques and their applications in plant sciences. **Plant cell reports**, v. 27, n. 4, p. 617-631, 2008.

ALQUDAH, A. M., SALLAM, A., BAENZIGER, P. S., E BÖRNER, A. GWAS: fast-forwarding gene identification and characterization in temperate cereals: lessons from barley—a review. **Journal of advanced research**, v. 22, p. 119-135, 2020.

ANDERSON, E. J., ALI, M. L., BEAVIS, W. D., CHEN, P., CLEMENTE, T. E., DIERS, B. W., ... & TILMON, K. J. Soybean [*Glycine max* (L.) Merr.] breeding: history, improvement, production and future opportunities. In: **Advances in plant breeding strategies: Legumes**. Springer, Cham, 2019. p. 431-516.

ARELLANO-VILLAGÓMEZ, F. C., GUEVARA-OLVERA, L., ZUÑIGA-MAYO, V. M., E CERBANTEZ-BUENO, V., VERDUGO-PERALES, M., R. MEDINA, H., ... E ACOSTA-GARCÍA, G. Arabidopsis cysteine-rich receptor-like protein kinase CRK33 affects stomatal density and drought tolerance. **Plant signaling & behavior**, v. 16, n. 6, p. 1905335, 2021.

AZEVEDO, C. F.; NASCIMENTO, M.; FONTES, V. C. F.; SILVA, F. F.; RESENDE, M. D. V.; CRUZ, C. D. (2019) GenomicLand: Software for genome-wide association studies and genomic prediction. **Acta Scientiarum. Agronomy**, v. 41, e45361.

BENJAMINI, Y., AND YOSEF H. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society. Series B (Methodological)**, vol. 57, no. 1, 1995, pp. 289–300.

BENNEWITZ, J., EDEL, C., FRIES, R. et al. Application of a Bayesian dominance model improves power in quantitative trait genome-wide association analysis. **Genet Sel Evol** 49, 7 (2017).

BERGER, D. E ALTMANN, T. A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arabidopsis thaliana*. **Genes and development**, v. 14, n. 9, p. 1119-1131, 2000.

BEZERRA, A. R. G., SEDIYAMA, T., DA SILVA, F. L., BORÉM, A., DA SILVA, A. F., SILVA, F. C. S. Atributos Agronômicos no Desenvolvimento de Cultivares. In: *Melhoramento da Soja*. Editora UFV. 1ª Ed. p. 507-528.

BHAT, J. A., ADEBOYE, K. A., GANIE, S. A., BARMUKH, R., HU, D., VARSHNEY, R. K., E YU, D. Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (*Glycine max* L.). **Frontiers in Genetics**, v. 13, p. 953833, 2022.

BILYEU, K., RATNAPARKHE, M. B., E KOLE, C. (1ª Ed.). Genetics, genomics, and breeding of soybean. **CRC Press**, 2017. 388 p.

CAMPOY, J.A., LERIGOLEUR-BALSEMIN, E., CHRISTMANN, H. et al. Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. **BMC Plant Biol** 16, 49 (2016).

CHEN, Q., ZHANG, Z. C., LIU, C. Y., XIN, D. W., QIU, H. M., SHAN, D. P., ... E HU, G. H. QTL analysis of major agronomic traits in soybean. **Agricultural Sciences in China**, v. 6, n. 4, p. 399-405, 2007.

CHEN, H., PAN, X., WANG, F. F., LIU, C., WANG, X., LI, Y., E ZHANG, Q. Novel QTL and Meta-QTL Mapping for Major Quality Traits in Soybean. **Frontiers in plant science**, p. 2773, 2021.

CHU S., WANG J., ZHU Y., LIU S., ZHOU X., ZHANG H., WANG C.E., YANG W., TIAN Z., CHENG H., YU D. An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. **PLoS Genet**. 2017.

