

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE CIÊNCIAS AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

ANTONIO HENRIQUE BOZI

**ASSOCIAÇÃO GENÔMICA AMPLA E PRINCIPAIS GENES DE
MATURIDADE EM SOJA**

DISSERTAÇÃO

PATO BRANCO

2021

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE CIÊNCIAS AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA**

ANTONIO HENRIQUE BOZI

**ASSOCIAÇÃO GENÔMICA AMPLA E PRINCIPAIS GENES DE
MATURIDADE EM SOJA**

DISSERTAÇÃO

PATO BRANCO

2021

ANTONIO HENRIQUE BOZI

**ASSOCIAÇÃO GENÔMICA AMPLA E PRINCIPAIS GENES DE
MATURIDADE EM SOJA**

Genome wide association studies and major genes of maturity in soybean

Dissertação apresentada ao Programa de Pós-Graduação em Agronomia da Universidade Tecnológica Federal do Paraná, *Campus* Pato Branco, como requisito parcial à obtenção do título de Mestre em Agronomia - Área de Concentração: Produção Vegetal.

Orientador: Prof. Dr. Giovani Benin

PATO BRANCO

2021



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite o download e o compartilhamento da obra desde que sejam atribuídos créditos ao(s) autor(es), sem a possibilidade de alterá-la ou utilizá-la para fins comerciais.



Ministério da Educação
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO-PB
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA



ANTONIO HENRIQUE BOZI

ASSOCIAÇÃO GENÔMICA AMPLA E PRINCIPAIS GENES DE MATURIDADE EM SOJA

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Agronomia da Universidade Tecnológica Federal do Paraná (UTFPR).
Área de concentração: Produção Vegetal.

Data de aprovação: 23 de Março de 2021

Prof Giovani Benin, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Eduardo Beche, Doutorado - University Of Missouri, Columbia

Prof Salvador Lima Brito Junior, Doutorado - Gdm Genética do Brasil Ltda

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 23/03/2021.

Dedico a realização deste trabalho aos meus pais Dirceu Antonio Bozi e Jucelia Arcari Bozi; as minhas irmãs Luana Bozi e Luiza Bozi; A minha namorada Angel Dalmas Petroski que estiveram sempre presentes e me apoiaram em todos os momentos difíceis.

AGRADECIMENTOS

Primeiramente a Deus por ter me dado saúde e força para superar todas as dificuldades encontradas.

A minha família: pais Dirceu e Jucelia e irmãs Luana e Luiza, e minha namorada Angel, que me deram apoio, incentivo nas horas difíceis, de desânimo e cansaço e que me incentivaram na conclusão do mestrado.

Ao meu orientador, Prof. Dr. Giovani Benin, pelos seus aconselhamentos, sua importante orientação e todos os conhecimentos transmitidos durante minha formação como Eng. Agrônomo e Mestre em Agronomia.

Aos colegas e amigos da UTFPR, em especial aos meus colegas de trabalho do grupo de melhoramento genético de soja e trigo, que me auxiliaram na execução desse e de inúmeros outros trabalhos, além de proporcionarem momentos felizes em minha vida.

À Universidade Tecnológica Federal do Paraná – UTFPR e ao Programa de Pós Graduação em Agronomia, pela oportunidade de aprendizado e crescimento pessoal e profissional, durante toda minha graduação e aperfeiçoamento com o mestrado.

À empresa GDM Seeds e a todas as pessoas que fazem parte da mesma, pelo provimento dos dados para a realização desta pesquisa, pelo apoio e incentivo a elaboração dos trabalhos e por acreditar no grupo de pesquisa coordenado pelo Professor Giovani Benin.

Agradeço a CAPES, visto que o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Por fim, agradeço a todos aqueles que de alguma forma contribuíram para minha formação como mestre.

“A imaginação é mais importante que a ciência, porque a ciência é limitada, ao passo que a imaginação abrange o mundo inteiro”.

(Albert Einstein)

RESUMO

BOZI, Antonio Henrique. Associação genômica ampla e principais genes de maturidade em soja. 61 f. Dissertação (Mestrado em Agronomia) – Programa de Pós-Graduação em Agronomia (Área de Concentração: Produção vegetal), Universidade Tecnológica Federal do Paraná (UTFPR). Pato Branco, 2021.

A soja (*Glycine max*) é uma espécie sensível ao fotoperíodo e uma leguminosa de grande interesse mundial. O período de floração e maturidade são fatores importantes que controlam a produtividade da cultura e a sua adaptação latitudinal. Entender os mecanismos moleculares que regulam o período de floração e maturidade, se torna necessário a fim de, melhorar a adaptabilidade e produtividade da soja. Embora quatro genes tenham sido utilizados em programas de melhoramento de maturidade precoce, novos genes e QTLs que determinam a maturidade estão continuamente sendo identificados, sugerindo que existem loci desconhecidos e que governam tais características de interesse. Dessa forma, o objetivo do presente estudo foi identificar novas regiões genômicas (QTLs) e marcadores SNP associados a grupos de maturação relativa e identificar genes de resposta ao fotoperíodo/indução floral da soja em linhagens, via análise de associação genômica ampla. Um total de 5428 linhagens de soja, oriundas do programa de melhoramento genético da GDM Seeds, foram avaliadas em três países (Argentina, Estados Unidos e Brasil) quanto ao grupo de maturidade (GM). Posteriormente, as linhagens foram genotipadas, utilizando marcadores moleculares SNPs, através da tecnologia de genotipagem baseada na competição alelo específica (KASP – *Kompetitive Allele Specific PCR*). Um total 2669 marcadores SNP polimórficos foram utilizados para analisar a estrutura das populações e realizar o estudo de associação genômica ampla (GWAS). Os resultados da análise GWAS revelaram um total de 29 SNPs significativos ($p < 0,001$) associados aos diferentes grupos de maturação, distribuídos em 13 cromossomos, incluindo os cromossomos 1, 2, 3, 6, 7, 8, 9, 12, 13, 17, 18, 19 e 20. Nove SNPs foram identificados como novos loci e 20 SNPs foram localizados 1,5Mb próximos de QTLs conhecidos. Quatro genes candidatos (Glyma.03g169700, Glyma19g.13800, Glyma.06g216400 e Glyma.06g207800) foram associados a grupos de maturação precoce, sendo identificado marcadores significativos até GMs 5.0. Outros seis genes (Glyma.06g204100, Glyma.06g057900, Glyma.12g089100, Glyma.12g103800, Glyma.12g073900 e Glyma.08g236300) foram candidatos, por estarem associados a grupos de maturação mais tardios, sendo significativos em GMs acima de 5.0. Os genes candidatos, dentre eles enzimas e fatores de transcrição, encontram-se relacionados direta ou indiretamente com o caractere analisado. Os loci e marcadores associados a grupos de maturação, identificados neste estudo, podem ser usados para o melhoramento genético da soja, por meio da seleção assistida por marcadores (SAM), propiciando a seleção de genótipos adaptados e com previsibilidade de comportamento a diferentes regiões Edafoclimáticas de cultivos, de forma ágil e precisa.

Palavras-chave: *Glycine max*. Melhoramento de soja. GWAS. Genes candidatos. Grupos de maturação.

ABSTRACT

BOZI, Antonio Henrique. Genome wide association and major soybean maturity genes. 61f. Dissertation (Masters in Agronomy) - Graduate Program in Agronomy (Concentration Area: Crop), Federal University of Technology - Paraná (UTFPR). Pato Branco, 2020.

Soybean (*Glycine max*) is a sensitive species to the photoperiod and a legume of great interest worldwide. The flowering time and maturity are important factors that control the productivity of the crop and its latitudinal adaptation. Understanding the molecular mechanism that regulate the flowering time and maturity is necessary in order, to improve adaptability and productivity. Although four genes have been used in early maturity breeding programs, new genes and QTLs that determine maturity are continually being identified, suggesting that unknown loci exist and that govern such characteristics of interest. Thus, the objective of the present study was to identify new genomic regions (QTLs) and SNP markers associated with groups of relative maturity and to identify response genes to the soybean photoperiod/ floral induction in strains via genome wide association. A total of 5428 soybean strains from the GDM Seeds genetic improvement program, were evaluated in three countries (Argentina, United States and Brazil) regarding the maturity group (MG). Subsequently, the strains were genotyped using SNP molecular markers, using genotyping technology based on specific allele competition (KASP- Kompetitive Allele Specific PCR). In total 2669 polymorphic SNP marker were used to analyze the structure of populations and carry out the genome wide association studies (GWAS). The results of the GWAS analysis revealed a total of 29 significant SNPs ($p < 0,001$) associated with the different maturity groups, which were distributed in 13 chromosomes, including chromosomes 1, 2, 3, 6, 7, 8, 9, 12, 13, 17, 18, 19 e 20. Nine SNPs were new loci and 20 SNPs were located within or 1.5Mb close to know QTLs. Four candidate genes (Glyma.03g169700, Glyma19g.13800, Glyma.06g216400 e Glyma.06g207800) were associated with groups of early maturation, being represented by significant markers until MGs 5.0. Another six genes (Glyma.06g204100, Glyma.06g057900, Glyma.12g089100, Glyma.12g103800, Glyma.12g073900 e Glyma.08g236300) were candidate, as they were associated with later maturity groups, being significant in MGs above 5.0. Candidate genes, including enzymes and transcription factors, are directly and directly related to the character being analyzed. The loci and markers associated with maturity groups, identified in this study can be used for the genetic improvement of soybean, through marker-assisted selection (MAS), providing genotypes with desirable maturation groups in a more agile and precise manner.

Keywords: *Glycine max*. Soybean breeding. GWAS. Candidate genes. Maturation groups.

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1 -- Mapa Americano com o melhor desempenho dos grupos de maturidade de soja nos EUA. Adaptado de: (MOURTZINIS E CONLEY 2017). | 9 |
| Figura 2 -- Mapa brasileiro com as localizações das macrorregiões (M1, M2, M3 e M4) e regiões edafoclimáticas (101–503). Modificado de: (ZDZIARSKI et al, 2018). 10 | 10 |
| Figura 3 Mapa Argentino com o melhor desempenho dos grupos de maturidade de soja na Argentina. Adaptado de: (BAIGORRI, 2009). | 11 |
| Figura 4 -- Locais de avaliação genotípica e fenotípica das linhagens de soja (UTFPR- Campus Pato Branco, 2020). UTFPR, <i>Campus Pato Branco</i> , 2020. | 15 |
| Figura 5 – Distribuição da frequência do conjunto de linhagens de soja para grupos de maturação. Conjunto de linhagens (A) geral, (B) argentinas (C) brasileiras, (D) norte americanas. UTFPR, <i>Campus Pato Branco</i> , 2020. | 20 |
| Figura 6 – Distribuição dos marcadores SNPs em 20 cromossomos de soja, avaliados em 5428 linhagens de soja. UTFPR, <i>Campus Pato Branco</i> , 2020. | 21 |
| Figura 7 – Análise de componentes principais dos três primeiros componentes de 5428 linhagens. Conjunto de linhagens argentinas, brasileiras e norte americanas. UTFPR, <i>Campus Pato Branco</i> , 2020. | 22 |
| Figura 8 -- Análise de componentes principais (PCA) dos três primeiros componentes de 5428 linhagens. Conjunto de linhagens (A) GM 0.3 a 3.5, (B) GM 3.6 a 5.0, (C) 5.1 a 6.0, (D) GM 6.1 a 7.0 e (E) GM 7.1 a 8.5. UTFPR, <i>Campus Pato Branco</i> , 2020. | 23 |
| Figura 9 – Decaimento do desequilíbrio de ligação em todo o genoma das 5428 linhagens de soja determinado pelas medidas de desequilíbrio de ligação (LD) contra a distância entre marcadores SNPs: r^2 vs as distancias físicas entre os pares de marcadores. UTFPR, <i>Campus Pato Branco</i> , 2020. | 24 |
| Figura 10- Manhattan Plot dos valores de $-\log_{10}(P)$ obtidos para cada SNP vs a posição no cromossomo para grupos de maturação (GM) geral e para Argentina, Brasil e Estados Unidos em soja. A linha paralela indica valor de LOD score = 3, correspondendo ao limite de p-value < 0,001. UTFPR, <i>Campus Pato Branco</i> , 2020. | 26 |
| Figura 11 -- Manhattan Plot dos valores de $-\log_{10}(P)$ obtidos para cada SNP vs a posição no cromossomo para grupos de maturação (GM) em soja. A linha paralela indica valor de LOD score = 3, correspondendo ao limite de p-value < 0,001. UTFPR, <i>Campus Pato Branco</i> , 2020. | 30 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 - Marcadores SNPs significativos na análise de associação genômica para grupos de maturação e países em soja, com a identificação dos alelos, cromossomos, posição, valor $-\text{Log}_{10}(p)$ e R^2 e MAF. UTFPR, <i>Campus</i> Pato Branco, 2020. | 25 |
| Tabela 2 - Marcadores SNPs significativos na análise de associação genômica para grupos de maturação em soja, com a identificação dos alelos, cromossomos, posição, valor $-\text{Log}_{10}(p)$ e R^2 , gene candidato e proteína correspondente. UTFPR, <i>Campus</i> Pato Branco, 2020. | 28 |
| Tabela 3- Distribuição e frequência alélica dos principais genes relacionados a maturidade em 5428 linhagens de soja, provenientes da Argentina, Estados Unidos e Brasil. UTFPR, <i>Campus</i> Pato Branco, 2020..... | 36 |

LISTA DE SIGLAS E ACRÔNIMOS

| | |
|----------|---|
| BLUP | <i>Best Linear Unbiased Prediction</i> |
| CONAB | Companhia Nacional de Abastecimento |
| DNA | Deoxyribonucleic acid |
| DTM | Dias totais até maturação |
| EUA | Estados Unidos da América |
| FAO | <i>Food and Agriculture Organization of the United Nations</i> |
| FAS/USDA | <i>Foreign Agricultural Service/United States Department of Agriculture</i> |
| FT | <i>Flowering Time</i> |
| GDM | Grupo Don Mario |
| GLM | Grupo de ligação molecular |
| GM | Grupo de maturação |
| GMR | Grupo de maturação relativa |
| GWAS | Genome Wide Association Studies |
| KASP | <i>Kompetitive Allele Specific</i> |
| Kb | Kilobase |
| LD | Linkage Disequilibrium |
| LOD | <i>Logarithm of odds</i> |
| Mb | Megabase |
| MLMM | <i>Multi-locus Mixed Model</i> |
| PCA | <i>Principal Component Analysis</i> |
| PJL | Período Juvenil longo |
| QTL | <i>Quantitative Trait Loci</i> |
| SAM | Seleção assistida por marcadores moleculares |
| SNP | <i>Single Nucleotide Polymorphism</i> |
| UTFPR | Universidade Tecnológica Federal do Paraná |

LISTA DE SÍMBOLOS

| | |
|----|-------------------|
| % | Porcentagem |
| °C | Graus centígrados |
| μ | Microlitro |
| cm | centímetro |

SUMÁRIO

| | |
|---|-----------|
| 1 INTRODUÇÃO | 4 |
| 2 REVISÃO DE LITERATURA..... | 6 |
| 2.1 A CULTURA DA SOJA | 6 |
| 2.2 FOTOPERÍODO DA SOJA E PERÍODO JUVENIL | 7 |
| 2.3 GRUPOS DE MATURIDADE RELATIVA DA SOJA..... | 8 |
| 2.4 GENES DA SÉRIE ALÉLICA E..... | 12 |
| 2.5 ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)..... | 14 |
| 3 MATERIAL E MÉTODOS..... | 15 |
| 3.1 MATERIAL VEGETAL | 15 |
| 3.2 AVALIAÇÕES FENOTÍPICAS | 16 |
| 3.3 AVALIAÇÕES GENOTÍPICAS..... | 16 |
| 3.4 ESTRUTURA POPULACIONAL E DESEQUILÍBRIO DE LIGAÇÃO | 17 |
| 3.5 ANÁLISE DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)..... | 18 |
| 3.6 GENES CANDIDATOS..... | 18 |
| 4 RESULTADOS E DISCUSSÕES | 20 |
| 5 CONCLUSÕES | 38 |
| 6 CONSIDERAÇÕES FINAIS | 39 |
| REFERÊNCIAS..... | 40 |

1 INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) é uma das culturas mais importantes econômica e nutricionalmente do mundo (SINGER et al., 2019), a qual fornece mais de um quarto de proteína mundial para ração animal e alimentos (GRAHAM, 2003). É uma das principais oleaginosas produzidas no Brasil e no mundo, sendo a principal *commodity* agrícola do país (ZANON et al., 2015; CONAB, 2019). O Brasil, na safra de 2019/20, foi responsável pela produção de 133,6 milhões de toneladas, em uma área plantada de aproximadamente 37,8 milhões de hectares (CONAB, 2020). Esse montante corresponde a cerca de 35% da produção mundial da cultura, que foi de 336,7 milhões de toneladas (FAS/USDA, 2020).

A soja é uma planta de dias curtos, sendo altamente sensível ao fotoperíodo (WATANABE et al. 2012). É cultivada em todo o mundo, em uma faixa ampla de latitude estendendo-se de 50° N a 22° S. No Brasil, o cultivo estava restrito do Rio Grande do Sul até o norte do estado do Paraná até 1970. Atualmente, apresenta cultivos abrangendo latitudes de 30° S a 2° S (ABRAHÃO e COSTA, 2018). Esta expansão foi devido ao desenvolvimento de cultivares adaptadas aos curtos fotoperíodo da região central e norte do Brasil (GAVIOLI, 2013).

No contexto adaptativo, os grupos de maturação relativa (GMR) foram sugeridos como um sistema para a classificação do ciclo total da cultura, em diferentes faixas de latitude nos EUA (SCOTT e ALDRICH, 1970), no Brasil (ALLIPRANDINI et al., 2009) e na Argentina (BAIGORRI, 2009). Em soja, os GMR variam de GM-000 (triplo zero) para as variedades de maturação precoces, até GM-10 para as variedades de grupo de maturação mais tardio (COBER et al., 2001). Para cada zona latitudinal, em seus respectivos países, há GMR com melhor adaptação ao fotoperíodo, este ajuste do GMR maximiza o rendimento de grãos (ZANON et al., 2016; ZDZIARSKI et al., 2018).

Há uma grande variabilidade de genes e QTLs (*Quantitative Trait Loci*) responsáveis pela resposta ao fotoperíodo ou indução floral da cultura da soja (WANG et al. 2019; LIU et al., 2020; RAVELOMBOLA et al., 2020). Esses genes podem sofrer variações e formar diferentes variações alélicas (XU et al. 2013). Até o presente momento, foram identificados 12 genes principais controlando a resposta ao

fotoperíodo/ciclo da soja: E1 e E2 (BERNARD 1971), E3 (BUZZELL 1971), E4 (BUZZEL & VOLDENG 1980), E5 (MCBLAIN & BERNARD 1987), E6 (BONATO & VELLO 1999), E7 (COBER et al., 2001), E8 (COBER et al. 2010), E9 (KONG et al. 2014), E10 (SAMANFAR et al. 2017), E11 (WANG et al.,2019) e J (RAY et al. 1995).

