

DALTON DE OLIVEIRA FERREIRA

PROSPECÇÃO DE GENES CANDIDATOS AO DEFICIT HÍDRICO EM SOJA

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

Orientador: Felipe Lopes da Silva

**VIÇOSA – MINAS GERAIS
2022**

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

T

F383p
2021 Ferreira, Dalton de Oliveira, 1987-
 Prospecção de genes candidatos ao deficit hídrico em soja /
 Dalton de Oliveira Ferreira. – Viçosa, MG, 2021.
 1 tese eletrônica (103 f.): il. (algumas color.).

Inclui anexo.

Orientador: Felipe Lopes da Silva.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Agronomia, 2021.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2022.737>

Modo de acesso: World Wide Web.

1. Glycine max. 2. Soja - Melhoramento genético.
3. Estresse hídrico. 4. Germinação. 5. Estudo de associação
genômica ampla. I. Silva, Felipe Lopes da, 1981-.
II. Universidade Federal de Viçosa. Departamento de
Agronomia. Programa de Pós-Graduação em Genética e
Melhoramento. III. Título.

CDD 22. ed. 633.343

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

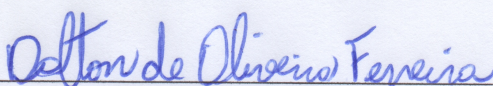
DALTON DE OLIVEIRA FERREIRA

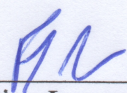
PROSPECÇÃO DE GENES CANDIDATOS AO DEFICIT HÍDRICO EM SOJA

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

APROVADA: 09 de setembro de 2022.

Assentimento:


Dalton de Oliveira Ferreira
Autor


Felipe Lopes da Silva
Orientador

À minha família, dedico.

AGRADECIMENTOS

Primeiramente a Deus, por me iluminar e conceder força, sabedoria e esperança ao longo desta jornada.

À toda a minha família, em especial aos meus pais, Dalva e Dimas e minha esposa Tamiris. Serei eternamente grato pelo amor, apoio, conselhos, paciência, incentivo, orações e por sempre terem acreditado nas minhas decisões.

Ao meu orientador Felipe Lopes da Silva pela amizade, apoio, ensinamentos e por confiar a mim o desenvolvimento desta tese. Será uma amizade para toda a vida.

Ao meu coorientador, Moysés Nascimento, por toda ajuda e ensinamentos no desenvolvimento deste trabalho.

À Universidade Federal de Viçosa, ao Programa de Pós-Graduação em Genética e Melhoramento e a todos os professores, pela oportunidade de aprender e viver coisas incríveis.

Aos membros da banca de qualificação e defesa por todas as contribuições no meu trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de Doutorado.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro ao Programa de Melhoramento da Soja – UFV.

Aos pós-graduandos e estagiários do Programa de Melhoramento Genético de Soja por todo apoio e suporte na condução dos experimentos e análises.

Aos meus grandes amigos, pelos momentos de descontração e por serem a minha família em Viçosa: Anunciene, Lucas, Martha, Polyana, Cleiton, Edna, Samyra, Samíria, Vítor, João Marcos, Gabriela, Ana Laura, Fernanda, Lorena Moura, Bruno, Guilherme, Francisco, Bruno Soares, Heloísa, Isabela, Uilton, Bruna, Nathan, Lorena, Maikon e Murilo.

À todos que contribuíram de alguma maneira para que este trabalho fosse realizado.

Muito obrigado!

RESUMO

FERREIRA, Dalton de Oliveira, D.Sc., Universidade Federal de Viçosa, agosto de 2022. **Prospecção de genes candidatos ao déficit hídrico em soja.** Orientador: Felipe Lopes da Silva.

A soja é a principal oleaginosa cultivada mundialmente, e contribui tanto para a alimentação humana quanto para a animal. Devido a essa grande importância, grandes esforços são realizados para o desenvolvimento de novas cultivares. Contudo, condições adversas, como períodos de escassez hídrica, têm se tornado cada vez mais frequentes com o avanço das mudanças das condições climáticas. Esses períodos causam grandes perdas nas culturas, na soja perdas de 24 - 50% têm sido registrados. Assim, muitos estudos têm focado nas fases críticas da soja, no entanto, o período de estiagem pode ocorrer durante qualquer fase da cultura. Assim no intuito de ajudar a compreender esse processo complexo e auxiliar no desenvolvimento de cultivares mais tolerantes, esse estudo tem os seguintes objetivos: (i) identificar os marcadores moleculares e os genes candidatos associados a estes marcadores no processo de tolerância ao déficit hídrico em soja durante o estágio de germinação; (ii) identificar os marcadores moleculares e os genes candidatos associados a estes marcadores no processo de tolerância ao déficit hídrico em soja durante o estágio vegetativo. Em cada estágio foram realizados dois experimentos em delineamento em blocos casualizados em esquema fatorial 100 x 2 (Genótipos x Disponibilidade hídrica) com 3 repetições. Os cultivares utilizados, foram provenientes de diversas empresas comerciais a fim de representar um painel de diversidade de soja no país. As disponibilidades hídricas foram: a condição controle (Capacidade de campo) e a condição estresse (-0,2 MPa (germinação) ou -1,2 MPa (estádio vegetativo)) com duração de 20 dias. A parcela experimental no estágio de germinação foi uma linha contendo dez sementes tratadas e no estágio vegetativo foi uma planta em um vaso de 3 litros de substrato (solo:areia). O início do período de estresse na germinação foi no ato do plantio das sementes, enquanto no vegetativo foi quando as plantas entraram no estágio vegetativo 2 (V2). Para cada característica foram obtidos os valores genotípicos (gBLUP) utilizando duas abordagens para o conceito de tolerância. Na primeira abordagem utilizou-se a diferença entre a condição estresse e a controle para obter o valor genotípico. Na segunda abordagem os gBLUPs foram obtidos utilizando os dados da condição estresse e controle como tratamentos. Os valores de gBLUP foram então utilizados no estudo de associação genômica ampla para a obtenção dos marcadores significativos. No estágio vegetativo foram identificados 25 marcadores do tipo SNP utilizando

a primeira abordagem e 18 marcadores utilizando a segunda. Dentre esses, 3 marcadores estavam presentes em ambas as abordagens. Foram encontrados 5 genes na região próxima a esses marcadores em que: um relacionado a parede celular, dois relacionados a transporte de proteínas e dois fatores de transcrição. No estágio vegetativo, foram encontrados 114 marcadores na primeira abordagem e 68 na segunda. Em ambas as abordagens foram encontrados 18 marcadores moleculares. Na região próxima a esses marcadores foram encontrados 15 genes: um fator de transcrição, quatro que já são relacionados à tolerância ao déficit hídrico, três relacionados ao crescimento de raízes e seis que são relacionados a outros tipos de estresse.

Palavras-chave: *Glycine max*. Estresse abiótico. GWAS. Germinação. Estádio vegetativo. Melhoramento de plantas.

ABSTRACT

FERREIRA, Dalton de Oliveira, D.Sc., Universidade Federal de Viçosa, August 2022.
Prospection of candidate genes for drought in soybean. Adviser: Felipe Lopes da Silva.

Soybean is the major oilseed crop grown worldwide and contributes to both human and animal nutrition. Due to its great importance, great efforts are made to develop new productive cultivars. However, though conditions, such as drought, have become more frequent with the climatic changes. These periods have caused great losses in crops, in soybean losses of 24 - 50% had recorded. Thus, many studies have focused on the critical phases of soybeans, however, the dry period can occur during any phase of the crop. Thus, for understanding this complex process and assist the development of tolerant cultivars, this study aims: (i) identify molecular markers and candidate genes associated with these markers in the process of tolerance to water deficit in soybean during the germination stage; (ii) identify molecular markers and candidate genes associated with these markers in the process of tolerance to water deficit in soybean during the vegetative stage. At each stage, it carried out two experiments in a randomized block design with 3 replications in a 100 x 2 factorial (Genotypes x Water availability). The cultivars, became from different companies, representing country's soybean diversity panel. The water availability was: the control condition (Field capacity) and the stress condition (-0.2 MPa (germination) and -1.2 MPa (vegetative stage)) with a duration of 20 days. The experimental plot at the germination stage was a row containing ten treated seeds and at the vegetative stage, it was a plant in a 3-liter pot of substrate. The substrate used in germination was sand, and in the vegetative stage, it was a mixture of soil and sand. For the germination stage, the stress period began at the time of planting the seeds, while the vegetative period was when the plants entered the vegetative stage 2 (V2). For each trait, genotypic values (gBLUP) were obtained using two approaches to the concept of tolerance. The first approach used the difference between the stress condition and the control to obtain the genotypic value. In the second approach, gBLUP values were obtained using data from stress and control conditions as treatments. The gBLUP values were then used in genome-wide association studies to obtain significant markers. At the germination stage, 25 SNP-type markers were significant using the first approach and 18 markers using the second. Among these, three markers were present in both approaches. In the region close to those three markers, are five genes: one related to the cell wall, two related to protein transport, and two transcription factors. At the vegetative stage,

114 markers were significant in the first approach and 68 in the second approach. In both approaches, it found 18 molecular markers in common. In the region close to these markers are 15 genes: one transcription factor, four that have been reported in other studies of tolerance to water deficit, three related to root growth, and six that have been reported in the literature related to other types of stress.

Keywords: *Glycine max*. Abiotic stress. GWAS. Germination. Vegetative stage. Plant breeding.

SUMÁRIO

INTRODUÇÃO GERAL	11
REFERÊNCIAS BIBLIOGRÁFICAS	14
CAPÍTULO 1 – ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA EM SOJA NO ESTÁDIO DE GERMINAÇÃO REVELA NOVOS GENES PARA A TOLERÂNCIA À SECA.....	18
1. INTRODUÇÃO	20
2. MATERIAL E MÉTODOS.....	21
2.1. <i>Condução dos experimentos</i>	21
2.2. <i>Caracterização fenotípica</i>	22
2.3. <i>Análise dos dados</i>	23
2.4. <i>Genotipagem com marcadores SNP</i>	24
2.5. <i>Estrutura de populações</i>	25
2.6. <i>Análise de associação genômica (GWAS)</i>	25
2.7. <i>Busca dos genes candidatos</i>	26
3. RESULTADOS	26
4. DISCUSSÃO	38
5. CONCLUSÕES	42
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	42
7. MATERIAL SUPLEMENTAR	50
CAPÍTULO 2 – ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS) PARA A SOJA DURANTE O ESTÁDIO VEGETATIVO REVELA NOVOS GENES ENVOLVIDOS NA TOLERÂNCIA AO DÉFICIT HÍDRICO.....	58
1. INTRODUÇÃO	60
2. MATERIAL E MÉTODOS.....	61
2.1. <i>Condução dos experimentos</i>	61
2.2. <i>Avaliação dos caracteres da parte aérea</i>	62
2.3. <i>Estimação dos valores genotípicos</i>	63
2.4. <i>Genotipagem com marcadores SNP, associação genômica ampla e pós-GWAS</i>	64
3. RESULTADOS	65
4. DISCUSSÃO	77

5. CONCLUSÃO	81
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	82
7. MATERIAL SUPLEMENTAR.....	91
CONCLUSÃO GERAL	102

INTRODUÇÃO GERAL

A soja (*Glycine max* (L) Merrill) é uma planta herbácea, da classe Rosidae, ordem Fabales, família Fabaceae, subfamília Papilionoideae, tribo Phaseoleae, gênero *Glycine* L., epíteto *max*. Ela é uma planta autógama e possui o fruto tipo legume onde se encontram os grãos. Atualmente, a soja é um dos principais grãos produzidos no país e constitui a principal fonte de renda de muitas propriedades brasileiras. Essa importância se dá pelo retorno econômico e versatilidade do grão, que pode ser utilizado pela indústria, como fonte de proteína para produção animal, produção de óleo vegetal e produção de biocombustíveis, entre outras utilidades (Borém et al., 2022). Por esses motivos e dentre outros, fazem com que a soja seja uma cultura amplamente difundida, responsável por alavancar o PIB do país, juntamente com as demais cadeias do agronegócio.

Além da importância econômica, a cadeia produtiva da soja tem um papel social essencial, pois emprega, direta ou indiretamente, milhares de pessoas. A cadeia produtiva da soja reúne no país cerca de 243 mil produtores e um mercado de 1,4 milhões de empregos (Aprosoja Brasil, 2022). O papel econômico da soja para o país fica ainda mais pronunciado quando se verifica que, de quatro dólares exportados pelo complexo agroindustrial brasileiro, um dólar provém do complexo da soja (Embrapa, 2022a). Essa cultura é responsável por alavancar o PIB do país, principalmente devido as grandes adversidades vindas com a pandemia global.

A cultura é tradicionalmente cultivada no verão devido a suas características fisiológicas e exigências climáticas, sendo a principal cultura de verão e ganhando mais espaço a cada ano. Na safra 2021/2022, a oleaginosa ocupou uma área (40,8 milhões ha) 4,1% maior que safra anterior. Apesar do aumento na área plantada, a recente onda de eventos climáticos, com destaque para o déficit hídrico no Rio Grande do Sul, Paraná e Mato Grosso do Sul, fez com que a produção tivesse redução de 11,4% em relação à safra anterior (Conab, 2022).

O estresse por déficit hídrico em plantas é caracterizado pela redução do potencial de água e a pressão de turgor na folha, fechamento estomático e redução do crescimento e expansão celular (Farooq et al., 2009). A redução do crescimento é um reflexo da influência de vários fatores tanto fisiológicos e bioquímicos como fotossíntese, síntese de clorofila, metabolismo de nutriente, translocação e absorção de íons, respiração e metabolismo de carboidratos (Jaleel et al., 2008; Farooq et al., 2009; Li et al., 2011).

Estudos com a cultura da soja têm demonstrado que o impacto do déficit hídrico tem levado a redução da produtividade (24 - 50%) tanto em ambientes controlados quanto em

campo (Frederick et al., 2001; Sadeghipour & Abbasi 2012). A compreensão de respostas das plantas em face aos estresses ambientais é fundamental para o desenvolvimento das plantas em condições de campo (Pandley et al., 2015).

Os prejuízos causados pelo déficit hídrico na cultura se inicia desde a primeira fase de implantação da cultura, o plantio. Durante a germinação a falta de água irá impactar no número de plantas germinadas e consequentemente no estande final de plantas. Isto pois cada semente requer uma quantidade de água para que indução do processo de germinação. Além disso, a limitação de água no solo reduz também o vigor da semente (Kaya et al., 2006; Farooq et al., 2009). Dentre os processos envolvidos, podemos citar a formação de espécies reativas de oxigênio (ROS) que causam alterações no DNA, RNA, estruturas das proteínas, danos na membrana, e redução da respiração e produção de ATP (Priestley, 1986).

Durante a fase vegetativa, o déficit hídrico causa a redução de altura de planta, área foliar, diâmetro de hipocótilo e matéria seca em diversas culturas (Apel & Hirt, 2004; Farooq et al., 2009; Zheng et al., 2016). O crescimento foliar é basicamente reduzido por causa da redução do potencial hídrico foliar (Hsiao & Xu, 2000). Nesse caso, ocorre a interrupção de água no xilema para os tecidos levando a redução da pressão de turgor e consequentemente a redução do enlogamento celular e a redução do tamanho das plantas (Nonami, 1998; Schuppler et al., 1998; Taiz & Zeiger, 2006). No entanto, algumas cultivares, em condições de estresse menos severo, utilizam como estratégia a manutenção da área foliar. Dessa forma, quando a condição ótima é retomada, fim do estresse, elas reestabelecem o crescimento. Logo, cultivares que possuem uma manutenção de área foliar, em condições de estresse são favorecidas, quando o estresse é brando ou têm duração curta (Stolf-Moreira et al., 2010; Benjamin & Nielsen, 2006). Tal estratégia é associada com a redução da condutância estomática e por consequência na taxa fotossintética (Stolf-Moreira et al., 2010), resultando assim em uma redução da perda de água e minimizando o custo da redução da fotossíntese.

Dentre os ajustes morfológicos para a tolerância ao déficit hídrico, uma série de características relacionadas às raízes têm sido propostas (Jia et al., 2006; Liu et al., 2005; Wang et al., 2004; Lu, 2000; Garay & Wilhelm, 1983). A distribuição das raízes no perfil do solo, em relação a comprimento e/ou matéria seca irá diferir em cultivares de soja mais tolerantes em face ao déficit hídrico (Benjamin & Nielsen, 2006; Tzenova et al., 2008). Em condição de seca, as raízes tendem a aprofundar no solo, aumentando a densidade nas camadas mais profundas e apresentando baixa densidade nas camadas superficiais, onde o teor de água é menor (Garay & Wilhelm, 1983). Cultivares tolerantes apresentam uma correlação positiva entre a tolerância e

as razões: peso seco da raiz/peso da planta, comprimento total da raiz/peso da planta e volume da raiz/peso da planta (Liu et al., 2005).

A relação raiz-parte área aumenta em condições de déficit hídrico (Wu & Cosgrove, 2000). O maior crescimento da raiz em detrimento da parte área tem sido proposto por essa última ser mais sensível a condição de seca (Wu & Cosgrove, 2000). Esse crescimento está intimamente relacionado a mudanças na composição da parede celular, que envolve o espessamento da parede celular da parte aérea e o relaxamento da expansão da parede celular radicular por enzimas específicas (Wu & Cosgrove, 2000). Estudo da proteína GmRD22 sugeriu uma relação entre estresse osmótico e metabolismo da parede celular (Wang et al., 2012). Ela é uma proteína contendo um domínio BURP localizada no apoplasto, que pode desempenhar um papel na tolerância ao estresse, regulando o conteúdo de lignina da parede celular sob estresse.

O déficit hídrico afeta drasticamente os componentes de produtividade: reduzindo a produção de flores e vagens, aumentando o aborto de flores e vagens, e reduzindo o tamanho e número de sementes em leguminosas, limitando significativamente a produção (Davies et al., 1999; Fang et al., 2010). Na cultura da soja, além desses componentes, o déficit hídrico pode reduzir o engalhamento da planta afetando diretamente no número de ramos da planta, dividido a paralisação do crescimento e expansão celular (Frederick et al., 2001).

Atualmente, somente a tecnologia HB4, está licenciada para ser cultivada pelos sojicultores e apresenta tolerância ao déficit hídrico. Esse gene descoberto em girassol e inserido em diversas culturas como soja e trigo confere essa tolerância e mantém um alto teto produtivo (Ribichich et al., 2020; González et al., 2019; Gago et al., 2002). Outra tecnologia que vem sendo desenvolvida pela Embrapa é através da inserção de um gene de *Arabidopsis thaliana*, o gene *Y*, que é patenteado pelo Japan International Research Center for Agricultural Sciences (Jircas) (Embrapa, 2022b). A chaperona molecular BiP (*soyBiPD*) em soja é outro gene que em condições de casa de vegetação demonstrou conferir tolerância a soja (Valente et al., 2009). Outros genes como *GmSYP24* (Chen et al., 2019), *GmFDL19* (Li et al., 2017) e *GmMYB84* (Wang et al., 2017) também demonstraram conferir tolerância em plantas transformadas em ambientes controlados. Assim a inserção de genes que conferem resistência ao déficit hídrico nos programas de melhoramento da soja, se torna essencial para a manutenção das altas produtividades da cultura.

Apesar desses esforços, o desenvolvimento de cultivares tolerantes é complexo por envolver uma característica quantitativa. Estudos tem se concentrado nas fases críticas de demanda de água da soja, no entanto o déficit hídrico pode ocorrer em qualquer fase do ciclo

da cultura. Além disso, se faz necessário identificar e entender os genes envolvidos nesse processo de tolerância e suas interações. Neste trabalho para acessar os genes que conferem resistência ao déficit hídrico, iremos utilizar duas abordagens para o conceito de tolerância. Na primeira abordagem as plantas que tiverem menor redução da característica em relação ao controle são consideradas as mais tolerantes. Na segunda abordagem as plantas que possuem maior média na condição estresse serão as plantas consideradas mais tolerantes.

O estudo associação genômica ampla (GWAS) é uma ferramenta primordial nesse processo pois ela permite que as características fenotípicas sejam associadas às regiões no genoma responsáveis pela variação observada. Nessas regiões genômicas é possível verificar quais genes estão próximos a elas e se estes possuem envolvimento no processo de tolerância. Somente após a identificação desses fatores essenciais, podemos através do processo de seleção ou mesmo de técnicas modernas como CRISPR-Cas9 utilizá-los para o desenvolvimento de cultivares mais tolerantes.

Assim o objetivo desse trabalho foi: Identificar os marcadores SNPs e os genes candidatos associados à estes marcadores no processo de tolerância ao déficit hídrico na cultura da soja. Com os objetivos específicos: (i) identificar os marcadores SNPs e os genes candidatos associados à estes marcadores no processo de tolerância ao déficit hídrico na cultura da soja durante o estágio de germinação e (ii) identificar os marcadores SNPs e os possíveis genes candidatos associados à estes marcadores no processo tolerância ao déficit hídrico na cultura da soja durante o estágio vegetativo.

Referências bibliográficas

- Apel, K.; Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annual Review of Plant Biology*, 55: 373–399. doi: 10.1146/annurev.arplant.55.031903.141701
- APROSOJA BRASIL. A SOJA. Associação Brasileira dos Produtores de Soja. Disponível em: <<https://aprosojabrasil.com.br/a-soja/>>, acesso em: 08/02/2022
- Benjamin, J.G.; Nielsen D.C. (2006). Water deficit effects on root distribution of soybean, field pea and chickpea. *Field Crops Research*, 97: 248-253. doi: 10.1016/j.fcr.2005.10.005
- Borém, A.; Silva, F.L.; Sediyaama, T.; Camara, G. (2022). *Soja: do plantio à colheita 2ªed.* Viçosa, MG: Oficina de Textos, 312.
- Chen, L.M.; Fang, Y.S.; Zhang, C.J.; Hao, Q.-N.; Cao, D.; Yuan, S.-L.; Chen, H.-F.; Yang, Z.-L.; Chen, S.-L.; Shan, Z.-H.; Liu, B.-H.; Wang, J.; Zhan, Y.; Zhang, X.J.; Qiu, D.-Z.; Li, W.-B.; Zhou, X.-A. 2019. GmSYP24, a putative syntaxin gene, confers osmotic/drought, salt stress tolerances and ABA signal pathway. *Scientific Reports*, 9: 5990. doi: 10.1038/s41598-019-42332-5
- Conab. Produção nacional de grãos é estimada em 269,3 milhões de toneladas na safra 2021/22. Companhia Nacional de Abastecimento. Disponível em: <

- <https://www.conab.gov.br/ultimas-noticias/4579-producao-nacional-de-graos-e-estimada-em-269-3-milhoes-de-toneladas-na-safra-2021-22/>>, acesso em: 10/04/2022
- Davies, S.L.; Turner, N.C.; Siddique, K.H.M.; Plummer, J.A.; Leport, L. (1999). Seed growth of desi and kabuli chickpea (*Cicer arietinum* L.) in a short season Mediterranean-type environment. *Australian Journal of Experimental Agriculture*, 39: 181–188. doi: 10.1071/EA98134
- Embrapa. Importância socioeconômica da soja. Agência Embrapa de Informação Tecnológica. Disponível em: <https://www.agencia.cnptia.embrapa.br/gestor/soja/arvore/CONTAG01_12_271020069131.html/>, acesso em: 07/05/2022a
- Embrapa. Uma soja para enfrentar a seca. Empresa Brasileira de Pesquisa Agropecuária. Disponível em: <<https://www.embrapa.br/en/busca-de-noticias/-/noticia/2271249/uma-soja-para-enfrentar-a-seca>>, acesso em: 10/04/2022b
- Fang, X., Turner, N. C., Yan, G., Li, F., and Siddique, K. H. M. (2010). Flower numbers, pod production, pollen viability, and pistil function are reduced and flower and pod abortion increased in chickpea (*Cicer arietinum* L.) under terminal drought. *Journal of Experimental Botany*, 61: 335–345. doi: 10.1093/jxb/erp307
- Farooq, M.; Wahid, A.; Kobayashi, N.; Fujita, D.; Basra, S.M.A. (2009). Plant drought stress: effects, mechanisms and management. *Agronomy for Sustainable Development*, 29: 185–212. doi: 10.1051/agro:2008021
- Frederick, J.R.; Camp C.R.; Bauer P.J. (2001). Drought-stress effects on branch and mainstem seed yield and yield components of determinate soybean. *Crop Science*, 41: 759-763. doi: 10.2135/cropsci2001.413759x
- Gago, G.M.; Almoguera, C.; Jordano, J.; Gonzalez, D.H.; Chan, R.L. (2002). Hahb-4, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower. *Plant, Cell and Environment*, 25: 633–640. doi: 10.1046/j.1365-3040.2002.00853.x
- Garay, A.F.; Wilhelm, W. (1983). Root system characteristics of two soybean isolines undergoing water stress condition. *Agronomy Journal*, 75: 973-977. doi: 10.2134/agronj1983.00021962007500060026x
- Gonzalez, F.G.; Capella, M.; Ribichich, K.F.; Curin, F.; Giacomelli, J.I.; Ayala, F.; Watson, G.; Otegui, M.E.; Chan, R.L. (2019). Field-grown transgenic wheat expressing the sunflower gene HaHB4 significantly outyields the wild type. *Journal of Experimental Botany*, 70: 1669–1681. doi: 10.1093/jxb/erz037
- Hsiao, T.C.; Xu, L.K. (2000). Sensitivity of growth of roots versus leaves to water stress: biophysical analysis and relation to water transport. *Journal of Experimental Botany*, 51: 1595-616. doi: 10.1093/jexbot/51.350.1595
- Jaleel, C.A.; Gopi, R.; Sankar, B.; Gomathinayagam, M.; Panneerselvam, R. (2008). Differential responses in water use efficiency in two varieties of *Catharanthus roseus* under drought stress. *Comptes Rendus Biologies*. 331, 42–47. doi: 10.1016/j.crv.2007.11.003
- Jia, Q.S.; Wei, L.; Yang, H.F. (2006). Primary report on the relationship between the root system and drought resistance in soybean seedlings. *Shaanxi Journal of Agricultural Sciences*, 2: 12-13.

- Kaya, M.D.; Okçu, G.; Ataka, M.; Çikili, Y.; Kolsarici, Ö. (2006). Seed treatments to overcome salt and drought stress during germination in sunflower (*Helianthus annuus* L.). *European Journal of Agronomy*, 24: 291–295. doi: 10.1016/j.eja.2005.08.001
- Li, Y.; Chen, Q.; Nan, H.; Li, X.; Lu, S.; Zhao, X.; Liu, B.; Guo, C.; Kong, F.; Cao, D. (2017). Overexpression of GmFDL19 enhances tolerance to drought and salt stresses in soybean. *Plos One* 12, e0179554. doi: 10.1371/journal.pone.0179554
- Li, Y.; Zhao, H.X.; Duan, B.L.; Korpelainen, H.; Li, C.Y. (2011). Effect of drought and ABA on growth, photosynthesis and antioxidant system of *Cotinus coggygria* seedlings under two different light conditions. *Environmental and Experimental Botany*, 71: 107–113. doi: 10.1016/j.envexpbot.2010.11.005
- Liu, Y.; Gai, J.Y.; Lu, H.N.; Wang, Y.J.; Chen S.Y. (2005). Identification of drought tolerant germplasm and inheritance and QTL mapping of related root traits in soybean (*Glycine max* (L.) Merr.). *Yi Chuan Xue Bao*, 32: 855-863.
- Lu, G.H. (2000). Studies on root properties and drought-resistance for different types of drought. *Journal of Shanxi Agricultural Sciences*, 28: 37-40.
- Nonami, H. (1998). Plant water relations and control of cell elongation at low water potentials, *Journal of Plant Research*, 111: 373–382. doi: 10.1007/BF02507801
- Pandey, P.; Ramegowda, V.; Senthil-Kumar, M. (2015). Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. *Frontiers in Plant Science*. 6:723. doi: 10.3389/fpls.2015.00723
- Priestley, D. A. (1986). *Seed Ageing: Implications for Seed Storage and Persistence in the Soil*. Ithaca, NY: Comstock Associates, 304.
- Ribichich, K.F.; Chiozza, M.; Ávalos-Britez, S.; Cabello, J.V.; Arce, A.L.; Watson, G.; Arias, C.; Portapila, M.; Trucco, F.; Otegui, M.E.; Chan, R.L. (2020). Successful field performance in warm and dry environments of soybean expressing the sunflower transcription factor HB4. *Journal of Experimental Botany*, 71: 3142–3156. doi: 10.1093/jxb/eraa064
- Sadeghipour, O.; Abbasi, S. (2012). Soybean response to drought and seed inoculation. *World Applied Sciences Journal*, 17: 55-60.
- Schuppler, U.; He, P.; John, P.C.L.; Munns, R. (1998). Effect of water stress on cell division and cell-division-cycle 2-like cell-cycle kinase activity in wheat leaves. *Plant Physiology*, 117: 667–678. doi: 10.1104/pp.117.2.667
- Stolf-Moreira, R.; Medri, M.E.; Neumaier, N.; Lemos, N.G.; Pimenta, J.A.; Tobita, S.; Brogin R.L.; Marcelino-Guimarães, F.C.; Oliveira, M.C.N.; Farias, J.R.; Abdelnoor, R.V.; Nepomuceno, A.L. (2010). Soybean physiology and gene expression during drought. *Genetics and Molecular Research*, 9: 1946-1956. doi: 10.4238/vol9-4gmr851
- Taiz, L.; Zeiger, E. (2006). *Plant Physiology*, 4th Edn. Sunderland, MA: Sinauer Associates Inc.
- Tzenova, V.; Kirkova, Y.; Stoimenov, G. (2008). *Methods for plant water stress evaluation of soybean canopy*, in Balwois 2008 - Water Observation and Information System for Decision Support 2008: Ohrid, Republic of Macedonia.
- Valente, M.A.; Faria, J.A.; Soares-Ramos, J.R.; Reis, P.A.; Pinheiro, G.L.; Piovesan, N.D.; Morais, A.T.; Menezes, C.C.; Cano, M.A.; Fietto, L.G.; Loureiro, M.E.; Aragão, F.J.; Fontes, E.P. (2009). The ER luminal binding protein (BiP) mediates an increase in

- drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco. *Journal of Experimental Botany*, 60: 533-46. doi: 10.1093/jxb/ern296
- Wang, H.; Zhou, L.; Fu, Y.; Cheung, M.Y.; Wong, F.L.; Phang, T.H.; Sun, Z.; Lam, H.M. (2012). Expression of an apoplast-localized BURP-domain protein from soybean (GmRD22) enhances tolerance towards abiotic stress. *Plant, Cell & Environment*, 35: 1932-1947. doi: 10.1111/j.1365-3040.2012.02526.x
- Wang, M.; Cheung, C.Y.; Ma, T.F.; Yao, W.C. (2004). Studies on the drought resistance of seedling in soybean. *Chinese Journal of Oil Crop Science*, 26: 29-32.
- Wang, N.; Zhang, W.; Qin, M.; Li, S.; Qiao, M.; Liu, Z.; Xiang, F. (2017). Drought tolerance conferred in soybean (*Glycine max* L.) by GmMYB84, a novel R2R3-MYB transcription factor. *Plant & Cell Physiology*, 58: 1764–1776. doi:10.1093/pcp/pcx111
- Wu, Y.; Cosgrove, D.J. (2000). Adaptation of roots to low water potentials by changes in cell wall extensibility and cell wall proteins. *Journal of Experimental Botany*, 51: 1543-1553. doi: 10.1093/jexbot/51.350.1543
- Zheng, M.; Tao, Y.; Hussain, S.; Jiang, Q.; Peng, S.; Huang, J.; KeHui, C.; LiXiao, N. (2016). Seed priming in dry direct-seeded rice: consequences for emergence, seedling growth and associated metabolic events under drought stress. *Plant Growth Regulation*, 78: 167–178. doi: 10.1007/s10725-015-0083-5

CAPÍTULO 1 – ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA EM SOJA NO ESTÁDIO DE GERMINAÇÃO REVELA NOVOS GENES PARA A TOLERÂNCIA À SECA

RESUMO

A soja é uma das culturas mais importantes mundialmente, em que os dois maiores produtores contribuem com mais de 60% da produção. Dentre estes países, se encontra o Brasil, como maior produtor mundial do grão. A produção desse grão tem crescido anualmente, com o aumento da produtividade e as áreas plantadas. No entanto, com o aumento da frequência de eventos climáticos como a seca, grandes perdas têm se tornado frequentes. Na busca de novas tecnologias para escape dessa condição e para um uso mais racional e produtivo das áreas agrícolas, a importância da época do plantio tem se destacado. A umidade das sementes é crucial para o processo de germinação, a falta de água no início da germinação compromete a planta morfológica, molecularmente e bioquimicamente, levando assim, a uma menor taxa de germinação e redução do vigor das plantas. Assim, cultivares tolerantes à seca durante o período de germinação tem se tornado cada vez mais importante. Dessa forma, este estudo tem o objetivo identificar os marcadores SNPs e os genes candidatos associados à estes marcadores no processo de tolerância ao déficit hídrico na cultura da soja durante o estágio de germinação. Foram conduzidos dois experimentos em delineamento em blocos casualizados em esquema fatorial 100 x 2 (genótipos x disponibilidade hídrica) com três repetições. Os cultivares utilizados são provenientes de diversas empresas representando um painel de diversidade. Os regimes hídricos foram a condição controle (-0,01 MPa) e a condição Estresse (-0,2 MPa). A parcela experimental constitui de uma linha de 10 sementes semeadas em areia, onde na condição controle utilizou-se apenas água e na condição estresse uma solução de polietilenoglicol para atingir o potencial desejado. A água evapotranspirada foi repostada diariamente e o número de plantas emergidas foi computado. Após 20 dias, as características fenotípicas desses cultivares foram avaliadas. Durante a análise dos dados foram consideradas duas abordagens para o conceito de cultivares tolerantes. Na primeira abordagem, a planta que teve menor redução da sua característica na condição de estresse em relação à condição controle foi considerada a mais tolerante. Na segunda abordagem, a planta mais tolerante foi aquela que teve maior produção levando em conta as condições estresse e controle. Foi realizada uma análise de PCA-biplot com a O valor genotípico para cada característica foi estimado via modelos mistos REML/gBLUP e realizado a correlação entre eles. Em ambas abordagens, foram selecionadas as características: conteúdo de água na folha (wc), taxa de emergência de

plantas (emer) e índice de velocidade de germinação (gsi). Essas características foram selecionadas para a realização do GWAS, onde se utilizou a abordagem bayesiana (Blink) e foram identificados ao total 25 marcadores moleculares do tipo SNP associados a tolerância ao déficit hídrico na primeira abordagem e 18 na segunda abordagem. Dentre esses, três marcadores apresentaram significância em ambas as abordagens e na região próxima a esses foram encontrados cinco genes. Destes genes, um está envolvido na constituição da parede celular, dois genes codificam para proteínas transportadoras entre membranas e os outros dois são fatores de transcrição.

