

UNIVERSIDADE FEDERAL DE VIÇOSA

ISABELLA CRISTINA CAVALLIN

**PREDIÇÃO GENÔMICA NO MELHORAMENTO DE SOJA VISANDO
TOLERÂNCIA AO DEFICIT HÍDRICO: TAMANHO POPULACIONAL E
SELEÇÃO DE CULTIVARES**

**VIÇOSA – MINAS GERAIS
2020**

ISABELLA CRISTINA CAVALLIN

**PREDIÇÃO GENÔMICA NO MELHORAMENTO DE SOJA VISANDO
TOLERÂNCIA AO DEFICIT HÍDRICO: TAMANHO POPULACIONAL E
SELEÇÃO DE CULTIVARES**

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

Orientador: Felipe Lopes da Silva

Coorientadores: Moysés Nascimento
Leonardo Lopes Bhering

**VIÇOSA – MINAS GERAIS
2020**

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

T

C377p
2020

Cavallin, Isabella Cristina, 1992-

Predição genômica no melhoramento de soja visando tolerância ao deficit hídrico : tamanho populacional e seleção de cultivares / Isabella Cristina Cavallin. – Viçosa, MG, 2020.

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

Orientador: Felipe Lopes da Silva.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Soja - Melhoramento genético. 2. Métodos de simulação.
3. Índice de seleção. 4. Valor genético. 5. Glycine max.

I. Universidade Federal de Viçosa. Departamento de Fitotecnia.

Programa de Pós-Graduação em Genética e Melhoramento.

II. Título.

CDD 22 ed. 633.342

ISABELLA CRISTINA CAVALLIN

**PREDIÇÃO GENÔMICA NO MELHORAMENTO DE SOJA VISANDO
TOLERÂNCIA AO DEFICIT HÍDRICO: TAMANHO POPULACIONAL E
SELEÇÃO DE CULTIVARES**

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

APROVADA: 12 de março de 2020.

Assentimento:



Isabella Cristina Cavallin
Autora



Felipe Lopes da Silva
Orientador

A Rosa, minha avó (in memoriam), que me deixou uma eterna saudade de sua alma jovem.

*Ao João Hilário, meu pai (in memoriam) de quem sinto uma enorme saudade, Deborah
minha mãe, meu exemplo de força, e aos meus irmãos, Bruno e Rafaela, pelo amor em seu
pleno significado*

DEDICO

AGRADECIMENTOS

Primeiramente agradeço a Deus, sem Ele nada seria possível.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade de realização do doutorado.

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

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento – 001, para realização desse trabalho.

A minha família Rocha e Cavallin, que sempre me apoiou incondicionalmente e esteve ao meu lado em todos os momentos, com muito amor, carinho e atenção.

A todos os professores da Universidade Federal de Viçosa pela oportunidade de adquirir conhecimento. Em especial ao professor Felipe, pela amizade, confiança e orientação durante todo o curso. E aos professores Moysés, Leonardo e Marcos Deon pela coorientação.

Ao Francisco, Isabela Sant’Anna, Camila e Carlos Nick pelas contribuições durante a execução do trabalho e nas bancas de qualificação e defesa.

Ao professor Marcos e Cacilda por todo o apoio e preocupação.

Aos secretários e funcionários do departamento de Genética, Marco Tulio e Odilon.

A todos os amigos do Programa Soja, que sem eles, esse trabalho não seria realizado com tanta garra e alegria.

Ao Maikon, Chico, Rafael Zina, Tiago Marçal e Ramon, por toda paciência e ajuda na realização das minhas análises.

Aos meus colegas e amigos do GenMelhor, das gestões 2017 e 2018, que contribuíram na minha formação acadêmica e crescimento pessoal.

Às minhas irmãs da república Amazonas, Clara, Floren, Gabriela, Jheniffer, Larissa, Leticia e Mayra, e aos meninos do 302, Edu, Lucas e Samuel, que foram minha família nesses anos, me acolhendo em todos os momentos.

A todos os amigos que fiz em Viçosa, em especial ao Alex, Pri, Teo, Thi, Julia, Leticia, Patrick, Herika, Aline, Hugo, Gabi, Jo, Lucas, Igor, Matheus, Bruno, Wylle.