A herança quantitativa, além da grande influência dos fatores ambientais, dificulta a seleção de linhagens a partir dos GMR para as diferentes faixas de latitude. A partir deste panorama, há necessidade de investir em técnicas que possibilitem identificar GMR, através de marcadores moleculares. Dentre os marcadores utilizados em programas de seleção assistida por marcadores moleculares, os SNPs (Single Nucleotide Polymorphism), apresentam a vantagem de serem codominantes (possibilitam identificar indivíduos heterozigotos), são encontrados em maior abundância no genoma e seus resultados são rapidamente obtidos, apresentando, portanto, níveis de rendimento mais elevados (ZHANG et al., 2016; TRAN et al.,2019; VINHOLES et al., 2019).

Uma das etapas mais importantes no uso de marcadores moleculares é o estabelecimento da relação entre um dado marcador e um loco de interesse. Neste contexto, a análise de associação genômica ampla (GWAS) é uma estratégia estabelecida para a identificação de regiões genômicas com características de interesse, em diferentes espécies de plantas cultivadas como trigo (CHAURASIA et al., 2020; SAFDAR et al., 2020), arroz (LONG et al., 2020),milho (RICE et al., 2020) e soja (MENGARELLI & ZANOR, 2021) . A análise de GWAS tem como pressuposição o desequilíbrio de ligação (LD) entre pelo menos um marcador e uma mutação de efeito maior ou QTLs, responsáveis pela variação fenotípica observada (LI et al., 2015; TRAN et al., 2019; WANG et al., 2019).

Neste contexto, este estudo se propôs avaliar o controles genéticos e moleculares associados à GMR em soja (LIU et al., 2018; MAO et al., 2017), a partir de um painel de associação com um grande número de genótipos. Assim, o objetivo deste trabalho foi (i) identificar novas regiões genômicas (QTLs) e marcadores SNP associados a grupos de maturação relativa, (ii) identificar genes de resposta ao fotoperíodo/indução floral da soja em linhagens brasileiras argentinas e americanas, via análise de associação genômica ampla. (iii) Propor marcadores para seleção de GMR para diferentes regiões edafoclimáticas brasileiras.

2 REVISÃO DE LITERATURA

2.1 A CULTURA DA SOJA

A soja (*Glycine max* (L.) Merrill) é a oleaginosa mais produzida no mundo. Por ser rica em óleo e proteína, é utilizada tanto para consumo humano quanto animal. As produções da cultura em larga escala ficaram restrita a China e Japão até o final do século XIX, quando teve início a expansão do cultivo para outras regiões do mundo.

No Brasil, a soja chegou no ano de 1908, introduzida por imigrantes japoneses, entretanto, devido à incentivo e importância do café, inicialmente a cultura não teve espaço significativo no mercado. Somente ao final da década de 60 foram implementadas as primeiras lavouras em larga escala, que se estabeleceram na região sul do país, como uma opção de cultura de verão (MUNDSTOCK & THOMAS, 2005).

Os genótipos introduzidos no Brasil eram de base genética norte-americana, ou seja, não tinham adaptação a climas tropicais. Tais genótipos foram introduzidos na região do sul do país, em áreas semelhantes ao lugar de origem dos materiais (SILVA, et al., 2015). Após a década de 1970, os programas de melhoramento genético introduziram ao mercado brasileiro genótipos portadores de genes com período juvenil longo, onde, obteve-se a possibilidade para o cultivo em baixas latitudes, e a expansão da cultura para as demais regiões do Brasil (SILVA, et al., 2015).

Devido a essa expansão de cultivo, a área cultivada passou de 6 milhões de hectares em 1970, para os 37,8 milhões na safra de 2019/20. O Brasil é o maior produtor mundial de soja, com produção de 124,8 milhões de toneladas na safra 2019/2020, seguido dos EUA e da Argentina, com 96 e 49 milhões de toneladas, respectivamente (FAO, 2020; CONAB, 2020). No Brasil, os estados de Mato Grosso, Rio Grande do Sul, Paraná e Goiás são os principais produtores.

2.2 FOTOPERÍODO DA SOJA E PERÍODO JUVENIL

A indução ao florescimento na soja ocorre através da interação entre fotoperíodo e temperatura, sendo considerada uma planta de dias curtos facultativa (MOURTZINIS & CONLEY, 2017). Fotoperíodo maior que o fotoperíodo crítico (cerca de 13 horas) atrasam o florescimento e fotoperíodo menor que o crítico, favorecem o florescimento (ZHANG et al., 2017). Esta interação entre temperatura e resposta ao fotoperíodo é o fator mais importante na adaptação da soja (BANDILLO et al., 2017).

Esta interação e sensibilidade fotoperiódica varia de acordo com os genótipos, restringindo a adaptação de cada cultivar a uma faixa mais restrita de latitude (RODRIGUES et al., 2001; YANG et al., 2019). Em função desta característica, cultivares apresentaram reduções significativas em seus rendimentos quando elas foram cultivadas 2° N além de suas latitudes de cultivos específicas (GAI & WANG, 2001). As cultivares, quando levadas para menores latitudes, podem apresentar redução da área foliar, resultando em plantas mais baixas e menor produtividade (SEDIYAMA et al., 1972; KONG et al., 2018)

O uso da característica período juvenil longo (PJL) foi a solução encontrada pelos melhoristas de soja para retardar o florescimento em condições de dias curtos (KIIHL et al., 1985; NEUMAIER & JAMES, 1993). Cultivares com PJL permanecem em estado vegetativo por um período mínimo de dias, mesmo sob condições de diminuição do período de luz (dias curtos), garantindo condições mais adequadas de desenvolvimento foliar e estruturas antes de iniciar o período reprodutivo (LU et al., 2015), e dessa forma, gerar sementes em escala adequada para a produção agrícola (LI & LAM, 2020). Tais cultivares apresentam estímulos para florescimento de forma mais tardia, a partir da emissão de 5ª folha trifoliolada, mesmo sendo cultivada em épocas de dias curtos, apresentando altura e produtividade adequadas em diversas épocas de semeadura (SEDIYAMA, 2009).

O PJL é determinado por dois genes recessivos E6 e J, os quais podem sofrer alterações condicionando floração tardia quando estes genes estão na forma recessiva, este fato foi encontrado na descendência das cultivares Bossier e Paraná (CARPENTIERI PÍPOLO et al., 2002). Cultivares com PJL apresentam crescimento vegetativo suficientes mesmo sob condições de dias curtos (SUN et al., 2019), desta maneira, podem ser cultivadas em regiões de baixa latitude (LU et al., 2017). A correta

combinação dos genes da série alélica “E” e “J”, controlando o comprimento do período juvenil, permite que a soja maximize o rendimento, dentro de faixas latitudinais muito estreitas (CARTER et al., 2004).

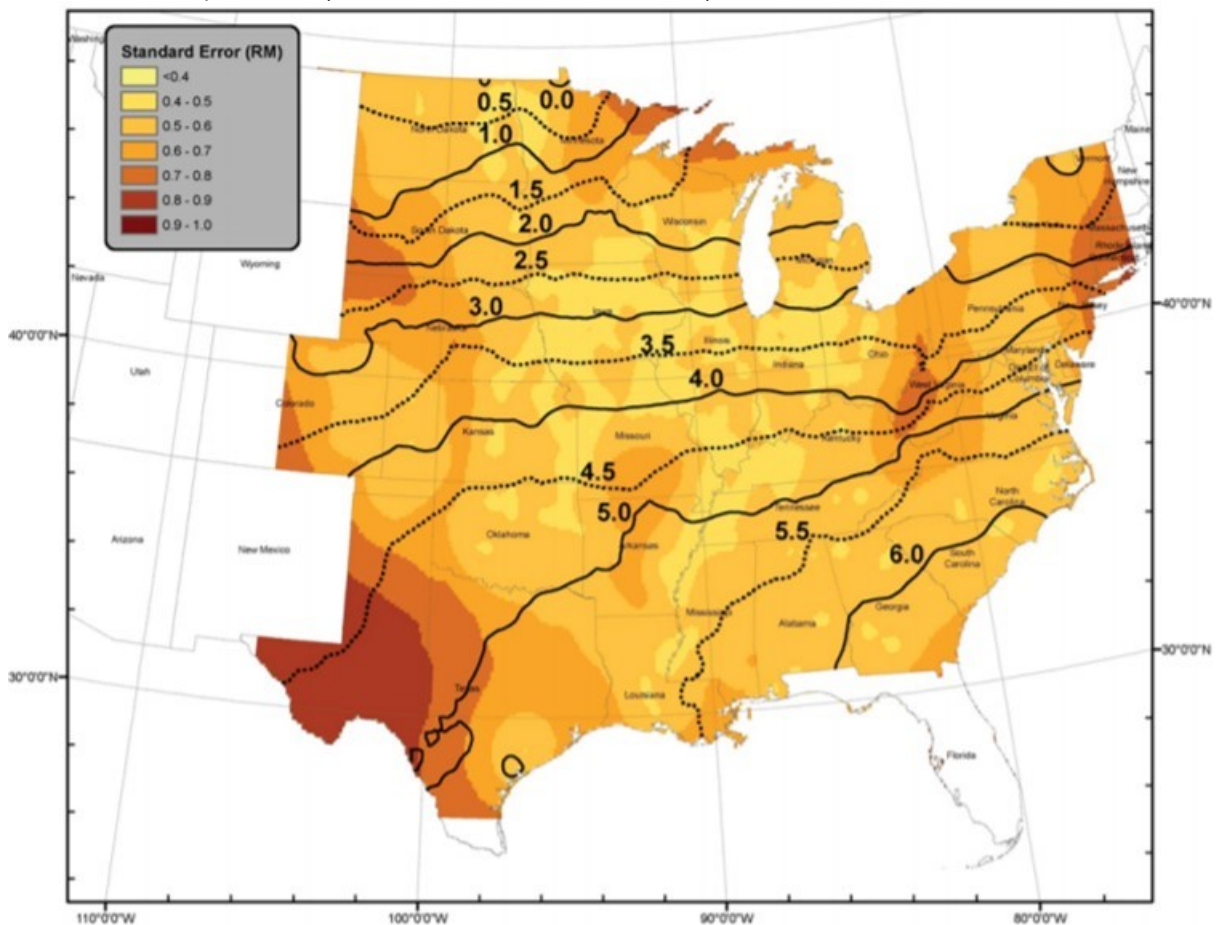
2.3 GRUPOS DE MATURIDADE RELATIVA DA SOJA

O ciclo da cultura da soja é classificado de acordo com o número de dias da emergência até a maturação, podendo variar de 77 a 200 dias, dependendo do local e época de semeadura (ALLIPRANDINI et al., 2009). Os termos tradicionalmente utilizados para classificar as cultivares de soja, com base no seu tempo de maturação, eram denominados de superprecoce, precoce, semiprecoce, médio e tardio. Entretanto, uma cultivar precoce em uma região específica poderia ser tardia em outra, e assim vice-versa. Desta maneira, para ajustar tais efeitos, ocorreu uma organização do ciclo das culturas em grupos de maturidade relativa (GMR), desenvolvido primeiramente nos Estados Unidos.

A dependência da cultura da soja na duração do dia, resultou em sua distribuição geográfica (GMR) (MILADINOVIC et al., 2018). Nos Estados Unidos, os GMRs da cultura da soja abrangem diferentes regiões, no total são 13 grupos que se iniciam pelas cultivares mais precoces 000 e se estendem até o grupo 6.

A diferença entre grupos de maturação em uma área pode ser causada por exigências fotoperiódicas de uma variedade, resultando em diferenças na data de maturidade, que pode ser de 10 a 18 dias, dependendo da latitude e do GMR em consideração (MILADINOVIC et al., 2006; JIA et al., 2014; SONG et al., 2019). Na maioria dos casos, as cultivares de cada grupo de maturação são adaptadas a uma faixa que pode variar de 160 a 240 km entre latitudes. Tal faixa de maturação dentro de cada grupo pode variar de 10 a 15 dias (SEDIYAMA, 2009). Nos EUA, Scott e Aldrich (1970), delinearam os GMRs hipotéticos com base no fotoperíodo. No entanto, mais recentemente, usando dados de experimentos de 1998 a 2003, Zhang et al (2007) redefiniram as zonas ótimas de GM. Esta nova definição foi atestada por Mourtzinis e Conley (2017) (Figura 1), com a indicação das localidades de melhor performance de cada GMR.

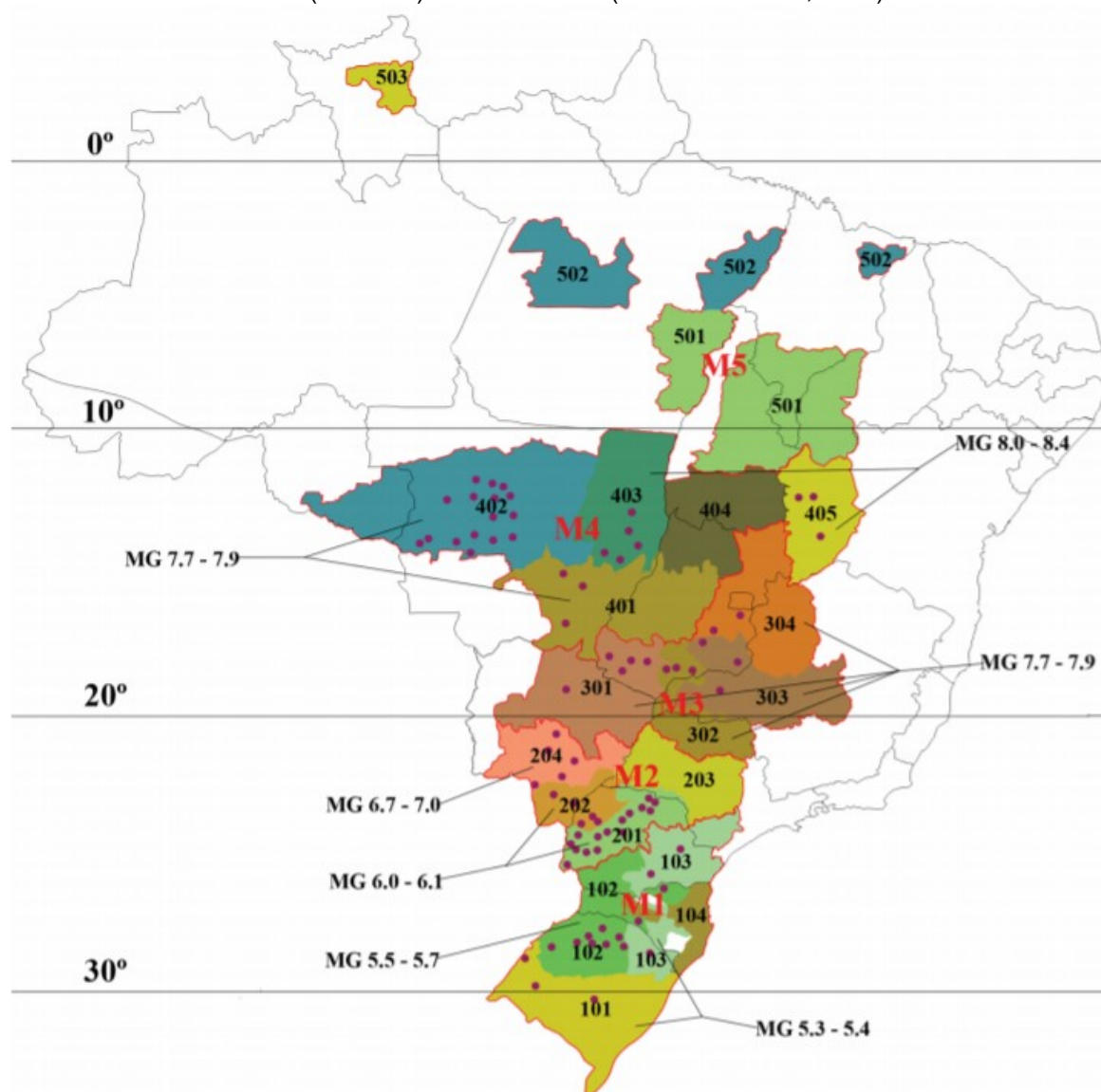
Figura 1 -- Mapa Americano com o melhor desempenho dos grupos de maturidade de soja nos EUA. Adaptado de: (MOURTZINIS E CONLEY 2017).



No Brasil, um sistema semelhante ao americano foi incorporado para a classificação de cultivares em GMs. Alliprandini et al (2009), realizaram os primeiros estudos sobre GMRs, utilizando um sistema que se assemelhava muito com o americano, através de comparação com testemunhas padronizadas de grupos de maturação conhecidos, e dessa forma permitiram a classificação dos genótipos de soja em GMRs em todo o país. Assim como nos EUA, as zonas ótimas de cada GMR estão sendo aprimoradas. De acordo com a terceira aproximação de Kaster e Farias (2012), o Brasil pode ser dividido em macrorregiões e dentro de cada macro são agrupadas as regiões edafoclimáticas de cultivo (REC). No Brasil, os GMs variam de 5 a 10. Para cultivo no Rio Grande do Sul o GM mais adequado é 5 chegando até GM 10 para cultivo em Roraima, portanto, quanto maior o número do GM, mais próximo a linha do Equador será sua região de adaptação (PENARIOL, 2000). Recentemente, Zdziarski et al. (2018) identificou o melhor desempenho de cultivares de distintos

grupos de maturação para todas as macro e microrregiões do Brasil (figura 2).

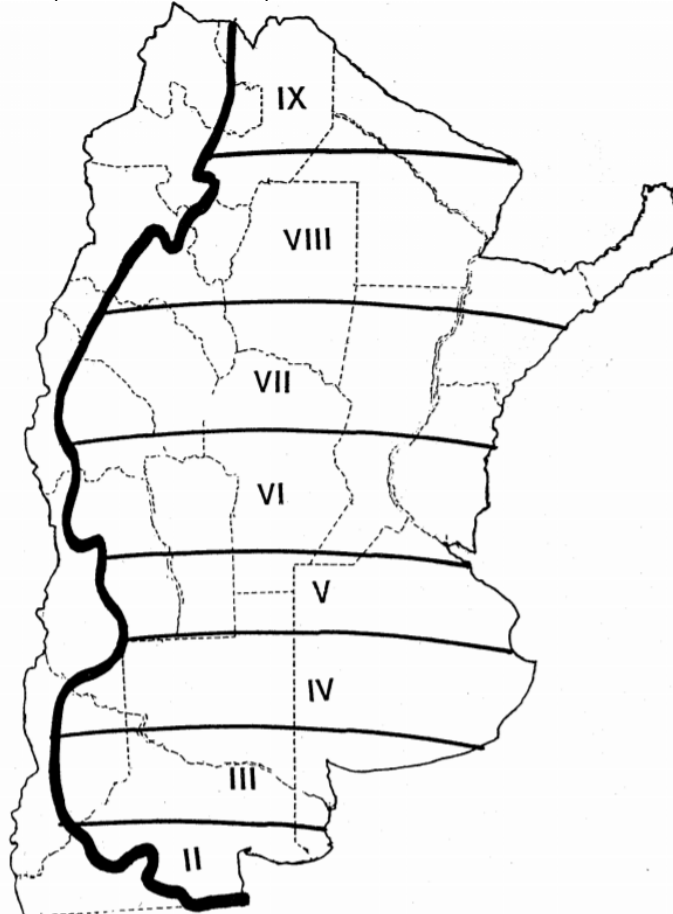
Figura 2 -- Mapa brasileiro com as localizações das macrorregiões (M1, M2, M3 e M4 e M5) e regiões edafoclimáticas (101–503). Modificado de: (ZDZIARSKI et al, 2018).