Palavras-chave: *Glycine max*. Deficit hídrico. Melhoramento de plantas

1. Introdução

A soja é uma das principais culturas plantadas atualmente, trata-se de um cultivo de verão independentemente da região cultivada no mundo. A produção mundial desse grão é de cerca de 395.370.000 toneladas por ano (USDA, 2022). Os dois maiores países produtores são o Brasil e Estados Unidos, que contribuem juntos com cerca de 60% da produção mundial.

Com o avanço da tecnologia, o uso racional das áreas e busca por alternativas mais lucrativas nos cultivos, o plantio cada vez mais cedo tem sido preconizado tanto nas regiões temperadas quanto nas tropicais (Maury et al., 2015). Contudo, devido a mudanças climáticas que vem ocorrendo nos últimos anos, períodos de seca tem sido cada vez comum, com previsão de aumento da ocorrência em 29% (UNCCD, 2022).

O deficit hídrico pode ocorrer em qualquer fase da cultura e afeta em diferente escala dependendo do estágio da planta: germinação, vegetativo ou reprodutivo (Anjum et al., 2017). Durante a germinação, o deficit hídrico afeta os atributos morfológicos (tamanho e biomassa de plantas) e moleculares (proteínas do estresse, aquaporinas e dehidrinas) e bioquímicos (atividade da amilase, protease e lipase) (Yigit et al., 2016). Além disso, esse estresse leva à redução da taxa de vigor e do índice de germinação (Tang et al., 2019), assim como alteração de características associadas ao desenvolvimento de raízes (Toscano et al., 2017).

Durante a germinação, os principais hormônios envolvidos são o ácido giberélico (GA), responsável pela ativação de enzimas, e o ácido abscísico (ABA) que é antagonico ao anterior. No período de deficit hídrico, os níveis de ABA são afetados e alteram o processo de germinação (Liu et al., 2019). Dentre os efeitos desse estresse, a produção de espécies reativas de oxigênio (ROS) é um dos primeiros processos afetados, e estes alteram as proteínas e os ácidos nucleicos por modificações por oxidação (Lofti et al., 2019). Para lidar com tais processos, as plantas possuem mecanismos antioxidantes enzimáticos e não enzimáticos. Outro processo comum durante o estresse é a produção de produtos osmoprotetores como prolina, glicina betaína, açúcar solúveis que ajudam na estabilização da estrutura celular, ajustes osmóticos e proteção contra danos oxidativos (Blum, 2017).

Em nível molecular, durante o processo de germinação são induzidos a síntese de proteínas MYB, fatores de transcrição responsivos a GA. Os fatores de transcrição MYB, NAC, MADS, WRKY, ZFHD, MYC e DREB induzem a produção de proteínas LEA (“Late Embryogenesis Abundant”), HSP (“Heat Shock Protein”), DHN (“Dehidrinas”) juntamente com o acúmulo de genes antioxidantes: superóxido dismutase (SOD), catalase (CAT),

peroxidase (POX), malondialdeído (MDA), polifenol oxidase (PPO), glutathione redutase (GR) e ascorbato peroxidase (APX) (Saha et al., 2022).

A regulação ocorre também nas proteínas que controlam a condutividade hidráulica como é o caso das aquaporinas (*PIP*). Os genes *PIP* aumentam sua expressão em condições de seca e controlam a condutividade de água nas raízes de forma a aumentar a absorção de água (Mukarram et al., 2021).

O processo de tolerância ao deficit hídrico é complexo controlado por diversos genes e envolve diversas características comuns (Fukao & Xiong, 2013). Para estudos de características complexas, uma ferramenta amplamente utilizada para identificar regiões e posteriormente os genes candidatos é o estudo de associação genômica ampla (GWAS) (Rathan et al., 2022; Cano-Gamez & Trynka, 2020; Korte & Farlow, 2013).

Outra questão relacionada à tolerância durante a germinação é que muitos estudos vêm sendo realizados em condições mais distantes das que ocorrem em campo. Por exemplo, em estudos de seleção de genótipos tolerantes, durante a fase de germinação, muitas vezes são realizados testes em laboratório com uso de papel germitest embebido com solução de Polietilenoglicol para simular as tensões desejadas (Kpoghomou et al. 1990; Khan et al., 2018; Liu et al., 2020; Begum et al., 2022). A utilização de areia tem sido preterida para estudos de germinação por proporcionar resistência ao processo de germinação simulando condições mais próximas às encontradas no campo.

Assim esse presente trabalho tem como objetivo identificar os marcadores SNPs e os genes candidatos associados à estes marcadores no processo de tolerância ao deficit hídrico na cultura da soja durante o estágio de germinação.

2. Material e Métodos

2.1. Condução dos experimentos

Foram conduzidos dois experimentos em casa de vegetação na Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil (latitude 20°45'17" S, longitude 42°52'57" O). O experimento 1 foi conduzido em abril de 2021 e o experimento 2 em junho de 2021.

O delineamento utilizado foi o de blocos casualizados em esquema fatorial 100 x 2 (genótipos x disponibilidade hídrica) com três repetições. Os genótipos utilizados foram oriundos de diversas empresas com ampla variabilidade com o objetivo de representar um painel de diversidade (Tabela S1). As disponibilidades hídricas utilizadas foram a condição estresse em que as plantas foram germinadas em uma tensão de -0,2 MPa e a condição controle em que as plantas foram germinadas na capacidade de campo para o material inerte (areia), ou

seja, na tensão de -0,01 MPa (Bernardo et al., 2019). A parcela experimental foi composta por dez sementes dispostas em uma linha continua em bandejas de plástico (42x28x8 cm) preenchidas com areia.

As sementes foram tratadas com as doses recomendadas pelo fabricante com o fungicida fludioxonil (200 mL/100Kg de sementes) (MAXIM®) e o inseticida Imidacloprido + Tiodicarbe (CROPSTAR®) (300 mL/100 Kg de sementes). Após a secagem, as sementes foram inoculadas com uma solução de *Bradyrhizobium japonicum* (RIZOKOP®) (100 mL/50 Kg de sementes) e semeadas.

Para a condição controle, foi utilizado apenas água destilada. Na condição estresse foi utilizado uma solução de Polietilenoglicol (PEG 6000) de acordo com Villela et al. (1991) para atingir a tensão de (-0,2 MPa). Para determinar o volume de solução utilizado em cada bandeja, foi adotado o método gravimétrico para determinação da capacidade de campo. Nesse método uma quantidade conhecida de areia (g) é umedecida até a sua saturação total com uma quantidade conhecida de água (mL). Após essa mistura é colocada em um funil de vidro com um tecido de organza, para impedir a passagem da areia. A água que não ocupa os macros e micropóros da areia é totalmente drenada. O volume de água que ficou retida, representa a quantidade de solução ou água por grama de areia que deve ser adicionada. As bandejas foram preenchidas com uma quantidade conhecida de areia (kg) e adicionado a solução ou água dependendo do tratamento.

O sistema então foi aferido com auxílio de uma balança e o peso inicial anotado e utilizado para a correção diária do volume de água perdido. Diariamente cada bandeja foi pesada e o seu volume de água perdido repostado de acordo com o seu respectivo peso inicial.

2.2. Caracterização fenotípica

Para a determinação dos parâmetros relacionados à emergência de plantas, o número de plantas emergidas foi anotado diariamente e no dia 20, correspondente ao final do experimento, foi contabilizado o número total de plantas emergidas. A planta foi considerada emergida a partir do momento em que o cotilédone não tocava mais o substrato. Os dados diários de emergência foram submetidos ao programa SeedCalc (Silva et al., 2019) e assim foi obtido o índice de velocidade de emergência (gsi). Após, a contagem final de plantas foi convertida em porcentagem de plantas emergidas para cada parcela (emer).

Na determinação do conteúdo de água nas folhas, foram coletados três discos foliares de 1 cm de diâmetro, sendo cada disco proveniente de uma planta diferente dentro da parcela experimental. Esses discos foram pesados em uma balança de precisão e depois colocados para

secar em estufa à uma temperatura de 65°C por 48 horas. Os discos secos foram pesados novamente e o conteúdo de água foi determinado por diferença. O conteúdo de água nas folhas (wc) (%) foi determinado pela divisão do conteúdo de água pela massa de matéria seca dos discos e, esse resultado, multiplicado por 100.

Para as características da altura de plantas, massa seca da parte aérea e da raiz foram utilizadas apenas as quatro plantas centrais da parcela por representarem uma boa medida de repetibilidade (Duarte et al., 2022). As plantas foram coletadas e a sua parte aérea separada da parte radicular. Na parte aérea foi medido a altura (cm) e após foi avaliado o conteúdo de massa seca (g) em estufa (65°C por 48h). Já as raízes foram lavadas em água para remoção de qualquer material inerte aderido e então levadas à estufa (65°C por 48h) para remoção do conteúdo de água e posterior aferição da massa seca das raízes (g). O valor médio de cada parcela foi utilizado para as características de altura (ph), massa seca da parte aérea (sdw) e massa seca das raízes (rdw).

Para as outras características relacionadas às raízes foram utilizadas apenas uma planta central por parcela. Essa planta removida do centro da parcela foi armazenada em álcool 70% e armazenada em 4 °C até o momento da avaliação. As raízes foram escaneadas isoladamente e a imagem obtida foi submetida ao software WinRHIZO® 2009. Assim foram obtidas as variáveis: comprimento total (cm) (lpvr), área superficial projetada (cm²) (apr), área superficial (cm²) (sar), volume (cm³) (rv), diâmetro médio (mm) (adr), comprimento da raiz por classes de diâmetros (cm) (rcl) e volume da raiz por classes de diâmetros (cm³) (rcv). As classes de diâmetro (D) foram: classe 1 ($0 < D \leq 0.50$), classe 2 ($0.50 < D \leq 1.00$), classe 3 ($1.00 < D \leq 1.50$), classe 4 ($1.50 < D \leq 2.00$), classe 5 ($2.00 < D \leq 2.50$), classe 6 ($2.50 < D \leq 3.00$), classe 7 ($3.00 < D \leq 3.50$), classe 8 ($3.50 < D \leq 4.00$), classe 9 ($4.00 < D \leq 4.50$), e classe 10 ($D > 4.50$).

2.3. Análise dos dados

Os valores para as características taxa de emergência e índice de velocidade de germinação, foram plotados em boxplot para cada condição de regime hídrico dentro de cada experimento. As médias foram comparadas com o teste t, $p < 0,01$. Foi realizado o teste de ANOVA para cada característica avaliada.

Foram utilizadas duas abordagens para o conceito de tolerância. Na primeira abordagem foi considerado que um genótipo é mais tolerante se na condição estresse a redução da característica em relação a condição controle é menor que de outro genótipo. Para isso, o valor de cada característica na condição estresse dentro de cada bloco foi subtraído pelo valor da característica na condição controle. Na segunda abordagem, uma planta foi considerada mais

tolerante se tiver um maior desempenho da característica tanto na condição estresse quanto na condição controle. Assim aquele genótipo que tiver melhor desempenho na condição controle e tiver melhor desempenho na condição estresse será selecionado.

Todas as análises foram realizadas no software R (R CORE TEAM, 2016) por meio dos pacotes específicos relacionados a cada análise realizada como se segue.

Os dados foram corrigidos a partir de modelos lineares mistos, via REML/gBLUP para cada característica utilizando o pacote “sommer” (Covarrubias-Pazaran, 2016). Para a determinação da matriz de parentesco foi utilizado a matriz dos marcadores utilizando o método proposto por Zhang et al. (2007). Neste modelo, cada experimento foi considerado uma época, e foi inserido como um fator aleatório. O modelo utilizado para cada abordagem se encontra representado abaixo:

1ª Abordagem:

$$y = \mu + Xr + Zg + Wp + e$$

2ª Abordagem:

$$y = \mu + Xr + Ti + Zg + Wp + e$$

onde “y” é o vetor de dados fenotípicos, “μ” é a média geral, “r” é o vetor dos efeitos de repetição (assumidos como fixos) dentro de cada época, “i” é o vetor dos efeitos de tratamento (assumidos como fixos), “g” é o vetor dos efeitos do i-ésimo genótipo (assumidos como aleatórios), “p” é o vetor dos efeitos de época (assumidos como aleatórios), e “e” é o vetor de erros ou resíduos (aleatórios) que segue $N(0, \sigma^2)$. As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

As características que apresentaram valores de herdabilidade menor que 1 % ($h^2 \leq 0,01$) foram descartadas dentro de cada abordagem. Para inferir as características mais relevantes ao processo de tolerância ao déficit hídrico foi realizado o PCA biplot das características, utilizando o pacote ggbiplot (Vu et al., 2015) dentro de cada abordagem. As características que se orientaram no sentido da separação dos grupos estresse e controle foram selecionadas para análises posteriores.

Foi verificado a correlação entre os valores de gBLUP das características selecionadas para cada abordagem por meio de uma análise de correlação de Pearson, utilizando o pacote “corrplot” (Wei & Simko, 2021). As características que não tiveram correlação com as características principais para essa fase, emergência e índice de velocidade de emergência, foram descartadas.

2.4. Genotipagem com marcadores SNP

A genotipagem com marcadores SNP nessas cultivares foi realizada previamente por Contreras-Soto et al. (2017a). Essa genotipagem foi realizada na Deoxi Biotecnologia Ltda® (Araçatuba, SP, Brazil) utilizando o “Infinium iSelect HD Custom Genotyping BARC-Soy

SNP6k” da plataforma iScan Illumina (Illumina, Inc., San Diego, CA, USA) customizada para soja. Marcadores SNP com mais de 10% de perda de dados e “minor allele frequency” menor que 1% foram descartados.

2.5. Estrutura de populações

Para evitar a identificação de associações falso-positivas entre os dados fenotípicos e moleculares, foi realizado uma análise prévia da estrutura de populações. Assim, a análise da estrutura genética da população (matriz Q), foi realizada por meio da Análise Discriminante de Componentes Principais (DPCA) da matriz de marcadores moleculares. Para isso, foi utilizado o pacote “adegenet” (Jombart & Ahmed, 2011) no software R (R CORE TEAM, 2016).

2.6. Análise de associação genômica (GWAS)

Os valores de gBLUP das características selecionadas para cada abordagem após a etapas descritas acima foram utilizados no estudo de associação genômica ampla (GWAS). A análise GWAS foi realizada utilizando o pacote GAPIT versão 3.0 (Wang & Zhang, 2021). Para análise foi considerado o modelo “Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway” (Blink) (Huang et al., 2019). Esse modelo conduz dois modelos de efeito fixos (FEM) e um processo de filtragem que seleciona um conjunto de pseudo QTNs que não estão em LD entre si como covariáveis. No primeiro modelo FEM, os marcadores são testados, um de cada vez, e pseudo QTNs são incluídos como covariáveis para controlar os falsos negativos e positivos:

$$y_i = S_{i1}b_1 + S_{i2}b_2 + \dots + S_{ik}b_k + S_{ij}d_j + e_i$$

onde y_i é a observação no i -ésimo indivíduo; $S_{i1}, S_{i2}, \dots, S_{ik}$ são os genótipos de k pseudo QTNs, iniciados como um conjunto vazio; $b_1, b_2, \dots, b_k$ são os efeitos correspondentes dos pseudo QTNs; S_{ij} é o genótipo do i -ésimo indivíduo e j -ésimo marcador genético; d_j é o efeito correspondente do j -ésimo marcador genético; e e_i é o resíduo com distribuição com média zero e variância σ_e^2 .

O segundo modelo FEM, é utilizado para otimizar a seleção de pseudo QTNs:

$$y_i = S_{i1}b_1 + S_{i2}b_2 + \dots + S_{ik}b_k + e_i$$

onde y_i é a observação no i -ésimo indivíduo; $S_{i1}, S_{i2}, \dots, S_{ik}$ são os genótipos de k pseudo QTNs, iniciados como um conjunto vazio; $b_1, b_2, \dots, b_k$ são os efeitos correspondentes dos pseudo QTNs; S_{ij} é o genótipo do i -ésimo indivíduo; e e_i é o resíduo com distribuição com média zero e variância σ_e^2 .

No segundo modelo, o termo marcador teste é removido e o número de pseudo QTNs covariáveis é variado, a fim de selecionar o conjunto ótimo do primeiro k de t pseudo QTNs. A otimização é realizada usando BIC, que é o dobro da probabilidade logarítmica negativa mais a penalidade no número de parâmetros, como segue:

$$BIC = -2LL + 2kLn(n)$$

onde LL é o log da probabilidade, k é o número de pseudo QTNs, Ln é o logaritmo natural e n é o número de indivíduos. Os pseudo QTNs disponíveis, t , são classificados com o mais significativo no

início e o menos significativo no final. Os primeiros k pseudo QTNs são selecionados para exame, com k variando de 1 a t .

No segundo modelo, os marcadores são filtrados com dois critérios: P-valor e correlação. Primeiramente, todos os marcadores do primeiro modelo são filtrados com correção de Bonferroni ($\alpha = 0,01$). Após os marcadores com correlação (correlação de Pearson) maior que 0,7 são eliminados. Então os primeiros conjuntos de k marcadores foram ajustados e os demais foram variados até que todas as possibilidades tenham sido examinadas. O conjunto de K marcadores com melhor BIC é usado como o conjunto de pseudo QTNs na primeira equação. O processo é repetido até que os pseudo QTNs permaneçam os mesmos. Ao final, os marcadores que apresentaram valores de $p < 0,01$ após a correção pela FDR foram considerados significativos.

2.7. Busca dos genes candidatos

A busca dos genes associados foi realizada na região genômica de 2000 bp à montante (“up”) e à jusante (“down”) de cada marcador significativo. A busca nessa região foi realizada através do “Genome Data Viewer” do NCBI utilizando o genoma da soja na versão 4.1 (*Glycine max* Wm82.a4.v1). Os genes encontrados dentro dessa região foram selecionados. Os marcadores que foram encontrados dentro da região codificadora do gene foram verificados para possíveis modificações na estrutura do gene via substituição alélica.

3. Resultados

A taxa de emergência de plantas na condição controle foi de 98,91% com erro padrão de 3,32 para o primeiro experimento e 98,31% com erro padrão de 3,75 para o segundo experimento (Figuras 1a e 1b). Nenhuma parcela no controle teve germinação menor que 80%. Na condição estresse, que diferiu significativamente da condição controle pelo teste t ($p < 0,01$), foram de 64,57 e 69,93 % para o primeiro e segundo experimento, respectivamente. As taxas de emergência na condição estresse variaram de 0 a 100% com desvio padrão da média de 22,89 nos dois experimentos. O índice de velocidade de emergência também diferiu significativamente entre os tratamentos em ambos os experimentos pelo teste t ($p < 0,01$) (Figuras 1c e 1d).

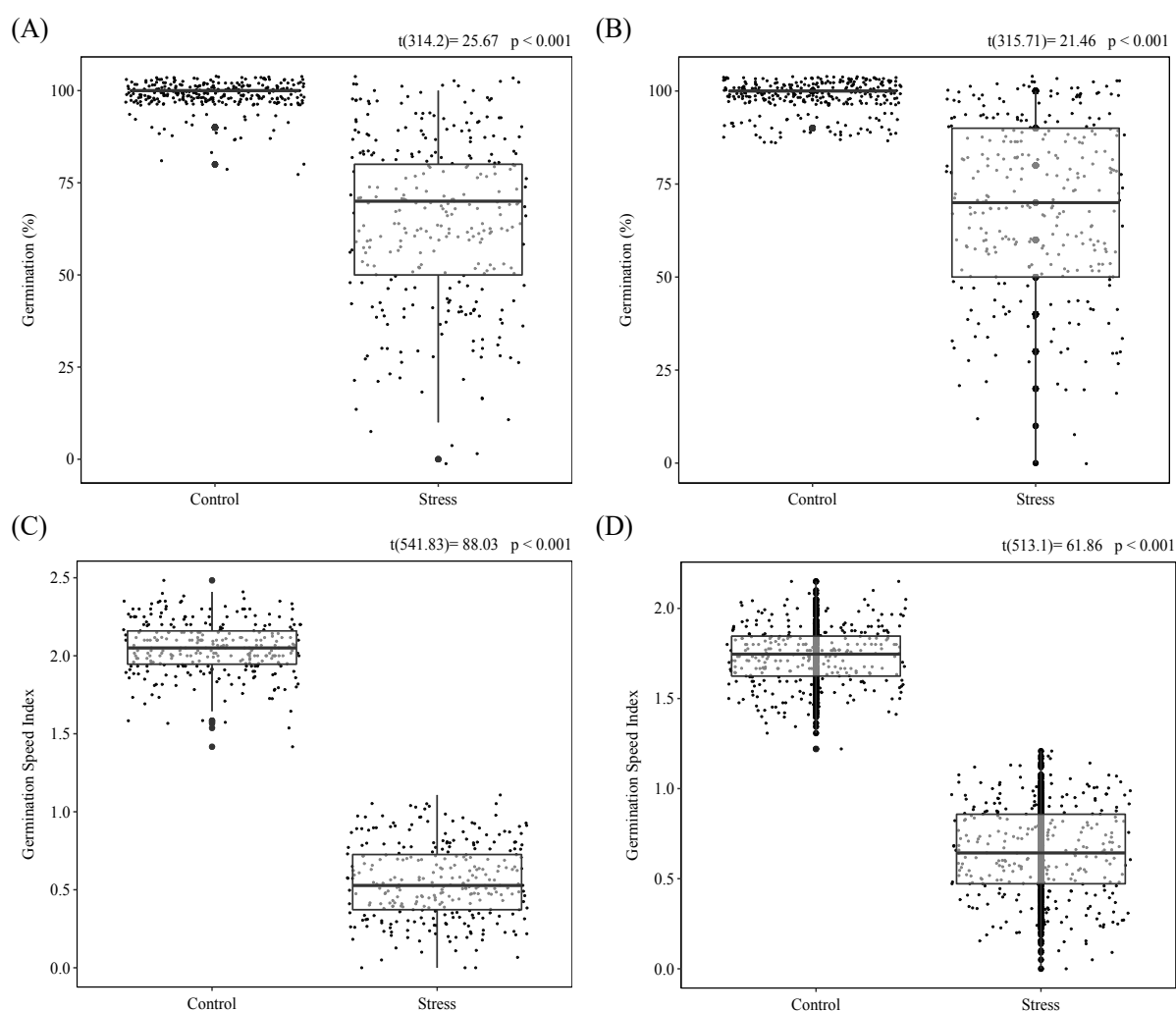


Figura 1 – Taxa de emergência e índice de velocidade de emergência nos dois experimentos. Média e dispersão dos dados de taxa de emergência no experimento 1 (A) e no experimento 2 (B). Média e dispersão do índice de velocidade de emergência no experimento 1 (C) e no experimento 2 (D). As médias foram significativamente diferentes pelo teste t ($p < 0,001$).

Considerando a primeira abordagem, em que é contabilizada a diferença entre tratamentos, apenas as variáveis: conteúdo de água nas folhas (wc), taxa de emergência de plantas (emer), índice de velocidade de emergência (gsi), comprimento da raiz por classes de diâmetros 2 (rc12), comprimento da raiz por classes de diâmetros 5 (rc15), volume da raiz por classes de diâmetros 2 (rcv2), volume da raiz por classes de diâmetros 5 (rcv5), não apresentaram herdabilidade menor que 1% (Tabela S2). A análise dos dados fenotípicos obtidos via PCA biplot, onde os dados brutos foram plotados junto com a análise dos componentes principais gerou dois grupos distintos (estresse e controle) demonstrando a eficácia dos experimentos (Figura 2).

Na análise de PCA, o primeiro componente explicou 43,6 % e o segundo 26,3 % da variância. Dentre os autovetores plotados, somente os correspondentes a comprimento por

classe de diâmetro 5 (rcl5) e o volume por classe de diâmetro 5 (rcv5) não se orientam na direção a nenhum grupo formado (controle ou estresse). Assim essas características foram descartadas para as próximas análises.

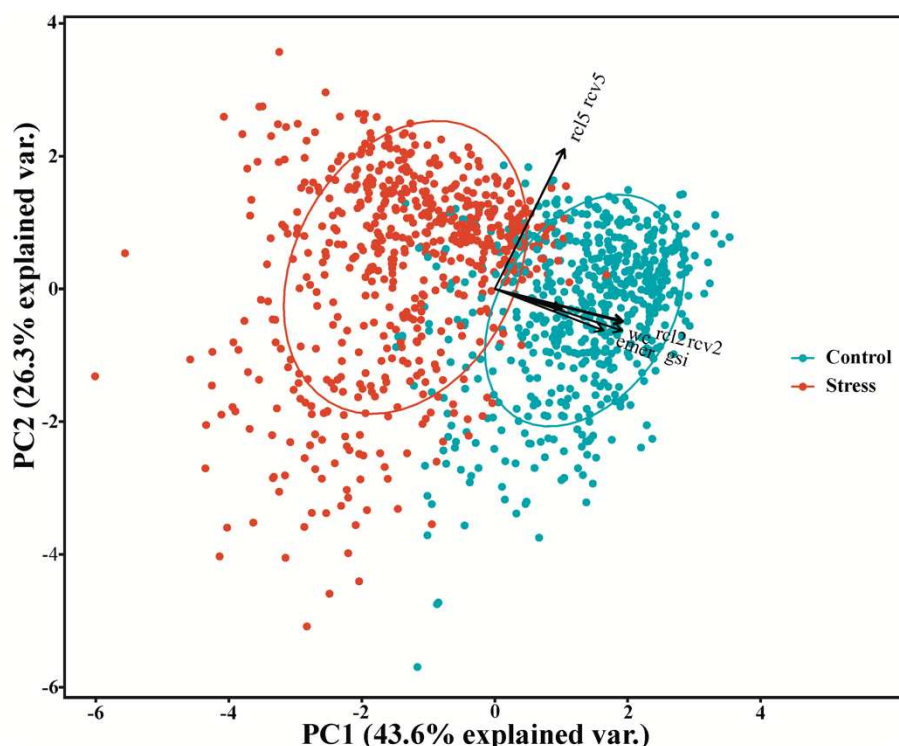


Figura 2 – PCA-biplot das variáveis selecionadas para a abordagem 1, que possuem herdabilidade maior que 1 %.

wc: Conteúdo de água nas folhas, emer: taxa de emergência de plantas, gsi: índice de velocidade de emergência, rcl2: comprimento da raiz por classes de diâmetros 2, rcl5: comprimento da raiz por classes de diâmetros 5, rcv2: volume da raiz por classes de diâmetros 2, rcv5: volume da raiz por classes de diâmetros 5.

Ao realizar a correlação dos valores de gBLUP das características selecionadas, verificou-se que a correlação foi alta e positiva (0,86) entre a taxa de emergência (emer) e o índice de velocidade de germinação (gsi) (Figura 3). A característica conteúdo de água foliar (wc) apresentou uma correlação moderada e negativa com emer (-0,47) e gsi (-0,41). As características comprimento por classe de diâmetro 2 (rcl2) e volume por classe de diâmetro 2 (rcv2) apresentaram correlação muito alta e positiva entre si (0,99) e não apresentaram correlação com as demais avaliadas.

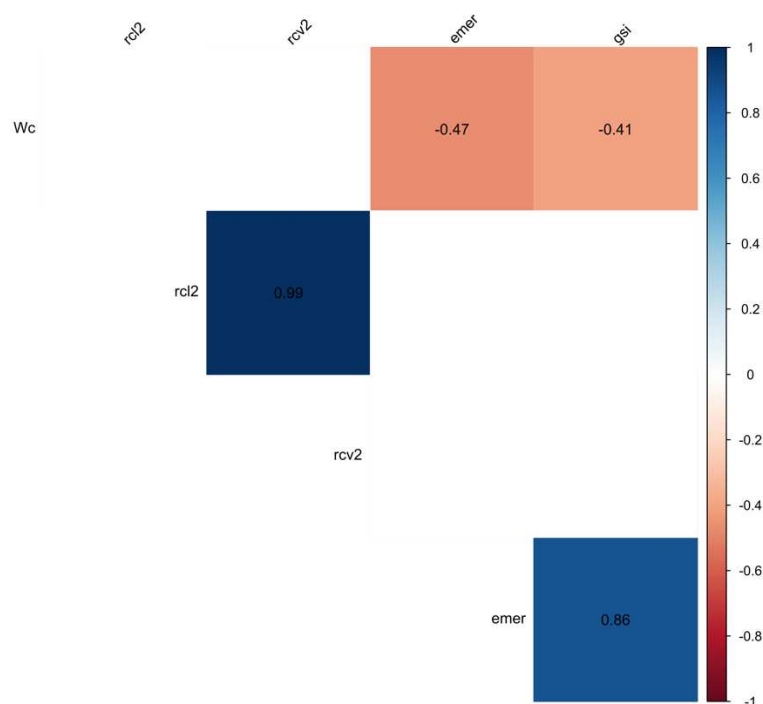


Figura 3 - Coeficientes da análise de correlação de Pearson entre os valores genotípicos preditos (BLUP) para a abordagem 1.

wc: Conteúdo de água nas folhas, emer: taxa de emergência de plantas, gsi: índice de velocidade de emergência, rcl2: comprimento da raiz por classes de diâmetros 2, rcv2: volume da raiz por classes de diâmetros 2.

Ao utilizar a segunda abordagem para o conceito de tolerância somente as variáveis taxa de emergência de plantas (emer), conteúdo de água nas folhas (wc), índice de velocidade de germinação (gsi), massa seca da raiz (rdw), comprimento por classe de diâmetro 2, 3, 4 e 5 (rcl2, rcl3, rcl4 e rcl5) e volume por classe de diâmetro 2, 3, 4 e 5 (rcv2, rcv3, rcv4 e rcv5) apresentaram herdabilidade maior que 1% (Tabela S3). Os tratamentos estresse e controle formaram grupos distintos ao realizar a análise de PCA-biplot com as variáveis selecionadas (Figura 4). O primeiro componente explicou 45,9 % da variância e o segundo explicou 17,5% da variância total. Dentre as variáveis utilizadas apenas as variáveis comprimento e volume por classe de diâmetro 5 (rcl5 e rcv5) não se orientaram em oposição aos grupos formados e assim não possuem grande contribuição para a separação dos grupos estresse e controle. Assim elas foram descartadas para a próxima análise.

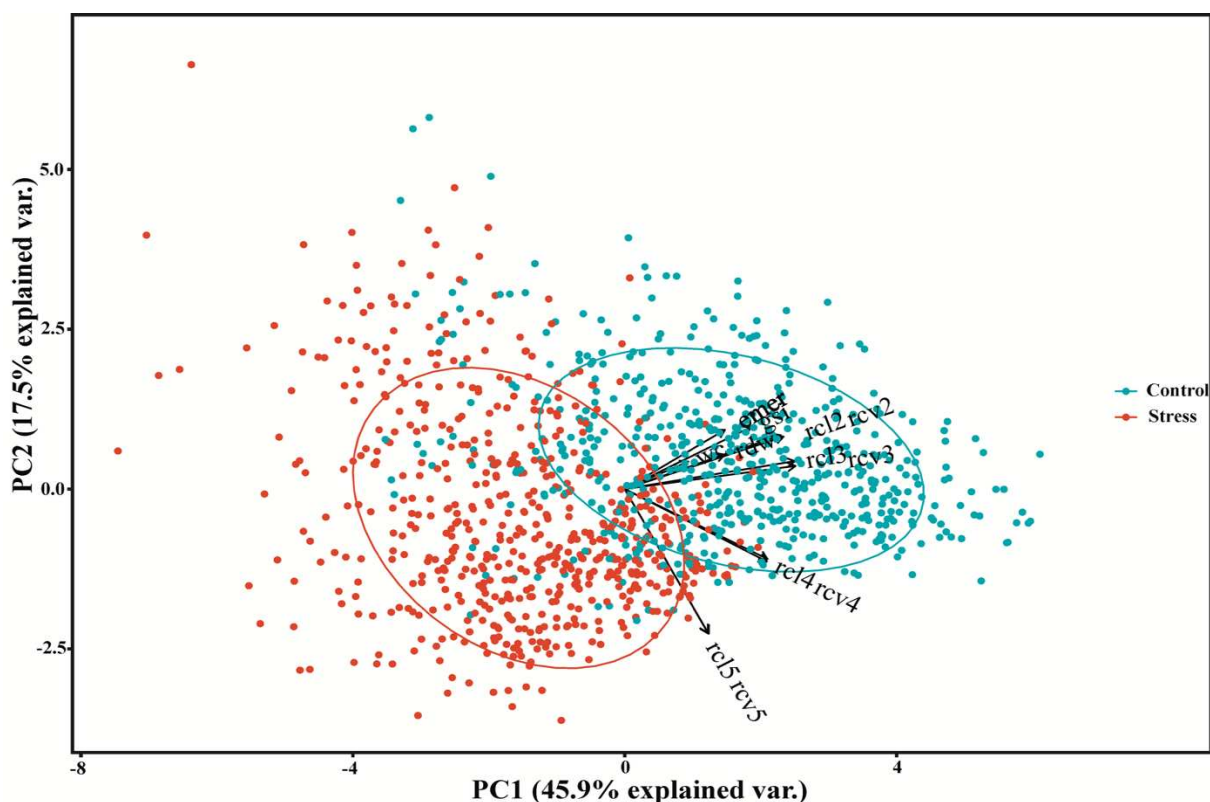


Figura 4 – PCA-biplot das variáveis selecionadas para a abordagem 2, que possuem herdabilidade maior que 1 %.

wc: Conteúdo de água nas folhas, rdw: massa seca da raiz, emer: taxa de emergência de plantas, gsi: índice de velocidade de emergência, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5.

A correlação entre os valores de gBLUP das características apresentou valor alto e positivo (0,88) entre taxa de emergência (emer) e índice de velocidade de germinação (gsi) (Figura 5). Essas características correlacionaram negativamente com os valores de conteúdo de água nas folhas (wc) (-0,43 e -0,27). As características relacionadas a parte radicular correlacionaram positivamente entre si com valores variando de moderado a alto. A característica conteúdo de água nas folhas correlacionou de maneira positiva e baixa com massa seca da raiz (rdw), comprimento e volume por classe de diâmetro 3 (rcl3 e rcv3). A partir disso as características wc, emer e gsi foram selecionadas para a realização do estudo de associação genômica ampla.