CONTRERAS-SOTO, R. I., DE OLIVEIRA, M. B., COSTENARO-DA-SILVA, D., SCAPIM, C. A., E SCHUSTER, I. Population structure, genetic relatedness and linkage disequilibrium blocks in cultivars of tropical soybean (*Glycine max*). **Euphytica**, v. 213, n. 8, p. 1-12, 2017.

COVARRUBIAS-PAZARAN G. Genome-Assisted Prediction of Quantitative Traits Using the R Package *sommer*. **PLOS ONE**. 2016. 11(6): e0156744.

CRUZ, C.D. Princípios de Genética Quantitativa. Viçosa, MG: **Editora UFV**. 2012. 394 p.

DALCHIAVON, F. C., DE PASSOS E CARVALHO, M. Correlação linear e espacial dos componentes de produção e produtividade da soja. **Semina: Ciências Agrárias**, p. 541-552, 2012.

DE CARVALHO, M. C. D. C. G., E SILVA, D. C. G. D. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, v. 40, p. 735-744, 2010.

DE FELIPE, M., GERDE, J. A., E ROTUNDO, J. L. Soybean genetic gain in maturity groups III to V in Argentina from 1980 to 2015. **Crop science**, v. 56, n. 6, p. 3066-3077, 2016.

DE ALMEIDA, L. A., KIIHL, R. D. S., DE MIRANDA, M. A. C., E CAMPELO, G. D. A. Melhoramento da soja para regiões de baixas latitudes. **Embrapa Meio-Norte-Capítulo em livro científico (ALICE)**, 1999.

DHUNGANA S.K., PARK J-H., OH J-H., KANG B.K., SEO J-H., SUNG J-S., KIM H-S., SHIN S-O., BAEK I-Y, JUNG C-S. Quantitative Trait Locus Mapping for Drought Tolerance in Soybean Recombinant Inbred Line Population. **Plants**. 2021; 10(9):1816.

DONG, Y., X., LIU, J., WANG, B. H., LIU, B. L., E WANG, Y. Z. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. **Nature communications**, v. 5, n. 1, p. 1-11, 2014.

DUBEY, A., KUMAR, A., ABD_ALLAH, E. F., HASHEM, A., & KHAN, M. L. Growing more with less: breeding and developing drought resilient soybean to improve food security. **Ecological Indicators**, v. 105, p. 425-437, 2019.

ELSHIRE, R. J., GLAUBITZ, J. C., SUN, Q., POLAND, J. A., KAWAMOTO, K., BUCKLER, E. S., & MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS one**, v. 6, n. 5, p. e19379, 2011.

FEHR, W.R., CAVINESS, C.E. **Stages on soybean development**. Ames: Iowa State University/Cooperative Extension Service, (Special Report, 80), p. 11, 1977.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER IV, E. S. Structure of linkage disequilibrium in plants. **Annu Rev Plant Biol**. 54:357-74. 2003.

FUNATSUKI, H., KAWAGUCHI, K., MATSUBA, S., SATO, Y., E ISHIMOTO, M. Mapping of QTL associated with chilling tolerance during reproductive growth in soybean. **Theoretical and Applied Genetics**, v. 111, n. 5, p. 851-861, 2005.

GRIFFITHS, A. J., WESSLER, S. R., LEWONTIN, R. C., GELBART, W. M., SUZUKI, D. T., & MILLER, J. H. Introdução à genética. In: **Introdução à genética**. 2006. p. 743-743.

GUO, X., Su, G., Christensen, O. F., Janss, L., E Lund, M. S. Genome-wide association analyses using a Bayesian approach for litter size and piglet mortality in Danish Landrace and Yorkshire pigs. **BMC genomics**, v. 17, n. 1, p. 1-12, 2016.

HIRAKURI, M. H. **Perdas econômicas geradas por estresses bióticos e abióticos na produção brasileira de soja no período 2016-2020**. 1. ed. Londrina, PR: Embrapa, 2021. 8 p. (Circular Técnica, 168).

HU, X. E ZUO, J. The CCCH zinc finger family of soybean (*Glycine max* L.): genome-wide identification, expression, domestication, GWAS and haplotype analysis. **BMC Genomics** 22, 511 (2021).