Na Argentina, os cultivos de soja abrangem uma ampla área latitudinal, variando de 23°S a 39°S, onde são cultivados genótipos de soja com GM-2.0 a GM-9.0 (DIAS et al., 2019). Assim como no Brasil, as primeiras variedades de soja registradas na Argentina eram originárias de base genética americana (ROSSI, 2012). Devido a isso, Baigorri (1997), propôs, o primeiro mapa de GMRs para soja na Argentina, com base nos cultivos americanos. Posteriormente, com o avanço do melhoramento, e a criação de um banco de germoplasma argentino, Baigorri (2009)

(Figura 3), aprimorou a classificação de cada GMR ideal para cada localidade, onde, os GMs 2.0 – 3.0 são cultivados, principalmente, ao norte da Patagônia ou ao sul de Buenos Aires, o GM-4.0 abrange quase toda a região Pampeana, e genótipos com GM- 6.0 a 9.0 se adaptam ao noroeste e nordeste da Argentina (DIAS et al., 2019).

Figura 3 Mapa Argentino com o melhor desempenho dos grupos de maturidade de soja na Argentina. Adaptado de: (BAIGORRI, 2009).



A soja possui grande adaptabilidade e grande diversidade de GMs, podendo ser cultivada em uma ampla faixa de latitudes (de 50° N a 35°S); este fato pode ser atribuído a variações naturais, grande diversidade e rica variabilidade genética nos principais genes (Série alélica “E”) e QTLs (Quantitative Trait Loci), responsáveis pela resposta ao fotoperíodo e que podem afetar a indução para floração e maturação (CAO et al. 2017). Além disso, os genes desta série alélica desempenham papéis distintos na floração e maturidade, sob diferentes fotoperíodos (JIANG et al., 2019).

2.4 GENES DA SÉRIE ALÉLICA E

Os genes da série alélica E determinam o número de dias da emergência a floração e ciclo total, possuindo diferentes papéis na maturidade e na sensibilidade da soja ao fotoperíodo (MILADINOVIC, 2018). Até o presente momento, 12 locus principais (série E1 a E11 e o locus J para “período juvenil longo”) foram identificados, determinando o número de dias da emergência a floração e ciclo total na soja (WANG, 2019). Quando os genes da série E apresentam-se em estado dominante, geralmente conferem florescimento e maturidade mais tardia, exceto os genes E6, E9 e J (BONATO & VELLO, 1999; KONG et al., 2014). Enquanto que, a sensibilidade do fotoperíodo é reduzida com maior quantidade de alelos recessivos.

Destes 12 principais loci, apenas E6, E9 e J foram identificados controlando a época de florescimento em condições de dias curtos (LI et al., 2017), os demais genes foram detectados em condições de dias longos (WANG et al., 2019). Além disso, foram identificadas e registradas no Soybase (04/03/2021) (<http://www.soybase.org>) um total de 313 QTLs para as características de floração e ciclo total em todos os 20 cromossomos

Destes genes, apenas E1, E2, E3, E4 E9, E10 e J foram identificados por mapeamento fino e abordagens de clonagem posicional ou genes candidatos (CAO et al., 2017; SAMANFAR, 2017). Tais genes estão localizados ao longo de todo o genoma da soja, em diferentes grupos de ligações moleculares (GLMs): O (Gm10) para E2 (AKKAYA et al. 1995, CREGAN et al. 1999), L (Gm19) para E3 (MOLNAR et al. 2003), I (Gm20) para E4 (ABE et al. 2003, MOLNAR et al. 2003), C2 (Gm06) para E1 e E7 (COBER & VOLDENG 2001a, MOLNAR et al. 2003) e C1 (Gm04) para E8 (COBER et al. 2010). Segundo Tsubokura et al. (2013), apesar da existência deste grande número de genes, os genes de maturidade E1-E4 contribuem com 62- 66% da variação para a época de florescimento em soja.

O gene E1 é considerado o de maior efeito na determinação da época de florescimento (UPADHYAY et al., 1994). Foi descrito no cromossomo 6 e identificado como um fator de transcrição das leguminosas em geral (XIA et al., 2012), sendo que o mesmo possui alelos funcionais e apresenta também variações não funcionais (XIA et al., 2012; MILADINOVIC et al., 2018). O gene E2 está localizado no cromossomo 10, sendo um ortólogo do gene GIGANTEA, pertencente a Arabidopsis.

Este gene é funcional quando apresenta a forma dominante E2, enquanto que, a sua forma recessiva e2 apresenta-se não funcional e com um genótipo diferente, devido a uma mutação sem sentido, de forma que, este controle sob o florescimento ocorre através de regulação do gene GmFT2a, atuando principalmente em condições de dias curtos (WATANABE et al., 2011).

Os genes E3 (GmPhyA3) e E4 (GmPhyA2) foram descritos nos cromossomos 19 e 20 (LIU et al., 2008; WATANABE et al., 2009). O gene E3 tem resposta em uma ampla gama de latitudes, enquanto que, o gene E4 é expresso somente em ambientes de altas latitudes (LU et al., 2015). Os genes E1 e E2 tem impacto significativo no período de desenvolvimento pré-reprodutivo, enquanto que os genes E3 e E4 podem afetar o período reprodutivo e pós floração, aumentando assim, a duração do período de diferenciação do número de vagens e enchimento das vagens (XU et al., 2013).

Os genes E6 e J estão envolvidos no controle da época de floração tardia sob dias curtos. Alelos dominantes nos dois genes resultam em floração antecipada (RAY et al., 1995; BONATO & VELLO, 1999). O locus J foi identificado como um ortólogo de *Arabidopsis thaliana* EARLY FLOWERING 3 (ELF3), controlando a indução a floração, através da capacidade de regulação da expressão de E1 (LU et al., 2017). Os genes E7 e E8 foram descritos em cultivares de soja de maturação precoce, respondendo a condições de dias longos (COBER & VOLDENG, 2001; COBER & MORRISON, 2010). O gene E7 está ligado ao gene E1 no cromossomo 6 (MOLNAR et al., 2003), e o E8 está localizado no cromossomo 4 (COBER et al., 2010). Os genes responsáveis por E6-E8 e J não foram isolados geneticamente até o momento (CAO et al., 2017).

O gene E9 foi identificado em populações segregantes a partir de um cruzamento entre soja selvagem (*Glycine soja*) e soja cultivada (*Glycine max*) (LIU et al., 2007; KONG et al., 2014). Os genes E9 (GmFT2a) e E10 (GmFT4) são homólogos ao gene FT (Flowering Time) de *Arabidopsis* (ZHAI et al., 2014; ZHAO et al., 2016). O gene E11 está localizado no cromossomo 7, semelhante ao gene E9, de tal forma que o alelo dominante promove o florescimento antecipado sob condições de dias longos (WANG et al., 2019).

2.5 ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)

Os estudos de associação genômica ampla (GWAS - Genome Wide Association Study) visam detectar SNPs, com alto desequilíbrio de ligação com o verdadeiro locus da mutação causal, podendo o marcador ser integrante do QTL ou ser a própria mutação causadora do fenótipo (ZHANG et al., 2019). Com isso, as análises GWAS possibilitam a identificação de correlações entre um loco marcador e uma característica fenotípica do caráter (BARZANOVA & NOVIKOVA, 2019).

A técnica GWAS baseia-se no desequilíbrio de ligação (DL) existente nos genomas, pois, para explicar toda a variação genética de uma característica, é necessário que se tenha um DL entre alelos dos marcadores e alelos dos genes que controlam o mesmo. Se a população estiver em DL, a associação entre um marcador com um alelo que confere resistência a uma característica fenotípica, será mantida (YOSHIDA et al., 2019). A capacidade da análise GWAS de melhor detectar uma característica fenotípica aliada à sua expressão genotípica, depende basicamente do contraste das avaliações fenotípicas na população, onde, quanto mais materiais tiverem e quanto mais contrastantes forem estes materiais, maior será a eficácia da análise (WANG et al., 2018).

Quando comparada com as análises de linhagem tradicionais, o estudo de GWAS aproveita todos os eventos recombinantes históricos que ocorreram em populações naturais (WANG et al., 2018). Além disso, permite prever ou identificar genes causais com maior facilidade (ZHANG et al., 2015a). Além da soja, o GWAS tem sido amplamente utilizado a culturas como milho (HAN et al., 2020) e arroz (LIU et al., 2020; MA et al., 2016).

Em soja, a técnica GWAS foi estudada em diversas características, tais como: fotossíntese e condutância estomática (WANG et al., 2020); resistência a nematóide do cisto (RAVELOMBOLA et al., 2020); conteúdo de ácido oleico (LIU et al., 2020); número de vagens (SONG et al., 2020); peso de sementes (QI et al., 2020) e teores de óleo e proteína (LI et al., 2019). Em relação aos grupos de maturidade e características relacionadas a sensibilidade ao fotoperíodo, há estudos para as características dias para a floração e dias para maturidade (ZHANG et al., 2015b; CONTRERAS-SOTO et al., 2017; MAO et al., 2017; PAN et al., 2018; LI et al., 2020; KIM et al., 2020; LU et al., 2020).

3 MATERIAL E MÉTODOS

3.1 MATERIAL VEGETAL

Um conjunto de 5428 linhagens homocigotas de soja, pertencentes ao programa de melhoramento genético da GDM Seeds, foi utilizado neste estudo. As linhagens foram conduzidas em formato de experimentos em blocos ao acaso, durante as safras de 2018 e 2019, sendo estes experimentos divididos em 3 países (Argentina, Brasil e Estados Unidos).

Foram utilizadas linhagens de grupos de maturidade precoce ao tardio (GM 0.5 a 8.5), de modo que, cada linhagem foi conduzida e avaliada no país respectivo onde foi desenvolvida. Desta maneira, foram avaliadas 2663 linhagens na Argentina, 1093 linhagens no Brasil e 1672 linhagens nos Estados Unidos (Figura 4). Além disso, os experimentos em cada país estavam divididos em grupos de maturação, de acordo com as zonas edafoclimáticas de melhor desempenho de cada material.

Figura 4 -- Locais de avaliação genotípica e fenotípica das linhagens de soja (UTFPR- Campus Pato Branco, 2020). UTFPR, *Campus Pato Branco*, 2020.



3.2 AVALIAÇÕES FENOTÍPICAS

As linhagens de soja foram classificadas fenotipicamente quanto ao grupo de maturidade (GM). Foi determinada a maturidade quando 95 % das vagens apresentaram maturidade fisiológica (R8) (FEHR E CAVINESS, 1977). Para a determinação do grupo de maturidade de cada linhagem, foram usados materiais com grupos de maturidade conhecidos, a fim de comparação entre ciclos, denominados de checks. De tal maneira que, o grupo triplo zero (GM-000), foi considerado o mais precoce e o grupo de maturação 10 (GM-10) foi considerado o mais tardio, de acordo com (COBER et al., 2001).

3.3 AVALIAÇÕES GENOTÍPICAS

O isolamento do DNA foi realizado a partir de um fragmento de semente. Após a amostragem, as placas foram armazenadas em ultrafreezer a -30°C até o momento da liofilização. A liofilização foi realizada em um liofilizador modelo LP820 – LIOTOP. Ao final do processo, as placas foram retiradas do equipamento e imediatamente fechadas com tampa de silicone, para posterior maceração das amostras. A maceração foi realizada de forma semi-automatizada, com o uso de um dispensador modelo 5G mixer - Fluid Management, que utiliza esferas metálicas para macerar as amostras dentro de cada compartimento da placa.

A extração de DNA foi realizada seguindo metodologias da AgriSeq. Os ensaios de genotipagem foram testados em uma placa de reação denominada MicroAmp™ EnduraPlate™ Optical 384 – Well Clear, com configurações de 10 μL (7 μL da mistura de amplificação e 3 μL de DNA normalizado).

Após isso foi realizado a amplificação das amostras através de um termociclador, onde foi utilizado o padrão de ciclos térmicos de PCR. A primeira etapa foi realizada em 99°C durante 2 minutos; seguido de treze ciclos Touchdown (99°C durante 15 segundos; 60°C durante 8 minutos). O ciclo final foi realizado em temperatura de 10°C por 60s (Collins et al., 2018). Em seguida foi preparada e misturada através da pipetagem a reação enzimática de pré-ligação em um tubo de microcentrifuga. Após o preparo das amostras, as mesmas foram adicionadas a placa

de 384 poços contendo 5 µL de cada poço de uma única coluna da placa de reação de 384 poços(16 amostras) em um único poço de um tubo de 8 tiras.

As bibliotecas de DNA foram preparadas usando enzimas de restrição seguindo protocolos de GBS. Posteriormente, o sequenciamento de DNA foi realizado usando o método de GBS (ELSHIRE et al., 2011). O sequenciamento de dupla extremidade foi realizado em cada genótipo de soja usando o protocolo GBS por um Illumina HiSeq. As leituras curtas foram alinhadas à sequência do genoma completo da soja (Wm82.a1.v1).

3.4 ESTRUTURA POPULACIONAL E DESEQUILÍBRIO DE LIGAÇÃO

A estrutura populacional foi realizada por meio da abordagem de *Principal Component Analysis* (PCA), através da matriz de marcadores moleculares, implementado no pacote GAPIT (LIPKA et al., 2012), no programa R (R Core Team, 2020).

O desequilíbrio de ligação (LD) foi utilizado estatística r^2 dos marcadores SNPs, através do pacote GAPIT-R (LIPKA et al., 2012). Posteriormente, para analisar o decaimento do LD, foram plotados os valores de r^2 vs as distâncias físicas entre os pares de marcadores.

A análise de QTL, na qual a ligação entre as características mendelianas (característica x marcador) foi realizada por meio do escore LOD (logaritmo das probabilidades). Este teste foi usado para estimar a frequência de recombinação entre dois locus (BHERING et al., 2008). O critério de ligação foi a razão de probabilidade LOD (logarithm of odds) score > 3,0 (lod threshold), sendo baseado em 1000 permutações. O cálculo do LOD foi realizado com base na seguinte equação:

$$LOD = \frac{(1 - \theta)^{NR} \times \theta^R}{0,5^{(NR+R)}}$$

Onde: NR representa o número de descendentes não recombinantes, e R o número de descendentes recombinantes. A análise será realizada utilizando pacote GAPIT-R (LIPKA et al., 2012).

3.5 ANÁLISE DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)

As associações entre dados genotípicos e fenotípicos de valores únicos (valores estimados por *Best Linear Unbiased Prediction* - BLUP) foram analisados utilizando o modelo *Multi-locus Mixed Model* (MLMM), proposto por Segura et al. (2012). Este método é baseado na inclusão (a cada etapa) do SNP com o menor valor- p como um regressor fixo em um modelo que contém um efeito poligênico aleatório, como no modelo clássico de GWAS de Yu et al. (2006). As variâncias poligênicas e residuais são reavaliadas em cada etapa, e uma nova varredura do genoma restante é realizada. Quanto mais integrados os regressores no modelo, menor é a variância atribuída ao termo poligênico aleatório. A análise de seleção direta para quando a proporção da variância explicada por esse efeito poligênico é próxima de zero. Este modelo foi utilizado para encontrar regiões cromossômicas ligadas aos grupos de maturação da soja.

Os resultados das análises de GWAS para cada grupo de maturação foram apresentados em gráficos denominados *Quantil-Quantil* (Q-Q plots) e *Manhattan Plots*. Os Q-Q plots foram utilizados a fim de visualizar os desvios dos logaritmos negativos dos p-valores obtidos na análise versus os valores esperados sob a hipótese de nulidade. Os Manhattan Plots foram plotados utilizando valores de $-\log_{10}(p)$ para cada SNP vs. a posição no genoma, para cada grupo de maturação. O gráfico de Manhattan consiste em um gráfico de dispersão em que o eixo Y apresenta os logaritmos negativos dos p-valores obtidos pela associação genômica ampla e o eixo X representa a posição genômica de cada SNP (por cromossomo) (LIPKA et al., 2012).

3.6 GENES CANDIDATOS

A partir das informações obtidas, foi realizada a busca por genes e regiões genômicas (QTLs) próximas aos marcadores SNPs significativos, a partir das sequências disponibilizadas no banco de dados Phytozome (<http://phytozome.jgi.doe.gov>) e SoyBase (<http://www.soybase.org>). A predição de tais genes candidatos foi referida seguindo as preferências: i) genes de função conhecida

em soja relacionados à característica em estudo e ii) genes com ortólogos de função conhecida em Arabidopsis relacionados à característica em estudo.

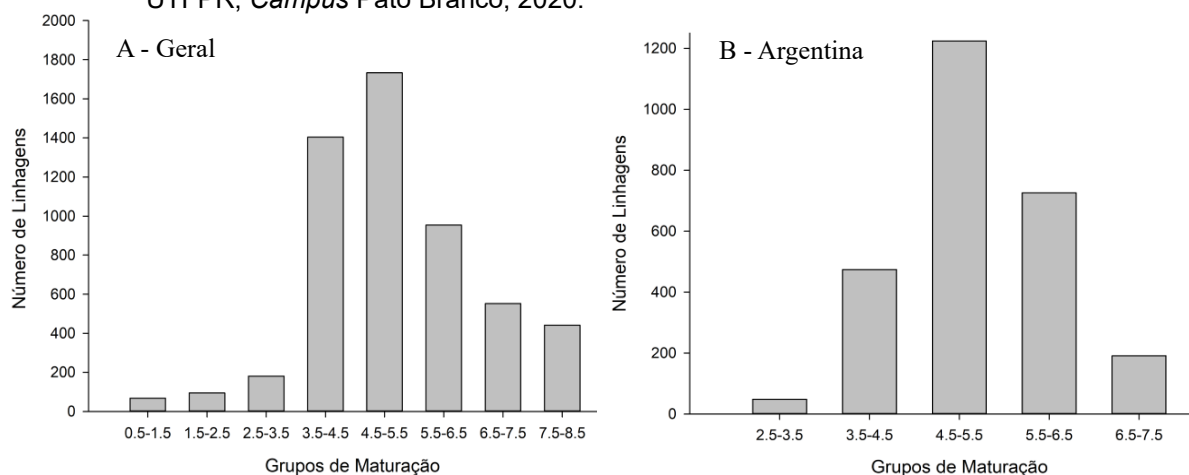
4 RESULTADOS E DISCUSSÃO

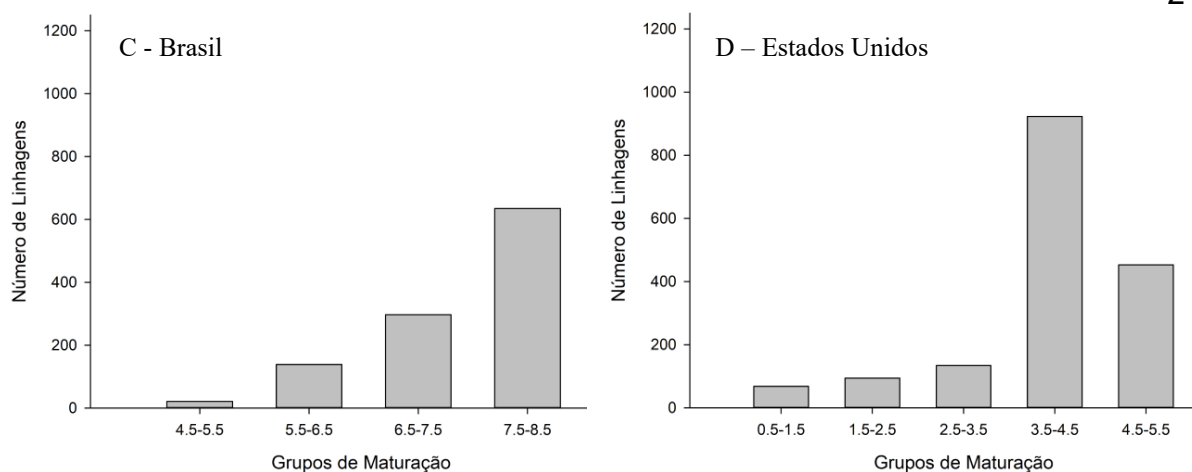
O conjunto de 5428 linhagens de soja avaliadas, apresentou distribuição variada entre os grupos de maturação, variando de 0.5 a 8.5 (Figura 5a). Os grupos de maturação mais precoce, entre 0.5 a 3.5, evidenciaram o menor número de linhagens, correspondendo a 6,5% do conjunto de linhagens avaliadas. A maioria das linhagens foram classificadas entre os GM de 3.5 a 6.5, com 75% das linhagens. Os GM tardios (de 6.5 a 8.5) reuniram 18,5% dos genótipos avaliados.