A todos os amigos que mantive a distância durante esse período, sempre me enviando forças e alegria.

Meus sinceros agradecimentos.

*“Agir, eis a inteligência verdadeira.
Serei o que quiser.
Mas tenho que querer o que for.
O êxito está em ter êxito, e não em ter condições de êxito.
Condições de palácio tem qualquer terra larga,
mas onde estará o palácio se não o fizerem ali?”*

Fernando Pessoa

RESUMO

CAVALLIN, Isabella Cristina, M.Sc., Universidade Federal de Viçosa, março de 2020. **Predição genômica no melhoramento de soja visando tolerância ao deficit hídrico: tamanho populacional e seleção de cultivares.** Orientador: Felipe Lopes da Silva. Coorientadores: Moysés Nascimento e Leonardo Lopes Bhering.

Características relacionadas com o deficit hídrico apresentam controle genético quantitativo e forte interação entre genótipos e ambientes. Por este fator, a genômica aparece ultimamente com um grande potencial para o estudo e entendimento destas características. No entanto, a acurácia da predição genômica depende do tamanho populacional, e existe um grande desafio na fenotipagem de uma população de grande tamanho e com precisão. Além disso, há necessidade de identificar cultivares tolerantes, por meio de metodologias de imposição de déficit hídrico confiáveis e com tamanho populacional suficiente. Nesse sentido, foram desenvolvidos dois capítulos. O primeiro teve como objetivo comparar a acurácia de duas populações simuladas de 100 e 500 indivíduos e 4000 marcadores através da predição genômica, tomando como referência a cultura da soja e características relacionadas ao déficit hídrico. Neste capítulo observamos que as acurácias foram consideradas boas a altas e os coeficientes de coincidência entre o valor genético genômico e os valores genéticos verdadeiros diferem pouco nos dois tamanhos populacionais. Sendo assim, não há necessidade de fenotipagem de uma população cinco vezes maior, visto que eleva o gasto de tempo e mão-de-obra. No segundo, objetivamos induzir o déficit hídrico em casa de vegetação e selecionar cultivares de soja tolerantes, através da metodologia de predição genômica e o índice de seleção de soma de “ranks”. Para a obtenção dos dados reais de imposição do déficit hídrico foram utilizados 91 cultivares, divididos em dois grupos de 47 cultivares, sendo três testemunhas comuns, em cada ensaio. A imposição do déficit hídrico ocorreu no estágio R5, durante 15 dias. Neste capítulo, obtivemos capacidades preditivas consideradas de moderadas a altas e foi possível identificar cultivares que apresentaram melhores ranqueamentos em ambos os ambientes avaliados (controle e sob deficit hídrico), consideradas tolerantes ao deficit hídrico e responsivas ao aumento da disponibilidade de água. Com isso, fica evidente o potencial do uso da predição genômica para características de difícil mensuração, mesmo para um tamanho populacional reduzido, tanto com a utilização de dados reais como simulados.

Palavras-chave: Simulação. Dados reais. Índice de seleção genômico. Valor genético genômico. *Glycine max*.

ABSTRACT

CAVALLIN, Isabella Cristina, M.Sc., Universidade Federal de Viçosa, March, 2020. **Genetic prediction in soybean breeding, tolerance to drought stress: population size and selection of cultivars.** Adviser: Felipe Lopes da Silva. Co-advisers: Moysés Nascimento and Leonardo Lopes Bhering.