HUSSAIN, R. M., ALI, M., FENG, X., E LI, X. The essence of NAC gene family to the cultivation of drought-resistant soybean (*Glycine max* L. Merr.) cultivars. **BMC plant biology**, v. 17, n. 1, p. 1-11, 2017.

JOMBART, T. *adegenet*: a R package for the multivariate analysis of genetic markers. **Bioinformatics**, v. 24, n. 11, p. 1403-1405, 2008.

JOMBART, T., DEVILLARD, S. E BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC Genet** 11, 94. 2010.

KALER A.S., GILLMAN J.D., BEISSINGER T., PURCELL L.C. Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. *Front. Plant Sci.* 2020.

KARASU, A. OZ, M. E. H. M. E. T., GÖKSOY, A. T., E TURAN, Z. M. Genotype by environment interactions, stability, and heritability of seed yield and certain agronomical traits in soybean [*Glycine max* (L.) Merr.]. ***African Journal of Biotechnology***, v. 8, n. 4, 2009.

KARIKARI, B., CHEN, S., XIAO, Y., CHANG, F., ZHOU, Y., KONG, J., ... E ZHAO, T. Utilization of interspecific high-density genetic map of RIL population for the QTL detection and candidate gene mining for 100-seed weight in soybean. ***Frontiers in plant science***, v. 10, p. 1001, 2019.

KAPPES, Claudinei. Sistemas de cultivo de milho safrinha no Mato Grosso. **XII Seminário Nacional Estabilidade e Produtividade**, Embrapa, Dourados, MS, p. 26-28, 2013.

KORIR, P. C., QI, B., WANG, Y., ZHAO, T., YU, D., CHEN, S., E GAI, J. A study on relative importance of additive, epistasis and unmapped QTL for aluminium tolerance at seedling stage in soybean. ***Plant Breeding***, v. 130, n. 5, p. 551-562, 2011.

KORTE, L. L., WILLIAMS, J. H., SPECHT, J. E., E SORENSEN, R. C. Irrigation of Soybean Genotypes During Reproductive Ontogeny. I. Agronomic Responses 1. ***Crop Science***, v. 23, n. 3, p. 521-527, 1983.

KUNERT, K. J., VORSTER, B. J., FENTA, B. A., KIBIDO, T., DIONISIO, G., FOYER, C. H. Drought stress responses in soybean roots and nodules. ***Frontiers in plant science***, v. 7, p. 1015, 2016.

KURODA, Y., KAGA, A., TOMOOKA, N., YANO, H., TAKADA, Y., KATO, S., E VAUGHAN, D. QTL affecting fitness of hybrids between wild and cultivated soybeans in experimental fields. ***Ecology and Evolution***, v. 3, n. 7, p. 2150-2168, 2013.

LEE, S., ABECASIS, G. R., BOEHNKE, M., E LIN, X. Rare-variant association analysis: study designs and statistical tests. ***The American Journal of Human Genetics***, v. 95, n. 1, p. 5-23, 2014.

LIANG, Q., CHENG, X., MEI, M., YAN, X., E LIAO, H. QTL analysis of root traits as related to phosphorus efficiency in soybean. ***Annals of Botany***, v. 106, n. 1, p. 223-234, 2010.

LIU, J., ZHANG, Y., JIANG, Y., SUN, H., DUAN, R., QU, J., E GUAN, S. Formation Mechanism and Occurrence Law of Pod Shattering in Soybean: A Review. ***Phyton***, v. 91, n. 7, p. 1327, 2022.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A. Finding the missing heritability of complex diseases. ***Nature***, v. 461, n. 7265, p. 747-753, 2009.

MAPHOSA, M., TALWANA, H., GIBSON, P., E TUKAMUHABWA, P. Combining ability for resistance to soybean rust in F2 and F3 soybean populations. **Field Crops Research**, v. 130, p. 1-7, 2012.