Considerando os países nos quais as linhagens foram desenvolvidas e avaliadas, na Argentina, os genótipos avaliados apresentaram GM variando de 2.5 a 7.5, demonstrando grande amplitude de ciclo, observando-se a maior concentração de genótipos entre os grupos 4.5 a 5.5 (45%) (Figura 5b). Por outro lado, as linhagens Brasileiras apresentaram grupos de maturação a partir de 4.5 e se estenderam até grupos de 8.5, apresentando a maior concentração de genótipos entre os grupos de 7.5 a 8.5 (58%) (Figura 5c). A distribuição das linhagens de acordo com o GM do Brasil e da Argentina apresentaram similaridade, devido à proximidade entre os dois países e a semelhança de latitude.

Os genótipos Norte americanos apresentaram grupos de maturação variando de 0.5 a 5.5 (Figura – 5d). Maior número de genótipos foi observado entre os grupos de 3.5 a 4.5 (55%), devido latitude dos cultivos americanos, que variam de 30° a 45° N (MOURTZINIS & CONLEY, 2017).

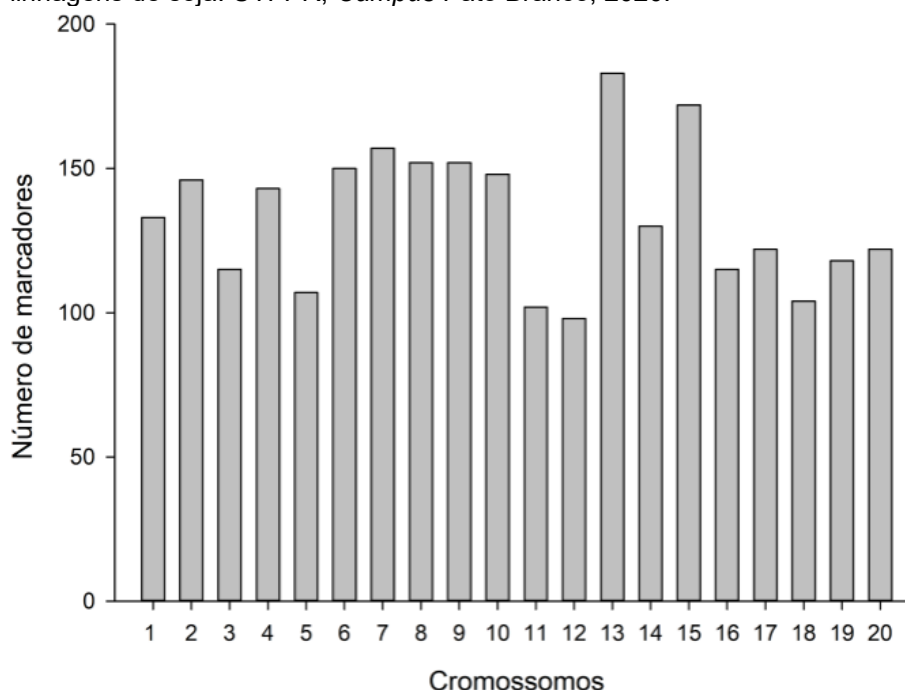
Figura 5 – Distribuição da frequência do conjunto de linhagens de soja para grupos de maturação. Conjunto de linhagens (A) geral, (B) argentinas (C) brasileiras, (D) norte americanas. UTFPR, *Campus Pato Branco*, 2020.





Um conjunto de dados de 2669 SNPs, cobrindo todos os 20 cromossomos, foi utilizado para a realização das análises GWAS. O maior número de SNPs foi identificado no cromossomo 13 (183 SNPs), seguido do 15 (172 SNPs) e do cromossomo 7 (157 SNPs) (Figura 6).

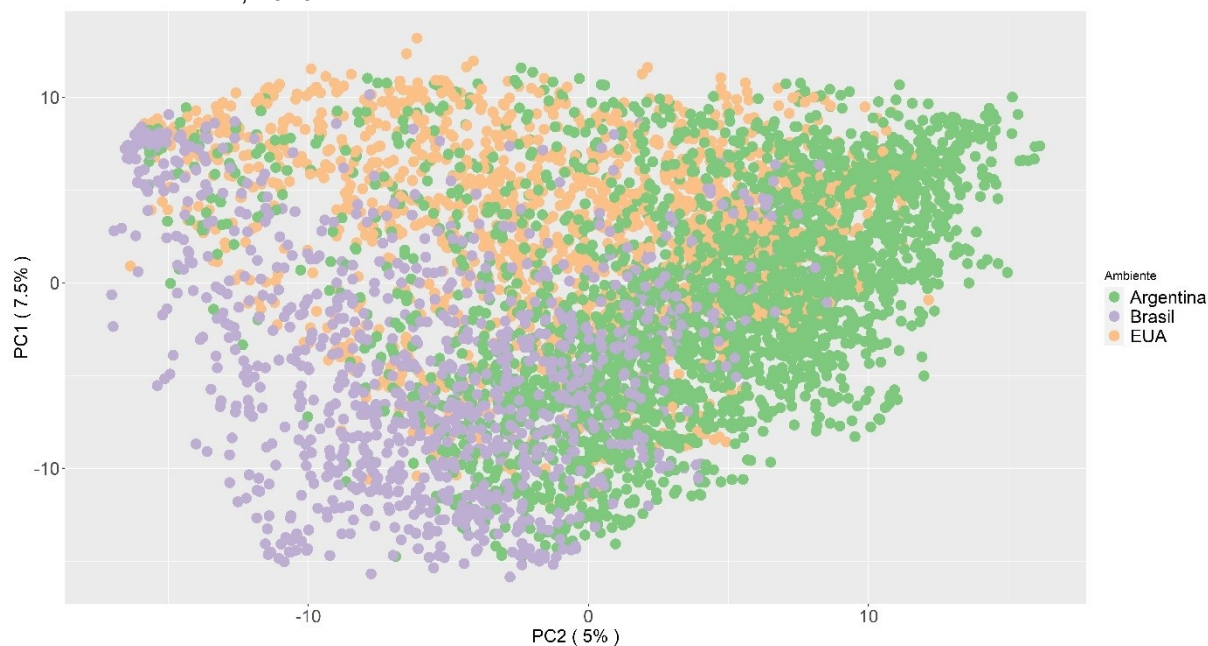
Figura 6 – Distribuição dos marcadores SNPs em 20 cromossomos de soja, avaliados em 5428 linhagens de soja. UTFPR, *Campus Pato Branco*, 2020.



A análise populacional, com o conjunto geral de 5248 linhagens, revelou formação de apenas um grupo, ressaltando não haver subpopulações dentro do conjunto (Figura 7) e nem dentro de cada país. Os dois primeiros componentes

principais explicaram 7.5 e 5%, da variância genética nas linhagens analisadas em conjunto, argentinas, brasileiras e norte americanas, respectivamente, e foram incluídos na análise GWAS como covariáveis, juntamente com a matriz de parentesco. Apesar do isolamento geográfico realizado, não se tornou necessária, a diferenciação genética dos grupos de maturação. Entretanto, Li et al. (2019) em estudo de associação genômica com uma coleção de 278 acessos de soja da China e Estados Unidos, concluíram que o isolamento geográfico entre as linhagens foi importante para a diferenciação genética das linhagens de soja, devido ao baixo parentesco genético entre os indivíduos perante os dois países.

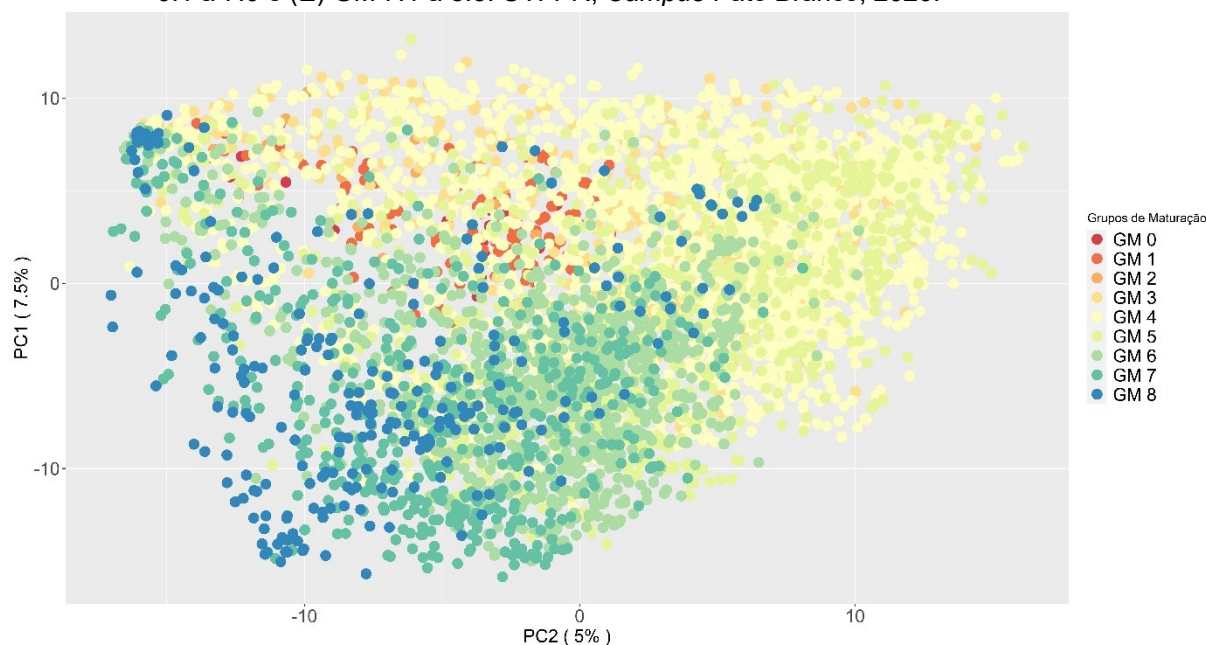
Figura 7 – Análise de componentes principais dos três primeiros componentes de 5428 linhagens. Conjunto de linhagens argentinas, brasileiras e norte americanas. UTFPR, *Campus Pato Branco*, 2020.



As linhagens foram agrupadas de acordo com a região de adaptação e com base nos grupos de maturação, em seus respectivos países, de acordo com classificação proposta por Baigorri (2009), Zdziarski et al (2018) e Mourtzinis & Conley (2017), na Argentina, Brasil e Estados Unidos, respectivamente. Assim, evidenciou-se a formação de grupos na análise populacional das linhagens, onde, observou-se diferenciação genética entre os grupos de maturação (Figura 8), principalmente entre os GM 0 a 4 e GM 5 a 8. Desta maneira, pode se afirmar que os grupos de maturação são diferentes geneticamente entre si. Estes resultados corroboram com Wang et al.

(2018), onde os autores revelaram estrutura populacional divergente entre linhagens de soja provenientes do Japão, da América do Norte e da região central da China para tempo de floração e maturidade.

Figura 8 -- Análise de componentes principais (PCA) dos três primeiros componentes de 5428 linhagens. Conjunto de linhagens (A) GM 0.3 a 3.5, (B) GM 3.6 a 5.0, (C) 5.1 a 6.0, (D) GM 6.1 a 7.0 e (E) GM 7.1 a 8.5. UTFPR, Campus Pato Branco, 2020.

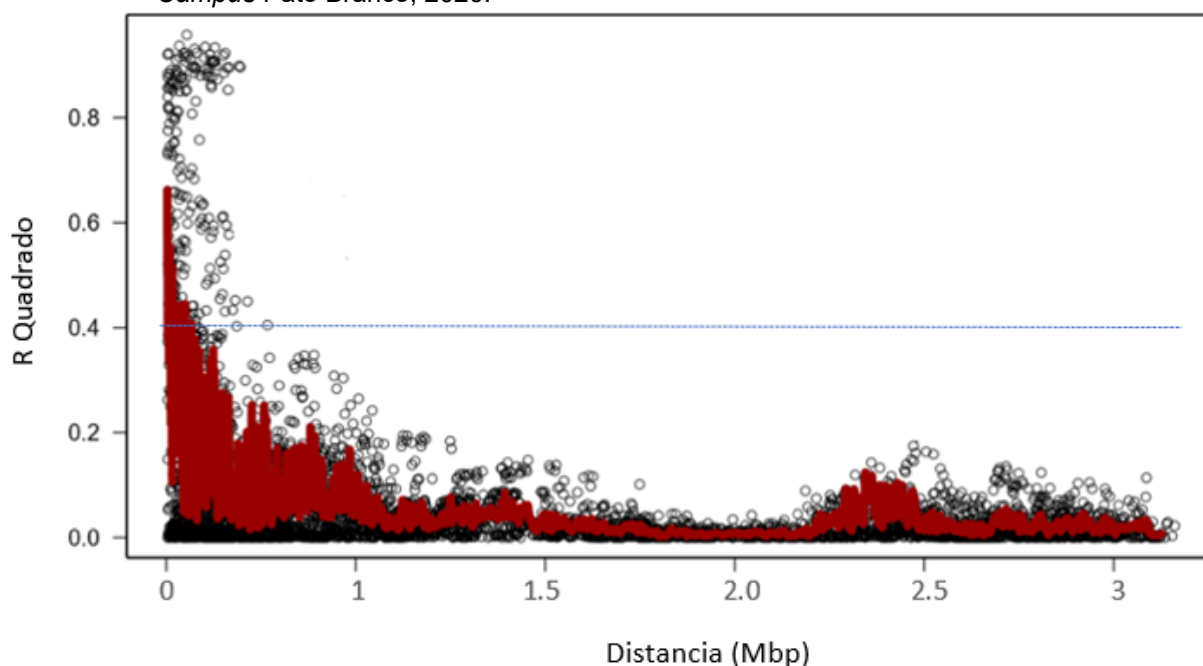


A análise de desequilíbrio de ligação (LD) da população, foi mensurada a partir do coeficiente de correlação ao quadrado r^2 (Figura 9). A extensão do LD considerando os 2669 marcadores SNPs foi de aproximadamente 550Kb quando r^2 atingiu 0,4. Outros estudos em soja têm relatado distâncias de decaimento entre 250 a 375 kb (WEN et al., 2014); 300 kb (LI et al., 2019; MAO et al., 2017); 447 kb (WANG et al., 2019) e 574 kb (HYTEN et al., 2007).

De maneira geral, a frequente seleção dos materiais e os cruzamentos com restrito número de cultivares nos programas de melhoramento, levam a redução da diversidade genética, podendo aumentar os padrões de LD (CONTRERAS-SOTO et al., 2017). Além disso, o sistema reprodutivo exerce influência sobre o desequilíbrio de ligação, de modo que espécies autógamas, recorrentemente, apresentam uma taxa de recombinação baixa, apresentando uma tendência a ter polimorfismos genéticos correlacionados e desequilíbrios extensos (GAUT & LONG, 2003). Song et al. (2015), estudando cultivares americanas, landraces e acessos de soja selvagem,

demonstrou que as populações que haviam passado por seleção natural e/ou artificial, apresentavam regiões extensas de desequilíbrio de ligação, sendo este, um dos indicadores da redução da diversidade genética.

Figura 9 – Decaimento do desequilíbrio de ligação em todo o genoma das 5428 linhagens de soja determinado pelas medidas de desequilíbrio de ligação (LD) contra a distância entre marcadores SNPs: r^2 vs as distancias físicas entre os pares de marcadores. UTFPR, Campus Pato Branco, 2020.