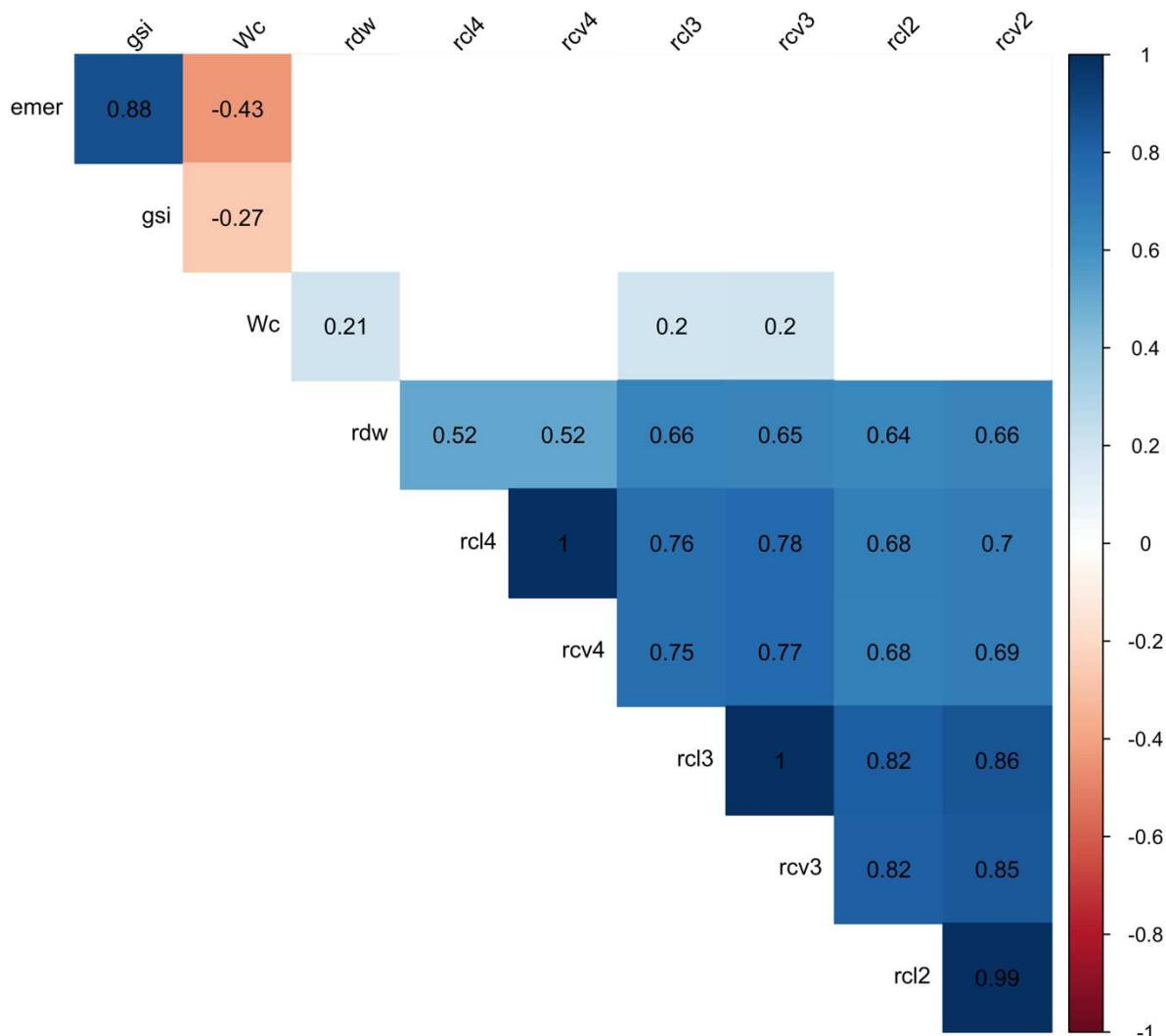


Figura 5 - Coeficientes da análise de correlação de Pearson entre os valores genotípicos preditos (gBLUP) para a abordagem 2.

wc: Conteúdo de água nas folhas, rdw: massa seca da raiz, emer: taxa de emergência de plantas, gsi: índice de velocidade de emergência, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4.

O processamento dos marcadores SNPs com $MAF \geq 0,01$ e call ratio $\geq 0,90$, gerou 4.835 marcadores do tipo SNP utilizáveis. Assim procedeu-se a análise de estrutura de populações.

Na análise discriminante por PCA os grupos são encontrados maximizando a diferença entre os clusters mais minimizando a variação dentro do grupo. Para diferenciar este conjunto de cultivares, optou-se por manter todos os PCAs na análise, devido a variância acumulativa explicada pelo PCA apresentar uma curva em que não se atinge um platô e não ser um número muito alto de PCAs (Figura S1a). Os valores de BIC não apresentaram uma curvatura do tipo

“cotovelo” então quaisquer valores de clusters menores que 10 satisfariam esse critério então optou-se também por reter esse valor (Figura S1b). Foram consideradas 5×10^8 interações e todos os autovalores foram retidos por se tratar de um número pequeno de clusters e assim as funções discriminantes foram avaliadas sem dificuldades (Figura S1c). Então esse conjunto de cultivares se agruparam em cinco grupos distintos (Figura 6a). Ao analisar os indivíduos observa-se que eles não possuem variação dentro do grupo como preconiza esse método (Figura 6b).

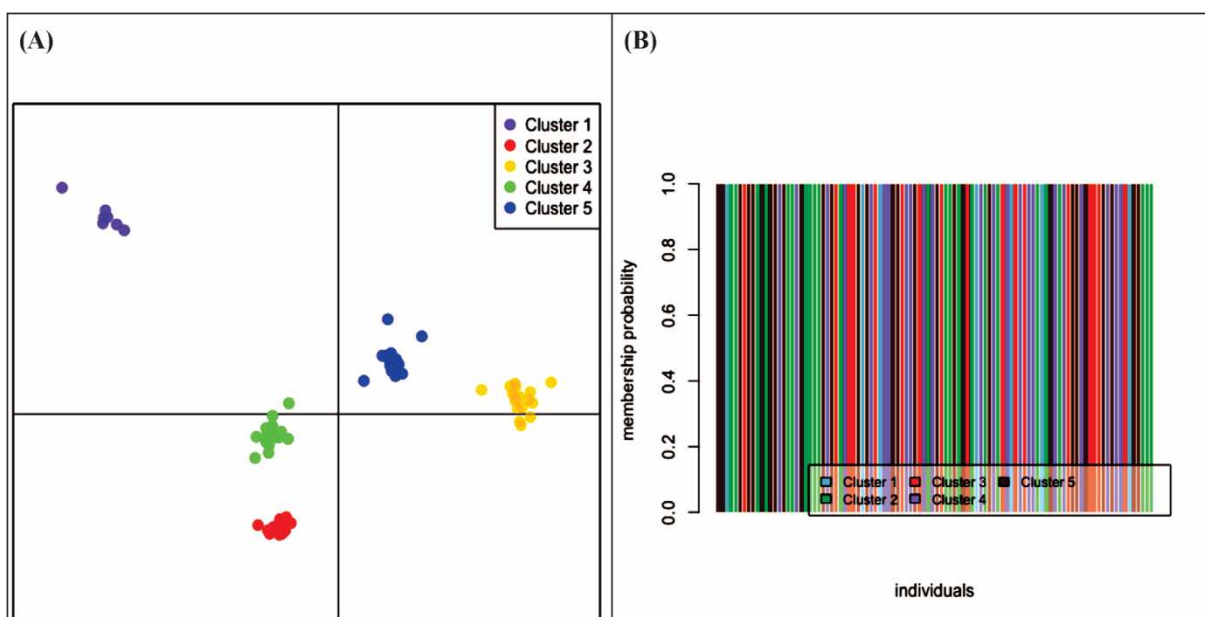


Figura 6 – Estrutura de população dos cultivares utilizados. Grupos formados pelo agrupamento em PCA dos cultivares (A) e Constituição dos indivíduos considerando os cinco grupos formados.

No estudo de associação genômica ampla para a abordagem 1 detectou 25 marcadores do tipo SNP significativamente associados com as características selecionadas (Figura 7 e Tabela S4). Para a característica wc foram encontrados 11 marcadores significativamente associados. Nas características emer e gsi foram encontrados 7 e 10 marcadores SNP associados, respectivamente. Os marcadores Gm10_39086213_C_A e Gm10_40258740_G_A foram significativamente associados à emer e à gsi. Já o marcador Gm17_36602853_A_G foi significativamente associado à emer e ao wc.

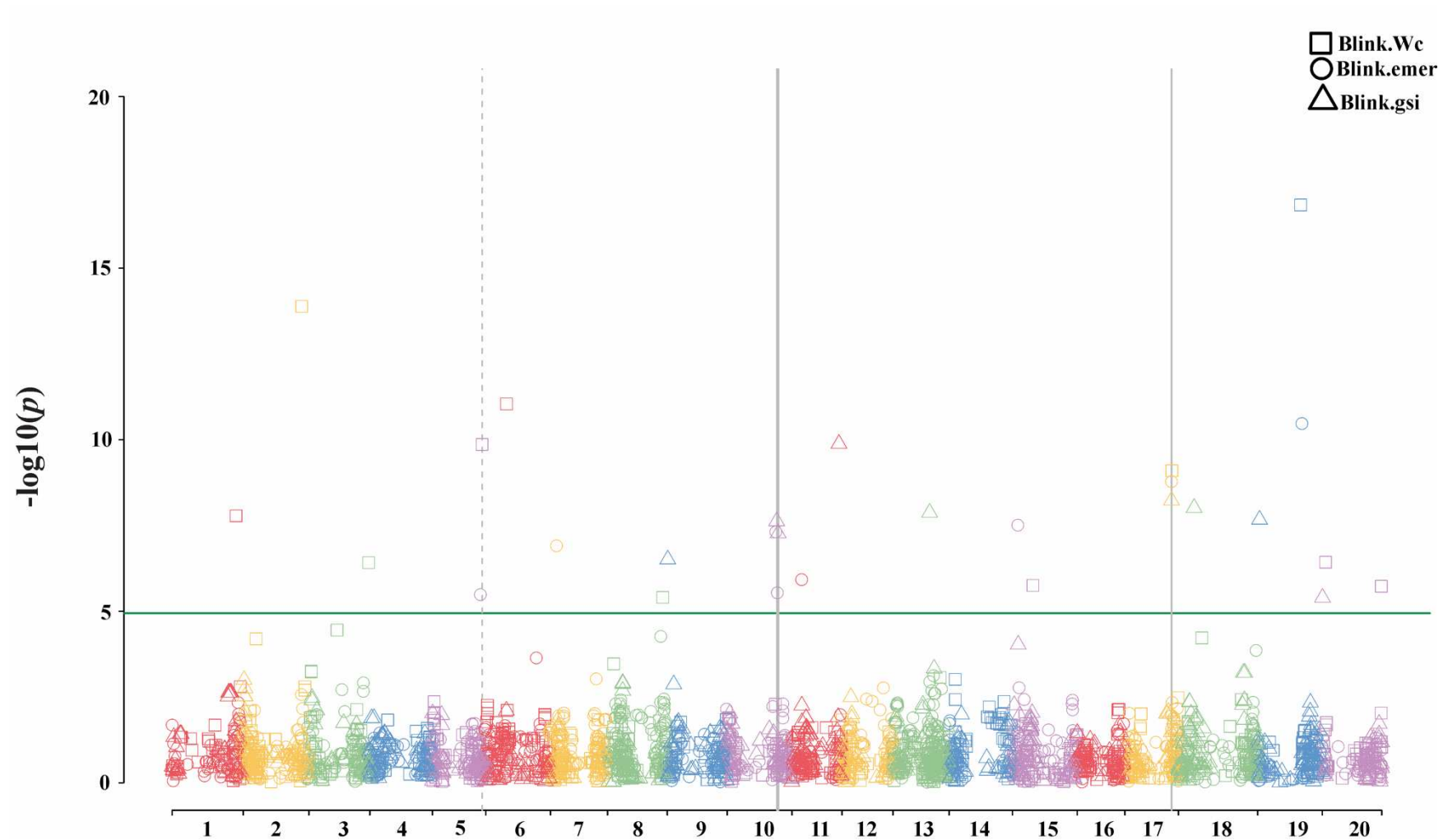


Figura 7 - Manhattan plot utilizando a abordagem 1.

SNPs acima da linha horizontal verde são significativos pela correção FDR.

emer: taxa de emergência de plantas, gsi: índice de velocidade de germinação, wc: conteúdo de água nas folhas.

A análise de associação genômica ampla utilizando a segunda abordagem encontrou um total de 18 marcadores do tipo SNPs significativamente associados com as características selecionadas (Figura 8 e Tabela S4). Dentre esses 13 marcadores foram significativamente associados com a característica wc. Outros cinco marcadores foram associados com a emer e dois com o gsi. Os marcadores Gm07_5989451_C_T e Gm10_39086213_C_A foram associados a emer e gsi simultaneamente.

O modelo utilizado nos dois estudos de associação genômica ampla, Blink, se ajustou nas duas abordagens com a maior parte dos marcadores do valor observado dentro da curva do modelo esperado (Figura S2).

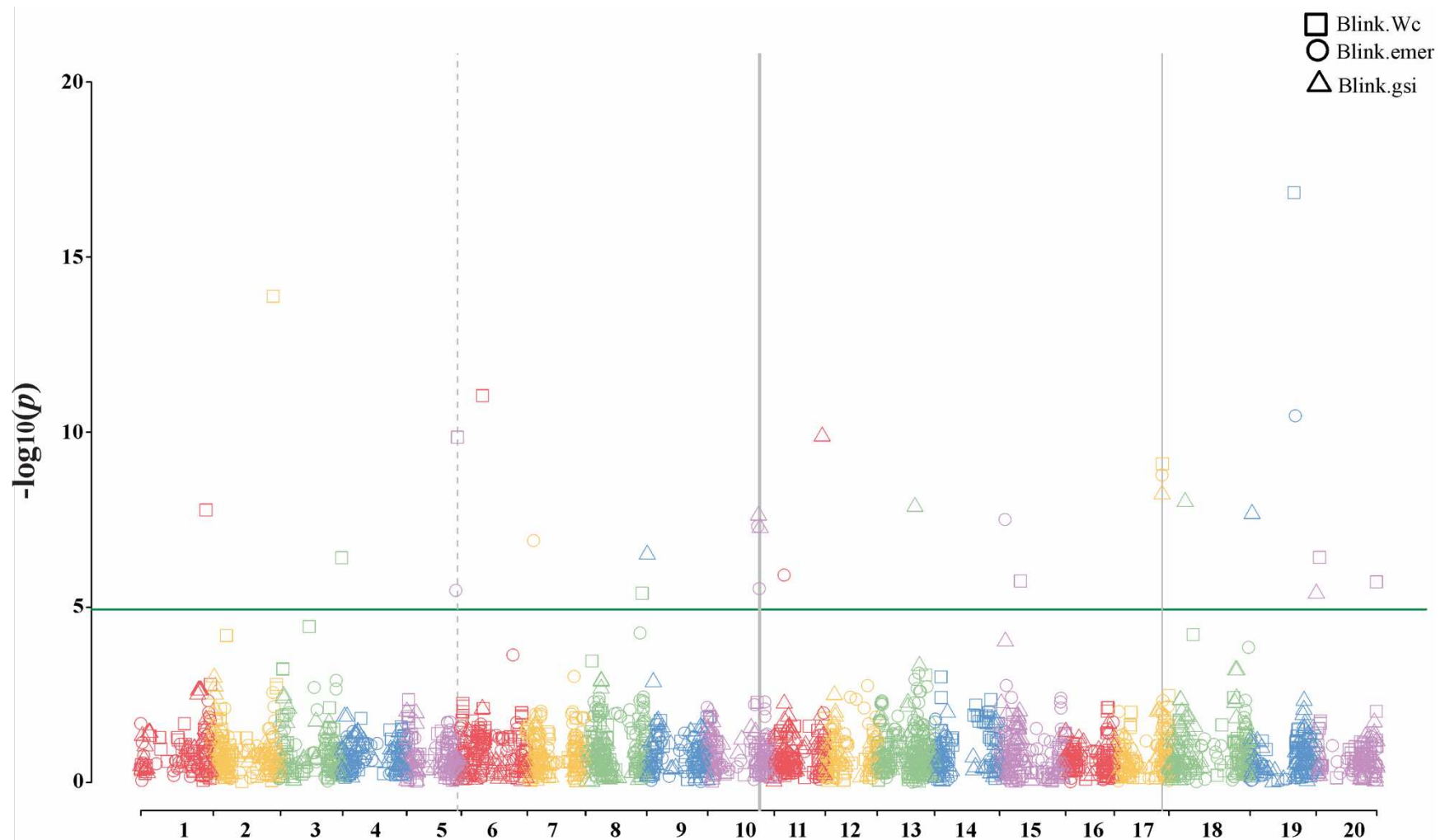


Figura 8 - Manhattan plot utilizando a abordagem 2.

SNPs acima da linha horizontal verde são significativos pela correção FDR.

emer: taxa de emergência de plantas, gsi: índice de velocidade de germinação, wc: conteúdo de água nas folhas.

Ao comparar os marcadores encontrados nas duas abordagens, apenas três deles estão presentes em ambas e foram então selecionados para a etapa de estudo pós-GWAS, sendo eles: Gm07_5989451_C_T, Gm10_39086213_C_A e Gm15_4613500_G_A (Tabela S4).

Na região que compreende o marcador Gm07_5989451 foi encontrado o gene GLYMA_07G066900, uma *Probable polygalacturonase*, em que o marcador se encontra localizado na região do íntron desse gene (Tabela 1). A região do segundo marcador Gm10_39086213_C_A contém dois genes: GLYMA_10G162100 (*bZIP transcription factor protein SBZ1*) e GLYMA_10G162200 (*mttA/Hcf106 domain-containing protein*). Em relação ao gene “*bZIP transcription factor protein SBZ1*” o marcador se encontra na região codante desse gene. A substituição alélica por esse marcador na posição 1082 modifica uma citosina para uma adenina, tal modificação leva a substituição do aminoácido prolina para uma glutamina na posição 361 da proteína (Tabela 1). Já em relação ao gene GLYMA_10G162200, o marcador se encontra na região intergênica. O terceiro marcador Gm15_4613500_G_A, também possui dois genes na região selecionada: GLYMA_15G059800 (*LOW QUALITY PROTEIN: oligopeptide transporter 4*) e GLYMA_15G059900 (*Heat shock 70 kDa protein, Hsp70*) em que o marcador se encontra na região intergênica em relação a ambos os genes.

Tabela 1 – Genes localizados em regiões próximas aos marcadores SNPs associados significativamente com a tolerância ao deficit hídrico. Localização dos marcadores em relação ao gene e seus efeitos na substituição alélica.

SNP	Position (WM82.a4) ^a	Gene annotation	SNP Location ^b	Gene description
Gm07_5989451_C_T	chr7:5879104	GLYMA_07G066900	intronic	Probable polygalacturonase
Gm10_39086213_C_A	chr10:39781572	GLYMA_10G162100	CDS CCG -> CAG Pro -> Gln (P361Q)	bZIP transcription factor protein SBZ1
Gm10_39086213_C_A	chr10:39781572	GLYMA_10G162200	intergenic	mttA/Hcf106 domain-containing protein
Gm15_4613500_G_A	chr15:4617772	GLYMA_15G059800	intergenic	LOW QUALITY PROTEIN: oligopeptide transporter 4
Gm15_4613500_G_A	chr15:4617772	GLYMA_15G059900	intergenic	Heat shock 70 kDa protein, Hsp70

a: posição do SNP no Genoma da soja versão 4 (*Glycine max* Wm82.a4.v1), **b:** posição onde se encontra o SNP em relação ao gene e a modificação levada pela substituição alélica para os marcadores dentro da região codante do gene

4. Discussão

A legislação brasileira de sementes e mudas estabelece pela Instrução Normativa MAPA 45/2013 que para fins de comercialização a germinação das sementes de soja devem estar acima de no mínimo 80% (MAPA, 2013). Dentre as sementes utilizadas todas atenderam a este requisito nos dois experimentos com germinação média de 96% na condição controle, demonstrando a alta qualidade das sementes utilizadas. Na condição estresse, essa germinação já reduziu bastante para cerca de 64 e 70 % no primeiro e segundo experimento respectivamente. A germinação das parcelas na condição estresse se demonstraram também bem dispersas, com parcelas atingindo próximo a 100 e outras próximo a zero. Isso demonstra que existe cultivares mais e menos tolerantes a dessecação nesse estágio e isto deve ser explorado.

O índice de velocidade de emergência de plantas (gsi) também foi grandemente afetado pelo tratamento. Na condição controle o gsi foi de cerca de 2,05 e 1,73 para o experimento 1 e 2 e na condição estresse foi de 0,55 e 0,65. Assim, o tratamento afetou também o vigor das plantas, apresentando valores mais dispersos na condição estresse. A velocidade de germinação das plantas é um fator crucial para a sobrevivência e estabelecimento da cultura, pois quanto mais cedo ocorre a germinação maior é a chance de iniciar a exploração de recursos do ambiente (Wijewardana et al., 2019; Singh et al., 2018; Reddy et al., 2017).

A indução de déficit hídrico em um sistema hidropônico, com a areia como o material inerte para a fixação das raízes, contribuiu para a discriminação dos genótipos entre os tratamentos controle e estresse. Esse sistema em bandejas contendo areia ainda proporcionou que a avaliação durante o estágio de germinação pudesse ocorrer durante um período maior que em outros sistemas como no papel germitest.

Em ambas as abordagens, a separação de dois grupos ao plotar os valores absolutos dos fenótipos de cada indivíduo no PCA-biplot demonstrou a eficácia do tratamento. Observa-se também em ambas as abordagens que os dados do grupo da condição estresse, apresentam-se mais dispersos, demonstrando a variabilidade entre os cultivares dentro dessa condição. Os autovetores apresentaram direcionamento para a condição controle demonstrando que aquela condição contribui para o aumento das características. Em ambas as abordagens não ocorreu nenhuma característica se orientando em direção a condição estresse. Alguns estudos demonstram que algumas características relacionadas ao desenvolvimento de raízes podem ser favorecidas em condições de estresse moderado quando aplicado em estágio mais avançado da cultura, com o aumento de raízes finas para absorção e exploração do perfil do solo (Turner et

al., 2001; Dayoub et al., 2021). No entanto, neste estudo foi adotado o estresse durante a germinação, e a planta necessitava ainda estabelecer o sistema radicular. Assim, as características que em outros estádios apresentam aumento com déficit hídrico, como raízes finas, não obedeceram a esse padrão. Tal fato pode ter contribuído para que nenhum valor genotípico (gBLUP) das características relacionadas à parte radicular correlacionasse com o valor genotípico das características taxa de emergência de plantas e índice de velocidade de germinação.

A análise de estrutura de populações favoreceu o agrupamento das cultivares em cinco grupos. As cinco populações encontradas no estudo se aproximam do número de populações que foram encontradas em estudos anteriores, quando utilizaram um número de cultivares maior (169 cultivares) que continham o painel de cultivares utilizado no presente estudo. Conteras-Soto et al. (2017b) utilizaram as 169 cultivares e identificaram nove grupos, em que utilizaram um MAF de 0,05. Outro fator que contribui para o baixo número de grupos é que a soja cultivada não possui uma diversidade tão ampla como em outras culturas (Conteras-Soto et al., 2017a). Tal fato pode ser percebido na Figura 1b, em que não se encontra indivíduo com dois ou mais fragmentos de populações distintas pela metodologia utilizada. A metodologia utilizada aqui pode ter contribuído para um menor número de populações já que ela tende a reduzir a diferença dentro dos grupos, mas possui a vantagem de não necessitar definir grupos a priori (Jombart & Collins, 2015; Jombart et al., 2010).

O processo de tolerância ao déficit hídrico, se trata de um processo complexo em que várias vias moleculares e genes estão envolvidos, assim a combinação de estratégias para a seleção de cultivares tolerantes é a mais indicada (Fritsche-Neto & Borém, 2012). Nesse sentido, foi verificado que três marcadores foram significativos em ambas as estratégias adotadas no presente estudo e estes foram selecionados para a etapa de pós-GWAS. As regiões genômicas contendo esses marcadores estão inseridas nos cromossomos 7, 10 e 15 da soja. Os principais genes presentes nestas são: GLYMA_07G066900 (cromossomo 7), GLYMA_10G162100 (cromossomo 10), GLYMA_10G162200 (cromossomo 10), GLYMA_15G059800 (cromossomo 15) e GLYMA_15G059900 (cromossomo 15).

O gene GLYMA_07G066900 uma “Probable polygalacturonase” que pertence a superfamília das “Pectin lyase-like”. Essa família de proteínas desempenha papel fundamental na composição da parede celular em plantas (Zhang et al., 2020). A parede celular desempenha funções tanto no estresse biótico quanto abiótico (Zhu, 2016). A sua integridade está intimamente associada ao estresse abiótico e a sua degradação pode funcionar como sinalização para condições adversas (Zhang et al., 2020).

A menor adesão de células por meio da redução de conteúdo de pectina na parede celular torna as plantas mais sensíveis às condições de estresse (Liu et al., 2014). Essa redução contribui para uma defesa das plantas através do enrolamento foliar (Zhang et al., 2020). Desse modo, a planta consegue reduzir a taxa de fotossíntese, a taxa de transpiração e perda de água, aumentando a eficiência do uso da água. Outra questão é que taxa de expansão celular é reduzida durante o estresse, assim a deposição de pectina menor na parede celular irá contribuir para que, quando houver a água suficiente, irrigação ou chuva, a célula consiga se expandir por não ter uma parede muito rígida (Clauw et al., 2015). Em estudos, a transformação de plantas de arroz com um gene pertencente a essa família levou a plantas mais sensíveis ao enrolamento foliar e consequentemente mais tolerante ao estresse hídrico (Zhang et al., 2020).

O gene GLYMA_10G162100 é um fator de transcrição da família bZIP (SBZ1) e seu homólogo em *Arabidopsis* é o bZIP63 (At5g28770). Os fatores de transcrição bZIP fazem parte de uma grande família de fatores de transcrição em plantas e são reconhecidos por possuírem um sítio de ligação ao DNA e outro chamado de zíper de leucina para a dimerização (Deppmann et al., 2006; Reinke et al., 2013). Esses fatores de transcrição estão envolvidos em diversos processos como proliferação e diferenciação celular, metabolismo, respostas a estresse e apoptose (Mayr & Montminy, 2001; Jakoby et al., 2002; Motohashi et al., 2002; Rodrigues-Pousada et al., 2010; Tsukada et al., 2011).

O fator de transcrição bZIP63 é conhecido por desempenhar papel importante no metabolismo de energia, maturação de sementes, senescência induzida, e germinação sobre estresse osmótico (Mair et al., 2015; Jakoby et al., 2002; Correa et al., 2008; Veerabagu et al., 2014). Esse fator de transcrição atua principalmente em respostas de “fome” das plantas, mas entre os seus inibidores se encontra o ácido abscísico (ABA), um dos principais hormônios envolvidos no déficit hídrico (Matiolli et al., 2011; Kunz et al., 2014). Assim, esse fator de transcrição pode estar atuando no déficit hídrico pela situação de privação de carbono que as plantas se encontram, pois, ao sofrer com esse estresse, as plantas reduziram o crescimento e acumularam menor matéria seca na parte aérea e radicular. No caso desse gene a substituição alélica leva a uma modificação na proteína na posição 361 em que uma prolina (aminoácido alifático, não polar e hidrofóbico) é substituída por uma glutamina (aminoácido polar, não carregado e hidrofílico). Dessa forma, essa parte da proteína que não faz interações com as moléculas de água e possivelmente está localizada no interior da proteína agora é capaz de fazer interações e estar localizada mais ao exterior da molécula de proteína. Isso pode causar modificações na estrutura da proteína impactando na sua função e interação com outras moléculas (Teng et al., 2010).

O gene GLYMA_10G162200 pertence a família de TatA/E (“twin-arginine protein translocase”) essa proteína transloca proteínas dobradas que possuem no peptídeo sinal um “twin-arginine” motif via membranas (Rollauer et al., 2012). Dentre as proteínas que possuem esse motif se encontra as proteínas “late embryogenesis abundant” (LEA) (Singh et al., 2020). As proteínas LEA são fatores de transcrição altamente expressos em sementes durante a germinação, mas também possuem papel importante em tecidos vegetativos em processos de tolerância à desidratação, baixa temperatura e estresse osmótico (Rodriguez-Salazar et al. 2017; Hundertmark & Hincha, 2008; Bies-Ethève et al., 2002). Diversos estudos têm comprovado a importância das proteínas LEA durante o déficit hídrico em plantas (Magwanga et al., 2018; Wang et al., 2014; Zhao et al. 2011; Xiao et al. 2007; Babu et al. 2004). Com isso, diversas propostas têm sido formuladas para explicar o papel crucial dessas proteínas como proteção enzimática (Hand et al. 2011), proteção molecular (Furuki et al. 2011), efeito tampão (Hundertmark et al. 2012), e interação de membrana (Olvera-Carrillo et al. 2011).

A “Oligopeptide transporter 4” (GLYMA_15G059800) é uma proteína envolvida com o transporte de peptídeos entre membranas celulares, e nesse caso está relacionada ao carregamento direto nos tecidos vasculares (Koh et al., 2002; Stacey et al., 2006). A sinalização em plantas através de oligopeptídeos está relacionada a diversos processos bioquímicos, fisiológicos e de desenvolvimento (Matsubayashi, 2014; Oh et al., 2018; Fletcher, 2020; Jeon et al., 2021; Kim et al., 2021). Durante o déficit hídrico, alguns peptídeos têm papel crucial de sinalização na tolerância como CLAVATA3(CLV)/EMBRYO-SURROUNDING REGION-RELATED 9 e 25 (CLE9 e CLE25) (Takahashi et al., 2018; Zhang et al., 2019), INFLORESCENCE DEFICIENT IN ABSCISSION (IDA) (Patharkar & Walker, 2016) e Phytosulfokine- α (PSK- α) (Stührwohltdt et al., 2021).

O GLYMA_15G059900 é uma “Heat shock 70 kDa protein” (Hsp70). As proteínas de heat-shock estão dentre os componentes moleculares mais comumente envolvidos em todos os tipos de estresse em plantas (Wahid et al. 2007; Priya et al. 2019). As Hsp70 são chaperonas moleculares que desempenham papel crucial em plantas como no dobramento correto de peptídeos, translocando proteínas através de membranas de organelas, montando ou desmontando complexo de proteínas, controlando a atividade de proteínas regulatórias e auxiliando na degradação de proteínas (Bukau & Horwich 1998; Leborgne-Castel et al. 1999; Sable & Agarwal 2018). Na tolerância ao déficit hídrico, as Hsp70 tem sido reportada em diversas culturas: cana-de-açúcar (Augustine et al. 2015), *Arabidopsis thaliana*, trigo (Duan et al. 2011) e tomate (Aghaie & Tafreshi, 2020).

A maior parte dos marcadores encontrados aqui se localizam fora da região codificadora dos genes. No entanto, não pode ser descartada a influência da substituição gênica na expressão desses genes. Regiões distantes do gene podem exercer grande influência por meio dos cis-elementos regulatórios conhecidos como enhancers (Ong & Corces, 2011). Esses são pequenas regiões de geralmente 20 à 400 bp que influenciam na velocidade da taxa de transcrição e pouco se sabe sobre essas regiões.

5. Conclusões

Neste trabalho, foram encontrados três marcadores nas duas abordagens comitaneamente, Gm07_5989451_C_T, Gm10_39086213_C_A e Gm15_4613500_G_A. Esses marcadores podem ser utilizados por programas de melhoramento para o desenvolvimento de cultivares mais tolerantes.

Nas regiões próximas a esses marcadores, foram encontrados cinco genes utilizando essas abordagens que estão envolvidos com o processo de tolerância ao déficit hídrico. O gene GLYMA_07G066900 está envolvido com a composição da parede celular, e os genes GLYMA_10G162200 e GLYMA_15G059800 são transportadores via membranas de proteínas e oligopeptídeos, respectivamente. Já o gene GLYMA_10G162100 (bZIP63) é um fator de transcrição e o GLYMA_15G059900 (Hsp70) é uma chaperona molecular que estão envolvidos em diversos processos biológicos.

Os genes encontrados nesse trabalho são excelentes candidatos a serem testados in vivo para verificar o papel no processo de tolerância. Assim, cultivares mais tolerantes a este tipo de estresse podem ser engenheiradas para sobrepor essa condição aversa e negativa para as culturas.