MASUDA, T.; GOLDSMITH, P. D. World soybean production: area harvested, yield, and long-term projections. **International food and agribusiness management review**, v. 12, n. 1030-2016-82753, p. 1-20, 2009.

MOURA, L. **IDENTIFICAÇÃO DE CARACTERES AUXILIARES PARA SELEÇÃO DE GENÓTIPOS DE SOJA TOLERANTES À ABERTURA DE VAGENS IMATURAS**. 2021. Tese (Doutorado) – Curso de Fitotecnia, Universidade Federal de Viçosa, Viçosa – MG, 2021.

MUIR, P., LI, S., LOU, S., WANG, D., SPAKOWICZ, D. J., SALICHOS, L., ... & GERSTEIN, M. The real cost of sequencing: scaling computation to keep pace with data generation. **Genome biology**, v. 17, n. 1, p. 1-9, 2016.

NAGAMINE Y., PONG-WONG R., NAVARRO P., VITART V., HAYWARD C., et al. Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping. **PLOS ONE** 7(10): e46501. 2012.

NASCIMENTO, H. R. D., MOURA, L. D. O., DUARTE, A. B., DANTAS, S. A. G., FERREIRA, D. D. O., ROSMANINHO, L. B. D. C., ... & DA SILVA, F. L. A new methodological approach for simulating water deficit in soybean genotypes. **Journal of Agronomy and Crop Science**, v. 207, n. 6, p. 946-955, 2021.

NAVARRO JÚNIOR, H. M., E COSTA, J. A. Expressão do potencial de rendimento de cultivares de soja. **Pesquisa Agropecuária Brasileira**, v. 37, p. 275-279, 2002.

OKEKE, U. G., AKDEMIR, D., RABBI, I., KULAKOW, P., E JANNINK, J. L. Regional heritability mapping provides insights into dry matter content in African white and yellow cassava populations. **The plant genome**, v. 11, n. 1, p. 170050, 2018.

OTYAMA, P.I., WILKEY, A., KULKARNI, R. et al. Evaluation of linkage disequilibrium, population structure, and genetic diversity in the U.S. peanut mini core collection. **BMC Genomics** 20, 481. 2019.

PALHARINI, W.G. **Influência do estresse hídrico sobre caracteres agronômicos, fisiológicos e abertura de vagens imaturas em soja**. Dissertação (Mestrado), Programa de Pós-Graduação em Fitotecnia, Universidade Federal de Viçosa, Viçosa-MG, 44 p., 2016. <https://locus.ufv.br/handle/123456789/9284>

QIN, Y., WANG, R., LYU, J., TIAN, Y., HUANG, T., XIE, H., E GUO, B. Transgenic identification of pod-shattering gene GmAGL8 in soybean. **Soybean Science**, v. 36, n. 3, p. 360-364, 2017.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2021.

RAY, D. K. MUELLER, N. D., WEST, P. C., & FOLEY, J. A. Yield trends are insufficient to double global crop production by 2050. **PloS one**, v. 8, n. 6, p. e66428, 2013.

RESENDE, R.T., RESENDE, M.D.V., SILVA, F.F., AZEVEDO, C.F., TAKAHASHI, E.K., SILVA-JUNIOR, O.B. AND GRATTAPAGLIA, D. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. **New Phytol**, **213**: 1287-1300. 2017.

RESENDE, R. T., RESENDE, M. D. V., AZEVEDO, C. F., SILVA, F. F., MELO, L. C., PEREIRA, H. S., SOUZA, T. L. P. O., VALDISSER, P. A. M. R., BRONDANI, C., E VIANELLO, R. P. Genome-wide association and regional heritability mapping of plant architecture, lodging and productivity in *Phaseolus vulgaris*. **G3: Genes, Genomes, Genetics**, **8**, 2841–2854. 2018.

REN, H., HAN, J., WANG, X., ZHANG, B., YU, L., GAO, H., QIU, L. J. QTL mapping of drought tolerance traits in soybean with SLAF sequencing. **The Crop Journal**, v. 8, n. 6, p. 977-989, 2020.