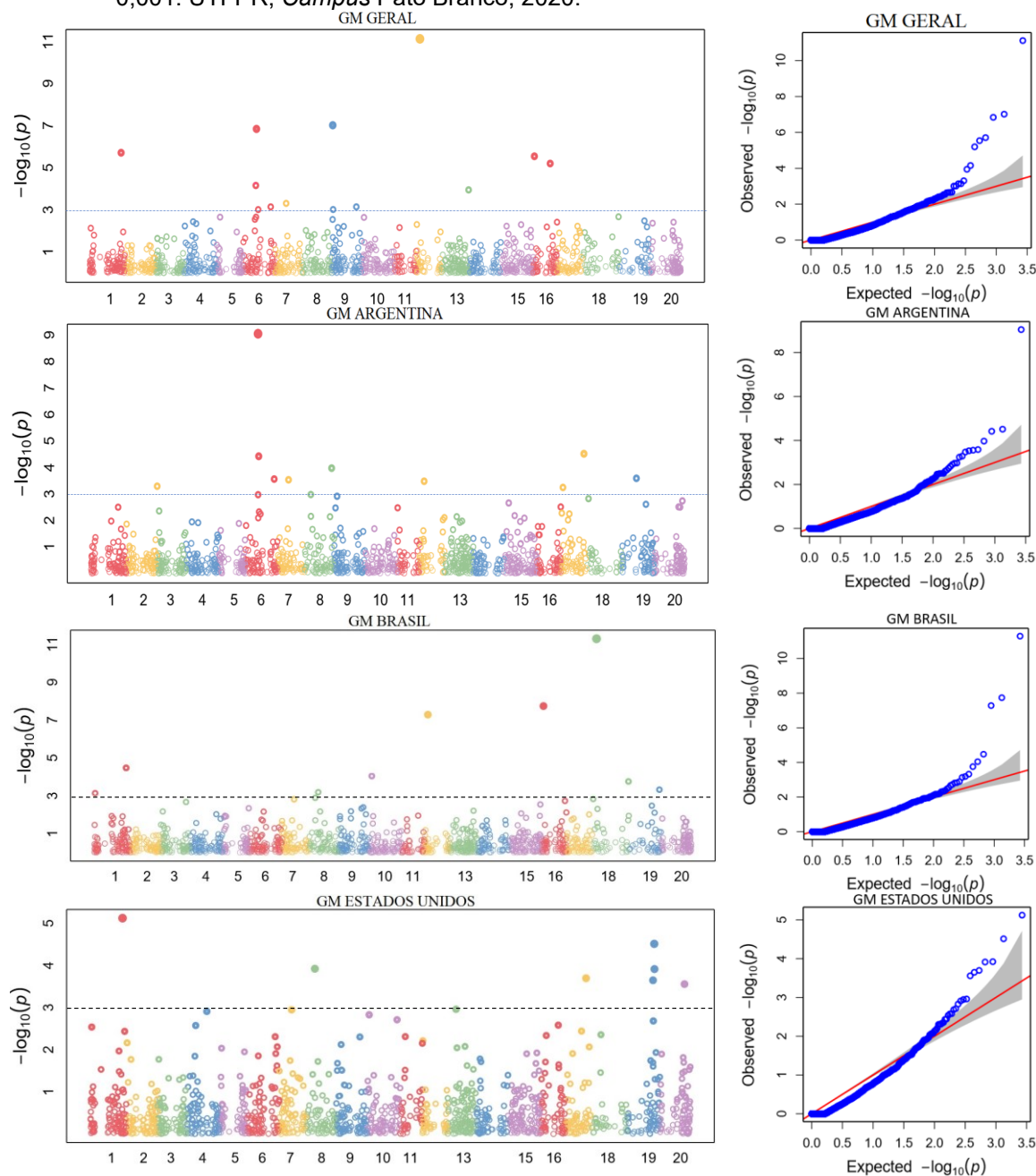
A análise GWAS foi realizada usando valores de BLUPs de todos os materiais avaliados, sob todos os ambientes, através do modelo denominado de *Multi-locus Mixed Model* (MLMM) (SEGURA et al., 2012). Este modelo utiliza os marcadores mais significativos como cofatores e tem demonstrado sucesso em estudos de mapeamento de associação (SAUVAGE et al., 2014; VAUGHN et al., 2014). No total, foram identificados 13 SNPs significativos ($p < 0,001$), associados a todos os grupos de maturação presentes nas análises. Quando tais materiais foram separados de acordo com o país de cultivo estes números foram diferentes, sendo, 10 SNPs associados aos materiais argentinos, 9 SNPs associados aos materiais brasileiros e 7 SNPs associados aos materiais americanos, localizados em diferentes cromossomos (Tabela 1). Além disso, os gráficos *QQ-plots* do modelo MLMM apresentaram um desvio de médio a acentuado na distribuição do valor de p -esperado na cauda, indicando que as associações falso-positivas foram

medianamente controladas pelo modelo MLM (Figura 10).

Tabela 1 - Marcadores SNPs significativos na análise de associação genômica para grupos de maturação e países em soja, com a identificação dos alelos, cromossomos, posição, valor $-\text{Log}_{10}(p)$ e R^2 e MAF. UTFPR, *Campus Pato Branco*, 2020.

| AMBIENTE | SNP | ALELOS | CROMOSSOMO | POSIÇÃO | LOD | R^2 | MAF |
|--------------|--------------|--------|------------|----------|-------|-------|------|
| GERAL | S12_5724762 | G/A | 12 | 5724762 | 11.11 | 52,9 | 0.05 |
| | S1_391761090 | A/G | 9 | 2062664 | 7.05 | 52,7 | 0.11 |
| | S1_267115233 | G/T | 6 | 21302179 | 6.8 | 53,1 | 0.36 |
| | S1_50403605 | C/A | 1 | 50403605 | 5.7 | 53,1 | 0.09 |
| | S16_1855885 | A/G | 16 | 1855885 | 5.5 | 52,9 | 0.07 |
| | S16_26910139 | T/A | 16 | 26910139 | 5.2 | 53,2 | 0.06 |
| | S1_266051587 | G/A | 6 | 20238533 | 4.1 | 53 | 0.36 |
| | S13_43326367 | G/A | 13 | 43326367 | 3.9 | 53 | 0.05 |
| | S1_314280574 | C/T | 7 | 17050934 | 3.3 | 53 | 0.28 |
| | S9_39492914 | C/A | 9 | 39492914 | 3.15 | 52,9 | 0.05 |
| | S1_289700405 | C/G | 6 | 43887351 | 3.14 | 52,7 | 0.29 |
| | S1_392321573 | G/A | 9 | 2623147 | 3.03 | 52,5 | 0.20 |
| | S1_269853445 | A/G | 6 | 24040391 | 3 | 52,5 | 0.35 |
| ARGENTINA | S1_266051587 | G/A | 6 | 20238533 | 9 | 75,5 | 0.30 |
| | S17_36601655 | C/G | 17 | 36601655 | 4.5 | 75,3 | 0.02 |
| | S1_267115233 | G/T | 6 | 21302179 | 4.4 | 75,2 | 0.29 |
| | S1_384948204 | C/G | 8 | 43087818 | 3.97 | 74,9 | 0.12 |
| | S1_872902103 | C/G | 19 | 22368015 | 3.6 | 74,5 | 0.48 |
| | S1_291549492 | G/A | 6 | 45736438 | 3.5 | 74,5 | 0.49 |
| | S1_314280574 | C/T | 7 | 17050934 | 3.5 | 74,1 | 0.23 |
| | S1_533812124 | C/T | 12 | 7589869 | 3.4 | 74 | 0.32 |
| | S2_48352600 | T/A | 2 | 48352600 | 3.3 | 73,9 | 0.10 |
| S1_754358518 | A/T | 17 | 3484738 | 3.2 | 73,9 | 0.19 | |
| BRASIL | S18_6728416 | T/C | 18 | 6728416 | 11.3 | 59,3 | 0.15 |
| | S16_1855885 | A/G | 16 | 1855885 | 7.7 | 59,0 | 0.09 |
| | S12_5724762 | G/A | 12 | 5724762 | 7.3 | 58,9 | 0.11 |
| | S1_50403605 | C/A | 1 | 50403605 | 4.4 | 57,8 | 0.15 |
| | S10_3816355 | A/G | 10 | 3816355 | 4.2 | 57,8 | 0.06 |
| | S18_56840751 | A/T | 18 | 56840751 | 3.7 | 57,2 | 0.06 |
| | S19_48070358 | T/C | 19 | 48070358 | 3.3 | 57,0 | 0.26 |
| | S8_14959118 | C/T | 8 | 14959118 | 3.2 | 56,9 | 0.07 |
| | S1_1701388 | G/A | 1 | 1701388 | 3.1 | 56,8 | 0.18 |
| EUA | S1_50403605 | C/A | 1 | 50403605 | 5.1 | 30,3 | 0.09 |
| | S1_892424673 | G/A | 19 | 41890585 | 4.5 | 30,0 | 0.09 |
| | S1_355472148 | G/A | 8 | 13611762 | 4 | 29,6 | 0.07 |
| | S1_892796220 | A/G | 19 | 42262132 | 3.9 | 29,6 | 0.12 |
| | S17_32451098 | T/C | 17 | 32451098 | 3.7 | 28,3 | 0.09 |
| | S1_890511653 | C/G | 19 | 39977565 | 3.6 | 27,5 | 0.08 |
| S1_939889624 | A/G | 20 | 38608520 | 3.5 | 27,3 | 0.14 | |

Figura 10- Manhattan Plot dos valores de $-\log_{10}(P)$ obtidos para cada SNP vs a posição no cromossomo para grupos de maturação (GM) geral e para Argentina, Brasil e Estados Unidos em soja. A linha paralela indica valor de LOD score = 3, correspondendo ao limite de p-value < 0,001. UTFPR, *Campus Pato Branco*, 2020.



Os SNPs significativos associados a maturidade foram encontrados em diferentes cromossomos (Tabela 1). De modo que, a análise geral de GM apresentou 13 marcadores significativos (S12_5724762; S1_391761090; S1_267115233; S1_50403605; S16_1855885; S16_26910139; S1_266051587; S13_43326367; S1_314280574; S9_39492914; S1_289700405; S1_392321573; S1_269853445)

distribuídos em 7 cromossomos, onde, os cromossomos 6, 9 e 16 apresentaram mais de um marcador significativo. Os materiais argentinos apresentaram 10 SNPs significativos associados à maturidade, sendo dois deles, os mesmos marcadores encontrados na análise geral (S1_266051587; S1_267115233), localizados no cromossomo 6.

Com relação os materiais brasileiros foram encontrados 9 SNPs significativos associados a maturidade, onde, 3 SNPs já tinham sido encontrados na análise geral (S1_50403605; S12_5724762; S16_1855885), localizados nos cromossomos 6, 12 e 16, respectivamente. A análise com os materiais americanos apresentou no total 7 SNPs significativos associados à maturidade, de modo que, apenas um marcador (S1_50403605), localizado no cromossomo 1, já estava relacionado com a análise geral.

A análise GWAS, quando foi realizada para os diferentes grupos de maturação com os materiais dos três países, Argentina, Brasil e Estados Unidos, identificou ao todo, 29 SNPs significativos ($p < 0,001$) associados aos diferentes grupos de maturação, localizados em diferentes cromossomos (Tabela 2). Além disso, os gráficos *QQ-plots* do modelo MLM apresentaram um desvio de médio a acentuado na distribuição do valor de p -esperado na cauda, indicando que as associações falso-positivas foram efetivamente controladas pelo modelo (Figura 10).

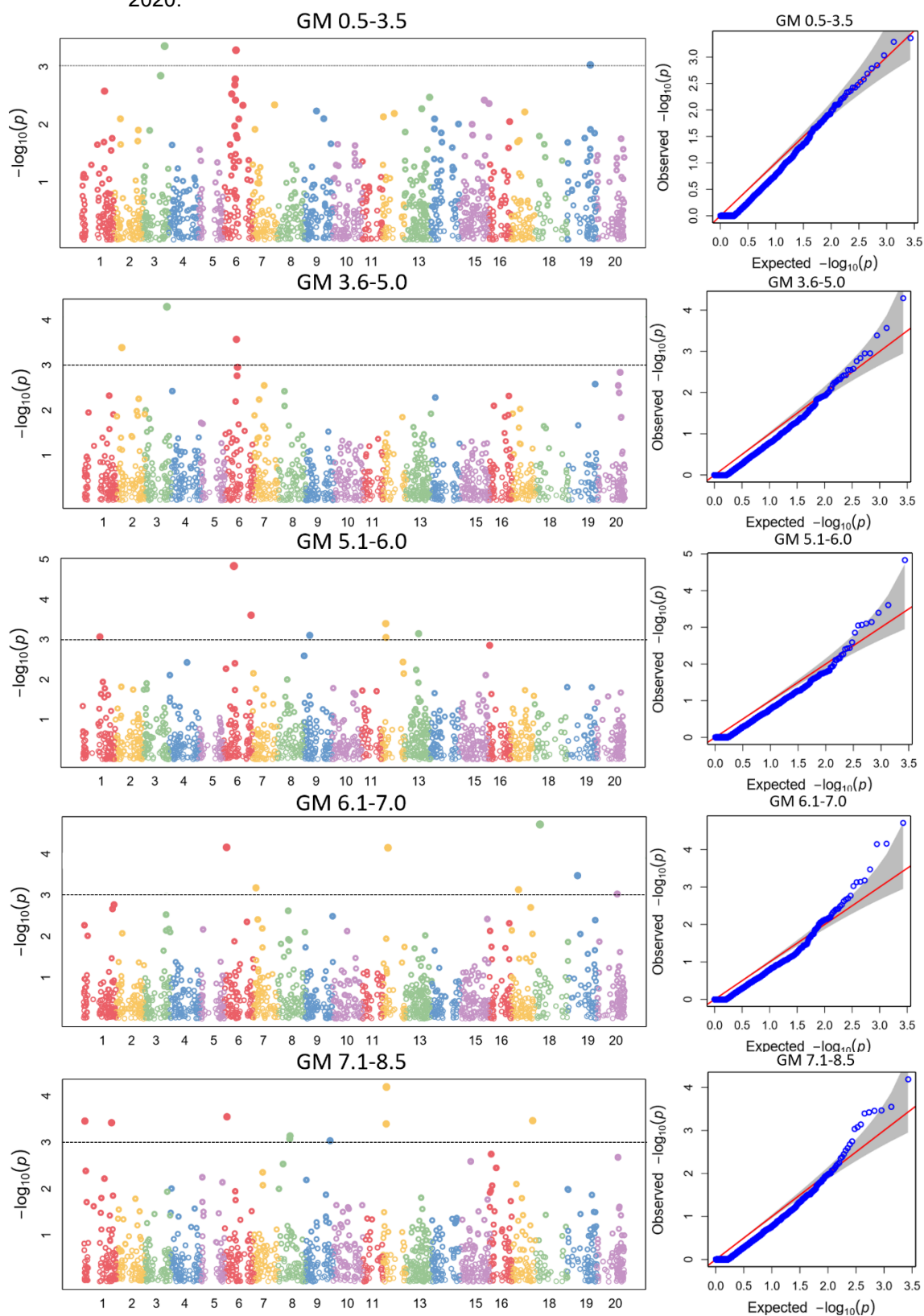
Tabela 2 - Marcadores SNPs significativos na análise de associação genômica para grupos de maturação em soja, com a identificação dos alelos, cromossomos, posição, valor $-\text{Log}_{10}(p)$ e R^2 , gene candidato e proteína correspondente. UTFPR, *Campus* Pato Branco, 2020.

| GM | SNP | Alelos | Crom | Posição | LOD | R^2 | Efeito aditivo | Gene candidato | Proteína codificante | Qtls Conhecidos/ Gene(1,5Mb) | Referência |
|---------|--------------|--------|------|----------|------|-------|----------------|------------------------|--|---|--------------------|
| | S3_38448188 | G/T | 3 | 38448188 | 3,36 | 74,3 | 0.21 | <i>Glyma.03g169700</i> | <i>Phosphatidylinositol N-Acetylglucosaminyl Transferase</i> | Reproductive stage length 4-g1; R8 full maturity 10-g1 | Copley et al. 2018 |
| 0.5-3.5 | S1_268375400 | T/C | 6 | 22562346 | 3,28 | 73,8 | 0.14 | <i>Glyma.06g216400</i> | myb-like HTH transcriptional regulator | First flower 6-g11; R8 full maturity 8-g16 | Zuo et al. 2013 |
| | S1_890511653 | C/G | 19 | 39977565 | 3,03 | 70,0 | 0.16 | <i>Glyma.19g138000</i> | Integrase-type DNA-binding | First flower 4-g76 | Mao et al. 2017 |
| | S1_65469593 | T/G | 2 | 8637869 | 3,39 | 28,4 | -0.03 | <i>Glyma.02g095400</i> | Cation efflux family | First flower 6-g14; R8 full maturity 8-g9; Reproductive period 3-g4 | Zuo et al. 2013; |
| 3.6-5.0 | S3_38448188 | G/T | 3 | 38448188 | 4,29 | 30,4 | -0.04 | <i>Glyma.03g169700</i> | <i>Phosphatidylinositol N-Acetylglucosaminyl Transferase</i> | Reproductive stage length 4-g1; R8 full maturity 10-g1 | Copley et al. 2018 |
| | S1_266051587 | G/A | 6 | 20238533 | 3,57 | 28,9 | -0.01 | <i>Glyma.06g207800</i> | AP2/B3-like transcriptional factor | R8 full maturity 10-g3.2; First flower 8-g3.2, 5-g7.3, 5-g23.3 | Copley et al. 2018 |
| | S1_264657337 | C/T | 6 | 18844283 | 4,83 | 33,1 | -0.03 | <i>Glyma.06g204100</i> | ABI five binding protein 3 | First flower 5-g23.1 | Fang et al. 2017 |
| | S6_48624734 | C/G | 6 | 48624734 | 3,60 | 33,5 | 0.03 | <i>Glyma.06g297300</i> | Core-2/I-branching beta-1,6-N-acetylglucosaminyl transferase | Reproductive period 3-g10 | Zuo et al. 2013 |
| | S1_533376320 | T/C | 12 | 7154065 | 3,40 | 32,9 | 0.02 | <i>Glyma.12g089100</i> | Homeodomain-like | First flower 4-g46 | Mao et al. 2017 |
| 5.1-6.0 | S13_24540749 | G/A | 13 | 24540749 | 3,14 | 32,7 | 0.06 | <i>Glyma.13g134500</i> | phosphatidic acid phosphohydrolase 2 | First flower 4-g51 | Mao et al. 2017 |
| | S1_399701818 | A/T | 9 | 10003392 | 3,10 | 30,0 | 0.01 | <i>Glyma.09g082000</i> | squamosa promoter-binding protein-like 12 | R8 full maturity 8-6 | Zuo et al. 2013 |
| | S1_30283734 | T/C | 1 | 30283734 | 3,06 | 25,4 | -0.01 | <i>Glyma.01g096000</i> | PLATZ transcription factor | N | N |
| | S1_533725911 | A/G | 12 | 7503656 | 3,04 | 25,3 | 0.03 | <i>Glyma.12g089100</i> | Homeodomain-like | First flower 4-g47 | Mao et al. 2017 |
| | S18_8861491 | C/A | 18 | 8861491 | 4,71 | 38,8 | -0.19 | <i>Glyma.18g089700</i> | lipid-binding serum glycoprotein | N | N |

| | | | | | | | | | | | |
|---------|--------------|-----|----|----------|------|------|-------|------------------------|---|--|------------------|
| | S1_250134260 | G/A | 6 | 4321206 | 4,15 | 38,6 | 0.08 | <i>Glyma.06g057900</i> | Tetratricopeptide repeat (TPR)-like | N | N |
| | S12_9282025 | A/C | 12 | 9282025 | 4,14 | 38,1 | -0.11 | <i>Glyma.12g103800</i> | B-box type zinc finger | First flower 4-g47 | Mao et al. 2017 |
| 6.1-7.0 | S1_867097159 | A/G | 19 | 16563071 | 3,47 | 37,3 | -0.11 | <i>Glyma.19g064700</i> | Chaperone DnaJ-domain | N | N |
| | S7_3896894 | C/A | 7 | 3896894 | 3,17 | 37,1 | 0.05 | <i>Glyma.07g047100</i> | Calcineurin-like metallo-phosphoesterase | First flower 4-g23 | Mao et al. 2017 |
| | S17_12416373 | C/T | 17 | 12416373 | 3,12 | 37,0 | -0.07 | <i>Glyma.17g150200</i> | RNA phosphodiester bond hydrolysis, endonucleolytic | First flower 6-g6 | Zuo et al. 2013 |
| | S20_34758376 | C/T | 20 | 34758376 | 3,02 | 36,5 | 0.08 | <i>Glyma.20g105900</i> | Major facilitator superfamily protein | First flower 5-g13, 5-g32 | Fang et al. 2017 |
| | S12_6176386 | C/T | 12 | 6176386 | 4,18 | 42,1 | 0.08 | <i>Glyma.12g073900</i> | pseudo-response regulator 7 | First flower 4-g46 | Mao et al. 2017 |
| | S1_250134260 | G/A | 6 | 4321206 | 3,54 | 41,8 | 0.12 | <i>Glyma.06g057900</i> | Tetratricopeptide repeat (TPR)-like superfamily protein | N | N |
| | S17_36601655 | C/G | 17 | 36601655 | 3,46 | 41,5 | 0.13 | <i>Glyma.17g218100</i> | Beta-1,3-N-Acetylglucosaminyltransferase | First flower 4-g67 | Mao et al. 2017 |
| 7.1-8.5 | S1_1701388 | G/A | 1 | 1701388 | 3,45 | 41,1 | -0.09 | <i>Glyma.01g018700</i> | hydroxyproline-rich glycoprotein | N | N |
| | S1_47938077 | T/C | 1 | 47938077 | 3,42 | 40,0 | 0.08 | <i>Glyma.01g145900</i> | glyoxal oxidase-related | N | N |
| | S12_5724762 | G/A | 12 | 5724762 | 3,39 | 38,3 | -0.12 | <i>Glyma.12g073900</i> | pseudo-response regulator 7 | First flower 5-g10.1, 5-g26.1, 5-g26.2, 5-g10.3, 5-g26.3 | Fang et al. 2017 |
| | S8_20380564 | C/G | 8 | 20380564 | 3,14 | 37,3 | 0.17 | <i>Glyma.08g236300</i> | SAUR-like auxin-responsive | N | N |
| | S8_19877904 | T/C | 8 | 19877904 | 3,08 | 37,1 | 0.16 | <i>Glyma.08g236300</i> | SAUR-like auxin-responsive | N | N |
| | S1_432713050 | G/C | 9 | 43014624 | 3,03 | 35,5 | 0.07 | <i>Glyma.09g206000</i> | SNARE associated Golgi | First flower 4-g28 | Mao et al. 2017 |

GM= Grupo de maturação, Crom= cromossomo, N=não identificado

Figura 11 -- Manhatan Plot dos valores de $-\log_{10}(P)$ obtidos para cada SNP vs a posição no cromossomo para grupos de maturação (GM) em soja. A linha paralela indica valor de LOD score = 3, correspondendo ao limite de p-value < 0,001. UTFPR, *Campus Pato Branco*, 2020.