6. Referências bibliográficas

- Aghaie, P.; Tafreshi, S.A.H. (2020). Central role of 70-kDa heat shock protein in adaptation of plants to drought stress. *Cell Stress Chaperones*, 25:1071-1081. doi: 10.1007/s12192-020-01144-7
- Anjum, S.A.; Ashraf, U.; Zohaib, A.; Tanveer, M.; Naeem, M.; Ali, I.; Tabassum, T.; Nazir, U. (2017). Growth and development responses of crop plants under drought stress: a review. *Zemdirbyste - Agriculture*, 104: 267–276. doi: 10.13080/z-a.2017.104.034.
- Augustine, S.M.; Narayan, J.A.; Syamaladevi, D.P.; Appunu, C.; Chakravarthi, M.; Ravichandran, V.; Subramonian, N. (2015). *Erianthus arundinaceus* HSP70 (EaHSP70)

- overexpression increases drought and salinity tolerance in sugarcane (*Saccharum* spp. hybrid). *Plant Science*, 232: 23–34. doi: 10.1016/j.plantsci.2014.12.012
- Babu, R.C.; Zhang, J.; Blum, A.; Ho, T.H.D.; Wu, R.; Nguyen, H.T. (2004). HVA1, a LEA gene from barley confers dehydration tolerance in transgenic rice (*Oryza sativa* L.) via cell membrane protection. *Plant Science*, 166: 855–862. doi: 10.1016/j.plantsci.2003.11.023
- Begum, N.; Hasanuzzaman, M.; Li, Y.; Akhtar, K.; Zhang, C.; Zhao, T. (2022). Seed Germination Behavior, Growth, Physiology and Antioxidant Metabolism of Four Contrasting Cultivars under Combined Drought and Salinity in Soybean. *Antioxidants*, 11: 498. doi: 10.3390/antiox11030498
- Bernardo, S.; Mantovani, C.E.; Silva, D.D.; Soares, A.A. (2019). Manual de irrigação. Viçosa: Editora UFV, 9a Edição, p. 545
- Bies-Ethève, N.; Gaubier-Comella, P.; Debures, A.; Lasserre, E.; Jobet, E.; Raynal, M.; Cooke, R.; Delseny, M. (2008). Inventory, evolution, and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Molecular Biology*, 67: 107–124. doi: 10.1007/s11103-008-9304-x
- Blum, A. (2017). Osmotic adjustment is a prime drought stress adaptive engine in support of plant production. *Plant, Cell & Environment*, 40: 4–10. doi: 10.1111/pce.12800
- Bukau, B.; Horwich, A.L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* 92: 351–366. doi: 10.1016/s0092-8674(00)80928-9
- Cano-Gamez, E.; Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics*, 11: 424. doi: 10.3389/fgene.2020.00424
- Clauw, P.; Coppens, F.; De Beuf, K.; Dhondt, S.; Van Daele, T.; Maleux, K.; Storme, V.; Clement, L.; Gonzalez, N.; Inzé, D. (2015). Leaf Responses to Mild Drought Stress in Natural Variants of *Arabidopsis*. *Plant Physiology*, 167: 800–816. doi: 10.1104/pp.114.254284
- Contreras-Soto, R.I.; Mora, F.; Oliveira, M.A.; Higashi, W.; Scapim, C.A.; Schuster, I. (2017a). A Genome-Wide Association Study for Agronomic Traits in Soybean Using SNP Markers and SNP-Based Haplotype Analysis. *Plos One*, 12: e0171105. doi: 10.1371/journal.pone.0171105
- Contreras-Soto, R.I.; Oliveira, M.B.; Costenaro-da-Silva, D.; Scapim, C.A.; Schuster, I. (2017b). Population structure, genetic relatedness and linkage disequilibrium blocks in cultivars of tropical soybean (*Glycine max*). *Euphytica*, 213: 173. doi: 10.1007/s10681-017-1966-5
- Correa, L.G.; Riano-Pachon, D.M.; Schrago, C.G.; Santos, R.V.; Mueller-Roeber, B.; Vincentz, M. (2008). The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *Plos One* 3: e2944. doi: 10.1371/journal.pone.0002944

- Covarrubias-Pazaran, G. (2016). Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. *Plos One*, 11: e0156744. doi: 10.1371/journal.pone.0156744
- Dayoub, E.; Lamichhane, J.R.; Schoving, C.; Debaeke, P.; Maury, P. (2021). Early-Stage Phenotyping of Root Traits Provides Insights into the Drought Tolerance Level of Soybean Cultivars. *Agronomy*, 11: 188. doi: 10.3390/agronomy 11010188
- Deppmann, C.D.; Alvania, R.S.; Taparowsky, E.J. (2006). Cross-species annotation of basic leucine zipper factor interactions: insight into the evolution of closed interaction networks. *Molecular Biology and Evolution*, 23: 1480–1492. doi: 10.1093/molbev/msl022
- Duan, Y.-H.; Guo, J.; Ding, K.; Wang, S.-J.; Zhang, H.; Dai, X.-W.; Chen, Y.-Y.; Govers, F.; Huang, L.-L.; Kang, Z.-S. (2011). Characterization of a wheat HSP70 gene and its expression in response to stripe rust infection and abiotic stresses. *Molecular Biology Reports*, 38: 301–307. doi: 10.1007/s11033-010-0108-0
- Duarte, A. B.; Ferreira, D. O.; Silva, F. L. (2022). Repeatability and the optimal number of measurements for screening of soybean cultivars under water deficit. *Revista Ciência Agronômica*, 53: e20218025. doi: 10.5935/1806-6690.20220028
- Fletcher, J. C. (2020). Recent advances in Arabidopsis CLE peptide signaling. *Trends in Plant Science*, 25: 1005–1016. doi: 10.1016/j.tplants.2020.04.014
- Fritsche-Neto, R; Borem, A. (2012). *Plant Breeding for Abiotic Stress Tolerance*. Springer Heidelberg New York Dordrecht London. doi: 10.1007/978-3-642-30553-5
- Fukao, T.; Xiong, L. (2013). Genetic mechanisms conferring adaptation to submergence and drought in rice: simple or complex?. *Current Opinion in Plant Biology*, 16: 196–204. doi: 10.1016/j.pbi.2013.02.003
- Furuki, T.; Shimizu, T.; Kikawada, T.; Okuda, T.; Sakurai, M. (2011) Salt effects on the structural and thermodynamic properties of a group 3 LEA protein model peptide. *Biochemistry*, 50: 7093–7103. doi: 10.1021/bi200719s
- Hand, S.C.; Menze, M.A.; Toner, M.; Boswell, L.; Moore, D. (2011). LEA proteins during water stress: not just for plants anymore. *Annual Review of Physiology*, 73: 115–134. doi: 10.1146/annurev-physiol-012110-142203
- Huang, M.; Liu, X.; Zhou, Y.; Summers, R. M.; Zhang, Z. (2018). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions, *GigaScience*, 8: 1-12. doi: 10.1093/gigascience/giy154
- Hundertmark, M.; Hinch, D.K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*, 9: 118. doi:10.1186/1471-2164-9-118
- Hundertmark, M.; Popova, A.V.; Rausch, S.; Seckler, R.; Hinch, D.K. (2012). Influence of drying on the secondary structure of intrinsically disordered and globular proteins.

- Biochemical and Biophysical Research Communications, 417: 122–128. doi: 10.1016/j.bbrc.2011.11.067
- Jakoby, M.; Weisshaar, B.; Droege-Laser, W.; Vicente-Carbajosa, J.; Tiedemann, J.; Kroj, T.; Parcy, F. (2002). bZIP transcription factors in Arabidopsis. Trends in Plant Science, 7:106–111. doi: 10.1016/S1360-1385(01)02223-3
- Jeon, B.W.; Kim, M.J.; Pandey, S.K.; Oh, E.; Seo, P.J.; Kim, J. (2021). Recent advances in peptide signaling during Arabidopsis root development. Journal of Experimental Botany, 72: 2889–2902. doi: 10.1093/jxb/erab050
- Jombart, T.; Ahmed, I. (2011). “adegenet 1.3-1: new tools for the analysis of genome-wide SNP data”. Bioinformatics, 27: 3070-3071. doi:10.1093/bioinformatics/btr521
- Jombart, T.; Collins, C. (2015). A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. London: Imperial College London, 43p.
- Jombart, T.; Devillard, S.; Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11: 94 (2010). doi: 10.1186/1471-2156-11-94
- Khan, M.A.; Tong, F.; Wang, W.; He, J.; Zhao, T.; Gai, J. (2018). Analysis of QTL–allele system conferring drought tolerance at seedling stage in a nested association mapping population of soybean [*Glycine max* (L.) Merr.] using a novel GWAS procedure. Planta, 248: 947–962. doi: 10.1007/s00425-018-2952-4
- Kim, M.J.; Jeon, B.W.; Oh, E.; Seo, P.J.; Kim, J. (2021). Peptide signaling during plant reproduction. Frontiers in Plant Science, 12: 704490. doi: 10.1016/j.tplants.2021.02.008.
- Koh, S., Wiles, A.M., Sharp, J.S., Naider, F.R., Becker, J.M., and Stacey, G. (2002). An oligopeptide transporter gene family in Arabidopsis. Plant Physiology, 128: 21–29. doi: 10.1104/pp.010332.
- Korte, A.; Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9: 29. doi: 10.1186/1746-4811-9-29
- Kpoghomou, B.K.; Sapra, V.T.; Beyl, C.A. (1990). Screening for Drought Tolerance: Soybean Germination and its Relationship to Seedling Responses. Journal of Agronomy and Crop Science, 164: 153–159. doi: 10.1111/j.1439-037x.1990.tb00801.x
- Kunz, S.; Pesquet, E.; Kleczkowski, L.A. (2014). Functional dissection of sugar signals affecting gene expression in Arabidopsis thaliana. Plos One 9: e100312. doi: 10.1371/journal.pone.0100312
- Leborgne-Castel, N.; Jelitto-Van Dooren, E.P.W.M.; Crofts, A.J.; Denecke, J. (1999) Overexpression of BiP in tobacco alleviates endoplasmic reticulum stress. Plant Cell, 11:459–470. doi: 10.1105/tpc.11.3.459

- Liu, H.; Ma, Y.; Chen, N.; Guo, S.; Liu, H.; Guo, X.; Chong, K.; Xu, Y. (2014). Overexpression of stress-inducible OsBURP16, the beta subunit of polygalacturonase 1, decreases pectin content and cell adhesion and increases abiotic stress sensitivity in rice. *Plant, Cell & Environment*, 37: 1144–1158. doi: 10.1111/pce.12223
- Liu, J.; Hasanuzzaman, M.; Wen, H.; Zhang, J.; Peng, T.; Sun, H.; Zhao, Q. (2019). High temperature and drought stress cause abscisic acid and reactive oxygen species accumulation and suppress seed germination growth in rice. *Protoplasma*, 256: 1217–1227. doi: 10.1007/s00709-019-01354-6
- Liu, Z.; Li, H.; Gou, Z.; Zhang, Y.; Wang, X.; Honglei Ren, H.; Wen, Z.; Kang, B.-K.; Li, Y.; Yu, L.; Gao, H.; Wang, D.; Qi, X.; Qiu, L. (2020). Genome-wide association study of soybean seed germination under drought stress. *Molecular Genetics and Genomics*, 295: 661–673. doi: 10.1007/s00438-020-01646-0
- Lotfi, N.; Soleimani, A.; Vahdati, K.; Çakmakçı, R. (2019). Comprehensive biochemical insights into the seed germination of walnut under drought stress. *Scientia Horticulturae*, 250: 329–343. doi: 10.1016/j.scienta.2019.02.060
- Magwanga, R.O.; Lu, P.; Kirungu, J.N.; Dong, Q.; Hu, Y.; Zhou, Z.; Cai, X.; Wang, X.; Hou, Y.; Wang, K.; Liu, F. (2018). Cotton Late Embryogenesis Abundant (LEA2) Genes Promote Root Growth and Confer Drought Stress Tolerance in Transgenic *Arabidopsis thaliana*. *G3 Genes|Genomes|Genetics*, 8: 2781–2803. doi: 10.1534/g3.118.200423.
- Mair, A.; Pedrotti, L.; Wurzinger, B.; Anrather, D.; Simeunovic, A.; Weiste, C.; Valerio, C.; Dietrich, K.; Kirchler, T.; Nägele, T.; Vicente Carbajosa, J.; Hanson, J.; Baena-González, E.; Chaban, C.; Weckwerth, W.; Dröge-Laser, W.; Teige, M. (2015). SnRK1-triggered switch of bZIP63 dimerization mediates the low-energy response in plants. *Elife*, 11: e05828. doi: 10.7554/eLife.05828
- MAPA. (2013). Instrução Normativa MAPA 45/2013 (D.O.U. 18/09/2013). Ministério da Agricultura, Pecuária e Abastecimento. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/sementes-e-mudas/publicacoes-sementes-e-mudas/copy_of_INN45de17desetembrode2013.pdf>, acesso em 05/07/2022.
- Matiolli, C.C.; Tomaz, J.P.; Duarte, G.T.; Prado, F.M.; Del Bem, L.E.; Silveira, A.B.; Gauer, L.; Correa, L.G.; Drumond, R.D.; Viana, A.J.; Di Mascio, P.; Meyer, C.; Vincentz, M. (2011). The *Arabidopsis* bZIP gene AtbZIP63 is a sensitive integrator of transient abscisic acid and glucose signals. *Plant Physiology*, 157:692–705. doi: 10.1104/pp.111.181743
- Matsubayashi, Y. (2014). Posttranslationally modified small-peptide signals in plants. *Annual Review of Plant Biology*, 65, 385–413. doi: 10.1146/annurev-arplant-050312-120122
- Maury, P.; Andrianasolo, F.N.; Alric, F.; Berger, M.; Beugniet, G.; Chambert, C.; Champolivier, L.; Doumenc, A.; Estragnat, A.; Gras, A.; Jeanson, P.; Jouffret, P.; Labalette, F.; Thomas, R.; Justes, E.E.; Debaeke, P.P. (2015). Le Semis Très Précoce: Une Stratégie Agronomique Pour Améliorer Les Performances Du Soja En France?. *Oilseeds Fats crops and lipids*, 22: 11. doi: 10.1051/ocl/2015028

- Mayr, B.; Montminy, M. (2001). Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nature Reviews Molecular Cell Biology*, 2: 599–609. doi: 10.1038/35085068
- Motohashi, H.; O'Connor, T.; Katsuoka, F.; Engel, J.D.; Yamamoto, M. (2002). Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. *Gene*, 294: 1–12. doi: 10.1016/S0378-1119(02)00788-6
- Mukarram, M.; Choudhary, S.; Kurjak, D.; Petek, A.; Khan, M.M.A. (2021). Drought: sensing, signaling, effects and tolerance in higher plants. *Physiologia Plantarum*, 172: 1291–1300. doi: 10.1111/ppl.13423
- Oh, E., Seo, P. J., and Kim, J. (2018). Signaling peptides and receptors coordinating plant root development. *Trends in Plant Science*, 23: 337–351. doi: 10.1016/j.tplants.2017.12.007
- Olvera-Carrillo, Y.; Reyes, J.L.; Covarrubias, A.A. (2011). Late Embryogenesis Abundant proteins: Versatile players in the plant adaptation to water limiting environments. *Plant Signaling & Behavior*, 6: 586–589. doi: 10.4161/psb.6.4.15042
- Ong, C. T.; Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, 12: 283–293. doi:10.1038/nrg2957
- Patharkar, O.R.; Walker, J.C. (2016). Core mechanisms regulating developmentally timed and environmentally triggered abscission. *Plant Physiology*, 172: 510–520. doi: 10.1104/pp.16.01004
- Priya, M.; Dhanker, O.P.; Siddique, K.H.; HanumanthaRao, B.; Nair, R.M.; Pandey, S.; Singh, S. Varshney, R.K.; Prasad, P.V.; Nayyar, H. (2019). Drought and heat stress-related proteins: an update about their functional relevance in imparting stress tolerance in agricultural crops. *Theoretical and Applied Genetics*, 132:1607–1638.
- R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rathan, N.D.; Krishna, H.; Ellur, R.K.; Sehgal, D.; Govindan, V.; Ahlawat, A.K.; Krishnappa, G.; Jaiswal, J.P.; Singh, J.B.; SV, S.; Ambati, D.; Singh, S.K.; Bajpai, K.; Mahendru-Singh, A. (2022). Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* L.). *Scientific Reports*, 12: 7037. doi: 10.1038/s41598-022-10618-w
- Reddy, K.R.; Brand, D.; Wijewardana, C.; Gao, W. (2017). Temperature effects on cotton seedling emergence growth and development. *Agronomy Journal*, 109: 1379–1387. doi: 10.2134/agronj2016.07.0439
- Reinke, A.W.; Baek, J.; Ashenberg, O.; Keating, A.E. (2013). Networks of bZIP protein-protein interactions diversified over a billion years of evolution. *Science*, 340: 730–734. doi: 10.1126/science.1233465
- Rodrigues-Pousada, C.; Menezes, R.A.; Pimentel, C. (2010). The Yap family and its role in stress response. *Yeast*, 27: 245–258. doi: 10.1002/yea.1752

- Rodriguez-Salazar, J.; Moreno S.; Espín, G. (2017). LEA proteins are involved in cyst desiccation resistance and other abiotic stresses in *Azotobacter vinelandii*. *Cell Stress and Chaperones*, 22: 397–408. doi: 10.1007/s12192-017-0781-1
- Rollauer, S.; Tarry, M.; Graham, J.E.; Jääskeläinen, M.; Jäger, F.; Johnson, S.; Krehenbrink, M.; Liu, S.-M.; Lukey, M.J.; Marcoux, J.; McDowell, M.A.; Rodriguez, F.; Roversi, P.; Stansfeld, P.J.; Robinson, C.V.; Sansom, M.S.P.; Palmer, T.; Högbom, M.; Berks, B.C.; Lea, S.M. (2012). Structure of the TatC core of the twin-arginine protein transport system. *Nature*, 492: 210–214. doi: 10.1038/nature11683
- Sable, A; Agarwal, S.K. (2018). Plant heat shock protein families: essential machinery for development and defense. *Journal of Biological Sciences and Medicine*, 4: 51–64.
- Saha, D.; Choyal, P.; Mishra, U.N.; Dey, P.; Bose, B.; Prathibha, M.D.; Gupta, N.K.; Mehta, B.K.; Kumar, P.; Pandey, S.; Chauhan, J.; Kumar, R.; Singhal, R. K. (2022). Drought stress responses and inducing tolerance by seed priming approach in plants. *Plant Stress*, 4: 100066. doi: 10.1016/j.stress.2022.100066
- Silva, L.J.; Medeiros, A.D.; Oliveira, A.M.S. (2019). SeedCalc, a new automated R software tool for germination and seedling length data processing. *Journal of Seed Science*, 41: 250–257. doi:10.1590/2317-1545v42n2217267
- Singh, B.; Norvell, E.; Wijewardana, C.; Wallace, T.; Chastain, D.; Reddy, K.R. (2018). Assessing morphological characteristics of elite cotton lines from different breeding programmes for low temperature and drought tolerance. *Journal of Agronomy and Crop Science*, 204: 467–476. doi: 10.1111/jac.12276
- Singh, K.K.; Graether, S.P. (2020) Conserved sequence motifs in the abiotic stress response protein late embryogenesis abundant 3. *Plos One*, 15: e0237177. doi: 10.1371/journal.pone.0237177
- Stacey, M.G.; Osawa, H.; Patel, A.; Gassmann, W.; Stacey, G. (2006). Expression analyses of *Arabidopsis* oligopeptide transporters during seed germination, vegetative growth and reproduction. *Planta*, 223: 291–305. doi: 10.1007/s00425-005-0087-x
- Stührwohltdt, N.; Bühler, E.; Sauter, M.; Schaller, A. (2021). Phytosulfokine (PSK) precursor processing by subtilase SBT3.8 and PSK signaling improve drought stress tolerance in *Arabidopsis*. *Journal of Experimental Botany*, 72: 3427–3440. doi: 10.1093/jxb/erab017
- Takahashi, F.; Suzuki, T.; Osakabe, Y.; Betsuyaku, S.; Kondo, Y.; Dohmae, N.; Fukuda, H.; Yamaguchi-Shinozaki, K.; Shinozaki, K. (2018). A small peptide modulates stomatal control via abscisic acid in long-distance signalling. *Nature* 556, 235–238. doi: 10.1038/s41586-018- 0009-2
- Tang, D.; Wei, F.; Qin, S.; Khan, A.; Kashif, M.H.; Zhou, R. (2019). Polyethylene glycol induced drought stress strongly influences seed germination, root morphology and cytoplasm of different Kenaf genotypes. *Industrial Crops and Products*, 137: 180–186. doi: 10.1016/j.indcrop.2019.01.019

- Teng, S.; Srivastava, A.K.; Schwartz, C.E.; Alexov, E.; Wang, L. (2010). Structural assessment of the effects of amino acid substitutions on protein stability and protein-protein interaction. *International Journal of Computational Biology and Drug Design*, 3: 334-349. doi: 10.1504/IJCBDD.2010.038396
- Toscano, S.T.; Romano, D.M.; Tribulato, A.A.F.; Patanè, C.H. (2017). Effects of drought stress on seed germination of ornamental sunflowers. *Acta Physiologiae Plantarum* 39, 184. doi: 10.1007/s11738-017-2484-8
- Tsukada, J.; Yoshida, Y.; Kominato, Y.; Auron, P.E. (2011). The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation. *Cytokine*, 54:6–19. doi: 10.1016/j.cyto.2010.12.019
- Turner, N.C.; Wright, G.C.; Siddique, K.H.M. (2001). Adaptation of Grain Legumes (Pulses) to Water-Limited Environments. *Advances in Agronomy*, 71: 193–231. doi: 10.1016/S0065-2113(01)71015-2
- UNCCD. DROUGHT IN NUMBERS 2022 - restoration for readiness and resilience. United Nations Convention to Combat Desertification. Disponível em: <<https://www.unccd.int/sites/default/files/2022-05/Drought%20in%20Numbers.pdf>>, acesso em 11/06/2022.
- USDA. (2022). Soybean Explorer. U. S. Department of Agriculture. Disponível em: <<https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=2222000>>, acesso em 11/06/2022.
- Veerabagu, M.; Kirchler, T.; Elgass, K.; Stadelhofer, B.; Stahl, M.; Harter, K.; Mira-Rodado, V.; Chaban, C. (2014). The interaction of the Arabidopsis response regulator ARR18 with bZIP63 mediates the regulation of PROLINE DEHYDROGENASE expression. *Molecular Plant*, 7: 1560–1577. doi: 10.1093/mp/ssu074
- Villela, F.A.; Filho, L.D.; Sequeira, E.L. (1991). Tabela de potencial osmótico em função da concentração de polietileno glicol 6000 e da temperatura (Table of osmotic potential as a function of polyethylene glycol 6000 concentration and temperature). *Pesquisa Agropecuária Brasileira*, 26: 1957-1968.
- Vu, V.Q.; Marwick, B.; Hester, J.; Held, M. (2015). R package ggbiplot. Disponível em: <<https://github.com/vqv/ggbiplot>>
- Wahid, A.; Gelani, S.; Ashraf, M.; Foolad, M.R. (2007) Heat tolerance in plants: an overview. *Environmental and Experimental Botany*, 61:199–223. doi: 10.1016/j.envexpbot.2007.05.011
- Wang, M.; Li, P.; Li, C.; Pan, Y.; Jiang, X.; Zhu, D.; Zhao, Q.; Yu, J. (2014). SiLEA14, a novel atypical LEA protein, confers abiotic stress resistance in foxtail millet. *BMC Plant Biology*, 14: 290. doi: 10.1186/s12870-014-0290-7
- Wei, T.; Simko, V. (2021). R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.92) Disponível em: <<https://github.com/taiyun/corrplot>>.

- Wijewardana, C.; Reddy, K.R.; Krutz, L.J.; Gao, W.; Bellaloui, N. (2019). Drought stress has transgenerational effects on soybean seed germination and seedling vigor. *Plos One*, 14: e0214977. doi: 10.1371/journal.pone.0214977
- Xiao, B.; Huang, Y.; Tang, N.; Xiong, L. (2007). Over-expression of a LEA gene in rice improves drought resistance under the field conditions. *Theoretical and Applied Genetics*, 115: 35–46. doi: 10.1007/s00122-007-0538-9
- Yigit, N.; Sevik, H.; Cetin, M.; Kaya, N. (2016). Determination of the effect of drought stress on the seed germination in some plant species, p. 43-62. In: Rahman, I.M.M.; Begum, Z.A.; Hiroshi Hasegawa, H. (eds). *Water Stress Plants*, London: IntechOpe, 128 p. doi: 10.5772/61897
- Zhang, G.; Hou, X.; Wang, L.; Xu, J.; Chen, J.; Fu, X.; Shen, N.; Nian, J.; Jiang, Z.; Hu, J.; Zhu, L.; Rao, Y.; Shi, Y.; Ren, D.; Dong, G.; Gao, Z.; Guo, L.; Qian, Q.; Luan, S. (2021). PHOTO-SENSITIVE LEAF ROLLING 1 encodes a polygalacturonase that modifies cell wall structure and drought tolerance in rice. *New Phytologist*, 229: 890-901. doi: 10.1111/nph.16899
- Zhang, L.; Shi, X.; Zhang, Y.; Wang, J.; Yang, J.; Ishida, T.; Jiang, W.; Han, X.; Kang, J.; Wang, X.; Pan, L.; Lv, S.; Cao, B.; Zhang, Y.; Wu, J.; Han, H.; Hu, Z.; Cui, L.; Sawa, S.; He, J.; Wang, G. (2018). CLE9 peptide-induced stomatal closure is mediated by abscisic acid, hydrogen peroxide, and nitric oxide in *Arabidopsis thaliana*. *Plant, Cell & Environment*, 42: 1033–1044. doi: 10.1111/pce.13475
- Zhang, Z.; Todhunter, R.J.; Buckler, E.S.; Van Vleck, L.D. (2007). Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *Journal of Animal Science*, 85: 881–885. doi:10.2527/jas.2006-656
- Zhao, P.; Liu, F.; Ma, M.; Gong, J.; Wang, Q.; Jiab, P.; Zhengb, G.; Liub, H. (2011). Overexpression of AtLEA3–3 confers resistance to cold stress in *Escherichia coli* and provides enhanced osmotic stress tolerance and ABA sensitivity in *Arabidopsis thaliana*. *Molecular Biology*, 45: 785-796. doi: 10.1134/ S0026893311050165
- Zhu, J.K. (2016). Abiotic stress signaling and responses in plants. *Cell*, 167: 313– 324. doi: 10.1016/j.cell.2016.08.029

7. Material Suplementar

Tabela S1. Cultivares de soja a serem utilizados nos experimentos e seus respectivos grupos de maturidade e a empresa detentora do registro da cultivar.

ID	Cultivar	Grupo Maturidade	Companhia
1	NA4990	4.9	Nidera
2	CD216	5.5	Coodetec
3	CD215	5.9	Coodetec
4	BMXENERGIARR	5	GDM
5	BMXTitanRR	5.3	GDM
6	NS4823	4.8	Nidera

7	BMXTURBORR	5.8	GDM
8	NK3358	6.4	Syngenta
9	CD250RRSTS	5.5	Coodetec
10	NIDERAA4725RG	4.7	Nidera
11	CD225RR	5.8	Coodetec
12	M-SOY6101	6.1	Monsanto
13	Bragg	6.6	Embrapa
14	CD5807	6	Coodetec
15	EMGOPA304	7.3	Embrapa
16	BRSMGRenascenca	7.9	Embrapa
17	OCEPAR3-PRIMAVERA	6.5	Embrapa
18	A8000	8	-
19	MERCEDES70A	6.6	-
20	A6001RR	6	-
21	CD224	6.9	Coodetec
22	BRS245RR	7.3	Embrapa
23	CD208	6.9	Coodetec
24	BRS282	6.5	Embrapa
25	BRS185	6.6	Embrapa
26	BRS133	6.6	Embrapa
27	CD206	6.8	Coodetec
28	TMG1067RR	6.7	TMG
29	EMBRAPA48	6.5	Embrapa
30	CD217	7.3	Coodetec
31	BRS257	6.4	Embrapa
32	CD229RR	7.3	Coodetec
33	Fundacep56RR	6.5	Bayer
34	ANTA82	7.4	TMG
35	BRS232	6.5	Embrapa
36	BRS283	6.5	Embrapa
37	5D711RR	7.1	DowAgroscience
38	BRS243RR	6.6	Embrapa
39	FUNDACEP57RR	6.2	Bayer
40	5D690RR	6.9	DowAgroscience
41	BMXPOTENCIARR	6.7	GDM
42	BRS246RR	7.3	Embrapa
43	5D660RR	6.6	DowAgroscience
44	FUNDACEP58RR	6.8	Bayer
45	FUNDACEP53RR	6.4	Bayer

46	CD231RR	7.3	Coodetec
47	CD232	7.3	Coodetec
48	M7211RR	7.2	Monsanto
49	BRS268	7.2	Embrapa
50	TMG1066RR	6.6	TMG
51	M6707RR	6.7	Monsanto
52	BMXMAGNARR	6.2	GDM
53	CD226RR	6.6	Coodetec
54	CD201	6.7	Coodetec
55	CD242RR	7.9	Coodetec
56	CD240RR	6.9	Coodetec
57	NK7059	7	Syngenta
58	BRS284	6.4	Embrapa
59	CD236RR	6.2	Coodetec
60	CD2630RR	6.3	Coodetec
61	CD202	6.4	Coodetec
62	IGRARA628RR	6.4	Igra
63	CD248RR	6.4	Coodetec
64	CD235RR	6.4	Coodetec
65	NA5909RR	5.9	Nidera
66	CD233RR	6.4	Coodetec
67	CD2721RR	7.2	Coodetec
68	CD2737RR	7.3	Coodetec
69	IAS5	6.4	Embrapa
70	CDFAPA	7.3	Coodetec
71	R7	7	-
72	FT-GUAIRA	6.4	Embrapa
73	GARIMPO	7.8	Embrapa
74	EMBRAPA59	6.5	Embrapa
75	5G770RR	7.7	DowAgroscience
76	BRS213	6.5	Embrapa
77	Fundacep33	8	Bayer
78	CD254RR	8.4	Coodetec
78	CD254RR	8.4	Coodetec
79	IGRARA518RR	6	Igra
80	OCEPAR14	5.8	Embrapa
81	MG/BR 46 Conquista	8.2	Embrapa
82	CD2840	8.4	Coodetec
83	CD2800	8	Coodetec

84	P98Y51	8.5	Pioneer
85	P98Y70	8.7	Pioneer
86	CD246	8.1	Coodetec
87	CD237RR	7.3	Coodetec
88	P98Y11	8.1	Pioneer
89	CD243RR	8	Coodetec
90	CD244RR	8	Coodetec
91	M-SOY7901	7.9	Monsanto
92	TMG115RR	8.6	TMG
93	M-SOY8001	8	Monsanto
94	M9144RR	9.1	Monsanto
95	CD234RR	8	Coodetec
96	BMXAPOLLORR	5.5	GDM
97	CD2792RR	7.9	Coodetec
98	M6009RR	6	Monsanto
99	CD253	8.7	Coodetec
100	BRSMT Pintado	8.5	Embrapa

(-) Mantenedor desconhecido

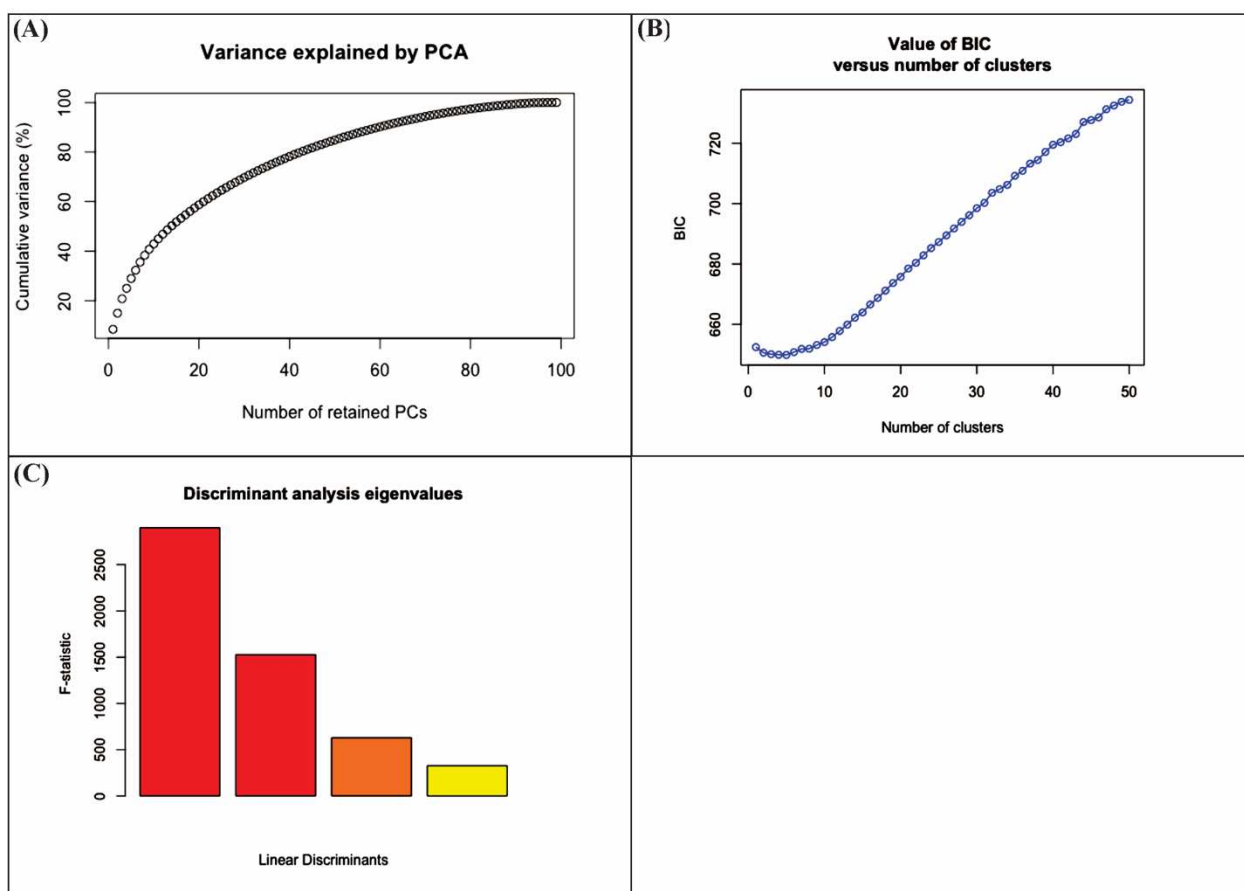


Figura S1. Critérios para clusterização dos cultivares. Variância acumulativa dos marcadores explicada pelos PCA (A), Valores de Critério de informação Bayesiana (BIC) versus o número de clusters (B), e os autovalores com sua respectiva significância (C).

Tabela S2 – Variâncias e herdabilidade das características fenotípicas utilizando a abordagem 1.