RIGGIO, V., PONG-WONG, R. Regional heritability mapping to identify loci underlying genetic variation of complex traits. In: **BMC proceedings. BioMed Central**, 2014. p. 1-4.

SALVIANO, Marcos Figueiredo; MOREIRA, Daniel Medeiros; PINTO, Eber José de Andrade. **Acompanhamento da estiagem nas regiões sudeste e centro oeste do Brasil: Acompanhamento das precipitações 2021**. CPRM, 2021.

SATBHAI, Santosh B.; RISTOVA, Daniela; BUSCH, Wolfgang. Underground tuning: quantitative regulation of root growth. **Journal of Experimental Botany**, v. 66, n. 4, p. 1099-1112, 2015.

SCHOLANDER, P.F.; HAMMEL, H.T.; BRADSTREET, E.D.; HEMMINGSEN, E.A. Sap pressure in vascular plants. **Science**, v. 148, p. 339-346, 1965.

SEHGAL, D., SINGH, R., E RAJPAL, V. R. Quantitative trait loci mapping in plants: concepts and approaches. **Molecular breeding for sustainable crop improvement**, p. 31-59, 2016.

SHIRALI, M., PONG-WONG, R., NAVARRO, P., KNOTT, S., HAYWARD, C., VITART, V., ... E HALEY, C. S. Regional heritability mapping method helps explain missing heritability of blood lipid traits in isolated populations. **Heredity**, v. 116, n. 3, p. 333-338, 2016.

SILVA, F., BORÉM, A., SEDIYAMA, T., & LUDKE, W. Melhoramento da soja. **Viçosa: Editora UFV**, 2017. 563p.

SILVA, F.; BORÉM, A.; SEDIYAMA, T.; G, E CÂMARA. Soja: do plantio à colheita. **Oficina de Textos**, 2022.

SIONIT, N., E KRAMER, P. J. Effect of Water Stress During Different Stages of Growth of Soybean 1. **Agronomy journal**, v. 69, n. 2, p. 274-278, 1977.

SOARES, B. A. **Associação genômica ampla para resistência de cultivares de soja à *Sclerotinia sclerotiorum***. Trabalho (dissertação), 2018.

SU, H. G., LI, B., SONG, X. Y., MA, J., CHEN, J., ZHOU, Y. B., ... E MA, Y. Z. Genome-wide analysis of the DYW subgroup PPR gene family and identification of GmPPR4 responses to drought stress. **International journal of molecular sciences**, v. 20, n. 22, p. 5667, 2019.

SUELA, M. M., AZEVEDO, C. F., NASCIMENTO, M., NASCIMENTO, A. C. C., E DE RESENDE, M. D. V. Regional heritability mapping and genome-wide association identify loci for rice traits. **Crop Science**, 62, 839– 858. 2022.

SUN, D., LI, W., ZHANG, Z., CHEN, Q., NING, H., QIU, L., E SUN, G. Quantitative trait loci analysis for the developmental behavior of soybean (*Glycine max* L. Merr.). **Theoretical and Applied Genetics**, v. 112, n. 4, p. 665-673, 2006.

TEODORO, L. P. R., BHERING, L. L., GOMES, B. E. L., CAMPOS, C. N. S., BAILO, F. H. R., GAVA, R., ... E TEODORO, P. E. Understanding the combining ability for physiological traits in soybean. **PLoS one**, v. 14, n. 12, p. e0226523, 2019.

USAI, M. G., GASPA, G., MACCIOTTA, N. P., CARTA, A., E CASU, S. XVth QTLMAS: simulated dataset and comparative analysis of submitted results for QTL mapping and genomic evaluation. In: **BMC proceedings**. BioMed Central, 2014. p. 1-9.

VANN, R., EVERMAN, W., HOLSHOUSER, D., HUSETH, A., REISIG, D., & THIESSEN, L. **General Technical Report (1 of 23 volumes)**, v. 260, p. 1-56, 2021.