Os SNPs significativos associados a maturidade variaram entre os grupos de maturação (Tabela 2). O intervalo dos grupos de maturação de 0.5 a 3.5 apresentaram marcadores significativos nos cromossomos 3 (S3_38448188; S1_268375400), 6 (S1_268375400) e 19 (S1_890511653). Além disso, o intervalo de 3.6 a 5.0, apresentou três marcadores significativos, sendo um deles o mesmo marcador do grupo anterior localizado no cromossomo 3 e outros dois marcadores localizados nos cromossomos 6 (S1_266051587) e 2 (S3_38448188). Os grupos de maturação compreendidos entre GM 5.1 a 6.0 apresentaram sete marcadores significativos, localizados nos cromossomos 6 (S1_264657337 e S6_48624734), 12 (S1_53336320 e S1_533725911), 13 (S13_24540749), 9 (S1_399701818) e 1 (S1_30283734), respectivamente.

Os GM de 6.1 a 7.0 apresentaram 7 marcadores significativos, localizados nos cromossomos 18 (S18_8861491), 6 (S1_250134260), 12 (S12_9282025), 19 (S1_867097159), 7 (S7_3896894), 17 (S17_12416373) e 20 (S20_34758376), respectivamente. Os grupos de maturação de 7.1 a 8.5 apresentaram o maior número de marcadores significativos, sendo localizados nos cromossomos 12 (S12_6176386 e S12_5724762), 6 (S1_250134260), 17 (S17_36601655), 1 (S1_1701388 e S1_47938077), 8 (S8_20380564 e S8_19877904) e 9 (S1_432713050), respectivamente (Tabela 2).

Entre os SNPs significativos observados, 19 dos 27 estão localizados em regiões de QTL, ou próximo (1,5 Mb) a estas regiões previamente descritas associadas a indução da primeira flor, duração de período reprodutivo e/ou maturação plena em R8 (Tabela 2). Além disso, oito SNPs (S1_30283734, S1_1701388, S1_47938077, S1_020134260, S8_20380564, S8_19877904, S18_8861491, S1_867097159), que podem estar relacionadas na determinação de GMs da soja, foram encontrados nos cromossomos 1, 6, 8, 18 e 19.

Há relatos de mais de 60 loci controlando a maturidade em soja (<https://www.soybase.org/>). As informações referentes aos cromossomos, locus, posição e funções descritas para cada gene onde os SNPs significativos foram localizados, foram obtidas nos bancos de dados Phytozome 11.0 (<http://phytozome.jgi.doe.gov>) e SoyBase (<http://www.soybase.org>).

O SNP S3_38448188, localizado no cromossomo 3, significativo para dois intervalos de GM (0.5 a 3.5 e 3.6 a 5.0), encontra-se posicionado dentro do gene

candidato Glyma.03g169700 e possui QTLs conhecidas a 154 Kb que determinam dias até maturidade (COPLEY et al., 2018). Este gene possui proteínas homólogas identificadas em um gene no cromossomo 19 (Glyma.19g170900), indicando que eventos de duplicação no genoma da soja ocorreram (HWANG et al., 2017). O gene Glyma.19g170900 está a 1,4 Mb do gene Glyma.19g189100 que codifica a proteína ubiquitina proteína ligase E3 B, responsável pela degradação de diferentes proteínas PIF (fatores de interação do fitocromo) (LEIVAR & MONTE, 2014). Os PIFs atuam como inibidores das respostas induzidas pelo fitocromo de forma parcialmente redundante (LEIVAR et al., 2012), podendo assim, ativar o fitocromo A, posicionado a 2,5 Mb, locus este, denominado de E3 (WATANABE et al., 2012). O marcador SNP localizado no cromossomo 19 (S1_890511653) também foi significativo para este intervalo de GM. Tal marcador possui QTLs conhecidas a 6 Kb de distância e que influenciam no florescimento da soja (Tabela 2). O marcador está posicionado a 38kb do gene candidato Glyma.19g13800, gene que codifica a proteína Integrase-type DNA-binding superfamily protein, proteína homóloga a genes de Arabidopsis, e que está relacionada com genes de florescimento em soja (MAO et al., 2017; XING et al., 2014; LI et al., 2016).

O cromossomo 6 apresentou três marcadores significativos para os GM 0.5-3.5 (S1_268375400), GM 3.6-5.0 (S1_264657332) e GM 5.1-6.0 (S1_264657332). Estes marcadores estão localizados a 2,3 Mb, 30Kb e 1,3Mb do gene E1 (UPADHYAY et al., 1994), respectivamente. O gene E1 tem sido reportado como principal gene determinante da maturidade precoce e floração em condições de campo (ABE et al., 2003; XIA et al., 2012). O marcador S1_268375400 está à 6 Kb do gene candidato Glyma.06g216400, o qual codifica a proteína myb-like HTH transcriptional regulator family protein, proteína que regula o tempo de floração através da formação de um complexo de proteína com CO (Constans), proteína esta, que facilita a transição da fase vegetativa para a reprodutiva em condições de dias longos (SIM et al., 2019). O marcador S1_264657332 está posicionado a apenas 30 Kb do gene candidato Glyma.06g207800, também denominado de gene E1, sendo o principal gene associado ao período de floração e maturidade (ZHANG et al., 2016). O gene E1 é livre de intron e codifica uma proteína da família do fator de transcrição do tipo AP2/B3 (HAN et al., 2018). De acordo com Levy (2002), os genes da superfamília B3 interagem com o locus de floração T (FT) e o locus de floração C (FLC), controlando

assim a floração e demais processos de desenvolvimento através de outros reguladores. Além destes, o marcador S1_264657332 também se encontra posicionado no cromossomo 6, tendo como candidato o gene Glyma.06g204100 distantes 312Kb do marcador, codificando a proteína ABI five binding protein 3 (AFP3), esta proteína segundo Chang et al (2019), atua juntamente com outra proteína (AFP2), determinando o atraso na floração de arabis, via regulação negativa do ácido abscísico atrasando assim a expressão do locus de floração C.

Somado a estes marcadores, o cromossomo 6 apresentou outro marcador que foi significativo (S1_250134260). Entretanto, este marcador foi responsável por determinar grupos mais tardios e foi significativo para dois intervalos de GMs, sendo eles, GM 6.1 a 7.0 e GM 7.1 a 8.5, apresentando como gene candidato o gene Glyma.06g057900 que codifica a proteína Tetratricopeptide repeat (TPR)-like superfamily protein (TPRs). As proteínas TPR medeiam as interações proteína-proteína bem como a montagem de complexos multiproteicos (BLOM et al., 2004). Tseng e Salomé (2004), encontraram em seus estudos, que um regulador negativo de giberelina (GI) denominado de SPINDLY (SPY) interage com um gene de repressão de floração GIGANTICA (GI) através do auxílio de TPRs, sendo capaz de superar a repressão floral. Outro gene codificador de TPRs, denominado de ATPRP39-1, foi identificado por Wang et al. (2007), desempenhando um papel na via autônoma de floração em Arabidopsis, regulando negativamente o gene repressor floral FLC, retardando assim a floração. Da mesma maneira, Sun et al., (2012) utilizando plantas transgênicas de Arabidopsis, concluiu que quando esta proteína foi super expressa, a floração de dessas plantas transgênicas foi atrasada em 6 a 7 dias comparativamente com Arabidopsis do tipo selvagem.

O cromossomo 12 apresentou cinco marcadores significativos (S1_533376320; S1_533725911; S12_9282025; S12_6176386 e S12_5724762) em um intervalo de 3,5Mb, resultando na indicação de 3 genes candidatos (Tabela 2). O gene Glyma.12g089100 apresentou maior associação aos GMs de 5.1 a 6.0, o qual codifica a proteína Homeodomain-like superfamily protein também conhecida como MYB domain protein que é essencial para o ritmo circadiano, envolvido no florescimento (HAZEN et al., 2005). Fang et al. (2019) relataram que tal proteína associa-se a um gene candidato responsável pela QTL denominada qLJ11, sendo uma das QTLs responsáveis pelo período juvenil da soja. Além desta região descrita,

pode se determinar outro gene candidato através da significância de outro marcador no cromossomo 12, resultando no gene candidato Glyma.12g103800 que codifica a proteína denominada B-box type zinc finger family protein. Em *Arabidopsis* a zinc finger family foi identificada como uma proteína de ligação a um RNA participante do desenvolvimento de flores e resistência a estresses abióticos (LI et al., 2001), onde o membro mais caracterizado desta família é CO (CONTANS) (KUMAGAI et al., 2008), descrito como um regulador transcricional desempenhando um papel regulador importante na via fotoperiódica (KINMONTH et al., 2013). Estudos demonstraram que o CO se liga diretamente ao promotor do locus de floração T e, portanto ativa a transcrição deste locus, necessário para iniciar a floração (TIWARI et al., 2010).

O cromossomo 12 apresentou ainda, mais uma região significativa para grupos mais tardios de maturação, sendo o gene Glyma.12g073900 o gene candidato que codificou os GMs de 7.1 a 8.5. Este gene é homólogo ao gene APRR7, que foi confirmado como vital na via de floração controlada pelo relógio circadiano em *Arabidopsis* (YAMANOTO et al., 2003; NAKAMICHI et al., 2005), através da codificação de uma proteína reguladora de pseudo-resposta (PRR) no terminal N (LI et al., 2019). Os genes PRR são fatores proteicos associados ao ciclo circadiano, conhecidos por serem altamente responsivos e reguladores chave no processo foto periódico em muitas culturas (MURAKAMI et al., 2005; NAKAMICHI et al., 2010). De modo que, o gene Glyma.12g073900 foi descrito como sendo o gene candidato da QTL qFT12.1, responsável pela indução ao florescimento (LIU et al., 2018; LI et al., 2019). Além disso, em estudo recente, Wang et al (2020) utilizando mutação mediada por CRISPR/Cas9, concluiu que a superexpressão do gene GmPRR37 responsável por esta QTL, atrasou significativamente o florescimento de plantas de soja transgênicas comparativamente com o tipo selvagem em condições de fotoperíodo longo.

Os GMs mais tardios apresentaram outra região no cromossomo 8 que determina ciclos mais longos da soja, através da significância do marcador S8_20380564. Entretanto, a função de tal região não está devidamente esclarecida, apresentando 2 marcadores significativos em um intervalo de 502Kb, que determinaram como candidato o gene Glyma.08g236300 o qual codifica a proteína SAUR-like auxin-responsive protein family. SAURs (Small Auxin Up RNA) são um grupo de proteínas induzidas por auxinas e que foram identificadas inicialmente em

soja e posteriormente em outras culturas (MCCLURE & GUILFOYLE 1987; HAGEN & GUILFOYLE 2002; KANT et al., 2009). As auxinas regulam quase todos os aspectos de desenvolvimento e crescimento das plantas, incluindo formação dos primórdios florais (WEIJERS & WAGNER 2016) e desenvolvimento dos estames e do gineceu (SONG et al., 2013; MARSCH-MARTÍNEZ & DE FOLTER, 2016). Estes genes responsivos à auxina, regulam processos fisiológicos, celulares e de desenvolvimento mediados por auxinas (REN & GRAY, 2015).

Devido ao grande número de SNPs e genes que podem estar relacionados a maturidade em soja, diversos estudos tem sido realizados para examinar como a variação alélica dos principais genes afetam o tempo de floração e maturidade (LANGEWISCH et al., 2017; LI et al., 2017; WOLFGAND & AN, 2017). A fim, da compreensão desta diversidade alélica, 10 marcadores distribuídos em 3 cromossomos, que possuíam os maiores efeitos alélicos e apresentavam significância nas análises GWAS foram usados (tabela 3). Os marcadores S1_264657337; S1_266051587; S1_267115233; S1_268375400; S1_269853445 e S1_289700405, todos localizados no cromossomo 6, apresentaram diferentes frequências alélicas perante os GM, possibilitando assim, a diferenciação de grupos de maturação, com base nas frequências de seus alelos. Os demais marcadores, localizados nos cromossomos 1 e 12 não possibilitaram a diferenciação entre os grupos, pois, apresentaram em todos os grupos de maturação, uma maior frequência de um dos seus alelos.

O SNP S1_266051587, localizado no cromossomo 6, com distância de 225Kb do gene de maturidade E1, apresentou significância para os GM 3.6 a 5.0 e para os materiais argentinos. A análise de frequência e variação alélica deste marcador, apresentou uma distribuição e seleção destes alelos. Onde, o alelo G teve sua frequência prevalente nos GMs de 0.5 à 3.5, 3.6 a 5.0, nos materiais argentinos e americanos, podendo está, por apresentar grupos de maturação precoces, ser a versão não funcional do gene de maturidade E1. (LANGEWISCH et al., 2017). Por outro lado, grupos mais tardios acima de GM 5.0 e materiais brasileiros apresentaram a frequência do alelo A como sendo predominante. Zhai et al, (2014) utilizando cultivares provenientes da China, confirmou que cultivares que possuíam o alelo não funcional e1-nf exibiram um fenótipo de floração muito precoce, confirmando que E1

Tabela 3- Distribuição e frequência alélica dos principais genes relacionados a maturidade em 5428 linhagens de soja, provenientes da Argentina, Estados Unidos e Brasil. UTFPR, Campus Pato Branco, 2020.

| Marcador | Crom | Efeito | Posição | Alelos | Grupos de maturação | | | | | Países | | |
|--------------|------|--------|----------|--------|---------------------|---------|---------|---------|---------|-----------|-----|--------|
| | | | | | 0.5-3.5 | 3.6-5.0 | 5.1-6.0 | 6.1-7.0 | 7.1-8.5 | Argentina | Eua | Brasil |
| S1_50403605 | 1 | -0.19 | 50403605 | C A | | | | | | | | |
| S1_264657337 | 6 | 0.12 | 18844283 | C T | | | | | | | | |
| S1_266051587 | 6 | -0.17 | 20238533 | G A | | | | | | | | |
| S1_267115233 | 6 | -0.29 | 21302179 | G T | | | | | | | | |
| S1_268375400 | 6 | -0.10 | 22562346 | T C | | | | | | | | |
| S1_269853445 | 6 | -0.10 | 24040391 | A G | | | | | | | | |
| S1_289700405 | 6 | -0.05 | 43887351 | C G | | | | | | | | |
| S12_5724762 | 12 | -0.19 | 5724762 | G A | | | | | | | | |
| S12_6176386 | 12 | 0.054 | 6176386 | C T | | | | | | | | |
| S1_533376320 | 12 | 0.10 | 7154065 | T C | | | | | | | | |

* GM ou país significativo a 0.01% na análise GWAS.

é o locus mais importante para esta característica. Somado a isso, Langewisch et al. (2017), utilizando linhagens americanas de GMs 0 a 9, também identificou variação dos alelos E, de tal modo que, as linhagens de maturação muito precoce apresentaram os alelos não funcionais e1, e2, e3-tr, e as linhagens de maturação mais tardia eram E1, E2, E3 com uma clara separação ocorrendo entre os GMs 4 e 5 para e1-as e E1, respectivamente.

A análise GWAS realizada neste estudo, possibilitou a compreensão da característica complexa, envolvendo grupos de maturação em soja, através da identificação de genes candidatos. Nove novos SNPs e 20 SNPs localizados dentro de QTLs conhecidos ou próximos a eles foram encontrados. Vinte e três genes candidatos relacionados aos períodos de crescimento em soja, podem servir como alvos promissores para estudos de mecanismos moleculares, determinando grupos de maturação.

5 CONCLUSÕES

A análise de associação genômica ampla para os diferentes grupos de maturação revelou vinte nove SNPs significativos localizados em treze cromossomos. Grande parte dos genes candidatos encontrados, como enzimas e fatores de transcrição, por exemplo, estão relacionados com funções que possuem relação direta ou indireta com o caractere analisado, e podem auxiliar numa melhor compreensão de sua arquitetura genética.

Com base nos resultados, a utilização de marcadores presentes no cromossomo 6 na região do gene E1, mostra-se como grande oportunidade para conhecimento e seleção de materiais com grupos de maturação menores que 5.0. contrapondo a este fato, a utilização de marcadores no cromossomo 12 mostra-se mais eficaz para o entendimento de grupos de maturação maiores que 5.0.

A explicação da variação fenotípica pelos marcadores significativos variou de 25,3 a 74,3%. As maiores explicações foram observadas para os marcadores S3_38448188 e S1_268375400, que estão localizados nas regiões codificadoras dos genes E1 e E3, dois dos principais genes envolvidos com grupos de maturação em soja.

Os marcadores significativos encontrados neste estudo, após validados, poderão ser empregados em programas de seleção assistida por marcadores moleculares (SAM), para a seleção precoce de genótipos com grupos de maturação desejáveis.

6 CONSIDERAÇÕES FINAIS

A identificação de marcadores moleculares associados a características de interesse tem sido o objetivo de programas de melhoramento ao redor do mundo, e tem contribuído significativamente para o sucesso no processo de seleção e obtenção de genótipos superiores

O conhecimento de QTLs, relacionadas tanto a florescimento como maturação, não só contribuirá para uma melhor compreensão da base molecular da regulação destes períodos, mas também pode ser potencialmente útil para melhorar a adaptabilidade e produtividade da soja. A melhor compreensão destas bases moleculares permite, o cultivo de variedades com floração ou maturidade ideal para regiões geográficas específicas

Os marcadores significativos encontrados no presente estudo, após validados, contribuirão para aumentar o conjunto de marcadores já disponíveis para a seleção de genótipos com grupos de maturação desejáveis, proporcionando a obtenção de genótipos ideais de maneira mais ágil e precisa.