Variável	σ^2_A	σ^2_F	h^2
wc	1.14E+00	1.64E+01	6.97E-02
rdw	5.83E-06	7.12E-04	8.20E-03
sdw	1.65E-05	2.98E-03	5.53E-03
ph	0.00E+00	5.34E+01	0.00E+00
emer	8.16E+01	4.84E+02	1.68E-01
gsi	1.19E-02	1.59E-01	7.46E-02
lpvr	0.00E+00	6.65E+03	0.00E+00
apr	0.00E+00	3.22E+01	0.00E+00
sar	0.00E+00	3.18E+02	0.00E+00
adr	0.00E+00	2.40E-02	0.00E+00
rv	0.00E+00	1.43E-01	0.00E+00
rcl1	0.00E+00	2.51E+03	0.00E+00
rcl2	8.88E+00	8.44E+02	1.05E-02
rcl3	0.00E+00	5.46E+01	0.00E+00
rcl4	0.00E+00	1.25E+01	0.00E+00
rcl5	1.91E-01	1.41E+01	1.35E-02
rcl6	0.00E+00	1.23E+01	0.00E+00
rcl7	0.00E+00	1.05E+01	0.00E+00
rcl8	9.68E-03	5.60E+00	1.73E-03
rcl9	6.81E-03	1.50E+00	4.53E-03
rcl10	1.77E-03	6.80E-01	2.60E-03
rcv1	0.00E+00	5.94E-04	0.00E+00
rcv2	2.06E-04	1.06E-02	1.95E-02
rcv3	0.00E+00	6.98E-03	0.00E+00
rcv4	0.00E+00	7.04E-03	0.00E+00
rcv5	3.57E-04	2.28E-02	1.57E-02
rcv6	0.00E+00	4.44E-02	0.00E+00
rcv7	0.00E+00	6.92E-02	0.00E+00
rcv8	1.60E-04	7.00E-02	2.28E-03
rcv9	1.30E-04	2.91E-02	4.46E-03
rcv10	0.00E+00	3.87E-02	0.00E+00

* σ^2_A – Variância aditiva, σ^2_F – Variância fenotípica, h^2 – herdabilidade no sentido restrito

wc: conteúdo de água nas folhas, rdw: massa seca da raiz, sdw: massa seca da parte aérea, ph: altura de plantas, emer: taxa de emergência de plantas, gsi: índice de velocidade de germinação, lpvr: comprimento total da raiz, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcl7: comprimento da raiz por classes de diâmetros 7, rcl8: comprimento da raiz por classes de diâmetros 8, rcl9: comprimento da raiz por classes de diâmetros 9, rcl10: comprimento da raiz por classes de diâmetros 10, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6, rcv7: volume da raiz por classes de diâmetros 7, rcv8: volume da raiz por classes de diâmetros 8, rcv9: volume da raiz por classes de diâmetros 9, rcv10: volume da raiz por classes de diâmetros 10.

Tabela S3 – Variâncias e herdabilidade das características fenotípicas considerando a abordagem 2.

Variável	σ^2_A	σ^2_F	h^2
wc	1.22E-01	9.54E+00	1.28E-02
rdw	1.00E-05	3.49E-04	2.87E-02
sdw	8.78E-06	2.26E-03	3.88E-03
ph	0.00E+00	1.43E+01	0.00E+00
emer	1.56E+01	2.57E+02	6.06E-02
gsi	1.67E-03	5.94E-02	2.81E-02
lpvr	2.74E+01	5.08E+03	5.39E-03
apr	2.43E-01	3.18E+01	7.63E-03
sar	2.40E+00	3.14E+02	7.63E-03
adr	1.04E-05	1.56E-02	6.66E-04
rv	1.00E-03	1.57E-01	6.37E-03
rcl1	0.00E+00	1.61E+03	0.00E+00
rcl2	1.12E+01	6.93E+02	1.61E-02
rcl3	6.51E-01	3.06E+01	2.12E-02
rcl4	1.93E-01	6.90E+00	2.80E-02
rcl5	8.31E-02	7.90E+00	1.05E-02
rcl6	0.00E+00	7.54E+00	0.00E+00
rcl7	0.00E+00	8.91E+00	0.00E+00
rcl8	6.99E-03	3.26E+00	2.14E-03
rcl9	0.00E+00	8.73E-01	0.00E+00
rcl10	1.24E-03	4.44E-01	2.78E-03
rcv1	1.75E-06	4.24E-04	4.11E-03
rcv2	1.48E-04	8.12E-03	1.82E-02
rcv3	8.56E-05	3.89E-03	2.20E-02
rcv4	1.12E-04	3.90E-03	2.88E-02
rcv5	1.36E-04	1.29E-02	1.06E-02

rcv6	0.00E+00	2.75E-02	0.00E+00
rcv7	0.00E+00	5.90E-02	0.00E+00
rcv8	7.91E-05	4.03E-02	1.96E-03
rcv9	0.00E+00	1.70E-02	0.00E+00
rcv10	2.23E-05	2.38E-02	9.36E-04

* σ^2_A – Variância aditiva, σ^2_F – Variância fenotípica, h^2 – herdabilidade no sentido restrito
wc: conteúdo de água nas folhas, rdw: massa seca da raiz, sdw: massa seca da parte aérea, ph: altura de plantas, emer: taxa de emergencia de plantas, gsi: índice de velocidade de germinação, lpvr: comprimento total da raiz, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcl7: comprimento da raiz por classes de diâmetros 7, rcl8: comprimento da raiz por classes de diâmetros 8, rcl9: comprimento da raiz por classes de diâmetros 9, rcl10: comprimento da raiz por classes de diâmetros 10, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6, rcv7: volume da raiz por classes de diâmetros 7, rcv8: volume da raiz por classes de diâmetros 8, rcv9: volume da raiz por classes de diâmetros 9, rcv10: volume da raiz por classes de diâmetros 10

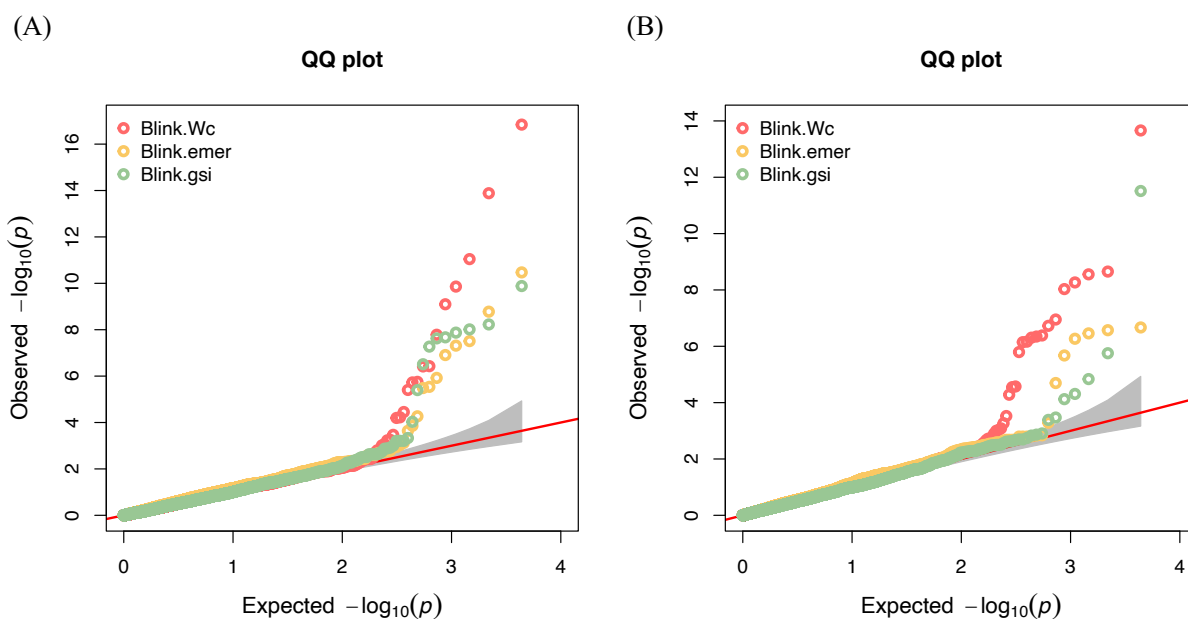


Figura S2 – Quantile-quantile (QQ) -plot dos P-valores esperado e observado utilizado na primeira abordagem (A) e na segunda abordagem (B) para as características fenotípicas selecionadas em cada situação.

wc: conteúdo de água na folha, emer: taxa de emergencia de plantas, gsi: índice de velocidade de germinação.

Tabela S4 – Marcadores SNPs significativamente associados a características em cada uma das duas abordagens utilizadas.

Approach associated	SNP	Associated Trait
1 & 2	Gm07_5989451_C_T	emer & gsi
1 & 2	Gm10_39086213_C_A	emer & gsi
1 & 2	Gm15_4613500_G_A	emer

1	Gm01_50423856_A_C	Wc
1	Gm02_46023727_T_C	Wc
1	Gm03_46998040_A_G	Wc
1	Gm05_39021151_C_A	Wc
1	Gm05_39054172_A_G	emer
1	Gm06_16315206_A_G	Wc
1	Gm08_43345646_C_T	Wc
1	Gm09_611923_A_G	gsi
1	Gm10_40258740_G_A	gsi & emer
1	Gm11_36931482_C_T	gsi
1	Gm11_8363320_T_C	emer
1	Gm13_28957669_T_C	gsi
1	Gm15_15896387_C_T	Wc
1	Gm17_36602853_A_G	emer & wc
1	Gm17_36610073_C_T	gsi
1	Gm18_12406839_A_G	gsi
1	Gm19_1489018_G_A	gsi
1	Gm19_33380479_G_T	Wc
1	Gm19_34472334_A_C	emer
1	Gm20_208975_C_T	gsi
1	Gm20_2372509_T_C	Wc
1	Gm20_45953874_G_A	Wc
2	Gm01_53365928_A_G	Wc
2	Gm01_55060537_G_T	Wc
2	Gm03_35516341_C_T	emer
2	Gm04_47092275_T_C	Wc
2	Gm04_47957714_A_C	Wc
2	Gm06_36128279_T_C	Wc
2	Gm07_14862607_C_T	Wc
2	Gm07_7618698_T_C	Wc
2	Gm08_11073425_T_G	Wc
2	Gm08_45270892_A_G	Wc
2	Gm10_40077126_A_G	emer
2	Gm11_30114058_A_G	Wc
2	Gm15_46014592_T_C	Wc
2	Gm16_30991544_A_G	Wc
2	Gm19_50322283_T_G	Wc

wc: conteúdo de água na folha, emer: taxa de emergencia de plantas, gsi: índice de velocidade de germinação.

CAPÍTULO 2 – ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS) PARA A SOJA DURANTE O ESTÁDIO VEGETATIVO REVELA NOVOS GENES ENVOLVIDOS NA TOLERÂNCIA AO DÉFICIT HÍDRICO

RESUMO

A cultura da soja é a principal oleaginosa produzida mundialmente devido a sua gama de utilização pela indústria. No entanto, nos últimos anos, devido as mudanças climáticas, eventos adversos como a seca estão cada vez mais frequente. Esse período pode ocorrer em qualquer estágio desenvolvimento da cultura e seu entendimento é dificultado pela complexidade dos fatores envolvidos. Assim, o objetivo desse trabalho é encontrar marcadores moleculares e os possíveis genes associados a estes marcadores na cultura da soja durante o déficit hídrico durante o estágio vegetativo. Foram conduzidos dois experimentos em casa de vegetação na Universidade Federal de Viçosa. Estes foram realizados em blocos casualizados em fatorial 100 x 2 (Genótipos x disponibilidade hídrica) com três repetições. A parcela experimental constituiu de uma planta em um vaso de 3 litros. As plantas foram mantidas na condição próximo a capacidade de campo até atingirem o estágio vegetativo 2 (V2), e a partir desse momento foi iniciado o período de estresse, que perdurou por 20 dias. A água evapotranspirada, foi repostada diariamente de acordo com cada condição (Controle (-0,01 MPa) ou estresse (-1,2 MPa)). Ao término do período de estresse foram obtidas as seguintes variáveis: massa seca da parte aérea (sdw), massa seca das raízes (rdw), diâmetro do hipocótilo (dhip), número de nós (nnos), altura de planta (ph), comprimento total da raiz (lpvr), área superficial da raiz (apr), comprimento da raiz por classes de diâmetros 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (rcl1, rcl2, rcl3, rcl4, rcl5, rcl6, rcl7, rcl8, rcl9, rcl10) e volume da raiz por classes de diâmetros 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (rcv1, rcv2, rcv3, rcv4, rcv5, rcv6, rcv7, rcv8, rcv9, rcv10). Para a análise foram consideradas duas abordagens para o conceito de tolerância. Na primeira, um genótipo é considerado mais tolerante se a diferença entre estresse e controle for menor que de outro genótipo. Na segunda, um genótipo é considerado tolerante se possui uma alta produtividade em ambas condições. Os dados então foram corrigidos a partir de modelos lineares mistos, via REML/BLUP para cada característica, sendo descartadas as com herdabilidade menor que 1%. Realizou-se uma análise de PCA-biplot com as características selecionadas dentro de cada abordagem, e as variáveis que não contribuíram muito para a diferenciação entre os grupos (estresse e controle) foram descartadas. Foi realizado também uma análise de correlação com os valores de gBLUP em cada uma das abordagens. Os valores de gBLUP das características foram utilizados para a análise de estudo de associação genômica ampla (GWAS) dentro de

cada abordagem. Foram encontrados 18 marcadores significativos simultâneos à ambas as abordagens, e nas regiões próximas a esses marcadores foram encontrados 15 genes candidatos. Dentre esses genes, encontra-se uma proteína chaperona Hsp40s, uma proteína “Ring finger”, quatro que são conhecidas em atuar em estresse hídrico, três que atuam no desenvolvimento das plantas, principalmente em raízes e outros seis que atuam na tolerância a outros tipos de estresse.

Palavras-chave: *Glycine max*. Seca. Melhoramento de plantas. Herança quantitativa.

1. Introdução

Dentre as culturas oleaginosas do mundo, a soja pode ser considerada a principal, fornecendo óleo e proteína para o consumo humano e animal, além de diversos usos na indústria (Graham & Vance, 2003). No entanto, inúmeras condições ambientais adversas restringem a produtividade agrícola global (Eftekhari et al., 2017; Muscolo et al., 2014). O déficit hídrico é uma dessas condições que restringem severamente o desenvolvimento e a produção das culturas, e que deve se tornar ainda mais intenso com as mudanças climáticas atuais (Shin et al., 2015). Na cultura da soja, segundo estimativas, essas condições podem comprometer a produtividade em cerca de 40% (Specht et al., 1999).

Estudos em plantas têm sido realizados para tentar entender o processo de tolerância ao déficit hídrico (Meeks et al., 2019; Jumrani & Bhatia, 2018; Wei et al., 2018; Gallardo et al., 2006; Goldhamer et al., 2001). Contudo, esse processo é muito complexo porque envolve adaptações morfológicas, fisiológicas, bioquímicas e metabólicas específicas que afetam tanto o crescimento quanto o desenvolvimento das plantas (Luo et al., 2020; Joshi et al., 2016; Easwar et al., 2016; Wang et al., 2015). Isso se deve fundamentalmente pela característica multigênica e quantitativa do processo de tolerância, e com isso a utilização de métodos convencionais para o melhoramento de plantas tem sido um desafio (Sun et al., 2022).

Em relação ao déficit hídrico, os estudos concentram-se principalmente em características ligadas à produtividade (Du et al., 2009), raízes fibrosas (Abdel-Haleem et al., 2011) e atributos relacionados às folhas (Hwang et al., 2016; Specht et al., 2001; Mian et al., 1998) nas fases de germinação e na fase reprodutiva. Porém, o déficit hídrico pode ocorrer em qualquer uma das fases de desenvolvimento da cultura. Quando, esse estresse ocorre nas fases de crescimento irão afetar o acúmulo de matéria seca e consequentemente afetar negativamente a produção (Jumrani & Bhatia, 2018; Wei et al., 2018).

Sabe-se que as plantas frente ao déficit hídrico ativam seus mecanismos de respostas como as mudanças morfológicas e estruturais, síntese de hormônios e produção de produtos osmorreguladores através da expressão de genes de tolerância (Luo et al., 2020). As plantas produzem moléculas sinalizadoras como ácido abscísico (ABA), Ca^{2+} , Inositol-1,2,5-trifosfato (IP3), adenosina 5'-difosfato ribose cíclica (cADP), NO, em resposta a sinais de estresse hídrico sob condições de seca. Como resposta a produção dessas moléculas alguns genes são ativados gerando respostas a essas moléculas. Assim, alguns produtos gênicos funcionais como prolina, glicina betaína, açúcar solúvel, proteínas LEA e aquaporinas são produzidos, e influenciam no processo de defesa da planta. (Yang et al. 2021). Outros produtos

atuam controlando as vias de transdução de sinal ou funcionam como fatores de transcrição, como as proteínas quinases dependentes de cálcio (CDPKs), proteínas quinases ativadas por mitógeno (MAPKs), HD-zip/bZIP, AP2/ERF, NAC, MYB e WRKY alterando a morfologia ou fisiologia das plantas induzindo o processo de tolerância (Yang et al. 2021).

Conhecer e entender a relação dos genes que atuam nesse processo é de fundamental importância. Dentre as ferramentas utilizadas para a busca de genes relacionados a fenótipos, principalmente de característica quantitativa, o estudo de associação genômica ampla tem auxiliado ao detectar regiões genômicas associadas a estas características (Uffelmann et al., 2021). Dessa forma, esse trabalho tem como objetivo encontrar os marcadores moleculares e os possíveis genes associados a estes marcadores na cultura da soja durante o déficit hídrico imposto no estágio vegetativo.

2. Material e Métodos

2.1. Condução dos experimentos

Foram conduzidos dois experimentos em casa de vegetação na Universidade Federal de Viçosa, Viçosa – Minas Gerais (20°45'17" S, 42°52'57" O). O experimento 1 foi realizado entre outubro e dezembro de 2019. Já o experimento 2, foi realizado entre setembro e novembro de 2021.

O delineamento utilizado foi em blocos casualizados com três repetições em um fatorial 100 x 2 (genótipos x disponibilidade hídrica). Os genótipos utilizados foram cultivares oriundas de diferentes empresas e com ampla variação nas características, a fim de representar um painel de diversidade da cultura (Tabela S1). Os regimes de disponibilidade hídrica foram a condição controle e a condição estresse.

Na condição controle, o solo foi mantido próximo a capacidade de campo, de acordo com a tensão para um solo franco-arenoso (-10 kPa), em todo o período do experimento (Bernardo et al., 2019). Já na condição estresse, o solo foi mantido na capacidade de campo até que as plantas atingissem o estágio vegetativo V2 (Fehr & Caviness, 1977), após esse período foram mantidas em uma tensão de -1.200 kPa por 20 dias. A metodologia para a indução do estresse foi de acordo com Nascimento et al. (2021), utilizando as pesagens diárias para reposição do volume de água evapotranspirado. No quinto dia, após o início do período de estresse, o potencial hídrico do solo foi verificado de maneira indireta com a bomba de pressão do tipo Scholander na antemanhã, na terceira folha completamente expandida (Scholander et al., 1965).

A parcela experimental constituiu de um vaso contendo uma planta cultivada em um substrato contendo uma mistura de solo e areia na proporção respectiva de 3:2 para o experimento 1 e uma mistura de solo e areia na proporção respectiva de 2:1 no experimento 2. Esses substratos foram classificados com uma textura franco-arenosa nos dois experimentos. A adubação utilizada foi de acordo com a correção da análise química para uma produção de 3.000 kg/ha (Ribeiro et al., 1999).

As sementes foram tratadas com o fungicida MAXIM® XL com a dose recomendada pelo fabricante. No sulco de plantio foi aplicado uma dose de *Bradyrhizobium japonicum* cepas SEMIA 5079 e 5080 da RIZOKOP®, utilizando a dose recomendada pelo fabricante. Os demais tratos culturais realizados foram de acordo com a cultura da soja (Silva et al., 2022). A temperatura e a umidade de todo o ambiente foram aferidas diariamente com auxílio de um termo-higrômetro digital durante todo o período em que as plantas se encontravam em período de estresse (Figura S1).

2.2. Avaliação dos caracteres da parte aérea

Após o término do período de estresse as plantas foram avaliadas para obter os caracteres agrônômicos da parte aérea e da parte radicular.

Na parte área, os caracteres avaliados foram: altura das plantas (ph) em centímetros, medida com o auxílio de uma régua; e, diâmetro do hipocótilo (dhip) em milímetros, medido com o auxílio de um paquímetro digital modelo 316119 - MTX. Após as medições a parte aérea foi colocada em estufa a 65°C por um período de 48 horas e a sua massa seca (sdw) foi aferida em gramas com auxílio de uma balança de precisão.

As raízes foram removidas e lavadas para remoção do material inerte (areia e solo). Após a lavagem cada raiz foi armazenada separadamente em um pote contendo álcool 70% e mantidas em temperatura de 4°C até a sua avaliação. Para a avaliação as raízes foram escaneadas e a imagem submetida ao software WinRHIZOPro® 2009, para a obtenção das seguintes características: comprimento total (cm) (lpvr), área superficial projetada (cm²), área superficial (cm²) (sar), volume (cm³) (rv), diâmetro médio (mm) (adr), comprimento da raiz por classes de diâmetros (cm) (rcl) e volume da raiz por classes de diâmetros (cm³) (rcv). As classes de diâmetro (D) foram: classe 1 ($0 < D \leq 0.50$), classe 2 ($0.50 < D \leq 1.00$), classe 3 ($1.00 < D \leq 1.50$), classe 4 ($1.50 < D \leq 2.00$), classe 5 ($2.00 < D \leq 2.50$), classe 6 ($2.50 < D \leq 3.00$), classe 7 ($3.00 < D \leq 3.50$), classe 8 ($3.50 < D \leq 4.00$), classe 9 ($4.00 < D \leq 4.50$), e classe 10 ($D > 4.50$).

Após serem escaneadas, as raízes foram acondicionadas em estufa de ar forçado por 48 h à 65 °C. As raízes secas foram pesadas em balança analítica digital de precisão, obtendo a massa seca das raízes em gramas (rdw).

2.3. Estimação dos valores genotípicos

Inicialmente, foi realizado uma análise de boxplot para as características de massa seca da parte aérea e da raiz, considerando somente o tratamento dentro de cada experimento e desconsiderando o efeito dos genótipos. As médias foram comparadas com o teste t, $P < 0,01$.

Para considerar uma planta mais tolerante, consideramos duas abordagens. Na primeira abordagem, foi considerado a diferença entre estresse e controle pois utilizamos o conceito de que uma planta pode ser mais tolerante, se a redução da sua característica na condição estresse em relação a controle for menor que de outro genótipo (Miti et al., 2010). Na segunda abordagem, utilizamos uma abordagem em que a planta tolerante é aquela que possui uma produção elevada nas duas condições. Assim, nessa segunda abordagem o genótipo mais tolerante seria aquele que sobre a condição de estresse conseguir ter a maior produtividade. Contudo, a sua produção dentro da condição controle também foi contabilizada.

Na primeira abordagem, os dados das plantas dentro de cada bloco em cada ano, da condição estresse foram subtraídos da condição controle. Assim obtemos o valor que aquela planta aumentou ou reduziu em relação a condição controle. Na segunda abordagem foram utilizados os dados brutos contabilizando os tratamentos, os experimentos (épocas) e os blocos dentro de cada época.

Utilizou-se a metodologia de modelos lineares mistos, via REML/BLUP em cada característica visando a estimativa de valores genotípicos e dos parâmetros genéticos. Para tanto foi utilizado o pacote “sommer” (Covarrubias-Pazaran, 2016) no software R (R CORE TEAM, 2016). A matriz de parentesco utilizada nesse modelo foi calculada a partir da matriz numérica dos marcadores utilizados nesse trabalho (informações sobre os marcadores serão detalhadas posteriormente) usando a metodologia de Zhang et al. (2007). Neste modelo, cada época (experimento) foi inserida como um fator aleatório, assim como o fator bloco. O modelo utilizado para cada abordagem se encontra representado abaixo:

1ª Abordagem:

$$y = \mu + Xr + Zg + Wp + e$$

2ª Abordagem:

$$y = \mu + Xr + Ti + Zg + Wp + e$$

onde “y” é o vetor de dados fenotípicos, “ μ ” é a média geral, “r” é o vetor dos efeitos de repetição (assumidos como fixos), “i” é o vetor dos efeitos de tratamento (assumidos como fixos), “g” é o vetor dos efeitos do i-ésimo genótipo (assumidos como aleatórios), “p” é o vetor dos efeitos de época (assumidos como aleatórios), e “e” é o vetor de erros ou resíduos (aleatórios) que segue $N(0, \sigma^2)$. As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

As características que apresentaram valores de herdabilidade menores que 1% foram descartadas para as próximas análises. A fim de selecionar as características mais importantes para o déficit hídrico, após selecionar as características através da herdabilidade, foi realizado uma análise com o pacote ggbiplot (Vu et al., 2015), em que foi plotado a análise de componentes principais e os valores de para cada parcela dentro de cada tratamento em cada uma das abordagens. As variáveis que se orientaram no sentido de separação dos grupos em estresse e controle foram selecionadas. Após essa seleção de características, foram realizadas as análises de correlação de Pearson, utilizando os valores de gBLUP, via o pacote “corrplot” (Wei & Simko, 2021). Os valores de gBLUP das características que correlacionaram com os valores de BLUP da massa seca de raiz e/ou massa seca da parte área, nas diferentes abordagens, foram selecionadas para a realização da associação genômica ampla. Todas as análises foram realizadas no software R (R CORE TEAM, 2016).

2.4. Genotipagem com marcadores SNP, associação genômica ampla e pós-GWAS

Os marcadores SNPs foram genotipados previamente por Contreras-Soto et al. (2017). Para essa genotipagem foi utilizado a plataforma iScan Illumina (Illumina, Inc., San Diego, CA, USA) customizada para soja, utilizando o “Infinium iSelect HD Custom Genotyping BARC-Soy SNP6k” na Deoxi Biotecnologia Ltda® (Araçatuba, SP, Brasil). A filtragem dos SNP foi realizada com um “call rate” > 90 e um MAF (“Minor Allele Frequency”) > 0,01.

Os valores de gBLUP das características selecionadas para cada abordagem após a etapas descritas acima foram utilizados no estudo de associação genômica ampla (GWAS). A análise GWAS foi realizada utilizando o pacote GAPIT versão 3.0 (Wang & Zhang, 2021). Para análise foi considerado o modelo “Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway” (Blink) (Huang et al., 2019). Esse modelo conduz dois modelos de efeito fixos (FEM) e um processo de filtragem que seleciona um conjunto de pseudo QTNs que não estão em LD entre si como covariáveis. No primeiro modelo FEM, os marcadores são testados, um de cada vez, e pseudo QTNs são incluídos como covariáveis para controlar os falsos negativos e positivos:

$$y_i = S_{i1}b_1 + S_{i2}b_2 + \dots + S_{ik}b_k + S_{ij}d_j + e_i$$

onde y_i é a observação no i -ésimo indivíduo; $S_{i1}, S_{i2}, \dots, S_{ik}$ são os genótipos de k pseudo QTNs, iniciados como um conjunto vazio; $b_1, b_2, \dots, b_k$ são os efeitos correspondentes dos pseudo QTNs; S_{ij} é o genótipo do i -ésimo indivíduo e j -ésimo marcador genético; d_j é o efeito correspondente do j -ésimo marcador genético; e e_i é o resíduo com distribuição com média zero e variância σ_e^2 .

O segundo modelo FEM, é utilizado para otimizar a seleção de pseudo QTNs:

$$y_i = S_{i1}b_1 + S_{i2}b_2 + \dots + S_{ik}b_k + e_i$$

onde y_i é a observação no i -ésimo indivíduo; $S_{i1}, S_{i2}, \dots, S_{ik}$ são os genótipos de k pseudo QTNs, iniciados como um conjunto vazio; $b_1, b_2, \dots, b_k$ são os efeitos correspondentes dos pseudo QTNs; S_{ij} é o genótipo do i -ésimo indivíduo; e e_i é o resíduo com distribuição com média zero e variância σ_e^2 .

No segundo modelo, o termo marcador teste é removido e o número de pseudo QTNs covariáveis é variado, a fim de selecionar o conjunto ótimo do primeiro k de t pseudo QTNs. A otimização é realizada usando BIC, que é o dobro da probabilidade logarítmica negativa mais a penalidade no número de parâmetros, como segue:

$$BIC = -2LL + 2k \ln(n)$$

onde LL é o log da probabilidade, k é o número de pseudo QTNs, $\ln$ é o logaritmo natural e n é o número de indivíduos. Os pseudo QTNs disponíveis, t , são classificados com o mais significativo no início e o menos significativo no final. Os primeiros k pseudo QTNs são selecionados para exame, com k variando de 1 a t .

No segundo modelo, os marcadores são filtrados com dois critérios: P-valor e correlação. Primeiramente, todos os marcadores do primeiro modelo são filtrados com correção de Bonferroni ($\alpha = 0,01$). Após os marcadores com correlação (correlação de Pearson) maior que 0,7 são eliminados. Então os primeiros conjuntos de k marcadores foram ajustados e os demais foram variados até que todas as possibilidades tenham sido examinadas. O conjunto de K marcadores com melhor BIC é usado como o conjunto de pseudo QTNs na primeira equação. O processo é repetido até que os pseudo QTNs permaneçam os mesmos. Ao final, os marcadores que apresentaram valores de $p < 0,01$ após a correção pela FDR foram considerados significativos.

Os marcadores que foram significativos nas duas abordagens simultaneamente foram selecionados para o estudo de pós-GWAS. Foi realizado uma busca dos genes associados aos marcadores, utilizando uma região de 2000 pares de bases (bp) à montante (“up”) e à jusante (“down”) de cada marcador. Para realizar essa busca foi usado o “Genome Data Viewer” do NCBI utilizando a versão mais recente do genoma da soja (*Glycine max* Wm82.a4.v1). Nos genes em que os marcadores se encontravam na região codante da proteína foi verificado o efeito da substituição alélica na codificação dos aminoácidos.

3. Resultados

A imposição de déficit hídrico se demonstrou eficiente nos dois anos. Na comparação do acúmulo da matéria seca, tanto na parte aérea quanto na parte radicular, verificou-se que os valores obtidos para os genótipos nos diferentes tratamentos de disponibilidade hídrica foram diferentes para os dois experimentos conduzidos (Figura 1).

A média da massa seca da parte aérea (sdw) de todos os genótipos na condição controle no experimento 1 foi 1,9 vezes maior que na condição estresse ($t(451,17)=23,54$; $p < 0,001$) (Figura 1A). Já no experimento 2, a média na condição controle foi de 5,4 vezes maior que na condição estresse ($t(290,85)=27,46$; $p < 0,001$) (Figura 1B). A média da massa seca das raízes (rdw) também diferiu significativamente entre a condição estresse e controle dentro dos experimentos. A média da massa seca das raízes (rdw) de todos os genótipos, no experimento 1, foi 1,6 vezes maior na condição controle em relação à condição estresse ($t(430,26)=18,28$; $p < 0,001$) (Figura 1C). No experimento 2, a média de todos os genótipos foi de 4,1 vezes maior na condição controle do que na condição estresse ($t(301,08)=22,13$; $p < 0,001$) (Figura 1D).

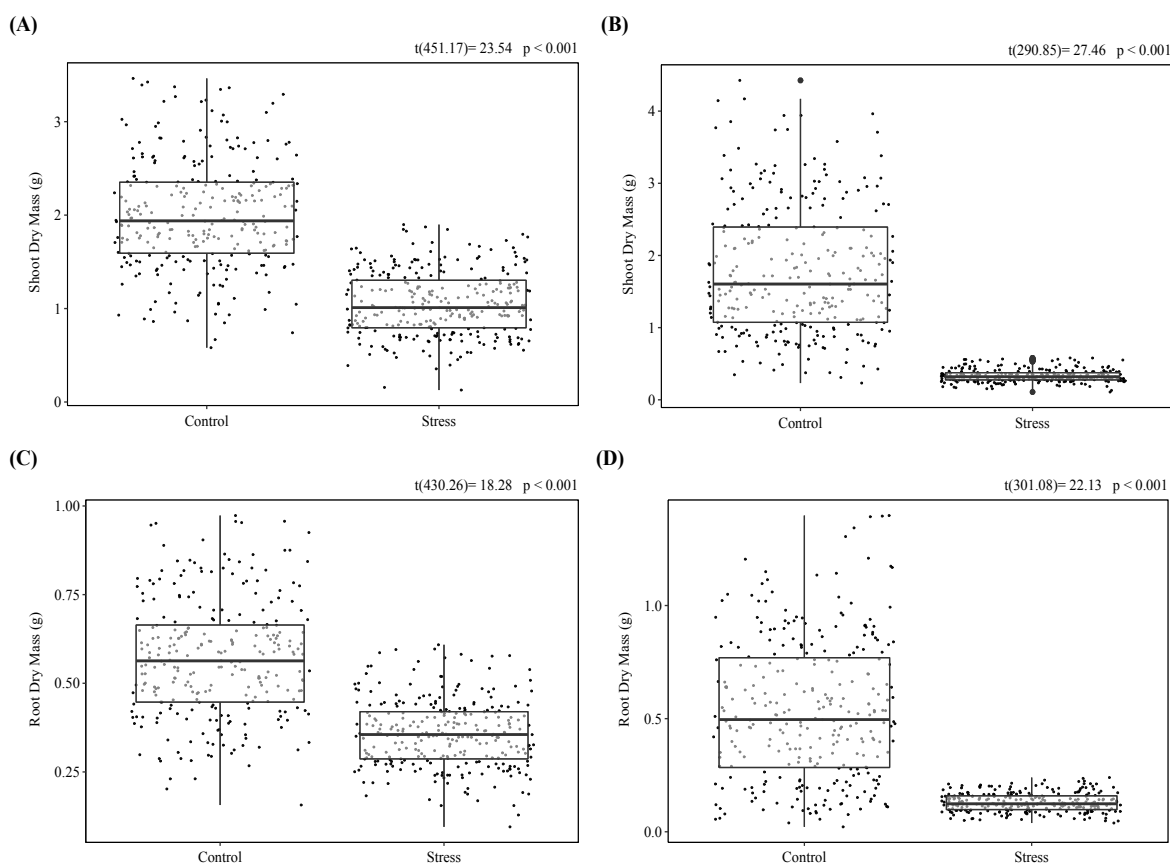
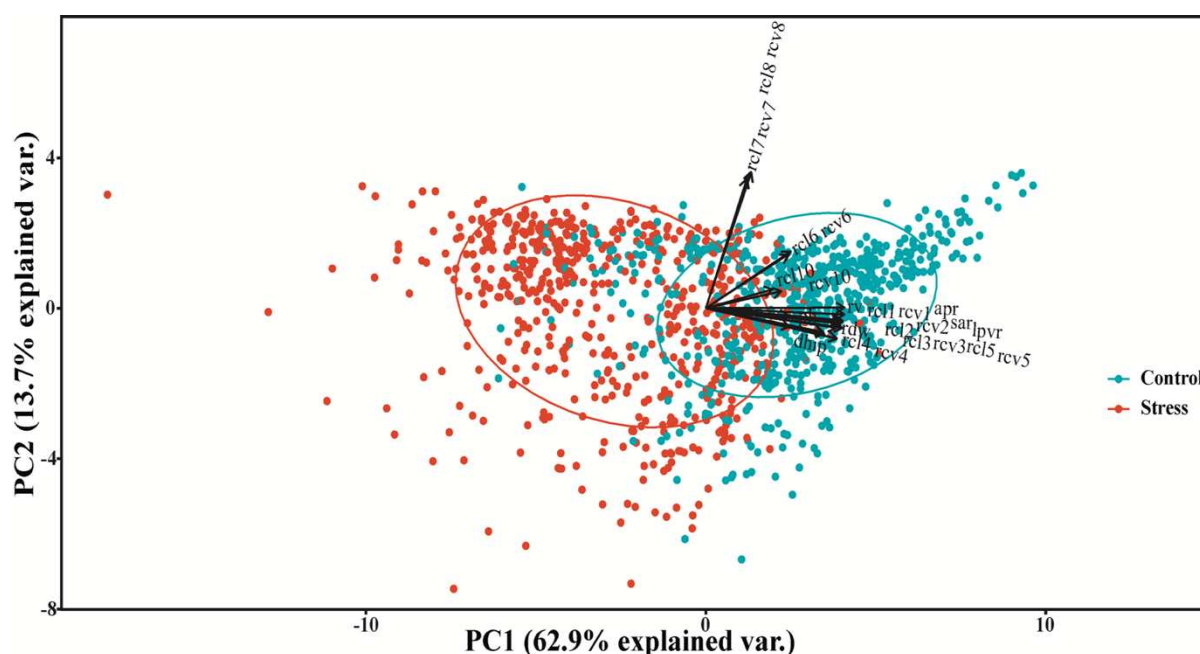


Figura 1 – Conteúdo de matéria seca das raízes e da parte aérea na condição controle (1) e na condição estresse (2) nos dois experimentos. Média e dispersão dos dados de Massa seca da parte aérea no experimento 1 (A) e no experimento 2 (B). Média e dispersão Massa seca das raízes no experimento 1 (C) e no experimento 2 (D).