VANRADEN P.M. Efficient methods to compute genomic predictions. **J Dairy Sci**. Nov; 91(11): 4414-23. 2008.

VOS, P.G., PAULO, M.J., VOORRIPS, R.E. et al. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. **Theor Appl Genet** **130**, 123–135 (2017).

WANG, N., ZHANG, W., QIN, M., LI, S., QIAO, M., LIU, Z., E XIANG, F. Drought tolerance conferred in soybean (*Glycine max*. L) by GmMYB84, a novel R2R3-MYB transcription factor. **Plant and Cell Physiology**, v. 58, n. 10, p. 1764-1776, 2017.

YU, J. PRESSOIR, G., BRIGGS, W. H., VROH BI, I., YAMASAKI, M., DOEBLEY, J. F., ... E BUCKLER, E. S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature genetics**, v. 38, n. 2, p. 203-208, 2006.

YU, Z., CHANG, F., LV, W., SHARMIN, R. A., WANG, Z., KONG, J., ... E ZHAO, T. Identification of QTN and candidate gene for seed-flooding tolerance in soybean [*Glycine max* (L.) Merr.] using genome-wide association study (GWAS). **Genes**, v. 10, n. 12, p. 957, 2019.

ZHANG, T., WU, T., WANG, L., JIANG, B., ZHEN, C., YUAN, S., ... E SUN, S. A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. **International journal of molecular sciences**, v. 20, n. 23, p. 5915, 2019.

ZHAO, H., NETTLETON, D., DEKKERS, J. C.M. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. **Genetics Research**, v. 89, n. 1, p. 1-6, 2007.

ZHOU, L., LIU, S., WU, W., CHEN, D., ZHAN, X., ZHU, A., ... E XU, H. Dissection of genetic architecture of rice plant height and heading date by multiple-strategy-based association studies. **Scientific reports**, v. 6, n. 1, p. 1-10, 2016.

Material suplementar

Tabela suplementar 1. Genótipos de Soja utilizados na condução e análises do experimento.

Genótipo	Grupo de Maturidade Relativa	Evento transgênico	Tipo de crescimento
54152 RSF IPRO	5.4	Intacta	Indeterminado
5855 RSF IPRO	5.5	Intacta	Indeterminado
58160 RSF IPRO	5.8	Intacta	Indeterminado
5953 RSF	5.0	Roundup Ready	Indeterminado
61159 IPRO	6.1	Intacta	-
6211 IPRO	-	Intacta	-
6262 RSF IPRO	-	Intacta	-
63164 RSF IPRO	6.4	Intacta	Indeterminado
6563 RSF IPRO	6.3	Intacta	Indeterminado
68170 RSF IPRO	6.8	Intacta	Indeterminado
7166 RSF IPRO (Ponta)	6.6	Intacta	Indeterminado
73170 RSF (BMX Power IPRO)	7.3	Intacta	Indeterminado
75177 RSF (BMX Ultra IPRO)	7.6	Intacta	Indeterminado
8075 RSF (Prisma)	7.5	Intacta	Indeterminado
8077 RSF (Forte)	7.7	Roundup Ready	Indeterminado
8579 RSF (Bônus IPRO)	7.9	Intacta	Indeterminado
9086 RSF (Opus IPRO)	8.6	Intacta	Determinado
ANTA 82	7.4	Roundup Ready	Semi determinado
AS 3570 IPRO	5.7	Intacta	Semi determinado
AS 3810	8.1	Intacta	Determinado
AS 3820 IPRO	8.2	Intacta	Determinado
BMX Potência RR	6.7	Roundup Ready	Indeterminado
CD 2728 IPRO	7.2	Intacta	Determinado
CD 2737	7.3	Roundup Ready	Indeterminado
DM 4338	-	Intacta	-
DM 4347	-	Intacta	-
DM 5.8i	5.5	Roundup Ready	Indeterminado