Entretanto, em função da complexidade da característica, e de sua grande interação com o ambiente, conseqüentemente, existe a dificuldade na obtenção de genótipos superiores, sendo necessários mais estudos nesta área a fim de identificar marcadores e genes úteis associados aos grupos de maturação.

REFERÊNCIAS

- ABE, J.; XU, D.; MIYANO, A.; KOMATSU, K.; KANAZAWA, A.; SHIMAMOTO, Y. Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. **Crop science**, v. 43, n. 4, p. 1300-1304, 2003.
- ABRAHÃO, Gabriel M.; COSTA, Marcos H. Evolution of rain and photoperiod limitations on the soybean growing season in Brazil: The rise (and possible fall) of double-cropping systems. **Agricultural and Forest Meteorology**, v. 256, p. 32-45, 2018.
- AKKAYA, M. S.; SHOEMAKER, R. C.; SPECHT, J. E.; BHAGWAT, A. A.; CREGAN, P. B. Integration of simple sequence repeat DNA markers into a soybean linkage map. **Crop Science**, v. 35, n. 5, p. 1439-1445, 1995.
- ALLIPRANDINI, L. F.; ABATTI, C.; BERTAGNOLLI, P.F.; CAVASSIM, J.E.; GABE, H.L.; KUREK, A.; MATSUMOTO, M. N.; OLIVEIRA, M. A. R.; PITOL, C. PRADO, L. C. STECKLING, C. Understanding soybean maturity groups in Brazil: environment, cultivar classification, and stability. **Crop Science**, v. 49, n. 3, p. 801-808, 2009.
- BAIGORRÍ H.E.. Manejo del Cultivo. En: **El Cultivo de la Soja en Argentina**. Ediciones INTA. p. 125-138 1997
- BAIGORRI, H.; CIAMPITTI, I.; GARCIA, F. **Manejo del cultivo de soja. Manual de manejo del cultivo de soja. IPNI (International Plant Nutrition Institute), Buenos Aires, Argentina**, p. 17-32, 2009.
- BANDILLO, N. B.; ANDERSON, J. E.; KANTAR, M. B.; STUPAR, R.M.; SPECHT, J. E.; GRAEF, G. L.; LORENZ, A. J. Dissecting the genetic basis of local adaptation in soybean. **Scientific reports**, v. 7, n. 1, p. 17195, 2017.
- BARZANOVA, V. V.; NOVIKOVA, A. A. Genomic analysis in soybean breeding. **Current Challenges in Plant Genetics, Genomics, Bioinformatics, and Biotechnology**, v. 24, p. 232, 2019.
- BERNARD, R. L. Two major genes for time of flowering and maturity in soybeans 1. **Crop Science**, v. 11, n. 2, p. 242-244, 1971.
- BHERING, Leonardo Lopes; CRUZ, Cosme Damião; GOD, Pedro Ivo Vieira Good. Estimativa de frequência de recombinação no mapeamento genético de famílias de irmãos completos. **Pesquisa Agropecuária Brasileira**, v. 43, n. 3, p. 363-369, 2008.
- BLOM, Eric et al. Multiple TPR motifs characterize the Fanconi anemia FANCG protein. **DNA repair**, v. 3, n. 1, p. 77-84, 2004.
- BONATO, Emídio Rizzo; VELLO, Natal Antonio. E6, a dominant gene conditioning early flowering and maturity in soybeans. **Genetics and Molecular Biology**, v. 22, n. 2, p. 229-232, 1999.
- BUZZELL, R. I. Inheritance of a soybean flowering response to fluorescent-daylength conditions. **Canadian Journal of Genetics and Cytology**, v. 13, n. 4, p. 703-707, 1971.

- BUZZELL, R. I.; VOLDENG, H. D. Inheritance of insensitivity to long daylength. **Soybean Genetics Newsletter**, v. 7, n. 1, p. 13, 1980.
- CAO, D.; TAKESHIMA, R.; ZHAO, C.; LIU, B.; JUN, A.; KONG, F. Molecular mechanisms of flowering under long days and stem growth habit in soybean. **Journal of experimental botany**, v. 68, n. 8, p. 1873-1884, 2017.
- CARPENTIERI-PIPOLO, V.; ALMEIDA, L. A.; KIIHL, R. A. S.; ROSOLEM, C. A. Inheritance of long juvenile period under short day conditions for the BR80-6778 soybean (*Glycine max* (L.) Merrill) line. **Euphytica**, v. 112, p. 203-209, 2002.
- Carter, T. E., Nelson, R. L., Sneller, C. H. & Cui, Z. Genetic Diversity in Soybean. **Soybeans: Improvement, Production, and uses**, 303–416, 2004.
- CHANG, Guanxiao et al. ABI5-BINDING PROTEIN2 coordinates CONSTANS to delay flowering by recruiting the transcriptional corepressor TPR2. **Plant physiology**, v. 179, n. 2, p. 477-490, 2019.
- CHAURASIA, Shiksha et al. Multi-locus genome-wide association studies reveal novel genomic regions associated with vegetative stage salt tolerance in bread wheat (*Triticum aestivum* L.). **Genomics**, v. 112, n. 6, p. 4608-4621, 2020.
- COBER, E. R.; MOLNAR, S. J.; CHARETTE, M.; VOLDENG, H. D. A new locus for early maturity in soybean. **Crop Science**, v. 50, n. 2, p. 524-527, 2010.
- COBER, E. R.; MORRISON, M. J. Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. **Theoretical and applied genetics**, v. 120, n. 5, p. 1005-1012, 2010.
- COBER, E. R.; STEWART, D. W.; VOLDENG, H. D. Photoperiod and temperature responses in early-maturing, near-isogenic soybean lines. **Crop science**, v. 41, n. 3, p. 721-727, 2001.
- COBER, E. R.; VOLDENG, H. D. A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. **Crop Science**, v. 41, n. 3, p. 698-701, 2001b.
- COLLINS, D.; EMEBIRI, L.; TAN, M. K.; EL BOUHSSINI, M.; WILDMAN, O. Association of KASP markers with Hessian fly resistance in wheat of diverse origin. **Euphytica**, v. 214, n.8, p. 1-8, 2018.
- CONAB-COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira de grãos**, v. 8 safra 2020/21 – n. 8 – Oitavo Levantamento. 2020.
- CONTRERAS-SOTO, R. I.; MORA, F.; LAZARRI, F.; OLIVEIRA, M. A.; SCAPIM, C. A.; SCHUSTER, I. Genome-wide association mapping for flowering and maturity in tropical soybean: implications for breeding strategies. **Breeding science**, p. 17024, 2017.
- CONTRERAS-SOTO, Rodrigo Iván et al. Population structure, genetic relatedness and linkage disequilibrium blocks in cultivars of tropical soybean (*Glycine max*). **Euphytica**, v. 213, n. 8, p. 173, 2017.
- COPLEY, Tanya R.; DUCEPPE, Marc-Olivier; O'DONOUGHUE, Louise S. Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. **BMC genomics**, v. 19, n. 1, p. 167, 2018.

CREGAN, P. B.; JARVIK, T.; BUSH, A. L.; SHOEMAKER, R. C.; LARK, K. G.; KAHLER, A. L.; NAYA, N.; VAN TOAI, T. T.; LOHNES, D. G.; CHUNG, J.; SPECHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Science**, v. 39, n. 5, p. 1464-1490, 1999.

DIAS, Moab D. et al. Current Status of Soybean Anthracnose Associated with *Colletotrichum truncatum* in Brazil and Argentina. **Plants**, v. 8, n. 11, p. 459, 2019.

ELSHIRE, Robert J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS one**, v. 6, n. 5, p. e19379, 2011.

FANG, Chao et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. **Genome biology**, v. 18, n. 1, p. 1-14, 2017.

FANG, Chao et al. Rapid identification of consistent novel QTLs underlying long-juvenile trait in soybean by multiple genetic populations and genotyping-by-sequencing. **Molecular Breeding**, v. 39, n. 6, p. 80, 2019.

FAO, 2020. FAOSTAT. Food and Agriculture Organization of the United Nations. Disponível em: <http://www.fao.org/faostat/en/#home> Acesso em 26 oct 2020.

FAS/USDA. Foreign Agricultural Service - United States Department of Agriculture. **World Agricultural Production**. Circular Series WAP 10-20, Oct, 2020. Disponível em: <https://apps.fas.usda.gov/psdonline/circulars/production.pdf>. Acesso em 26 out. 2020.

FEHR, Walter R.; CAVINESS, Charles E. **Stages of soybean development**. 1977.

GAI, J. Y.; WANG, Y. S. A study on the varietal eco-regions of soybeans in China. **Scientia Agricultura Sinica**, v. 34, n. 2, p. 139-145, 2001.

GAUT, Brandon S.; LONG, Anthony D. The lowdown on linkage disequilibrium. **The Plant Cell**, v. 15, n. 7, p. 1502-1506, 2003.

GAVIOLI, Eduardo Antonio. Explanations for the rise of soybean in Brazil. In: **Comprehensive Survey of International Soybean Research - Genetics, Physiology, Agronomy and Nitrogen Relationships**. InTechOpen, 2013.

GRAHAM, P.H. & VANCE, C.P. Legumes: importance and constraints to greater use. **Plant Physiol**. V. 131, 872–877, 2003.

HAGEN, Gretchen; GUILFOYLE, Tom. Auxin-responsive gene expression: genes, promoters and regulatory factors. **Plant molecular biology**, v. 49, n. 3-4, p. 373-385, 2002.

HAN, Guomin et al. Genome-wide association study leads to novel genetic insights into resistance to *Aspergillus flavus* in maize kernels. **BMC Plant Biology**, v. 20, p. 1-11, 2020.

HAN, Shaohuai et al. Genome-wide analysis of DNA methylation to identify genes and pathways associated with male sterility in soybean. **Molecular Breeding**, v. 38, n. 10, p. 118, 2018.

HAZEN, Samuel P. et al. LUX ARRHYTHMO encodes a Myb domain protein

essential for circadian rhythms. **Proceedings of the National Academy of Sciences**, v. 102, n. 29, p. 10387-10392, 2005.

HWANG, Won Joo et al. A candidate flowering gene in mungbean is homologous to a soybean Phytochrome A gene. **Euphytica**, v. 213, n. 4, p. 79, 2017

HYTEN, David L. et al. Highly variable patterns of linkage disequilibrium in multiple soybean populations. **Genetics**, v. 175, n. 4, p. 1937-1944, 2007

JIA, Hongchang et al. Maturity group classification and maturity locus genotyping of early-maturing soybean varieties from high-latitude cold regions. **PLoS One**, v. 9, n. 4, p. e94139, 2014.

JIANG, Bingjun et al. Natural variations of FT family genes in soybean varieties covering a wide range of maturity groups. **BMC genomics**, v. 20, n. 1, p. 1-16, 2019.

JUNG, C.; WONG, C. E.; SINGH, M.; BHALLA, P. Comparative genomic analysis of soybean flowering genes. **PLoS one**, v. 7, n. 6, p. e38250, 2012.

KANT, Surya et al. SAUR39, a small auxin-up RNA gene, acts as a negative regulator of auxin synthesis and transport in rice. **Plant physiology**, v. 151, n. 2, p. 691-701, 2009.

Kaster, M; Farias. J.R.B. Regionalização dos testes de valor de cultivo e uso e da indicação de cultivares de soja: Terceira aproximação. **Embrapa Soja**, João Strass, Paraná, Brazil, 2012.

KIIHL, R.A.S.; ALMEIDA, L.A.; DALL'AGNOL, A. Strategies for cultivar development in the tropics. **World Soybean Research Conference III**, p.301-304, 1985.

KIM, Kyoung Hyoun et al. Genome-wide association and epistatic interactions of flowering time in soybean cultivar. **Plos one**, v. 15, n. 1, p. e0228114, 2020.

KINMONTH-SCHULTZ, Hannah A.; GOLEMBESKI, Greg S.; IMAIZUMI, Takato. Circadian clock-regulated physiological outputs: dynamic responses in nature. In: **Seminars in cell & developmental biology**. Academic Press. p. 407-413, 2013.

KONG, F.; NAN, H.; CAO, D.; LI, Y.; WU, F.; WANG, J.; LU, S.; YUAN, X.; COBER, E.; ABE, J.; LIU, B. A new dominant gene E9 conditions early flowering and maturity in soybean. **Crop Science**, v. 54, n. 6, p. 2529-2535, 2014.

KONG, L.; LU, S.; WANG, Y.; FANG, C.; WANG, F.; NAN, H.; SU, T.; LI, S.; ZHANG, F.; LI, X.; ZHAO, X.; YUAN, X.; LIU, B.; KONG, F. Quantitative trait locus (QTL) mapping of flowering time and maturity in soybean using next generation sequencing (NGS)-based analysis. **Frontiers in plant science**, v. 9, p. 995, 2018.

KUMAGAI, Takeshi et al. The common function of a novel subfamily of B-Box zinc finger proteins with reference to circadian-associated events in Arabidopsis thaliana. **Bioscience, biotechnology, and biochemistry**, v. 72, n. 6, p. 1539-1549, 2008.

LANGEWISCH, Tiffany et al. The development and use of a molecular model for soybean maturity groups. **BMC plant biology**, v. 17, n. 1, p. 1-13, 2017.

LEIVAR, Pablo et al. Dynamic antagonism between phytochromes and PIF family

basic helix-loop-helix factors induces selective reciprocal responses to light and shade in a rapidly responsive transcriptional network in Arabidopsis. **The Plant Cell**, v. 24, n. 4, p. 1398-1419, 2012.

LEIVAR, Pablo; MONTE, Elena. PIFs: systems integrators in plant development. **The Plant Cell**, v. 26, n. 1, p. 56-78, 2014.

LEVY, Yaron Y. et al. Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. **Science**, v. 297, n. 5579, p. 243-246, 2002.

LI, D.; ZHAO, X.; HAN, Y.; LI, W.; XIE, F. Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. **Genomics**, v. 111, n. 1, p. 90-95, 2019.

LI, Jicun et al. Genetic variation of maturity groups and four E genes in the Chinese soybean mini core collection. **PLoS One**, v. 12, n. 2, p. e0172106, 2017.

LI, Junjie; JIA, Dongxuan; CHEN, Xuemei. HUA1, a regulator of stamen and carpel identities in Arabidopsis, codes for a nuclear RNA binding protein. **The Plant Cell**, v. 13, n. 10, p. 2269-2281, 2001.

LI, Man-Wah; LAM, Hon-Ming. The Modification of Circadian Clock Components in Soybean During Domestication and Improvement. **Frontiers in Genetics**, v. 11, p. 1081, 2020.

LI, Minmin et al. Identification of genetic loci and candidate genes related to soybean flowering through genome wide association study. **BMC genomics**, v. 20, n. 1, p. 987, 2019.

LI, X.; FANG, C.; XU, M.; ZHANG, F.; LU, S.; NAN, H.; SU, T.; LI, S.; ZHAO, X.; KONG, L.; YUAN, X.; LIU, B.; ABE, J.; COBER, E. R.; KONG, F. Quantitative trait locus mapping of soybean maturity gene E6. **Crop science**, v. 57, n. 5, p. 2547-2554, 2017.

LI, Y.; REIF, J.; MA, Y.; HONG, H.; LIU, Z.; CHANG, R.; QIU, L. Targeted association mapping demonstrating the complex molecular genetics of fatty acid formation in soybean. **BMC genomics**, v. 16, n. 1, p. 841, 2015.

LI, Ying-hui et al. Identification of loci controlling adaptation in Chinese soya bean landraces via a combination of conventional and bioclimatic GWAS. **Plant biotechnology journal**, v. 18, n. 2, p. 389-401, 2020.

LI, Yuqiu et al. Positional cloning of the flowering time QTL qFT12-1 reveals the link between the clock related PRR homolog with photoperiodic response in soybeans. **Frontiers in Plant Science**, v. 10, p. 1303, 2019.

LI, Zhiying et al. Transcriptome sequencing determined flowering pathway genes in *Aechmea fasciata* treated with ethylene. **Journal of Plant Growth Regulation**, v. 35, n. 2, p. 316-329, 2016.

Lipka A.E., Tian F., Wang Q., Peiffer J., Li M., Bradbury P.J. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*. 28(18): 2397-2399

LIU, B.; FUJITA, T.; YAN, Z.; SAKAMOTO, S.; XU, D.; ABE, J. QTL mapping of domestication-related traits in soybean (*Glycine max*). **Annals of botany**, v. 100, n.

5, p. 1027-1038, 2007.

LIU, B.; KANAZAWA, A.; MATSUMURA, H.; TAKAHASHI, R.; HARADA, K.; ABE, J. Genetic redundancy in soybean photoreponses associated with duplication of the phytochrome A gene. **Genetics**, v. 180, n. 2, p. 995-1007, 2008.

LIU, D.; YAN, Y.; FUJITA, Y.; XU, D. A major QTL (qFT12. 1) allele from wild soybean delays flowering time. **Molecular breeding**, v. 38, n. 4, p. 45, 2018.

LIU, Dequan et al. A major QTL (qFT12. 1) allele from wild soybean delays flowering time. **Molecular Breeding**, v. 38, n. 4, p. 45, 2018.

LIU, Luping et al. Allele combinations of maturity genes E1-E4 affect adaptation of soybean to diverse geographic regions and farming systems in China. *Plos one*, v. 15, n. 7, p. e0235397, 2020.

LIU, Ming-Hao et al. Genome-wide association study identifies an NLR gene that confers partial resistance to *Magnaporthe oryzae* in rice. **Plant Biotechnology Journal**, v. 18, n. 6, p. 1376-1383, 2020.

LIU, Xiaoyi et al. Genome-wide association study identifies candidate genes related to oleic acid content in soybean seeds. **BMC plant biology**, v. 20, n. 1, p. 1-14, 2020.

LONG, Weixiong et al. Genome-wide association analysis of resistance to rice false smut. **Molecular Breeding**, v. 40, n. 5, 2020.

LU, S.; LI, W.; WANG, J.; SRINIVES, P.; NAN, H.; CAO, D.; WANG, Y.; LI, J.; LI, X.; FANG, C.; SHI, X.; YUANG, X.; WATANABE, S.; FENG, X.; LIU, B.; ABE, J.; KONG, F. QTL mapping for flowering time in different latitude in soybean. **Euphytica**, v. 206, n. 3, p. 725-736, 2015.

LU, S.; ZHAO, X.; HU, Y.; LIU, S.; NAN, H.; LI, X.; FANG, C.; CAO, D.; SHI, X.; KONG, L.; SU, T.; ZHANF, F.; LI, S.; WANG, Z.; YUAN, X.; COBER, E.; WELLWE, J. Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. **Nature genetics**, v. 49, n. 5, p. 773, 2017.

LU, Sijia et al. Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. **Nature genetics**, v. 49, n. 5, p. 773-779, 2017

LU, Sijia et al. Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. **Nature Genetics**, v. 52, n. 4, p. 428-436, 2020.