As médias foram significativamente diferentes pelo teste t ($p < 0,001$).

Os valores genotípicos dos genótipos e os parâmetros genéticos e ambientais em cada abordagem para as características avaliadas se encontram nas Tabelas S2 e S3. As variáveis que apresentaram herdabilidade menor que 1% ($h^2 \leq 0,01$) foram descartadas das próximas análises. Na primeira abordagem foram descartadas as variáveis número de nós (nnos), massa seca da parte aérea (sdw), diâmetro médio das raízes (adr), comprimento da raiz por classes de diâmetros 9 (rcl9), e volume da raiz por classes de diâmetros 9 (rcv9). Na segunda abordagem foram descartadas as variáveis número de nós (nnos), diâmetro médio das raízes (adr), comprimento da raiz por classes de diâmetros 7, 8, 9 e 10 (rcl7, rcl8, rcl9 e rcl10), e volume da raiz por classes de diâmetros 7, 8, 9 e 10 (rcv7, rcv8, rcv9 e rcv10).

Ao realizar a análise de PCA biplot para a abordagem 1, observa-se que os dados do tratamento controle e estresse formam grupos distintos (Figura 2). O primeiro componente contribuiu com 62,9% e o segundo com 13,7%. A maior parte das variáveis contribuem para essa separação. Apenas as variáveis comprimento da raiz por classes de diâmetros 7 e 8 (rcl7 e



rcl8), e volume da raiz por classes de diâmetros 7 e 8 (rcv7 e rcv8) não se orientam na direção da separação dos grupos e foram eliminadas das próximas análises (Figura 2).

Figura 2 – Relação entre as variáveis e sua contribuição para diferenciação entre os grupos de estresse e controle na abordagem 1. PCA biplot, plotando as variáveis de acordo com os dois primeiros componentes principais e os dados de estresse e controle (A).

ph: altura de planta, dhpr: diâmetro do hipocótilo, rdw: massa seca da raiz, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcl7: comprimento da raiz por classes de diâmetros 7, rcl8: comprimento da raiz por classes de diâmetros 8, rcl10: comprimento da raiz por classes de diâmetros 10, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2,

rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6, rcv7: volume da raiz por classes de diâmetros 7, rcv8: volume da raiz por classes de diâmetros 8, rcv10: volume da raiz por classes de diâmetros 10.

Ao realizar uma análise de correlação de Pearson com os valores de gBLUP das variáveis restantes da abordagem 1, verifica-se que todas as características possuem correlação positiva entre si (Figura3). A variável massa seca da raiz (rdw), apresenta uma correlação moderada e positiva com as variáveis altura de planta (0,36) (ph) e comprimento da raiz por classes de diâmetros 10 (0,68) (rcv10) e apresenta uma correlação forte e positiva com as demais variáveis. A partir disso todas as variáveis foram selecionadas para participar do estudo de associação genômica ampla na abordagem 1.

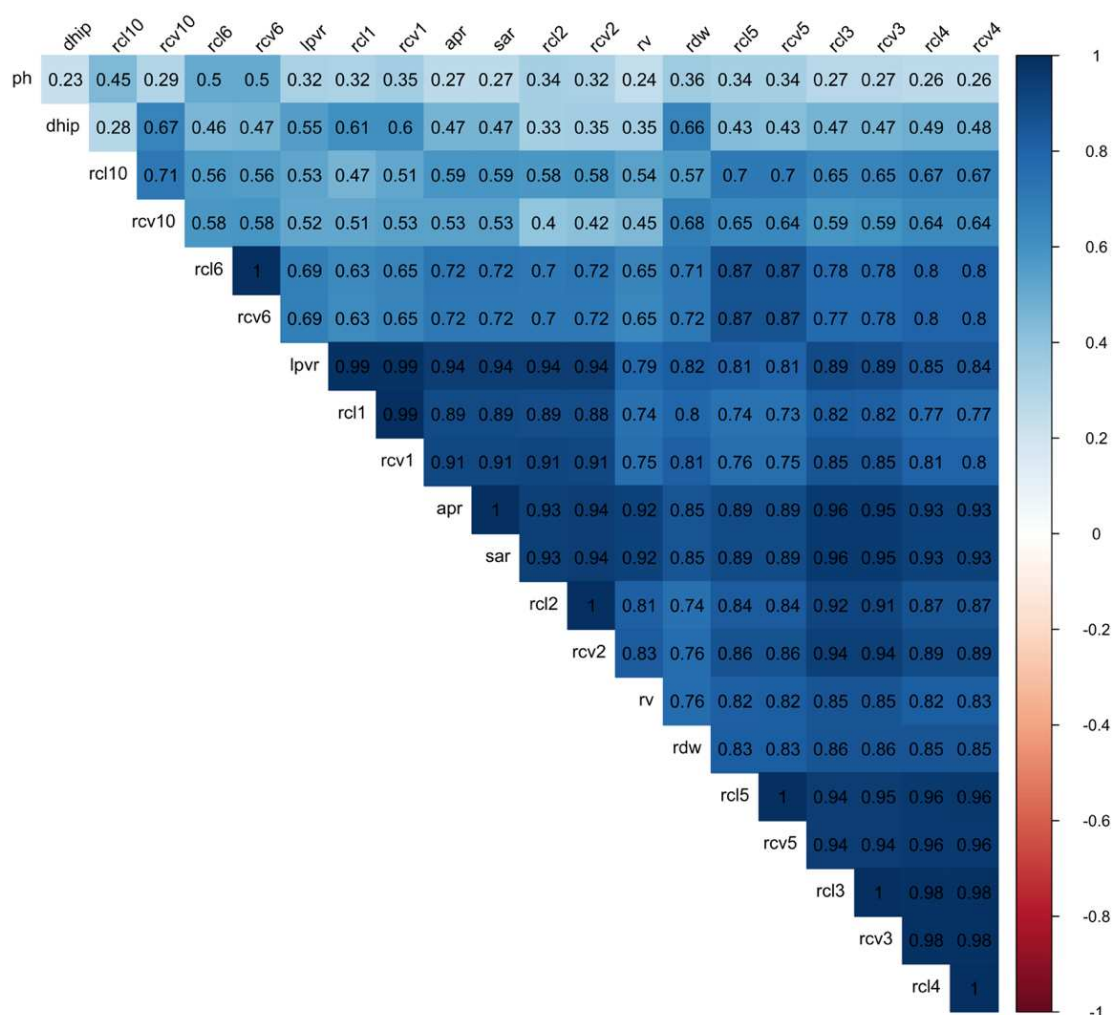


Figura 3 - Coeficientes da análise de correlação de Pearson entre os valores genotípicos preditos (BLUP) das variáveis avaliados para a abordagem 1.

ph: altura de planta, dhip: diâmetro do hipocótilo, rdw: massa seca da raiz, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de

ph: altura de planta, dh1p: diâmetro do hipocótilo, rdw: massa seca da raiz, sdw: massa seca da parte aérea, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz

por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6

Ao se realizar a correlação entre os valores genotípicos preditos dos genótipos para as variáveis, a maior parte das variáveis apresentaram correlação positiva entre si (Figura 5). Apenas a variável altura de planta (ph) não apresentou correlação com a massa seca da raiz (rdw) e apresentou correlação moderada e positiva com a massa seca da parte aérea (sdw). As demais variáveis apresentaram uma correlação forte e positiva com sdw e rdw. Assim, todas essas características foram selecionadas para a etapa de GWAS.

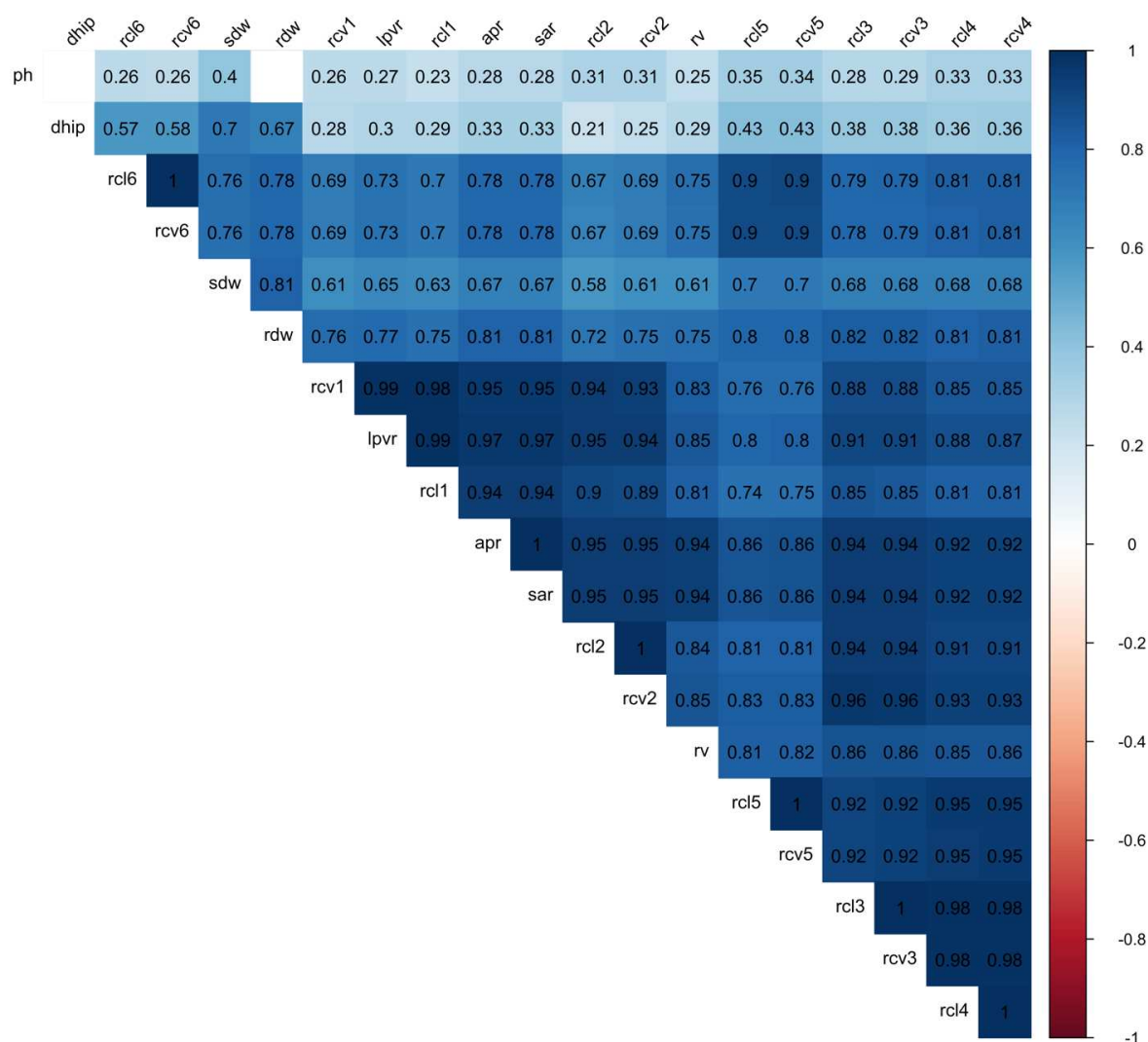


Figura 5 - Coeficientes da análise de correlação de Pearson entre os valores genotípicos preditos (BLUP) das variáveis avaliados para a abordagem 2.

ph: altura de planta, dhip: diâmetro do hipocótilo, rdw: massa seca da raiz, sdw: massa seca da parte aérea, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5:

comprimento da raiz por classes de diâmetros 5, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5.

No estudo de associação genômica utilizando o modelo Blink, foram encontrados 114 SNPs significativos unindo todas as características utilizadas (Figura 6). O modelo utilizado se ajustou bem para todas as variáveis (Figura S3). As linhas verticais no manhattan plot indicam que o SNP foi significativamente em mais de uma característica. No total foram 29 SNPs que apresentaram significativos em mais de uma característica, como o SNP Gm09_34354926_C_T que foi significativo em 11 características diferentes.

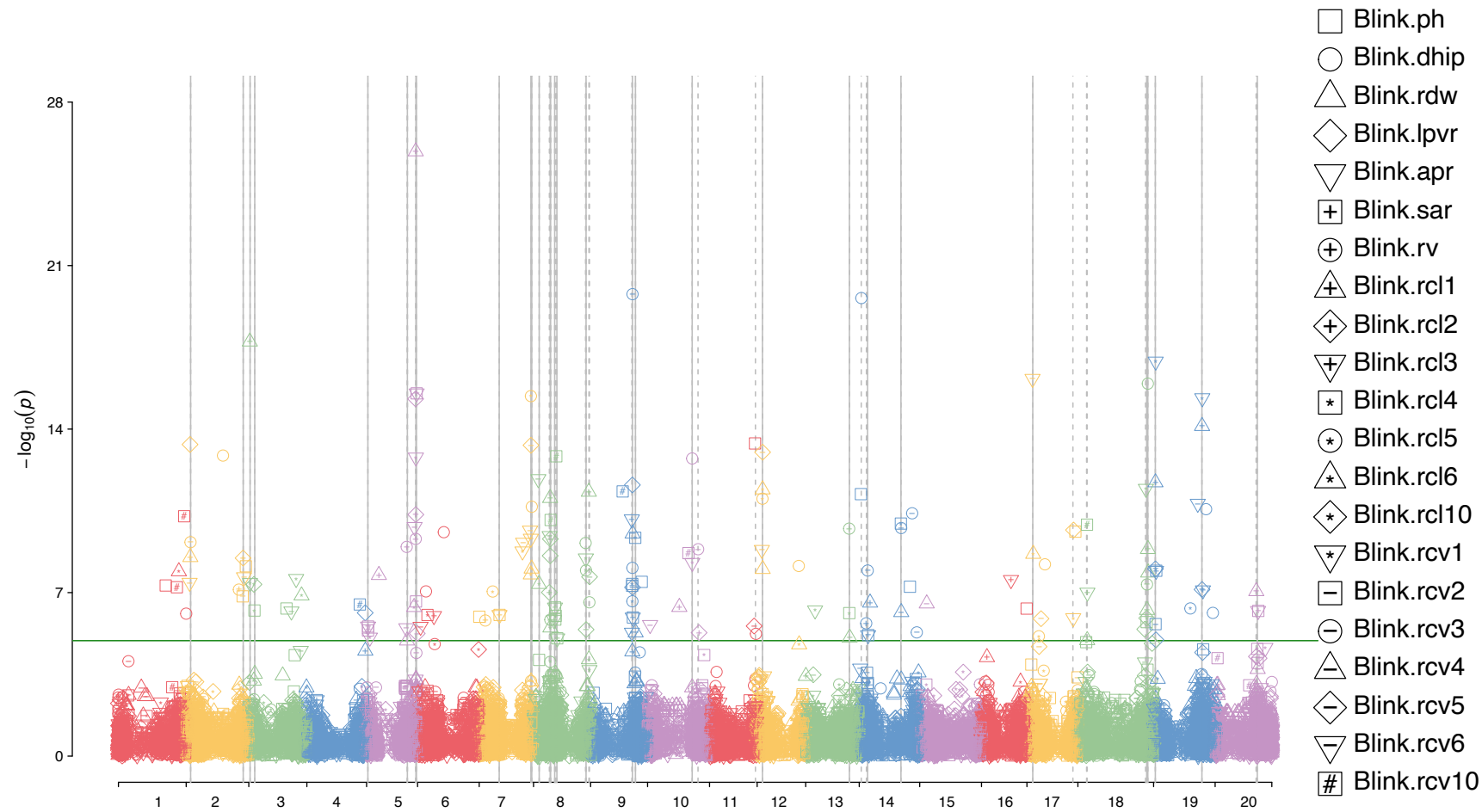


Figura 6 – Manhattan plot utilizando a abordagem 1. SNPs acima da linha horizontal verde são significativos pela correção FDR.

ph: altura de planta, dhip: diâmetro do hipocótilo, rdw: massa seca da raiz, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcl10: comprimento da raiz por classes de diâmetros 10, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6, rcv10: volume da raiz por classes de diâmetros 10.

Após realizou-se o estudo de associação genômica para a abordagem 2, o modelo utilizado se ajustou para todas as características utilizadas (Figura S4). Nessa abordagem foram identificados 68 SNPs significativos (Figura 7). Dos quais, 29 deles apresentaram significativos em mais de uma característica.

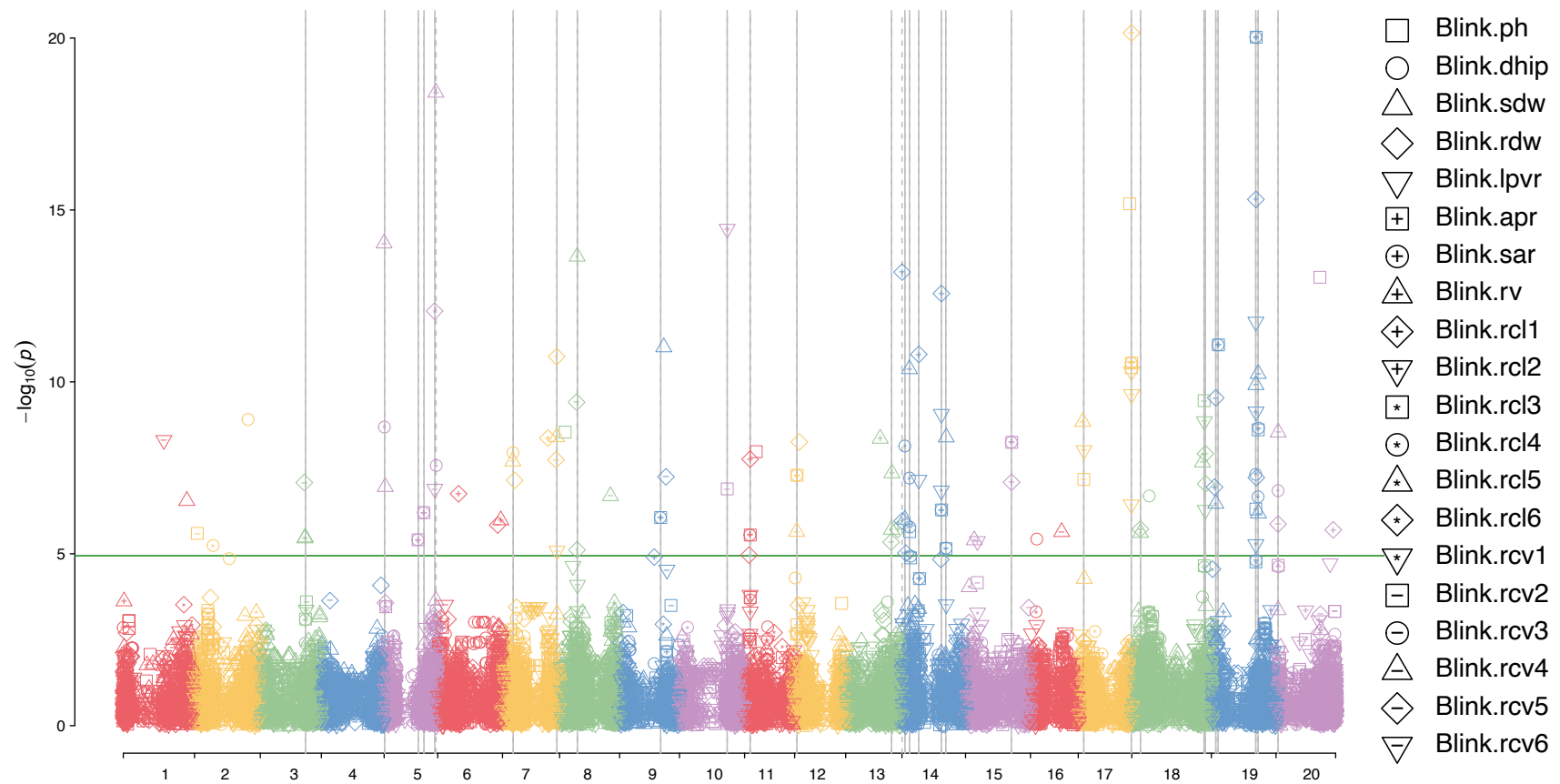


Figura 7 – Manhattan plot utilizando a abordagem 2.

SNPs acima da linha horizontal verde são significativos pela correção FDR.

ph: altura de planta, dhip: diâmetro do hipocótilo, rdw: massa seca da raiz, sdw: massa seca da parte aérea, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5.

Após foi verificado dentre os SNPs significativos, quais estão presentes nas duas abordagens. Foram encontrados 18 SNPs significativos à ambas as abordagens (Figura 8). Assim 96 foram significativos apenas na abordagem 1 e 50 foram significativos na abordagem 2 (Tabela S4). Esses 18 SNPs então foram selecionados para um estudo de Pós-GWAS.

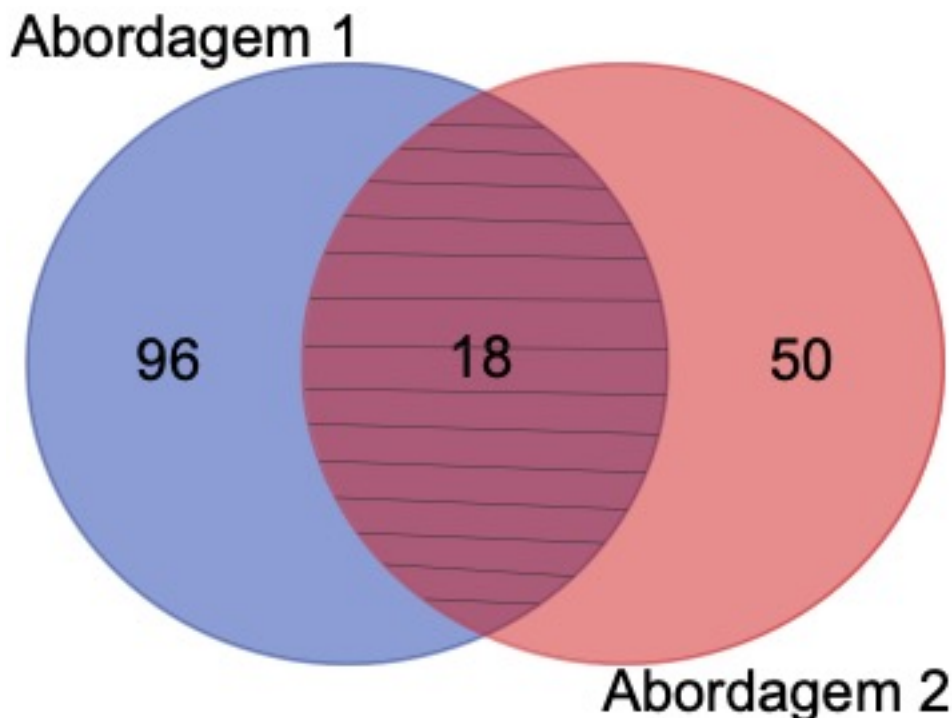


Figura 8 – Diagrama de Venn para os SNPs significativos nas duas abordagens utilizadas.

Ao utilizar a versão mais recente do genoma da soja, versão 4 (Wm82.a4), apenas para o SNP Gm12_4363882_T_C não foi encontrada uma região que possuísse similaridade. Assim esse SNP foi descartado para a busca gênica (Tabela 1). Dentre os outros 17 SNPs restantes, apenas seis não apresentaram genes na região. Entre os 11 SNPs restantes, quatro deles apresentaram dois genes na região de busca, o que totalizou 15 genes encontrados. Em sete desses genes, o marcador SNP apresenta-se localizado na região intergênica. Enquanto cinco genes apresentam o marcador dentro da região codificadora para a proteína (CDS). E outros três genes apresentam o marcador em regiões do transcrito primário desses genes, mas não em região codante, em que dois estão localizados em região intrônica e um na região 3'UTR.

Ao se testar o efeito da substituição genica, causada pelo marcador SNP, dentre os cinco genes que apresentam marcadores na região CDS, apenas um, GLYMA_07G245400, apresenta uma mudança no aminoácido codificado pelo códon.

Tabela 1 - Genes localizados em regiões próximas aos marcadores SNPs associados significativamente com a tolerância ao deficit hídrico. Localização dos marcadores em relação ao gene e seus efeitos na substituição alélica.

SNP	Position (Wm82.a4) ^a	Gene annotation	SNP location ^b	Gene description
Gm05_40528969_T_G	chr5:39858251	GLYMA_05G218100	intergenic	Transcriptional coactivator PC4/P15 protein KELP-like protein
Gm05_40528969_T_G	chr5:39858251	GLYMA_05G218200	intergenic	DnaJ homolog subfamily B member 1-like
Gm05_40919406_C_T	chr5:39467821	GLYMA_05G212700	CDS GAC -> GAU (Asp -> Asp)	Probable membrane-associated kinase regulator 2
Gm07_35547398_T_C	chr7:35769123	GLYMA_07G186600	intergenic	Sulfite exporter TauE/SafE family protein
Gm07_35547398_T_C	chr7:35769123	GLYMA_07G186700	intronic	Kinesin-like protein KIN-14C
Gm07_42514160_T_G	chr7:42780085	GLYMA_07G245400	CDS AAC -> ACC (Asn -> Thr)	RING-H2 finger protein ATL2
Gm08_13782412_C_T	chr8:13798142	Not found		
Gm08_13914377_A_G	chr8:13935746	Not found		
Gm08_4667338_C_T	chr8:4671165	GLYMA_08G060600	CDS GUC -> GUU (Val -> Val)	Pentatricopeptide repeat-containing protein At2g17140
Gm09_34354926_C_T	chr9:37821865	GLYMA_09G149300	intergenic	Carboxylesterase 1-like
Gm09_36844563_T_C	chr9:40332402	GLYMA_09G169500	3'UTR	Serine/Threonine Kinase-Related
Gm12_4363882_T_C	Not similarity found			
Gm13_35909612_G_A	chr13:36449668	GLYMA_13G267851	intronic	Gamma-glutamyl hydrolase precursor
Gm13_35909612_G_A	chr13:36449668	GLYMA_13G268000	intergenic	Gamma-glutamyl hydrolase 2
Gm14_34265654_T_C	chr14:34395136	Not found		
Gm14_5908034_T_C	chr14:5915210	GLYMA_14G071600	CDS UUU -> UUC (Phe -> Phe)	Alcohol acyltransferase 9
Gm16_24389823_A_C	chr16:24938583	Not found		
Gm17_39891056_A_G	chr17:39717783	GLYMA_17G240700	intergenic	NAC domain-containing protein 71
Gm18_56486524_T_C	chr18:52491885	Not found		
Gm18_57446256_C_T	chr18:53450519	GLYMA_18G243800	CDS GUC -> GUU (Val -> Val)	Protein TIC 20-II, chloroplastic
Gm18_57446256_C_T	chr18:53450519	GLYMA_18G243900	intergenic	Putative RNA-binding protein Luc7-like 2-like
Gm19_36313784_T_C	chr19:36969747	Not found		

a: posição do SNP no Genoma da soja versão 4 (*Glycine max* Wm82.a4.v1), **b:** posição onde se encontra o SNP em relação ao gene e a modificação levada pela substituição alélica para os marcadores dentro da região codante do gene.

4. Discussão

O período de estresse severo de 20 dias causou a redução do incremento da matéria seca nas plantas da condição estresse em relação a condição controle, em ambos os experimentos, como já bem conhecido na literatura (Jumrani & Bahatia, 2018). Essa redução demonstra a eficiência da imposição das diferentes disponibilidades hídricas, como relatado por Nascimento et al., (2021).

No segundo ano, a média da matéria seca em ambas as partes da planta foram menores que o primeiro ano, isto pode ser explicado pelas temperaturas do ambiente que no segundo ano foram mais elevadas (Figura S1). A combinação de temperaturas mais altas e o déficit hídrico afeta mais severamente as plantas levando a reduções mais drásticas de suas características como maior redução acúmulo de massa seca (Dreesen et al., 2012).

A separação dos grupos estresse e controle quando plotados em um PCA biplot, nos dois experimentos, demonstra clara diferença entre plantas estressadas e plantas em condições ideais. No grupo formado por plantas da condição controle, os dados se apresentam mais agregados demonstrando que a diferença entre as características mensuradas na condição controle diferem entre os cultivares, mas não de maneira tão ampla, essas pequenas diferenças vem das características intrínsecas de cada cultivar. Em contrapartida, no grupo da condição estresse, os dados apresentam mais esparsados, isto demonstra que os genótipos se comportam de maneira diferente nessa condição. Existindo plantas mais e menos tolerante nessa condição.

A correlação na maior parte entre os valores genéticos foi forte e positiva, principalmente com as características de acúmulo de matéria seca que em outros trabalhos correlacionam com produtividade final esperada (Koutroubas et al., 1998). Isso reforça que as características selecionadas no presente estudo para associação genômica ampla foram as melhores dentro de cada abordagem.

A natureza quantitativa da tolerância ao déficit hídrico pode explicar o grande número de SNPs significativos encontrados nesse estudo, 114 na primeira abordagem e 68 na segunda abordagem. Essa diferença no número de marcadores significativos se deve ao fato de que na segunda abordagem existe a eficiência agrônômica em que o objetivo é selecionar um genótipo que seja bom nos dois ambientes. Nessa abordagem o genótipo considerado tolerante é aquele que possui um bom desempenho nas duas condições. Assim a combinação dessas estratégias irá fornecer genes que contribuem para tolerância para o deficit hídrico e que possuem pouco ou nenhum efeito negativo no desempenho dos genótipos em condição controle.

O gene GLYMA_05G218100, “Transcriptional coactivator PC4/P15 protein KELP-like protein”, está envolvido diretamente com atividades de regulação de transcritos. Os genes KELP já foram reportados com envolvimento de interação planta-patógeno atuando contra vírus e bactérias em, principalmente, movimento de proteínas (Mukherjee et al., 2009; Sasaki et al., 2009, Matsushita et al., 2001). Em condição de estresse, algumas proteínas sinalizadoras necessitam se movimentar nos sistemas de endomembranas como a *carbonic anhydrase 1* que se movimenta do cloroplasto para o citosol (Li et al., 2020).

O “DnaJ homolog subfamily B member 1-like”, GLYMA_05G218200 também conhecidos como proteínas do tipo Hsp40s são essenciais para a atuação das proteínas Hsp70 (Kampinga & Craig, 2010; Kampinga et al., 2019). As proteínas Hsp70 são chaperonas responsáveis pelo correto dobramento das proteínas, previnem o dobramento incorreto e o enovelamento, auxiliam no transporte de proteínas através das membranas, mediando interações proteína-proteína e removem agregados de proteínas citotóxicas, sequestrando os agregados ou direcionando-os para vias de degradação (Verma et al., 2019). Então a função das proteínas DnaJ é entregar alvos específicos ou recrutar alvos para uma localização celular específica para ação da chaperona Hsp70 (Verma et al., 2019). As proteínas DnaJ em soja são induzidas por uma variedade de estresses como injúria mecânica, calor, desidratação e alta salinidade (So et al., 2013). Vale ressaltar que algumas proteínas chaperonas já foram identificadas como responsáveis pela tolerância ao déficit hídrico em soja (Valente et al., 2009).

Em Arabidopsis o homólogo do gene GLYMA_05G212700, “Probable membrane-associated kinase regulator 2”, é o AT1G64080 (MPK2). Esse gene possui um padrão de expressão nas células da epiderme e no córtex do meristema da raiz (Novikova et al., 2022; Marquès-Bueno et al., 2021). Ele é responsável por regular respostas para o gravitropismo, as plantas com perda de função para esse gene respondem muito mais rápido ao estímulo de gravitropismo (Novikova et al., 2022; Armengot et al., 2016). Uma das primeiras respostas ao déficit hídrico em plantas é o crescimento da parte radicular, a fim de explorar mais solo para encontrar água (Souza et al., 2021). A substituição alélica nesse caso, pode levar a uma menor tradução da proteína, já que existem códons que são mais preferidos que outros e impactam assim na estabilidade do transcrito (Oliveira et al., 2021; Duret, 2002; Chen et al. 2004).

O gene GLYMA_07G186600 é um “Sulfite exporter TauE/SafE family protein” que possui papel no metabolismo de taurina e na exportação de ânions através da membrana. Em estudos de associação genômica, o homólogo desse gene já foi reportado para tolerância ao déficit hídrico em trigo (Shokat et al., 2020). Já o seu homólogo em arabidopsis AT1G61740

foi identificado dentre os 20 genes mais expressos para estresses com metais pesados como As, Au, Cd, Pb e Cu (Niu et al., 2019).

O “Kinesin-like protein KIN-14C”, GLYMA_07G186700, pertence à família das Cinesinas 14. Essa família é responsável pela montagem dos fusos meióticos e mitóticos (Quan et al., 2008; Chen et al., 2002; Marcus et al., 2002, 2003). Em arroz o homólogo desse gene, LOC_Os04g53760 tem papel na tolerância ao estresse salino (Saha et al., 2020). Nesse caso o marcador se encontra na região do intrón, essa região apesar de ser descartada no processamento do mRNA, possui importantes pontos de regulação. A substituição genica causada pelo SNP, pode modificar esses pontos o que irá impactar diretamente na proteína codificada (Szczésniak et al., 2013).