DM 7.8i	-	Roundup Ready	Indeterminado
GDM 15A004	6.1	Roundup Ready	-
GDM 15C004	8.0	Convencional	-
GDM 15I009	7.2	Intacta	-
GDM 15I010	7.5	Intacta	-
GDM 15I011	7.6	Intacta	-
GDM 15I012	7.8	Intacta	-
GDM 15I013	7.8	Intacta	-
GDM 15I014	7.8	Intacta	-
GDM 15I015	8.6	Intacta	-
GDM 15I016	8.5	Intacta	-
GDM 15I017	7.5	Intacta	-
GDM 15I018	5.3	Intacta	-
GDM 15I019	6.2	Intacta	-
GDM 15I020	6.0	Intacta	-
GDM 15I022	6.1	Intacta	-
GDM 15I024	6.2	Intacta	-
GDM 15I027	6.0	Intacta	-
GDM 15I028	6.0	Intacta	-
GDM 15I029	8.0	Intacta	-
GDM 15I032	9.0	Intacta	-
GDM 15I033	7.5	Intacta	-
GDM 15I035	8.4	Intacta	-
GDM 15I037	6.1	Roundup Ready	-
GDM 15I038	8.0	Convencional	-
GDM 15I041	7.2	Intacta	-
GDM 15I042	7.5	Intacta	-
GDM 15I048	7.6	Intacta	-
GDM 15I049	7.8	Intacta	-
GDM 15I050	7.8	Intacta	-
GDM 15I051	7.8	Intacta	-
GDM 15I052	8.6	Intacta	-
GDM 15I053	8.5	Intacta	-
GDM 15I054	7.5	Intacta	-
GDM 15I055	5.3	Intacta	-
GDM 15I057	6.2	Intacta	-
GDM 15I075	6.0	Intacta	-
GDM 15I079	6.1	Intacta	-
IGRA 645 TR	6.2	Roundup Ready	-
LG 60163 IPRO	6.3	Roundup Ready	Semi determinado
M 5947 IPRO	5.9	Intacta	Indeterminado
M 7739 IPRO	7.7	Intacta	Semi determinado
M 8133 IPRO	8.1	Intacta	Determinado
M 8349 IPRO	8.3	Intacta	Determinado
M 8372 IPRO	8.3	Intacta	Determinado
M 8644 RR	8.6	Intacta	Determinado
M 8808	8.8	Intacta	Determinado
M-SOY 7639	7.6	Roundup Ready	Indeterminado

M-SOY 5730 IPRO	5.7	Intacta	Indeterminado
NA 7337 RR	7.6	Roundup Ready	Indeterminado
NS 5727 IPRO	5.7	Intacta	Indeterminado
NS 6006 IPRO	5.7	Intacta	Indeterminado
NS 6909 IPRO	6.3	Intacta	Indeterminado
NS 7300 IPRO	7.0	Intacta	Indeterminado
NS 7490	7.4	Roundup Ready	Indeterminado
NS 7901	7.9	Roundup Ready	Indeterminado
P 95Y21	5.2	Convencional	-
P 95Y72	5.0	Roundup Ready	Indeterminado
P 98Y30	8.3	Roundup Ready	-
ROOS CAMINO	5.3	Convencional	Indeterminado
SYN 1080	8.0	Roundup Ready	-
SYN 1285	8.5	Roundup Ready	-
TMG 1179	7.9	Roundup Ready	Determinado
TMG 1180 RR	8.0	Roundup Ready	Semi determinado
TMG 1288 RR	8.8	Roundup Ready	Determinado
TMG 2181 IPRO	8.1	Intacta	Determinado
TMG 4182	8.2	Convencional	Determinado
TMG 4185	8.5	Convencional	Determinado
TMG 4190	9.0	Convencional	Determinado
TMG 7062 IPRO	6.2	Intacta	Semi determinado
TMG 7262	6.2	Roundup Ready	Semi determinado
TMG 7363 RR	6.3	Roundup Ready	Indeterminado