MA, X.; FENG, F.; WEI, H.; MEI, H.; XU, K.; CHEN, S.; LI, T.; LIANG, X.; LIU, H.; LUO, L. Genome-wide association study for plant height and grain yield in rice under contrasting moisture Regimes. **Frontiers in plant science**, v. 7, p. 1801, 2016.

MAO T.; LI, J.; WEN Z.; WU, T.; WU, C.; SUN, S.; JIANG B.; HOU W.; LI, W.; SONG, Q.; WANG, D.; HAN, T. Association mapping of loci controlling genetic and environmental interaction of soybean flowering time under various photo-thermal conditions. **BMC Genomics**, v.18, n1, p.415, 2017.

MARSCH-MARTÍNEZ, Nayelli; DE FOLTER, Stefan. Hormonal control of the development of the gynoecium. **Current opinion in plant biology**, v. 29, p. 104-114,

2016.

MCBLAIN, B. A.; BERNARD, R. L. A new gene affecting the time of flowering and maturity in soybeans. **Journal of Heredity**, v. 78, n. 3, p. 160-162, 1987.

MCCLURE, Bruce A.; GUILFOYLE, Tom. Characterization of a class of small auxin-inducible soybean polyadenylated RNAs. **Plant molecular biology**, v. 9, n. 6, p. 611-623, 1987.

MENGARELLI, Diego A.; ZANOR, María Inés. Genome-wide characterization and analysis of the CCT motif family genes in soybean (*Glycine max*). **Planta**, v. 253, n. 1, p. 1-17, 2021.

Miladinović D. Allelic variation and distribution of the major maturity genes in different soybean collections. **Frontiers in plant science**, v. 9, p. 1286, 2018.

MILADINOVIC, J.; KUROSAKI, H.; BURTON, J. W.; HRUSTIC, M.; MILADINOVIC, D. The adaptability of short season soybean genotypes to varying longitudinal regions. **European journal of agronomy**, v. 25, n. 3, p. 243-249, 2006.

MOLNAR, S. J.; RAI, S.; CHARETTE, M.; COBER, E. R. Simple sequence repeat (SSR) markers linked to E1, E3, E4, and E7 maturity genes in soybean. **Genome**, v. 46, n. 6, p. 1024-1036, 2003.

MOURTZINIS, S.; CONLEY, S. P. Delineating soybean maturity groups across the United States. **Agronomy Journal**, v. 109, n. 4, p. 1397-1403, 2017.

MUNDSTOCK, C.; Thomas, A. L. Soja: Fatores que afetam o crescimento e o rendimento de grãos. Porto Alegre: Departamento de Plantas de Lavoura da Universidade Federal do Rio Grande do Sul, **Evangraf**. 31p. 2005.

MURAKAMI, Masaya et al. Circadian-associated rice pseudo response regulators (Os PRRs): insight into the control of flowering time. **Bioscience, Biotechnology, and Biochemistry**, v. 69, n. 2, p. 410-414, 2005.

NAKAMICHI, Norihito et al. PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the Arabidopsis circadian clock. **The Plant Cell**, v. 22, n. 3, p. 594-605, 2010.

NAKAMICHI, Norihito et al. PSEUDO-RESPONSE REGULATORS, PRR9, PRR7 and PRR5, together play essential roles close to the circadian clock of *Arabidopsis thaliana*. **Plant and Cell Physiology**, v. 46, n. 5, p. 686-698, 2005.

NEUMAIER, N.; JAMES, A. T. Exploiting the long-juvenile trait to improve adaptation of soybeans to the tropics. **Acia Food Legume Newslett**, 1993.

PAN, L.; HE, J.; ZHAO, T.; XING, G.; WANG, Y.; YU, D.; CHEN, S.; GAI, J. Efficient QTL detection of flowering date in a soybean RIL population using the novel restricted two-stage multi-locus GWAS procedure. **Theoretical and applied genetics**, v. 131, n. 12, p. 2581-2599, 2018.

PENARIOL, A. Soja: **cultivares no lugar certo**. **Informações agrônômicas** nº 90, 2000.

QI, Zhongying et al. Identification of QTNs controlling 100-seed weight in soybean

using multilocus genome-wide association studies. **Frontiers in genetics**, v. 11, p. 689, 2020.

RAVELOMBOLA, Waltram et al. Genome-wide association study and genomic selection for plant height, maturity, seed weight, and yield in soybean. **Research Square**. 2020.

RAVELOMBOLA, Waltram Second et al. Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation. **Plos one**, v. 15, n. 7, p. e0235089, 2020.

RAY, J. D.; HINSON, K.; MANKONO, J.; MALO, M. F. Genetic control of a long-juvenile trait in soybean. **Crop Science**, v. 35, n. 4, p. 1001-1006, 1995.

REN, Hong; GRAY, William M. SAUR proteins as effectors of hormonal and environmental signals in plant growth. **Molecular plant**, v. 8, n. 8, p. 1153-1164, 2015.

RICE, Brian R.; FERNANDES, Samuel B.; LIPKA, Alexander E. Multi-Trait Genome-wide Association Studies Reveal Loci Associated with Maize Inflorescence and Leaf Architecture. **Plant and Cell Physiology**, 2020.

ROSSI, R.; BAIGORRÍ, H. E. J.; SALADO-NAVARRO, L. S. Historia del mejoramiento genético de la soja en la Argentina. **El cultivo de Soja en Argentina, 1st edn. Agroeditorial, Vicente López**, p. 33-50, 2012.

RODRIGUES, Osmar et al. Resposta quantitativa do florescimento da soja à temperatura e ao fotoperíodo. **Pesquisa Agropecuária Brasileira**, v. 36, n. 3, p. 431-437, 2001.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

SAFDAR, Luqman Bin et al. Genome-wide association study identifies five new cadmium uptake loci in wheat. **The Plant Genome**, v. 13, n. 2, p. e20030, 2020.

SAMANFAR, B.; MOLNAR, S. J.; CHARETTE, M.; SCHOENROCK, A.; DEHNE, F.; GOLSHANI, A.; BELZILE, F. Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. **Theoretical and applied genetics**, v. 130, n. 2, p. 377-390, 2017.

SAUVAGE, C.; SEGURA, V.; BAUCHET, G.; STEVENS, R.; DO, P.; NIKOLOSKI, Z. et al. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. **Plant Physiol.** 165:1120–1132, 2014.

Scott, W.O.; ALDRICH, S. R. Modern soybean production. S & A Publ. **Inc, Champaign, IL**, 1970.

SEDIYAMA, T. Tecnologia de produção da soja. **Editora Mecenias LTDA**. 2009.

SEDIYAMA, Carlos S. et al. Influência do retardamento da colheita sobre a deiscência das vagens e sobre a qualidade e poder germinativo das sementes da soja. **Experientiae**, 1972.

SEGURA, V., VILHJÁLMSSON, B.J.; PLATT, A.; KORTE, A.; ÜSEREN Long. et al.

An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. **Nat. Genet.** 44:825–830. doi:10.1038/ng.2314, 2012.

SILVA, A. F. D.; SEDIYAMA, T.; BORÉM, A.; In: SEDIYAMA, T.; SILVA, F.; BORÉM, A. A soja do plantio à colheita. Viçosa, MG: **Editora UFV**, 2015.

SIM, Soon Ae et al. FLOWERING HTH1 is involved in CONSTANS-mediated flowering regulation in Arabidopsis. **Applied Biological Chemistry**, v. 62, n. 1, p. 1-6, 2019.

SINGER, William Monte et al. Soybean Amino Acids in Health, Genetics, and Evaluation. **Soybean for Human Consumption and Animal Feed**. IntechOpen, 2019.

SONG, Jie et al. Identification of QTL and genes for pod number in soybean by linkage analysis and genome-wide association studies. **Molecular Breeding**, v. 40, n. 6, p. 1-14, 2020.

SONG, Qijian et al. Fingerprinting soybean germplasm and its utility in genomic research. **G3: Genes, genomes, genetics**, v. 5, n. 10, p. 1999-2006, 2015.

SONG, Susheng et al. Regulation of stamen development by coordinated actions of jasmonate, auxin, and gibberellin in Arabidopsis. **Molecular Plant**, v. 6, n. 4, p. 1065-1073, 2013.

SONG, Wenwen et al. Standard cultivar selection and digital quantification for precise classification of maturity groups in soybean. **Crop Science**, v. 59, n. 5, p. 1997-2006, 2019.

SUN, Fei et al. Characterization and quantitative trait locus mapping of late-flowering from a Thai soybean cultivar introduced into a photoperiod-insensitive genetic background. **Plos one**, v. 14, n. 12, p. e0226116, 2019.

SUN, Xiaoming et al. Cloning and characterization of a *Doritaenopsis* hybrid PRP39 gene involved in flowering time. **Plant Cell, Tissue and Organ Culture (PCTOC)**, v. 110, n. 3, p. 347-357, 2012.

TIWARI, Shiv B. et al. The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. **New Phytologist**, v. 187, n. 1, p. 57-66, 2010.

TRAN, D. T.; STEKETEE, C. J.; JUNIOR, J. D. B.; NOE, J.; LI, Z. Genome-Wide Association Analysis Pinpoints Additional Major Genomic Regions Conferring Resistance to Soybean Cyst Nematode (*Heterodera glycines* Ichinohe). **Frontiers in Plant Science**, v. 10, p. 401, 2019.

TSENG, Tong-Seung et al. SPINDLY and GIGANTEA interact and act in Arabidopsis thaliana pathways involved in light responses, flowering, and rhythms in cotyledon movements. **The Plant Cell**, v. 16, n. 6, p. 1550-1563, 2004.

TSUBOKURA, Y.; WATANABE, S.; XIA, Z.; KANAMORI, H.; YAMAGATA, H.; KAGA, A.; KATAYOSE, Y.; ABE, J.; ISHIMOTO, M.; HARADA, K. Natural variation in the genes responsible for maturity loci E1, E2, E3 and E4 in soybean. **Annals of botany**, v. 113, n. 3, p. 429-441, 2013.

- UPADHYAY, A. P.; ELLIS, R. H.; SUMMERFIELD, T. J.; ROBERTS, E. H. QI, A. Characterization of photothermal flowering responses in maturity isolines of soybean [*Glycine max* (L.) Merrill] cv. Clark. **Annals of Botany**, v. 74, n. 1, p. 87-96, 1994.
- VAUGHN, J.N.; NELSON, R.L.; SONG, Q.; CREGAN, P.B.; and Li, Z. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. **G3: GENES, GENOMES, GENETICS**(Bethesda) 4:2283–2294, 2014.
- VINHOLES, P.; ROSADO, R.; ROBERTS, P.; BORÉM, A.; SCHUSTER, I. Single Nucleotide Polymorphism-Based Haplotypes Associated with Charcoal Rot Resistance in Brazilian Soybean Germplasm. **Agronomy Journal**, v. 111, n. 1, p. 182-192, 2019.
- WANG, Cunxi et al. The *Arabidopsis thaliana* AT PRP39-1 gene, encoding a tetratricopeptide repeat protein with similarity to the yeast pre-mRNA processing protein PRP39, affects flowering time. **Plant cell reports**, v. 26, n. 8, p. 1357-1366, 2007
- WANG, F.; NAN, H.; CHEN, L.; FANG, C.; ZHANG, H.; SU, T.; LI, S.; CHENG, Q.; DONG, L. LIU, B.; KONG, F.; LU, S. A new dominant locus, E11, controls early flowering time and maturity in soybean. **Molecular Breeding**, v. 39, n. 5, p. 70, 2019.
- WANG, J.; ZHAO, X. WANG, W.; QU, Y.; TENG, W.; ZHENG, H.; HAN, Y.; LI, W. Genome-wide association study of inflorescence length of cultivated soybean based on the high-throughout single-nucleotide markers. **Molecular Genetics and Genomics**, v. 294, n. 3, p. 607-620, 2019.
- WANG, Jinyang et al. Genome-wide association study of inflorescence length of cultivated soybean based on the high-throughout single-nucleotide markers. **Molecular Genetics and Genomics**, v. 294, n. 3, p. 607-620, 2019.
- WANG, Li et al. GWAS reveals two novel loci for photosynthesis-related traits in soybean. **Molecular Genetics and Genomics**, p. 1-12, 2020.
- WANG, Liwei et al. Natural variation and CRISPR/Cas9-mediated mutation in GmPRR37 affect photoperiodic flowering and contribute to regional adaptation of soybean. **Plant Biotechnology Journal**, p. 1-13, 2020.
- WANG, Y.; LI, Y.; WU, H.; HU, B.; ZHENG, J.; ZHAI, H.; LV, S.; LIU, X.; CHEN, X.; QIU, H.; YANG, J. Genotyping of soybean cultivars with medium-density array reveals the population structure and QTNs underlying maturity and seed traits. **Frontiers in plant science**, v. 9, 2018.
- WATANABE, S.; HIDESHIMA, R.; XIA, Z.; TSUBOKURA, Y.; SATO, S.; NAKAMOTO, Y.; YAMANAKA, N.; TAKAHASHI, R. Map-based cloning of the gene associated with the soybean maturity locus E3. **Genetics**, v. 182, n. 4, p. 1251-1262, 2009.
- WATANABE, Satoshi; HARADA, Kyuya; ABE, Jun. Genetic and molecular bases of photoperiod responses of flowering in soybean. **Breeding science**, v. 61, n. 5, p. 531-543, 2012.
- WEIJERS, Dolf; WAGNER, Doris. Transcriptional responses to the auxin

hormone. **Annual review of plant biology**, v. 67, p. 539-574, 2016.

WEN Z.; TAN R.; YUAN J.; BALES C.; DU W.; ZHANG S.; CHILVERS MI.; SCHMIDT C.; SONG Q.; CREGAN Pb. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. **BMC Genomics**;15(1):809, 2014.

WOLFGANG, Goettel; AN, Yong-qiang Charles. Genetic separation of southern and northern soybean breeding programs in North America and their associated allelic variation at four maturity loci. **Molecular Breeding**, v. 37, n. 1, p. 8, 2017.

XIA, Z.; WATANABE, S.; YAMADA, T.; TSUKOBURA, Y.; NAKASHIMAN H.; ZHAI, H.; ANAI, T.; SATO, S.; YAMAZAKI, T.; LU, S.; WU, H.; TABATA, S.; HARADA, K. Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. **Proceedings of the National Academy of Sciences**, v. 109, n. 32, p. E2155-E2164, 2012.

XIA, Zhengjun et al. Molecular identification of genes controlling flowering time, maturity, and photoperiod response in soybean. **Plant Systematics and Evolution**, v. 298, n. 7, p. 1217-1227, 2012.

XING, Libo et al. Genome-wide identification of vegetative phase transition-associated microRNAs and target predictions using degradome sequencing in *Malus hupehensis*. **BMC genomics**, v. 15, n. 1, p. 1125, 2014.

XU, M.; XU, Z.; LIU, B.; KONG, F.; TSUBOKURA, Y.; WATANABE, S.; XIA, Z.; HARADA, K.; KANAZAWA, A.; YAMADA, T.; ABE, J. Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. **BMC plant biology**, v. 13, n. 1, p. 91, 2013.

YAMAMOTO, Yoko et al. Comparative genetic studies on the APRR5 and APRR7 genes belonging to the APRR1/TOC1 quintet implicated in circadian rhythm, control of flowering time, and early photomorphogenesis. **Plant and Cell Physiology**, v. 44, n. 11, p. 1119-1130, 2003.

YANG, Wenying et al. Critical Photoperiod Measurement of Soybean Genotypes in Different Maturity Groups. **Crop Science**, v. 59, n. 5, p. 2055-2061, 2019.

YOSHIDA, Grazyella M. et al. Genome-wide patterns of population structure and linkage disequilibrium in farmed Nile tilapia (*Oreochromis niloticus*). **Frontiers in genetics**, v. 10, p. 745, 2019.

YU, Jianming et al. 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.

ZANON, A.J; WINCK, J. E. M.; STRECK, N. A.; ROCHA, T. S. M.; CERA, J. C.; RITCHER, G.L.; LAGO, I.; SANTOS, P. M.; MACIEL, L. R.; GUEDES, J. V. C.; MARCHESAN, E. Desenvolvimento de cultivares de soja em função do grupo de maturação e tipo de crescimento em terras altas e terras baixas. **Bragantia**, v. 74, n. 4, 2015.

ZANON, Alencar Junior; STRECK, Nereu Augusto; GRASSINI, Patricio. Climate and management factors influence soybean yield potential in a subtropical environment. **Agronomy Journal**, v. 108, n. 4, p. 1447-1454, 2016.

ZDZIARSKI, A. D.; TODESCHINI, M. H.; MILIOLI, A. S.; WOYANN, L. G.; MADUREIRA, A.; STOCO, M. G.; BENIN, G. Key soybean maturity groups to increase grain yield in Brazil. **Crop Science**, v. 58, n. 3, p. 1155-1165, 2018.

ZHAI, H.; LU, S.; LIANG, S.; WU, H.; ZHANG, X.; LIU, B.; KONG, F.; YUAN, X.; LI, J.; XIA, Z. GmFT4, a homolog of FLOWERING LOCUS T, is positively regulated by E1 and functions as a flowering repressor in soybean. **PLoS One**, v. 9, n. 2, p. e89030, 2014.

ZHAI, Hong et al. Allelic variations at four major maturity E genes and transcriptional abundance of the E1 gene are associated with flowering time and maturity of soybean cultivars. **PloS one**, v. 9, n. 5, p. e97636, 2014.

ZHANG H.; HAO D.; SITO E. H. M.; YIN Z.; HU Z.; ZHANG G; YU D. Genetic dissection of the relationship between plant architecture and yield component traits in soybean (*Glycine max*) by association analysis across multiple environments. **Plant Breeding**, v. 134, n. 5, p. 564-572, 2015a

ZHANG, J.; SONG, Q.; CREGAN, P. B.; JIANG, G. L. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycinemax*). **Theoretical and Applied Genetics**, v. 129, n. 1, p. 117-130, 2016.

ZHANG, J.; SONG, Q.; CREGAN, P.; NELSON, R.; WANG, X.; WU, J.; JIANG, G. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. **BMC genomics**, v. 16, n. 1, p. 217, 2015b.

ZHANG, L. X.; BOAHEN, S. K.; ZHANG, J.; ZHANG, M. H.; FREELAND T. B.; WATSON, C. E.; LIU, X. Modifications of optimum adaptation zones for soybean maturity groups in the USA. **Crop Management**, v. 6, n. 1, p. 0-0, 2007.

ZHANG, S.; WANG, H.; WANG, S.; REN, Y.; NIU, L.; LIU, J.; LIU, B. Photoperiodism dynamics during the domestication and improvement of soybean. **Science China Life Sciences**, v.60, p.1416-1427, 2017

ZHANG, Tengfei et al. 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.

ZHANG, Xingzheng et al. Functional conservation and diversification of the soybean maturity gene E1 and its homologs in legumes. **Scientific reports**, v. 6, p. 29548, 2016.

ZHAO, C.; TAKESHIMA, R.; ZHU, J.; XU, M.; SATO, X.; WATANABE, S.; KANAZAWA, A.; LIU, B. A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog. **BMC plant biology**, v. 16, n. 1, p. 20, 2016.