O gene GLYMA_07G245400, “RING-H2 finger protein ATL2”, é um tipo especial de proteína de dedos de zinco. Estudos demonstram que essas proteínas possuem uma ampla gama de funções em vários processos biológicos incluindo a transdução de sinais, replicação viral e desenvolvimento (Tian et al., 2015). As proteínas do tipo RING fingers têm sido amplamente encontradas entre as plantas (Yuan et al., 2013). Dentre as diversas funções que essas proteínas atuam no reino vegetal, elas têm sido encontradas na tolerância ao déficit hídrico de diversas espécies (Song et al., 2016; Park et al., 2015; Ning et al., 2011; Ko et al., 2006). A substituição alélica nesse caso leva à substituição de um aminoácido na proteína. Nessa substituição a Arginina, um aminoácido polar e carregado positivamente é substituído por uma treonina, um aminoácido levemente polar e não carregado. As asparaginas preferem a superfície das proteínas e frequentemente participam de sítios de reconhecimento ou de atividade das proteínas (Betts & Russel, 2003).

A proteína “pentatricopeptide repeat-containing protein At2g17140-like”, codificada pelo gene GLYMA_08G060600, é altamente expressa nos nódulos das raízes, nas folhas e na raiz principal em soja (Mandozai et al., 2021; Libault et al., 2010; Severin et al., 2010). A família dessa proteína tem papel fundamental no crescimento e desenvolvimento e nos estresses bióticos e abióticos (Saidi & Hajibarat, 2020). A substituição alélica nesse caso leva apenas a mudança de códons, contudo a preferência de códons pode levar também a uma maior taxa de transcrição, o que contribuiria para uma maior expressão desse gene.

O homólogo do gene “Carboxylesterase 1-like” (GLYMA_09G149300) em *Arabidopsis AtCXE20* (At5g62180), possui ação de inibição as estrigolactonas (Roesler et al., 2021). As estrigolactonas são hormônios vegetais que influenciam a arquitetura da parte aérea inibindo ramificações, e influência no crescimento radicular (Kapulnik et al., 2011; Gomez-Roldan et al., 2008; Umehara et al., 2008). Esses hormônios têm papel na tolerância ao estresse

abiótico, sendo que durante o déficit hídrico, há aumento da concentração desses nas raízes (Bu et al., 2014; Ha et al., 2014; Haider et al., 2018).

O gene GLYMA_09G169500 é uma proteína não caracterizada que pertence à família das Serine/Threonine Kinases. Essa família desempenha papéis cruciais de transdução de sinais em diversos processos biológicos em plantas (Hardie, 1999). Dentre os diversos processos biológicos se encontram a tolerância ao déficit hídrico (Lim et al., 2020; Kulik et al., 2011). Nesse caso, o marcador SNP, que modifica na codificação do RNA de uma Uracila para uma Citosina, se encontra na região 3'UTR, essa região é importante para a regulação da proteína. Ela possui função de regular a estabilidade do mRNA via regiões ricas em AU, regula a localização do mRNA, regula a tradução, apresenta sítios de ligação de ribossomos, são regiões ativadas por tecido específico e a sua composição influencia na interação proteína-proteína nessa região (Mayr, 2019).

Os genes “Gamma-glutamyl hydrolase precursor” (GLYMA_13G267851) e “Gamma-glutamyl hydrolase 2” (GLYMA_13G268000) estão envolvidos na síntese *de novo* de folatos (Akhtar et al., 2010). Os folatos são cofatores indispensáveis no metabolismo de um carbono. Eles fornecem carbono para diversos tipos de oxidação (metil, metileno ou formil) em moléculas alvo na síntese de purinas, pantotenato, metionina, timidilato e formil-metionil tRNA (Scott et al., 2000). O gene GLYMA_13G268000 interage com GmMMK1, uma GmMAPK responsável pela sinalização durante o estresse salino (Liao et al., 2021). E em arroz, o homólogo do GLYMA_13G267851, LOC_Os05g44130, possui papel na tolerância ao déficit hídrico (Shaar-Moshe et al., 2015).

O “Alcohol acyltransferase 9” (GLYMA_14G071600) é uma Álcool aciltransferase, que são responsáveis pela produção de ésteres. Assim essa superfamília contribui na biossíntese de vários terpenóides, ésteres, alcalóides e flavonóides (Zhou et al., 2021; Kruse et al., 2020). Assim esse gene está envolvido com a arquitetura das plantas e com resistência à insetos (Porth et al., 2012).

O gene GLYMA_17G240700 é uma NAC domain-containing protein 71 ou GmNAC135 (Melo et al., 2018). Esse gene é classificado como relacionado à senescência (GmNAC-SAG) (Fraga et al., 2021). O seu homólogo em arabidopsis AT1G32510 (ANAC11) é conhecido como responsivo ao dano mecânico (Matsuoka et al., 2021). Apesar disso, a super expressão de ANAC11 mostrou ser um gene que confere a tolerância ao déficit hídrico e às baixas temperaturas (Pang et al., 2019).

A “Protein TIC 20-II, chloroplastic” (GLYMA_18G243800) é um componente do translocon do envelope interno do cloroplasto funcionando como um canal de translocação (Machettira et al., 2011).

A “Putative RNA-binding protein Luc7-like 2-like” (GLYMA_18G243900) é uma proteína acessória para o processo de splicing. Essa proteína, LUC7, em *arabidopsis* desempenha papel no desenvolvimento das plantas, gerando plantas com diversos defeitos quando realizado o seu “knockout”, como por exemplo o nanismo, e possui papel na sinalização de ácido abscísico e na tolerância ao frio e à salinização (Amorim et al., 2018). O ácido abscísico é o principal hormônio envolvido com a tolerância ao déficit hídrico em plantas (Zhou et al., 2019; Boominathan et al., 2004).

A maior parte dos marcadores moleculares se localizaram fora da região codificada do gene, na região entre genes. No entanto não pode ser totalmente desconsiderado o efeito da substituição gênica nesses casos. Essas podem ser regiões cis regulatórias de genes que influenciam na transcrição dos mesmos (Ong & Corces, 2011).

5. Conclusão

O estresse hídrico reduziu significativamente o acúmulo de matéria seca da parte aérea e da parte radicular das plantas. A seleção das características com base nas suas contribuições para diferenciar os grupos estresse e controle contribuiu para encontrar marcadores SNPs associados à tolerância ao déficit hídrico.

A utilização de duas abordagens empregadas para a tolerância ao déficit hídrico contribuíram para acessar de forma mais precisa genes que contribuem para ela. Dentre os 18 SNPs que apresentaram significativamente associados à tolerância nas duas abordagens, 11 deles apresentaram genes próximos.

Relata-se os genes encontrados: o gene GLYMA_05G218200 é uma chaperona molecular do tipo Hsp40s que auxilia a chaperona Hsp70, que ao serem superexpressadas conferem tolerância ao déficit hídrico; o gene GLYMA_07G245400 é um tipo especial de dedo de zinco responsável pela transdução de sinais. os genes GLYMA_09G169500, GLYMA_13G267851, GLYMA_17G240700, GLYMA_18G243900 são correlacionados com tolerância ao déficit hídrico; os genes GLYMA_05G212700, GLYMA_08G060600 e GLYMA_09G149300 estão envolvidos também com a regulação do crescimento das raízes; e, os genes GLYMA_05G218100, GLYMA_07G186600, GLYMA_07G186700, GLYMA_13G268000, GLYMA_14G071600 e GLYMA_18G243800 estão envolvidos em pelo menos um tipo de tolerância aos estresses bióticos e abióticos conhecidos.

6. Referências bibliográficas

- Abdel-Haleem, H.; Lee, G.-J.; Boerma, R.H. (2011). Identification of QTL for increased fibrous roots in soybean. *Theoretical and Applied Genetics*, 122: 935–946. doi: 10.1007/s00122-010-1500-9
- Akhtar, T.A.; Orsomando, G.; Mehrshahi, P.; Lara-Núñez, A.; Bennett, M.J.; Gregory, J.F.; Hanson, A.D. (2010). A central role for gamma-glutamyl hydrolases in plant folate homeostasis. *The Plant Journal*, 64: 256–66. doi: 10.1111/j.1365-3113X.2010.04330.x
- Amorim, M.F.; Willing, E.M.; Szabo, E.X.; Francisco-Mangilet, A.G.; Droste-Borel, I.; Maček, B.; Schneeberger, K.; Laubinger, S. (2018). The U1 snRNP Subunit LUC7 Modulates Plant Development and Stress Responses via Regulation of Alternative Splicing. *Plant Cell*, 30: 2838–2854. doi: 10.1105/tpc.18.00244
- Armengot, L.; Marquès-Bueno, M.M.; Jaillais, Y. (2016) Regulation of polar auxin transport by protein and lipid kinases. *Journal of Experimental Botany*, 67: 4015–4037. doi: 10.1093/jxb/erw216
- Benjamini, Y.; Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57: 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bennett, G.L.; Pollak, E.J.; Kuehn, L.A.; Snelling, W.M. (2014). Breeding: Animals. In: Alfen, N.K.V. (ed), *Encyclopedia of Agriculture and Food Systems*, Davis: Elsevier, 27141 p.
- Bernardo, S.; Mantovani, C.E.; Silva, D.D.; Soares, A.A. (2019). Manual de irrigação. Viçosa: Editora UFV, 9a Edição, p. 545.
- Betts, M.J.; Russell, R.B. (2007). Amino-acid properties and consequences of substitutions. In Barnes, M.R.; Gray I.C. eds., *Bioinformatics for Geneticists*, 578p. doi: 10.1002/0470867302.ch14.
- Boominathan, P.; Shukla, R.; Kumar, A.; Manna, D.; Negi, D.; Verma, P. K.; Chattopadhyay, D. (2004). Long Term Transcript Accumulation during the Development of Dehydration Adaptation in *Cicer arietinum*. *Plant Physiology*, 135: 1608–1620. doi: 10.1104/pp.104.043141
- Bu, Q.; Lv, T.; Shen, H.; Luong, P.; Wang, J.; Wang, Z., Huang, Z.; Xiao, L.; Engineer, C.; Kim, T.H.; Schroeder, J.I.; Huq, E. (2014). Regulation of drought tolerance by the f-box protein MAX2 in Arabidopsis. *Plant Physiology*, 164: 424–439. doi: 10.1104/pp.113.226837
- Chen, C.; Marcus, A.; Li, W.; Hu, Y.; Calzada, J.P.; Grossniklaus, U.; Cyr, R.J.; Ma, H. (2002) The Arabidopsis ATK1 gene is required for spindle morphogenesis in male meiosis. *Development*, 129: 2401–2409. doi: 10.1242/dev.129.10.2401

- Chen, S.L.; Lee, W.; Hottes, A.K.; Shapiro, L.; McAdams, H.H. (2004). Codon usage between genomes is constrained by genome-wide mutational processes. *Proceedings of the National Academy of Sciences*, 101: 3480–3485. doi: 10.1073/pnas.0307827100
- Contreras-Soto, R.I.; Oliveira, M.B.; Costenaro-da-Silva, D.; Scapim, C.A.; Schuster, I. (2017). Population structure, genetic relatedness and linkage disequilibrium blocks in cultivars of tropical soybean (*Glycine max*). *Euphytica*, 213: 173. doi: 10.1007/s10681-017-1966-5
- Covarrubias-Pazaran, G. (2016). Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. *Plos One*, 11: e0156744. doi: 10.1371/journal.pone.0156744
- Dreesen, F.E.; De Boeck, H.J.; Janssens, I.A.; Nijs, I. (2012). Summer heat and drought extremes trigger unexpected changes in productivity of a temperate annual/biannual plant community. *Environmental and Experimental Botany*, 79: 21–30. doi: 10.1016/j.envexpbot.2012.01.005
- Du, W.; Yu, D.; Fu, S. (2009). Detection of quantitative trait loci for yield and drought tolerance traits in soybean using a recombinant inbred line population. *Journal of Integrative Plant Biology*, 51: 868–878. doi: 10.1111/j.1744-7909.2009.00855.x
- Duret L. 2002. Evolution of synonymous codon usage in metazoans. *Current Opinion in Genetics & Development*, 12: 640–649. doi: 10.1016/S0959-437X(02)00353-2
- Easwar, D.R.; Chaitanya, K.V. (2016). Photosynthesis and antioxidative defense mechanisms in deciphering drought stress tolerance of crop plants. *Biologia Plantarum*, 60: 201–218. doi: 10.1007/s10535-016-0584-8
- Eftekhari, A.; Baghizadeh, A.; Yaghoobi, M.; Abdolshahi, R. (2017). Differences in the drought stress response of DREB2 and CAT1 genes and evaluation of related physiological parameters in some bread wheat cultivars. *Biotechnology & Biotechnological Equipment*, 31: 709–716. doi: 10.1080/13102818.2017.1316214
- Fehr, W. R.; Caviness, C. E. (1977). Stages on soybean development. Ames: Iowa State University/Cooperative Extension Service, (*Special Report*, 80), p. 11.
- Fraga, O.T.; Melo, B.P.; Quadros, I.P.S.; Reis, P.A.B.; Fontes, E.P.B. (2021). Senescence-Associated Glycine max (Gm)NAC Genes: Integration of Natural and Stress-Induced Leaf Senescence. *International Journal of Molecular Sciences*, 22: 8287. doi: 10.3390/ijms22158287
- Gallardo, M.; Thompson, R.B.; Valdez, L.C.; Fernández, M.D. (2006). Use of stem diameter variations to detect plant water stress in tomato. *Irrigation Science*, 24: 241–255. doi: 10.1007/s00271-005-0025-5
- Goldhamer, D.A.; Fereres, E. (2001). Irrigation scheduling protocols using continuously recorded trunk diameter measurements. *Irrigation Science*, 20: 115–125. doi: 10.1007/s002710000034

- Gomez-Roldan, V.; Fermas, S.; Brewer, P.B.; Puech-Pages, V.; Dun, E.A.; Pillot, J.P.; Letisse, F.; Matusova, R.; Danoun, S.; Portais, J.-C.; Bouwmeester, H.; Bécard, G.; Beveridge, C.A.; Rameau, C.; Rochange, S.F. (2008). Strigolactone inhibition of shoot branching. *Nature* 455, 189–194. doi: 10.1038/nature07271
- Graham, P.H.; Vance, C.P. (2003). Legumes: Importance and constraints to greater use. *Plant Physiology*, 131: 872–877. doi: 10.1104/pp.017004
- Ha, C.V.; Leyva-Gonzalez, M.A.; Osakabe, Y.; Tran, U.T.; Nishiyama, R.; Watanabe, Y.; Tanakae, M.; Sekie, M.; Yamaguchi, S.; Van Dongb, N.; Yamaguchi-Shinozakig, K.; Shinozakid, K.; Herrera-Estrellac, L.; Trana, L.-S.P. (2014). Positive regulatory role of strigolactone in plant responses to drought and salt stress. *Proceedings of the National Academy of Sciences*, 111: 851–856. doi: 10.1073/pnas.1322135111
- Haider, I.; Andreo-Jimenez, B.; Bruno, M.; Bimbo, A.; Flokova, K.; Abuauf, H.; Ntui, V.O.; Guo, X.; Charnikhova, T.; Al-Babili, A.; Bouwmeester, H.J.; Carolien Ruyter-Spira, C. (2018). The interaction of strigolactones with abscisic acid during drought response in rice. *Journal of Experimental Botany*, 69: 2403–2414. doi: 10.1093/jxb/ery089
- Hardie, D. G. (1999). Plant protein serine/threonine kinases: classification and functions. *Annual Review of Plant Physiology and Plant Molecular Biology*, 50: 97-131. doi: 10.1146/annurev.arplant.50.1.97
- Huang, M., Liu, X., Zhou, Y., Summers, R. M.; Zhang, Z. (2018). BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*, 8: 1-12. doi: 10.1093/gigascience/giy154
- Hwang, S.; King, C.A.; Chen, P.; Ray, J.D.; Cregan, P.B.; Carter, T.E.; Li, Z.; Abdel-Haleem, H.; Matson, K.W.; Schapaugh, W. (2016). Meta-analysis to refine map position and reduce confidence intervals for delayed-canopy-wilting QTLs in soybean. *Molecular Breeding*, 36: 91. doi: 10.1007/s11032-016-0516-5
- Joshi, R.; Wani, S.H.; Singh, B.; Bohra, A.; Dar, Z.A.; Lone, A.A.; Pareek, A.; Singla-Pareek, S.L. Transcription factors and plants response to drought stress: Current understanding and future directions. *Frontiers in Plant Science*, 7: 1029. doi: 10.3389/fpls.2016.01029
- Jumrani, K.; Bhatia, V.S. (2018). Impact of combined stress of high temperature and water deficit on growth and seed yield of soybean. *Physiology and Molecular Biology of Plants*, 24: 37–50. doi: 10.1007/s12298-017-0480-5
- Kampinga, H.H.; Andreasson, C.; Barducci, A.; Cheetham, M.E.; Cyr, D.; Emanuelsson, C.; Genevau, P.; Gestwicki, J.E.; Goloubinoff, P.; Huerta-Cepas, J.; Kirstein, J.; Liberek, K.; Mayer, M.P.; Nagata, K.; Nillegoda, N.B.; Pulido, P.; Ramos, C.; De Los Rios, P.; Rospert, S.; Rosenzweig, R.; Sahi, C.; Taipale, M.; Tomiczek, B.; Ushioda, R.; Young, J.C.; Zimmermann, R.; Zylicz, A.; Zylicz, M.; Craig, E.A.; Marszałek, J. (2019). Function, evolution, and structure of J-domain proteins. *Cell Stress Chaperones*, 24: 7–15. doi:10.1007/s12192-018-0948-4

- Kampinga, H.H.; Craig, E.A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nature Reviews Molecular Cell Biology*, 11: 579–592. doi:10.1038/nrm2941
- Kapulnik, Y.; Delaux, P.-M.; Resnick, N.; Mayzlish-Gati, E.; Wininger, S.; Bhattacharya, C.; et al. (2011). Strigolactones affect lateral root formation and root-hair elongation in *Arabidopsis*. *Planta*, 233: 209–216. doi: 10.1007/s00425-010-1310-y
- Ko, J.H.; Yang, S.H.; Han, K.H. (2006). Upregulation of an *Arabidopsis* RING- H2 gene, XERICO, confers drought tolerance through increased abscisic acid biosynthesis. *The Plant Journal*, 47: 343–355. doi: 10.1111/j.1365-313X.2006.02782.x
- Koutroubas, S.D.; Papakosta, D.K.; Gagianas, A.A. (1998). The importance of early dry matter and nitrogen accumulation in soybean yield. *European Journal of Agronomy*, 9: 1–10. doi: 10.1016/s1161-0301(97)00067-1
- Kruse, L.H.; Weigle, A.T.; Martínez-Gómez, J.; Chobirko, J.D.; Schaffer, J.E.; Bennett, A.A.; Specht, C.D.; Jez, J.M.; Shukla, D.; Moghe, G.D. (2020). Ancestral class-promiscuity as a driver of functional diversity in the BAHD acyltransferase family in plants. *bioRxiv*. doi: 10.1101/2020.11.18.385815
- Kulik, A.; Wawer, I.; Krzywińska, E.; Bucholc, M.; Dobrowolska, G. (2011). SnRK2 protein kinases-key regulators of plant response to abiotic stresses. *OMICS A Journal of Integrative Biology*, 15: 859-72. doi: 10.1089/omi.2011.0091
- Li, P.; Liu, H.; Yang, H.; Pu, X.; Li, C.; Huo, H.; Chu, Z.; Chang, Y.; Lin, Y.; Liu, L. Translocation of Drought-Responsive Proteins from the Chloroplasts. *Cells*, 9: 259. Doi: 10.3390/cells9010259
- Liao X, Shi M, Zhang W, Ye Q, Li Y, Feng X, Bhat JA, Kan G, Yu D. (2021). Association analysis of GmMAPKs and functional characterization of GmMMK1 to salt stress response in soybean. *Physiologia Plantarum*, 173: 2026-2040. doi: 10.1111/ppl.13549
- Libault, M.; Farmer, A.; Brechenmacher, L.; Drnevich, J.; Langley, R.J.; Bilgin, D.D.; Radwan, O.; Neece, D.J.; Clough, S.J.; May, G.D.; Stacey, G. (2010). Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiology*, 152: 541–552. doi: 10.1104/pp.109.148379
- Lim, J.; Lim, C.W.; Lee, S.C. (2020). Pepper Novel Serine-Threonine Kinase CaDIK1 Regulates Drought Tolerance via Modulating ABA Sensitivity. *Frontiers in Plant Science*, 11:1133. doi: 10.3389/fpls.2020.01133
- Luo, Y.Z.; Li, G.; Yan, G.; Liu, H.; Turner, N.C. (2020). Morphological features and biomass partitioning of lucerne plants (*Medicago sativa* L.) subjected to water stress. *Agronomy*, 10: 322. doi: 10.3390/agronomy10030322
- Machettira, A.B.; Gross, L.E.; Sommer, M.S.; Weis, B.L.; Englich, G.; Tripp, J.; Schleiff, E. (2011). The localization of Tic20 proteins in *Arabidopsis thaliana* is not restricted to

- the inner envelope membrane of chloroplasts. *Plant Molecular Biology*, 77: 381–90. doi: 10.1007/s11103-011-9818-5
- Mandozai, A.; Moussa, A.A.; Zhang, Q.; Qu, J.; Du, Y.; Anwari, G.; Al Amin, N.; Wang, P. (2021). Genome-Wide Association Study of Root and Shoot Related Traits in Spring Soybean (*Glycine max* L.) at Seedling Stages Using SLAF-Seq. *Frontiers in Plant Science*, 2021 Jul 28;12:568995. doi: 10.3389/fpls.2021.568995.
- Marcus, A.I.; Ambrose, J.C.; Blickley, L.; Hancock, W.O.; Cyr, R.J. (2002). *Arabidopsis thaliana* protein, ATK1, is a minus-end directed kinesin that exhibits non-processive movement. *Cell Motility and the Cytoskeleton*, 52: 144–150. doi: 10.1002/cm.10045
- Marcus, A.I.; Li, W.; Ma, H.; Cyr, R.J. (2003) A kinesin mutant with an atypical bipolar spindle undergoes normal mitosis. *Molecular Biology of the Cell*, 14: 1717–1726. doi: 10.1091/mbc.E02-09-0586
- Marquès-Bueno, M.M.; Armengot, L.; Noack, L.C.; Bareille, J.; Rodriguez, L.; Platre, M.P.; Bayle, V.; Liu, M.; Opdenacker, D.; Vanneste, S.; Möller, B.K.; Nimchuk, Z.L.; Beeckman, T.; Canó-Delgado A.I.; Frim, J.; Jaillais, Y. (2021) Auxin-regulated reversible inhibition of TMK1 signaling by MAK2 modulates the dynamics of root gravitropism. *Current Biology*, 31: 228–237. doi: 10.1016/j.cub.2020.10.011
- Matsuoka, K.; Sato, R.; Matsukura, Y.; Kawajiri, Y.; Iino, H.; Nozawa, N.; Shibata, K.; Kondo, Y.; Satoh, S.; Asahina, M. (2021). Wound-inducible ANAC071 and ANAC096 transcription factors promote cambial cell formation in incised *Arabidopsis* flowering stems. *Communications Biology*, 4: 369. doi: 10.1038/s42003-021-01895-8
- Matsushita, Y., Deguchi, M., Youda, M., Nishiguchi, M. and Nyunoya, H. (2001) The tomato mosaic tobamovirus movement protein interacts with a putative transcriptional coactivator KELP. *Molecules and Cells*, 12: 57–66.
- Mayr C. (2019). What Are 3' UTRs Doing? *Cold Spring Harbor Perspectives in Biology*, 11: a034728. doi: 10.1101/cshperspect.a034728
- Meeks, C.D.; Snider, J.L.; Babb-Hartman, M.E.; Barnes, T. (2019). Evaluating the mechanisms of photosynthetic inhibition under growth-limiting, early-season water deficit stress in cotton. *Crop Science*, 59: 1144–1154. doi: 10.2135/cropsci2018.07.0432
- Melo, B.P.; Fraga, O.T.; Silva, J.C.F.; Ferreira, D.O.; Brustolini, O.J.B.; Carpinetti, P.A.; Machado, J.P.B.; Reis, P.A.B.; Fontes, E.P.B. (2018). Revisiting the Soybean GmNAC Superfamily. *Frontiers in Plant Science*, 9: 1864. doi: 10.3389/fpls.2018.01864
- Mian, M.; Ashley, D.; Boerma, H. (1998). An additional QTL for water use efficiency in soybean. *Crop Science*, 38: 390–393. doi: 10.2135/cropsci1998.0011183X003800020020x
- Miti F, Tongoona P, Derera J (2010) S1 selection of local maize landraces for low soil nitrogen tolerance in Zambia. *African Journal of Plant Science*, 4: 67–81.

- Mukherjee, A.K.; Lev, S.; Gepstein, S.; Horwitz, B.A. (2009). A compatible interaction of *Alternaria brassicicola* with *Arabidopsis thaliana* ecotype DiG: evidence for a specific transcriptional signature. *BMC Plant Biology*, 9: 31. doi: 10.1186/1471-2229-9-31
- Muscolo, A.; Sidari, M.; Anastasi, U.; Santonoceto, C.; Maggio, A. (2014). Effect of PEG-induced drought stress on seed germination of four lentil genotypes. *Journal of Plant Interactions*, 9: 354–363. doi: 10.1080/17429145.2013.835880
- Nascimento, H.R.; Moura, L.D.O.; Duarte, A.B.; Dantas, S.A.G.; Ferreira, D.O.; Rosmaninho, L.B.D.C., Cavallin, I.C., Cunha, F.F., Silva, F.L. (2021). A new methodological approach for simulating water deficit in soybean genotypes. *Journal of Agronomy and Crop Science*, 207: 946–955. doi: 10.1111/jac.12543
- Ning, Y.; Jantasuriyarat, C.; Zhao, Q.; Zhang, H.; Chen, S.; Liu, J.; Liu, L.; Tang, S.; Park, C.H.; Wang, X.; Liu, X.; Dai, L.; Xie, Q.; Wang, G.-L. (2011). The SINA E3 ligase OsDIS1 negatively regulates drought response in rice. *Plant Physiology*, 157: 242–255. doi: 10.1104/pp.111.180893
- Niu, C.; Jiang, M.; Li, N.; Cao, J.; Hou, M.; Ni, D.A.; Chu, Z. (2019). Integrated bioinformatics analysis of As, Au, Cd, Pb and Cu heavy metal responsive marker genes through *Arabidopsis thaliana* GEO datasets. *PeerJ*, 18: e6495. doi: 10.7717/peerj.6495
- Novikova, D.D.; Korosteleva, A.L.; Mironova, V.; Jaillais, Y. (2022), Meet your MAKRR: the membrane-associated kinase regulator protein family in the regulation of plant development. *The FEBS Journal*. doi: 10.1111/febs.16132
- Oliveira, J.L.; Morales, A.C.; Hurst, L.D.; Urrutia, A.O.; Thompson, C.R.L.; Wolf, J.B. (2021). Inferring Adaptive Codon Preference to Understand Sources of Selection Shaping Codon Usage Bias. *Molecular Biology and Evolution*, 38: 3247–3266. doi: 10.1093/molbev/msab099
- Ong, C. T.; Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, 12: 283–293. doi:10.1038/nrg2957
- Pang, X.; Xue, M.; Ren, M.; Nan, D.; Wu, Y.; Guo, H. (2019). *Ammopiptanthus mongolicus* stress-responsive NAC gene enhances the tolerance of transgenic *Arabidopsis thaliana* to drought and cold stresses. *Genetics and Molecular Biology*, 42: 624–634. doi: 10.1590/1678-4685-GMB-2018-0101
- Park, C.; Lim, C.W.; Baek, W.; Lee, S.C. (2015). RING type E3 ligase CaAIR1 in pepper acts in the regulation of ABA signaling and drought stress response. *Plant & Cell Physiology*, 56, 1808–1819. doi: 10.1093/pcp/pcv103
- Porth, I.; White, R.; Jaquish, B.; Alfaro, R.; Ritland, C.; Ritland, K. (2012). Genetical genomics identifies the genetic architecture for growth and weevil resistance in spruce. *Plos One*, 7: e44397. doi: 10.1371/journal.pone.0044397
- Quan, L.; Xiao, R.; Li, W.; Oh, S.-A.; Kong, H.; Ambrose, J.C.; Malcos, J.L.; Cyr, R.; Twell, D.; Ma, H. (2008). Functional divergence of the duplicated AtKIN14a and AtKIN14b

- genes: critical roles in Arabidopsis meiosis and gametophyte development. *The Plant Journal*, 53: 1013-1026. doi: 10.1111/j.1365-313X.2007.03391.x
- R Core Team. (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ribeiro, A.C.; Guimarães, P.T.G.; Alvarez, V.H. (1999). *5ª Aproximação - Recomendações para o uso de corretivos e fertilizantes em Minas Gerais*. Viçosa: Sociedade Brasileira de Ciências do Solo, p.359.
- Roesler, K.; Lu, C.; Thomas, J.; Xu, Q.; Vance, P.; Hou, Z.; Williams, R.W.; Liu, L.; Owens, M.A.; Habben, J.E. (2021). Arabidopsis Carboxylesterase 20 Binds Strigolactone and Increases Branches and Tillers When Ectopically Expressed in Arabidopsis and Maize. *Frontiers in Plant Science*, 12: 639401. doi: 10.3389/fpls.2021.639401
- Saha, J.; Giri, K.; Roy, S. (2020). Identification and characterization of differentially expressed genes in the rice root following exogenous application of spermidine during salt stress. *Genomics*, 112: 4125-4136. doi: 10.1016/j.ygeno.2020.07.011
- Saidi, A.; Hajibarat, Z. (2020). Computational analysis for characterization and evaluation of pentatricopeptide repeat-containing protein (PPR) in *Arabidopsis thaliana*. *Polish Journal of Natural Sciences*, 35: 395 - 418.
- Sasaki, N.; Ogata, T.; Deguchi, M.; Nagai, S.; Tamai, A.; Meshi, T.; Kawakami, S.; Watanabe, Y.; Matsushita, Y.; Nyunoya, H. (2009). Over-expression of putative transcriptional coactivator KELP interferes with Tomato mosaic virus cell-to-cell movement. *Molecular Plant Pathology*, 10: 161-73. doi: 10.1111/j.1364-3703.2008.00517.x
- Scholander, P.F.; Hammel, H.T.; Bradstreet, E.D.; Hemmingsen, E.A. (1965). Sap pressure in vascular plants. *Science*, 148: 339-346. doi: 10.1126/science.148.3668.339
- Scott, J.; Rébeillé, F.; Fletcher, J. (2000). Folic acid and folates: the feasibility for nutritional enhancement in plant foods. *Journal of the Science of Food and Agriculture*, 80: 795–824. doi: 10.1002/(SICI)1097-0010(20000515)80:7<795::AID-JSFA599>3.0.CO;2-K
- Severin, A.J.; Woody, J.L.; Bolon, Y.-T.; Joseph, B.; Diers, B.W.; Farmer, A.D.; Muehlbauer, G.J.; Nelson, R.T.; Grant, D.; Specht, J.E.; Graham, M.A.; Cannon, S.B.; May, G.D.; Vance, C.P.; Shoemaker, R.C. (2010). RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biology*, 10:160. doi: 10.1186/1471-2229-10-160
- Shaar-Moshe, L.; Hübner, S.; Peleg, Z. (2015). Identification of conserved drought-adaptive genes using a cross-species meta-analysis approach. *BMC Plant Biology*, 15: 111. doi: 10.1186/s12870-015-0493-6
- Shin, J.H.; Vaughn, J.N.; Abdel-Haleem, H.; Chavarro, C.; Abernathy, B.; Kim, K.D.; Jackson, S.A.; Li, Z. (2015). Transcriptomic changes due to water deficit define a general soybean response and accession-specific pathways for drought avoidance. *BMC Plant Biology*, 15: 26.

- Shokat, S.; Sehgal, D.; Liu, F.; Singh, S. (2020). GWAS Analysis of Wheat Pre-breeding Germplasm for Terminal Drought Stress Using Next Generation Sequencing Technology. *Preprints*, 2020020272. doi: 10.20944/preprints202002.0272.v1
- Silva, F.L.; Borém, A.; Sedyama, T.; Camara, G. (2022). *Soja do plantio à colheita (2ªed.)*. Viçosa: Oficina de Textos, p. 312.
- So, H.-A.; Chung, E.; Lee, J.-H. (2013). Molecular characterization of soybean GmDjpl encoding a type III J-protein induced by abiotic stress. *Genes & Genomics*, 35: 247–256. doi: 10.1007/s13258-013-0078-4
- Song, J.; Xing, Y.; Munir, S.; Yu, C.; Song, L.; Li, H.; Wang, T.; Ye, Z. (2016). An ATL78-Like RING-H2 Finger Protein Confers Abiotic Stress Tolerance through Interacting with RAV2 and CSN5B in Tomato. *Frontiers in Plant Science*, 7: 1305. doi: 10.3389/fpls.2016.01305
- Souza, L.T., Castro, S.A.Q., Andrade, J.F.; Politano, A.A.; Meneghetti, E.C.; Favarin, J.L.; Almeida, M.; Mazzafera, P. (2021). Drought stress induces changes in the physiology and root system of soybean plants. *Brazilian Journal of Botany*, 44: 779–789 (2021). doi: 10.1007/s40415-021-00776-y
- Specht, J.; Chase, K.; Macrander, M.; Graef, G.; Chung, J.; Markwell, J.; Germann, M.; Orf, J.; Lark, K. (2001). Soybean response to water: A QTL analysis of drought tolerance. *Crop Science*, 41: 493–509.
- Specht, J.; Hume, D.; Kumudini, S. (1999). Soybean yield potential—A genetic and physiological perspective. *Crop Science*, 39: 1560–1570. doi: 10.2135/cropsci1999.3961560x
- Sun, M.; Li, Y.; Zheng, J.; Wu, D.; Li, C.; Li, Z.; Zang, Z.; Zhang, Y.; Fang, Q.; Li, W.; Han, Y.; Zhao, X.; Li, Y. (2022). A Nuclear Factor Y-B Transcription Factor, GmNFYB17, Regulates Resistance to Drought Stress in Soybean. *International Journal of Molecular Sciences*, 23:7242. doi: 10.3390/ijms23137242
- Szcześniak, M.W.; Kabza, M.; Pokrzywa, R.; Gudyś, A.; Makałowska, I. (2013). ERISdb: A Database of Plant Splice Sites and Splicing Signals. *Plant & Cell Physiology*, 54: e10. doi: 10.1093/pcp/pct001
- Tian, M.; Lou, L.; Liu, L.; Yu, F.; Zhao, Q.; Zhang, H.; Wu, Y.; Tang, S.; Xia, R.; Zhu, B.; Serino, G.; Xie, Q. (2015). The RING finger E3 ligase STRF1 is involved in membrane trafficking and modulates salt-stress response in *Arabidopsis thaliana*. *The Plant Journal*, 82: 81-92. doi: 10.1111/tpj.12797
- Uffelmann, E.; Huang, Q.Q.; Munung, N.S.; Vries, J.; Okada, Y.; Martin, A. R.; Martin, H. C.; Lappalainen, T.; Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1: 59. doi: 10.1038/s43586-021-00056-9
- Umehara, M.; Hanada, A.; Yoshida, S.; Akiyama, K.; Arite, T.; Takeda-Kamiya, N., et al. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature*, 455: 195–200. doi: 10.1038/nature07272

- Valente, M.A.; Faria, J.A.; Soares-Ramos, J.R.; Reis, P.A.; Pinheiro, G.L.; Piovesan, N.D.; Morais, A.T.; Menezes, C.C.; Cano, M.A.; Fietto, L.G.; Loureiro, M.E.; Aragão, F.J.; Fontes, E.P. (2009). The ER luminal binding protein (BiP) mediates an increase in drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco. *Journal of Experimental Botany*, 60: 533-46. doi: 10.1093/jxb/ern296
- Verma, A.K.; Tamadaddi, C.; Tak, Y.; Lal, S.S.; Cole, S.J.; Hines, J.K.; Sahi, C. (2019). The expanding world of plant J-domain proteins. *Critical Reviews in Plant Sciences*, 38: 382-400. doi: 10.1080/07352689.2019.1693716
- Vu, V.Q.; Marwick, B.; Hester, J.; Held, M. (2015). R package ggbiplot. Disponível em: <https://github.com/vqv/ggbiplot>
- Wang, J.; Zhang, Z. (2021). GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, Proteomics & Bioinformatics*, 19: 629-640. doi: 10.1016/j.gpb.2021.08.005
- Wang, X.; Marija, V.; Liu, F.; Susanne, J.; Dong, J.; Bernd, W. (2015). Drought priming at vegetative growth stages improves tolerance to drought and heat stresses occurring during grain filling in spring wheat. *Plant Growth Regulation*, 75: 677–687. doi: 10.1007/s10725-014-9969-x
- Wei, T.; Simko, V. (2021). R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.92) Disponível em: <<https://github.com/taiyun/corrplot>>.
- Wei, Y.Q.; Jin, J.L.; Jiang, S.M.; Ning, S.W.; Liu, L. (2018). Quantitative response of soybean development and yield to drought stress during different growth stages in the Huaibei Plain, China. *Agronomy*, 8: 97. doi: 10.3390/agronomy8070097
- Wu, Y., Liu, W., Li, X., Li, M., Zhang, D., Hao, Z., Weng, J.; Xu, Y.; Bai, L.; Zhang, S.; Xie, C. (2011). Low-nitrogen stress tolerance and nitrogen agronomic efficiency among maize inbreds: comparison of multiple indices and evaluation of genetic variation. *Euphytica*, 180: 281-290. doi: 10.1007/s10681-011-0409-y
- Yang, X.; Lu, M.; Wang, Y.; Wang, Y.; Liu, Z.; Chen, S. (2021). Response Mechanism of Plants to Drought Stress. *Horticulturae*, 7, 50. doi: 10.3390/horticulturae7030050
- Yuan, X.; Zhang, S.; Liu, S.; Yu, M.; Su, H.; Shu, H.; Li, X. (2013). Global analysis of ankyrin repeat domain C3HC4-type RING finger gene family in plants. *Plos One*, 8: e58003. doi: 10.1371/journal.pone.0058003
- Zhang, Y.; Liu, Z.; Wang, X.; Li Y.; Li, Y.; Gou, Z.; Zhao, X.; Hong, H.; Ren, H.; Qi, X.; Qiu, L. (2022). Identification of genes for drought resistance and prediction of gene candidates in soybean seedlings based on linkage and association mapping. *The Crop Journal*, 10: 830–839. doi: 10.1016/j.cj.2021.07.010
- Zhang, Z.; Todhunter, R.J.; Buckler, E.S.; Van Vleck, L. D. (2007). Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *Journal of Animal Science*, 85: 881–885. doi: 10.2527/jas.2006-656

Zhou, W.; Kong, W.; Yang, C.; Feng, R.; Xi, W. (2021). Alcohol Acyltransferase Is Involved in the Biosynthesis of C6 Esters in Apricot (*Prunus armeniaca* L.) Fruit. *Frontiers in Plant Science*, 12: 763139. doi: 10.3389/fpls.2021.763139

Zhou, Y.; He, R.; Guo, Y.; Liu, K.; Huang, G.; Peng, C.; Liu, Y.; Zhang, M.; Li, Z.; Duan, L.A. (2019). A novel ABA functional analogue B2 enhances drought tolerance in wheat. *Sci Rep* 9, 2887. doi: 10.1038/s41598-019-39013-8

7. Material Suplementar

Tabela S1. Cultivares de soja a serem utilizados nos experimentos e seus respectivos grupos de maturidade e a empresa detentora do registro da cultivar.

ID	Cultivar	Grupo Maturidade	Companhia
1	NA4990	4.9	Nidera
2	CD216	5.5	Coodetec
3	CD215	5.9	Coodetec
4	BMXENERGIARR	5	GDM
5	BMXTitanRR	5.3	GDM
6	NS4823	4.8	Nidera
7	BMXTURBORR	5.8	GDM
8	NK3358	6.4	Syngenta
9	CD250RRSTS	5.5	Coodetec
10	NIDERAA4725RG	4.7	Nidera
11	CD225RR	5.8	Coodetec
12	M-SOY6101	6.1	Monsanto
13	Bragg	6.6	Embrapa
14	CD5807	6	Coodetec
15	EMGOPA304	7.3	Embrapa
16	BRSMGRenascenca	7.9	Embrapa
17	OCEPAR3-PRIMAVERA	6.5	Embrapa
18	A8000	8	-
19	MERCEDES70A	6.6	-
20	A6001RR	6	-
21	CD224	6.9	Coodetec
22	BRS245RR	7.3	Embrapa
23	CD208	6.9	Coodetec
24	BRS282	6.5	Embrapa
25	BRS185	6.6	Embrapa
26	BRS133	6.6	Embrapa
27	CD206	6.8	Coodetec
28	TMG1067RR	6.7	TMG
29	EMBRAPA48	6.5	Embrapa

30	CD217	7.3	Coodetec
31	BRS257	6.4	Embrapa
32	CD229RR	7.3	Coodetec
33	Fundacep56RR	6.5	Bayer
34	ANTA82	7.4	TMG
35	BRS232	6.5	Embrapa
36	BRS283	6.5	Embrapa
37	5D711RR	7.1	DowAgroscience
38	BRS243RR	6.6	Embrapa
39	FUNDACEP57RR	6.2	Bayer
40	5D690RR	6.9	DowAgroscience
41	BMXPOTENCIARR	6.7	GDM
42	BRS246RR	7.3	Embrapa
43	5D660RR	6.6	DowAgroscience
44	FUNDACEP58RR	6.8	Bayer
45	FUNDACEP53RR	6.4	Bayer
46	CD231RR	7.3	Coodetec
47	CD232	7.3	Coodetec
48	M7211RR	7.2	Monsanto
49	BRS268	7.2	Embrapa
50	TMG1066RR	6.6	TMG
51	M6707RR	6.7	Monsanto
52	BMXMAGNARR	6.2	GDM
53	CD226RR	6.6	Coodetec
54	CD201	6.7	Coodetec
55	CD242RR	7.9	Coodetec
56	CD240RR	6.9	Coodetec
57	NK7059	7	Syngenta
58	BRS284	6.4	Embrapa
59	CD236RR	6.2	Coodetec
60	CD2630RR	6.3	Coodetec
61	CD202	6.4	Coodetec
62	IGRARA628RR	6.4	Igra
63	CD248RR	6.4	Coodetec
64	CD235RR	6.4	Coodetec
65	NA5909RR	5.9	Nidera
66	CD233RR	6.4	Coodetec
67	CD2721RR	7.2	Coodetec
68	CD2737RR	7.3	Coodetec

69	IAS5	6.4	Embrapa
70	CDFAPA	7.3	Coodetec
71	R7	7	-
72	FT-GUAIRA	6.4	Embrapa
73	GARIMPO	7.8	Embrapa
74	EMBRAPA59	6.5	Embrapa
75	5G770RR	7.7	DowAgroscience
76	BRS213	6.5	Embrapa
77	Fundacep33	8	Bayer
78	CD254RR	8.4	Coodetec
78	CD254RR	8.4	Coodetec
79	IGRARA518RR	6	Igra
80	OCEPAR14	5.8	Embrapa
81	MG/BR 46 Conquista	8.2	Embrapa
82	CD2840	8.4	Coodetec
83	CD2800	8	Coodetec
84	P98Y51	8.5	Pioneer
85	P98Y70	8.7	Pioneer
86	CD246	8.1	Coodetec
87	CD237RR	7.3	Coodetec
88	P98Y11	8.1	Pioneer
89	CD243RR	8	Coodetec
90	CD244RR	8	Coodetec
91	M-SOY7901	7.9	Monsanto
92	TMG115RR	8.6	TMG
93	M-SOY8001	8	Monsanto
94	M9144RR	9.1	Monsanto
95	CD234RR	8	Coodetec
96	BMXAPOLLORR	5.5	GDM
97	CD2792RR	7.9	Coodetec
98	M6009RR	6	Monsanto
99	CD253	8.7	Coodetec
100	BRSMT Pintado	8.5	Embrapa

(-) Mantenedor desconhecido

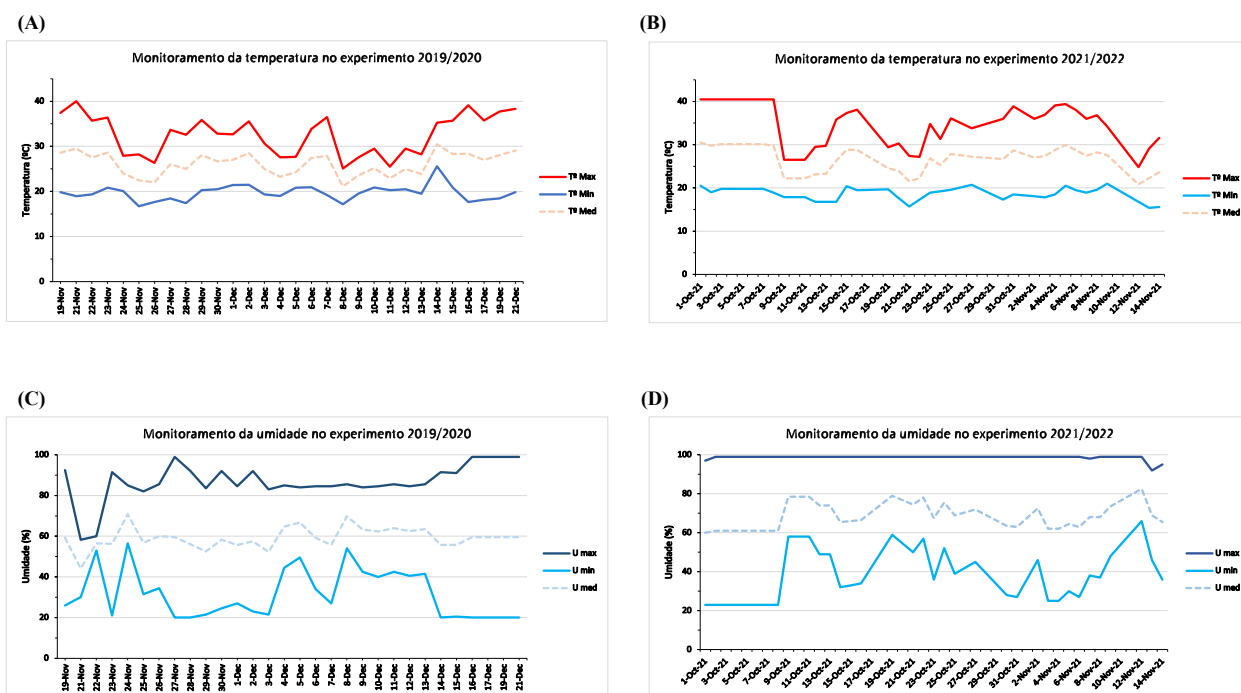


Figura S1 – Monitoramento da Temperatura e Umidade relativa do ambiente durante o período em que as plantas estavam no déficit hídrico. Monitoramento da Temperatura durante o período de estresse no experimento 1 (safra 2019/2020) (A) e no experimento 2 (safra 2021/2022) (B). Monitoramento da Umidade relativa durante. O período de estresse do experimento 1 (C) e no experimento 2 (D).

T° Máx: Temperatura máxima, T° Min: Temperatura mínima, T° Med: Temperatura média, U Máx: Umidade relativa máxima, U Min: Umidade relativa mínima, e U med: Umidade relativa média.

Tabela S2 – Variâncias e herdabilidade das características fenotípicas utilizando a abordagem 1.

Variável	σ^2_A	σ^2_F	h^2
ph	7.56E+00	1.62E+02	4.65E-02
dhip	2.16E-02	8.65E-01	2.49E-02
nнос	1.89E-02	4.70E+00	4.02E-03
sdw	8.15E-03	9.21E-01	8.85E-03
rdw	1.02E-02	7.94E-02	1.28E-01
lpvr	1.91E+05	3.20E+06	5.97E-02
apr	7.39E+02	8.94E+03	8.27E-02
sar	7.30E+03	8.82E+04	8.27E-02
adr	0.00E+00	6.44E-02	0.00E+00
rv	2.01E+00	2.46E+01	8.14E-02
rcl1	8.62E+04	1.78E+06	4.85E-02
rcl2	6.62E+03	1.06E+05	6.24E-02
rcl3	1.60E+03	1.09E+04	1.47E-01
rcl4	1.81E+02	1.34E+03	1.35E-01
rcl5	2.63E+01	3.06E+02	8.59E-02
rcl6	4.54E+00	8.94E+01	5.07E-02

rcl7	1.82E+00	4.51E+01	4.02E-02
rcl8	3.17E-01	2.22E+01	1.43E-02
rcl9	0.00E+00	9.84E+00	0.00E+00
rcl10	1.43E+00	4.77E+01	3.00E-02
rcv1	1.70E-02	3.19E-01	5.34E-02
rcv2	9.48E-02	1.27E+00	7.48E-02
rcv3	2.07E-01	1.41E+00	1.47E-01
rcv4	1.04E-01	7.50E-01	1.39E-01
rcv5	4.14E-02	4.83E-01	8.57E-02
rcv6	1.53E-02	3.12E-01	4.89E-02
rcv7	1.37E-02	2.96E-01	4.65E-02
rcv8	3.65E-03	2.65E-01	1.37E-02
rcv9	0.00E+00	1.98E-01	0.00E+00
rcv10	1.31E-01	4.37E+00	3.00E-02

* σ^2_A – Variância aditiva, σ^2_F – Variância fenotípica, h^2 – herdabilidade no sentido restrito

ph: altura de planta, dhip: diâmetro do hipocótilo, rdw: massa seca da raiz, sdw: massa seca da parte aérea, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcl7: comprimento da raiz por classes de diâmetros 7, rcl8: comprimento da raiz por classes de diâmetros 8, rcl9: comprimento da raiz por classes de diâmetros 9, rcl10: comprimento da raiz por classes de diâmetros 10, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6, rcv7: volume da raiz por classes de diâmetros 7, rcv8: volume da raiz por classes de diâmetros 8, rcv9: volume da raiz por classes de diâmetros 9, rcv10: volume da raiz por classes de diâmetros 10.

Tabela S3 – Variâncias e herdabilidade das características fenotípicas utilizando a abordagem 2.

Variável	σ^2_A	σ^2_F	h^2
ph	8.88E+00	8.60E+01	1.03E-01
dhip	4.35E-02	8.45E-01	5.15E-02
nнос	1.19E-02	1.96E+00	6.05E-03
sdw	2.39E-02	5.41E-01	4.41E-02
rdw	6.00E-03	4.58E-02	1.31E-01
lpvr	8.90E+04	1.48E+06	6.01E-02
apr	2.63E+02	4.02E+03	6.54E-02
sar	2.60E+03	3.97E+04	6.54E-02
adr	0.00E+00	3.22E-02	0.00E+00
rv	5.31E-01	1.13E+01	4.71E-02
rcl1	4.37E+04	7.97E+05	5.49E-02
rcl2	4.59E+03	5.80E+04	7.92E-02

rcl3	4.85E+02	5.36E+03	9.03E-02
rcl4	4.33E+01	6.48E+02	6.68E-02
rcl5	7.07E+00	1.53E+02	4.61E-02
rcl6	9.89E-01	4.36E+01	2.27E-02
rcl7	1.36E-01	2.59E+01	5.25E-03
rcl8	1.17E-01	1.18E+01	9.95E-03
rcl9	1.97E-02	6.02E+00	3.27E-03
rcl10	2.49E-02	2.64E+01	9.43E-04
rcv1	9.78E-03	1.54E-01	6.34E-02
rcv2	5.76E-02	6.81E-01	8.45E-02
rcv3	6.17E-02	6.97E-01	8.86E-02
rcv4	2.41E-02	3.63E-01	6.64E-02
rcv5	1.10E-02	2.42E-01	4.56E-02
rcv6	3.40E-03	1.51E-01	2.25E-02
rcv7	1.64E-03	1.70E-01	9.62E-03
rcv8	1.32E-03	1.42E-01	9.29E-03
rcv9	3.74E-04	1.21E-01	3.09E-03
rcv10	6.66E-03	2.21E+00	3.01E-03

* σ^2_A – Variância aditiva, σ^2_F – Variância fenotípica, h^2 – herdabilidade no sentido restrito

ph: altura de planta, dhip: diâmetro do hipocótilo, rdw: massa seca da raiz, sdw: massa seca da parte aérea, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcl7: comprimento da raiz por classes de diâmetros 7, rcl8: comprimento da raiz por classes de diâmetros 8, rcl9: comprimento da raiz por classes de diâmetros 9, rcl10: comprimento da raiz por classes de diâmetros 10, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6, rcv7: volume da raiz por classes de diâmetros 7, rcv8: volume da raiz por classes de diâmetros 8, rcv9: volume da raiz por classes de diâmetros 9, rcv10: volume da raiz por classes de diâmetros 10.

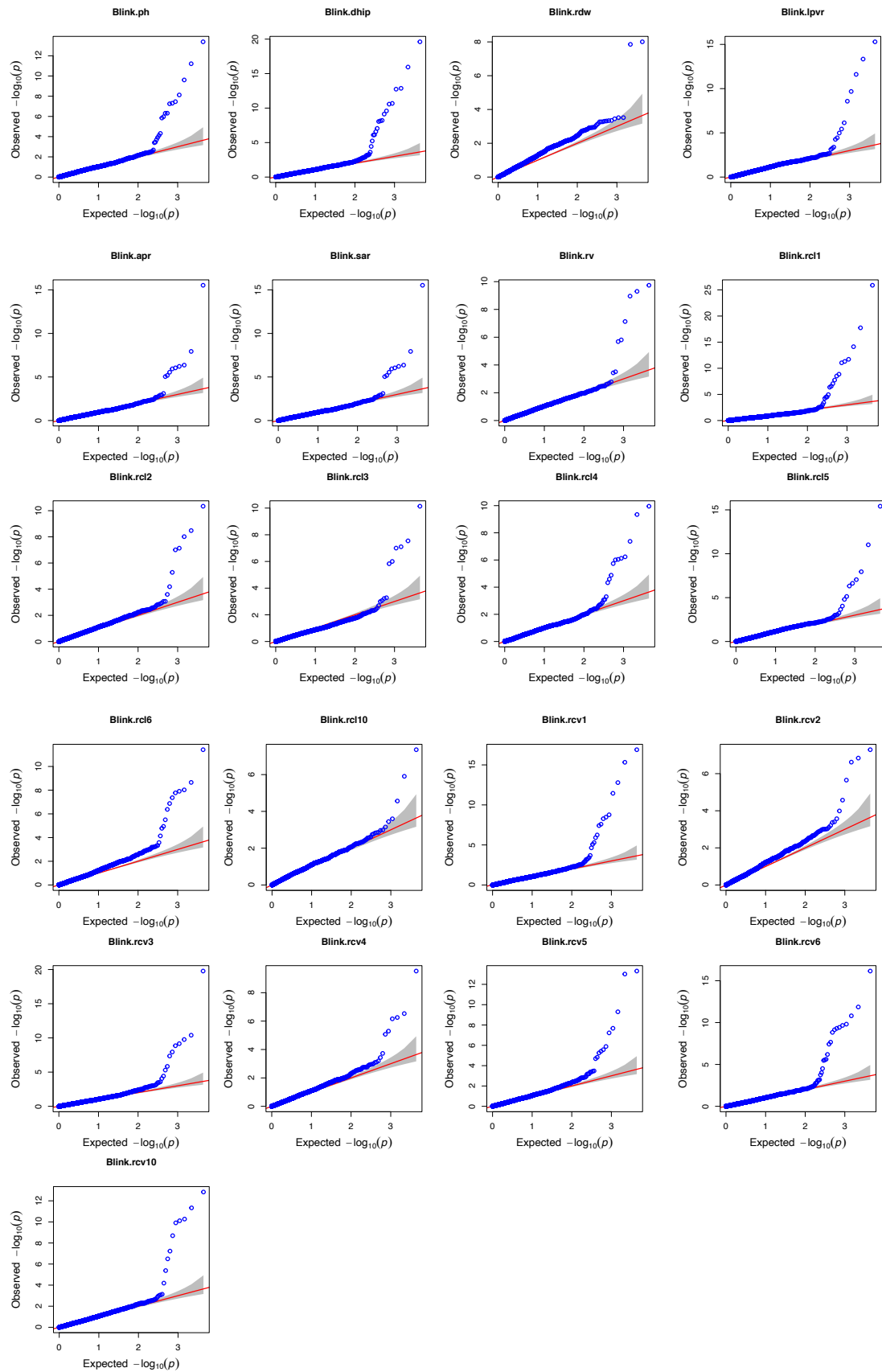


Figura S3- Quantile-quantile (QQ) -plot dos P-valores esperado e observado das características selecionadas utilizando na primeira abordagem.

ph: altura de planta, dhip: diâmetro do hipocótilo, rdw: massa seca da raiz, sdw: massa seca da parte aérea, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcf1:

comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3: comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcl10: comprimento da raiz por classes de diâmetros 10, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6, rcv10: volume da raiz por classes de diâmetros 10.

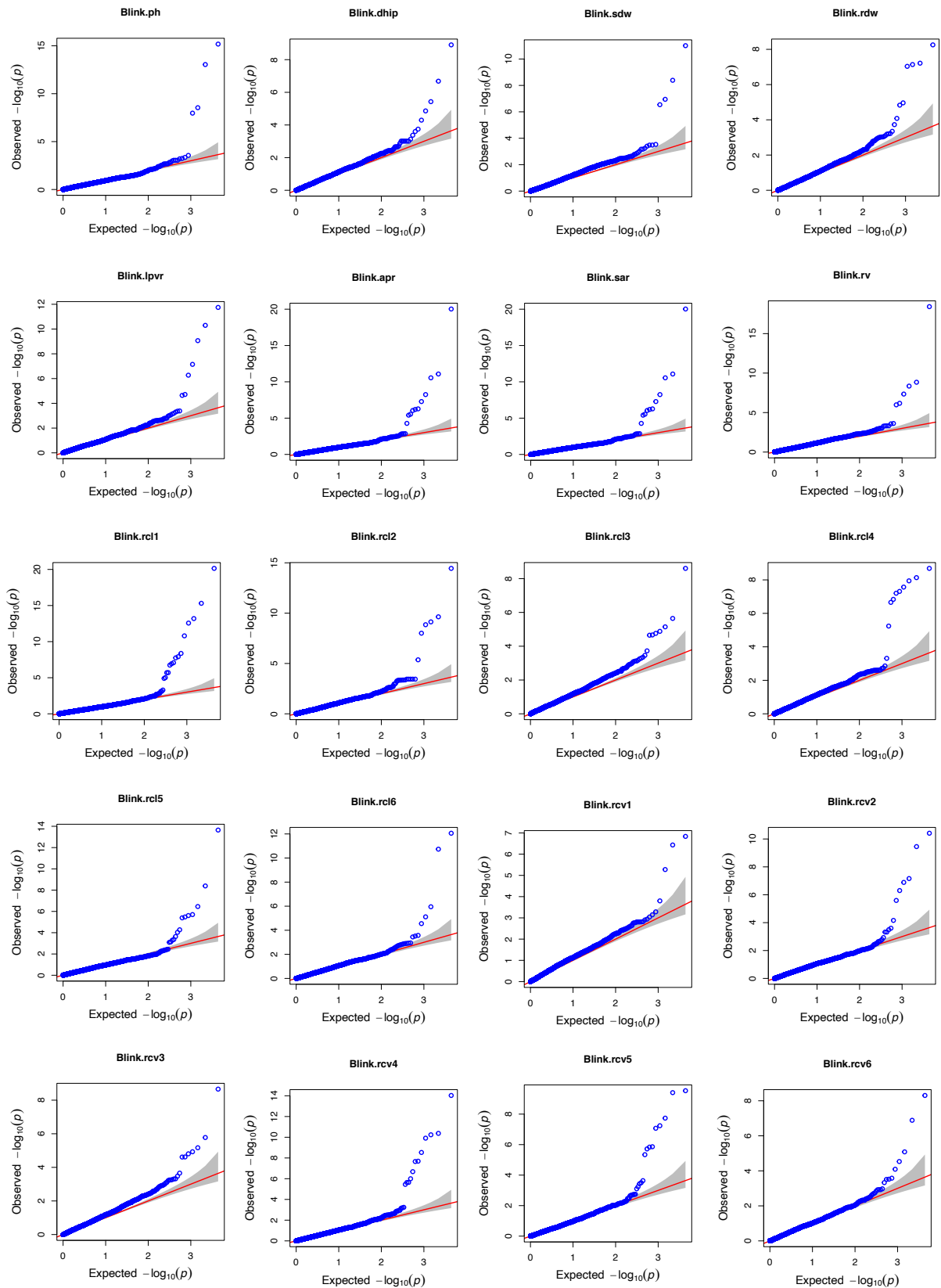


Figura S4- Quantile-quantile (QQ) -plot dos P-valores esperado e observado das características selecionadas utilizando a segunda abordagem.

ph: altura de planta, dhip: diâmetro do hipocótilo, rdw: massa seca da raiz, sdw: massa seca da parte aérea, apr: área projetada da raiz, sar: área superficial da raiz, rv: volume da raiz, lpvr: comprimento total da raiz, rcl1: comprimento da raiz por classes de diâmetros 1, rcl2: comprimento da raiz por classes de diâmetros 2, rcl3:

comprimento da raiz por classes de diâmetros 3, rcl4: comprimento da raiz por classes de diâmetros 4, rcl5: comprimento da raiz por classes de diâmetros 5, rcl6: comprimento da raiz por classes de diâmetros 6, rcv1: volume da raiz por classes de diâmetros 1, rcv2: volume da raiz por classes de diâmetros 2, rcv3: volume da raiz por classes de diâmetros 3, rcv4: volume da raiz por classes de diâmetros 4, rcv5: volume da raiz por classes de diâmetros 5, rcv6: volume da raiz por classes de diâmetros 6.

Tabela S4 - Marcadores SNP significativos em cada abordagem utilizada e na iteração entre as abordagens.

Nomes	Total	SNPs significativos	
Abordagem 1 e Abordagem 2	18	Gm09_36844563_T_C	Gm08_13914377_A_G
		Gm05_40528969_T_G	Gm08_4667338_C_T
		Gm18_56486524_T_C	Gm08_13782412_C_T
		Gm13_35909612_G_A	Gm07_42514160_T_G
		Gm14_5908034_T_C	Gm07_35547398_T_C
		Gm19_36313784_T_C	Gm05_40919406_C_T
		Gm14_34265654_T_C	Gm12_4363882_T_C
		Gm16_24389823_A_C	Gm17_39891056_A_G
		Gm18_57446256_C_T	Gm09_34354926_C_T
Abordagem 1	96	Gm06_14138713_C_A	Gm14_6770667_T_C
		Gm17_11359322_G_T	Gm05_39273308_T_G
		Gm08_18443609_T_C	Gm10_2714130_C_T
		Gm19_39686084_T_C	Gm02_3954305_C_T
		Gm07_675190_G_A	Gm09_42452071_T_G
		Gm15_5522858_T_C	Gm06_6914278_T_C
		Gm19_48707459_C_T	Gm01_39803394_C_A
		Gm10_36651216_A_C	Gm18_7149340_A_G
		Gm02_46855228_G_A	Gm17_4789679_C_A
		Gm11_36518427_A_C	Gm12_34217043_C_T
		Gm10_42296579_T_G	Gm07_43189903_A_G
		Gm02_30250662_T_C	Gm14_6829154_T_C
		Gm14_41890154_T_C	Gm05_10273133_T_C
		Gm10_33347216_A_G	Gm14_43270248_C_T
		Gm01_50262496_T_C	Gm07_5490895_G_T
		Gm02_49776196_A_G	Gm08_18550866_C_T
		Gm17_37843080_C_A	Gm03_1077329_C_T
		Gm08_18379278_T_C	Gm18_54390711_G_T
		Gm10_41445184_T_C	Gm17_9749711_A_G
		Gm14_8918021_T_G	Gm07_36875730_T_C
		Gm07_43182856_G_A	Gm05_33501321_C_T

		Gm14_47081002_T_C Gm19_1420943_T_C Gm18_55580646_T_C Gm08_17399376_A_G Gm08_42793022_A_G Gm01_55789200_A_C Gm05_33472263_A_C Gm08_45541906_A_C Gm08_13959439_T_C Gm13_7559523_G_A Gm07_11153441_A_G Gm08_45888869_C_A Gm05_41658399_C_T Gm17_37769727_A_G Gm18_7573213_G_A Gm14_1657815_T_G Gm08_13040141_A_G Gm16_37209075_T_G Gm03_43707104_A_G Gm17_14781024_A_C Gm07_16438076_A_G Gm19_30103637_T_C Gm02_3644248_T_C Gm04_48198636_G_A Gm06_21567013_G_A Gm01_54123492_T_C Gm20_33676755_C_T	Gm02_48583877_A_C Gm05_1062563_A_G Gm02_43952429_A_G Gm14_1723025_T_C Gm18_7895324_G_A Gm08_19189540_G_A Gm10_26188429_A_G Gm04_43390997_A_C Gm03_36018412_T_C Gm08_43146859_G_T Gm03_38976026_C_T Gm09_33894091_A_C Gm19_40508288_C_T Gm06_8338093_G_A Gm08_42937574_T_C Gm08_45981724_G_A Gm06_3335673_A_G Gm20_34626867_A_G Gm03_31711487_A_G Gm19_43117852_A_C Gm05_40299923_A_G Gm01_48119773_C_T Gm11_38136659_T_C Gm09_26158568_A_G Gm11_37710966_A_G Gm05_2708074_T_C Gm03_4901669_A_G
Abordagem 2	50	Gm20_1577990_G_T Gm14_13119138_G_A Gm18_6840290_T_C Gm12_1761905_G_T Gm19_3458700_A_G Gm17_41481512_T_C Gm06_47631544_T_C Gm11_4345198_G_A Gm07_8327392_G_A	Gm14_2311158_G_A Gm16_4950614_T_C Gm05_888436_A_G Gm05_26685967_G_A Gm19_36505181_A_G Gm06_16372276_T_C Gm20_44585208_A_G Gm05_394764_T_C Gm03_35330372_G_T

	Gm14_3042771_T_C	Gm13_35854636_G_A
	Gm14_226921_G_A	Gm02_14817982_A_C
	Gm19_2372133_C_T	Gm02_41931129_T_G
	Gm15_35867161_C_A	Gm19_5195925_C_T
	Gm05_39643904_T_C	Gm13_26749514_T_C
	Gm18_55167509_A_G	Gm18_13676905_G_A
	Gm15_9164446_A_G	Gm15_6925513_T_C
	Gm11_4329480_T_C	Gm14_30760829_T_C
	Gm01_49987647_C_T	Gm06_49343224_G_T
	Gm08_40012659_C_T	Gm07_10967802_C_T
	Gm17_3252095_T_C	Gm14_86112_G_A
	Gm09_32096458_C_T	Gm13_40715679_A_G
	Gm11_9176042_A_G	Gm10_37298266_A_G
	Gm20_33770596_G_A	Gm17_4312738_C_T
	Gm01_32070134_C_T	Gm19_34687566_A_G
	Gm02_2246479_A_G	Gm05_31160955_A_G

CONCLUSÃO GERAL

A restrição hídrica na cultura da soja afeta negativamente as plantas independentemente se ocorre na fase de germinação ou durante o estágio vegetativo. Na germinação, as características taxa de germinação de plantas e índice de velocidade de emergência de plantas são severamente afetadas. No estágio vegetativo a restrição hídrica afeta grandemente o acúmulo de matéria seca nas plantas tanto da parte aerea quanto da parte vegetativa.

Durante a germinação foram selecionados três marcadores moleculares SNPs que estão associados com a tolerância ao déficit hídrico e no estágio vegetativo foram selecionados 18 marcadores. Esses marcadores podem ser utilizados em programas de melhoramento para a seleção de cultivares mais tolerantes à seca.

Na germinação foram encontrados cinco genes putativos associados a tolerância ao déficit hídrico em que um está envolvido com a constituição da parede celular, dois estão envolvidos em transporte de proteínas via membranas e os outros dois são fatores de transcrição. No estágio vegetativo foram encontrados 15 genes putativos. Dentre desses quatro estão relacionados à tolerância ao déficit hídrico, três relacionados ao crescimento de raízes, um fator de transcrição e outros seis relacionados a outros estresses bióticos e abióticos.

Os genes aqui encontrados são possíveis fontes de resistência ao déficit hídrico em soja e necessitam estudos mais aprofundados para verificar o estudo funcional desses genes. O incremento aos programas de melhoramento para estresse hídrico é crucial para uma agricultura sustentável frente as mudanças climáticas e os marcadores aqui encontrados contribuem efetivamente para esse papel.