Characteristics related to drought stress have quantitative genetic control and interaction between genotypes and environments. For this reason, the genomic prediction has recently appeared with great potential for the study and understanding of these characteristics. However, the accuracy of the genomic prediction depends on population size, and there is a great challenge in the phenotyping of a large and accurate population. Besides, there is a need to identify tolerant cultivars, through reliable drought stress imposition methodologies with sufficient population size. In this sense, two chapters have been developed. At first, it aimed to compare the accuracy of two simulated populations of 100 and 500 individuals and 4000 markers through genomic prediction, taking as a reference to the soybean and characteristics related to water deficit. In this chapter, we observed that the accuracy was considered good to high and the coincidence coefficients between the genomic genetic value and the true genetic values differ little in the two population sizes. Thus, there is no need for phenotyping a five times larger population as it would increase both costs and time. Soon after, we aim to induce water deficit in a greenhouse and select tolerant soybean cultivars, through the genomic prediction methodology and the sum of ranks selection index. To obtain the real data on the imposition of the drought stress, 91 cultivars were used, divided into two groups of 47 cultivars, three common checks, in each trial. The imposition of water deficit occurred in stage R5, for 15 days. In this chapter, we obtained predictive capacities considered moderate to high and it was possible to identify cultivars that presented better rankings in both evaluated conditions (control and under drought stress), considered tolerant to drought stress and responsive to the increase in water availability. Thus, they demonstrate the potential of using genomic prediction for characteristics that are difficult to measure, even for reduced population size, using both real and simulated data.

Keywords: Simulation. Real data. Genomic selection index. Genomic estimated breeding values. *Glycine max*.

LISTA DE ILUSTRAÇÕES

Capítulo I

Figura 1 Cariograma: Posição dos *loci* controladores das características produtividade (h1), comprimento do sistema radicular (h2), murcha de dossel (h3) e discriminação isotópica de carbono (h4) em relação a cada grupo de ligação.....23

Figura 2. Acurácia realizada considerando o ajuste do G-BLUP para os dois tamanhos populacionais e as quatro características avaliadas ($h^2=0.2$, $h^2=0.4$, $h^2=0.6$ e $h^2=0.8$).....27

LISTA DE TABELAS

Capítulo I

Tabela 1 Número de *loci* (L) controladores das características produtividade (Prod.), comprimento do sistema radicular (CSR), murcha de dossel (MD) e discriminação isotópica de carbono (CID) e suas respectivas herdabilidades (h^2), valores máximos (max), valores mínimos (min), média (\bar{x}), variância (σ^2).....23

Tabela 2 Eficiência relativa (ER) do uso de seleção genômica ampla.....29

Tabela 3 Comparação entre os diferentes tamanhos populacionais, via análise de coincidência*, para a características simuladas: $h^2=0,20$, $h^2=0,40$, $h^2=0,60$ e $h^2=0,80$, com a seleção dos 10, 20, 30, 40 e 50 melhores indivíduos, os valores no corpo da tabela, fora do parênteses, utilizando os GEBVs obtidos pelo método G-BLUP e os valores genéticos obtidos pela simulação. Entre parênteses estão os valores da análise de coincidência*, utilizando os valores fenotípicos (vfen) e genéticos (vgen) obtidos pela simulação.....29

Capítulo II

Tabela 1 Quadro das cultivares de soja (*Glycine max* (L.) Merrill) utilizadas em cada experimento.....49

Tabela 2 Resumo da análise física do solo dos 4 experimentos.....51

Tabela 3 Média (μ), herdabilidade (h^2), acurácia seletiva (r_w), coeficiente de variação ambiental (CVE), coeficiente de variação genético (CVg), variância genética entre cultivares (σ_g^2), variância da interação cultivar por ambiente (σ_{ge}^2) e variância residual (σ_e^2) dos componentes de produção: produção de grãos por planta ($g.planta^{-1}$)(PG), número de grãos (NG), número de vagens (NV), número de grãos por vagens (NGV), peso de cem grãos (P100G), número de vagens sem grãos (VCH) e número de vagens totais (NVT) nos ambientes deficit hídrico e controle.....53

Tabela 4 Correlação entre produção de grãos por planta (g/planta) (PG), número de grãos (NG), número de vagens (NV), número de grãos por vagens (NGV), peso de 100 grãos (P100G), número de vagens sem grãos (VCH) e número de vagens totais (NVT), avaliados em cultivares de soja nos ambientes sob deficit hídrico (acima da diagonal) e controle (abaixo da diagonal).....54

Tabela 5 Capacidade preditiva (r_{gy}), viés e erro padrão (sd) das características peso de grãos por planta (PG), número de grãos por planta (NG), número de grãos por vagem (NGV), número de vagens com grãos (NV), peso de 100 grãos (P100G), número de vagens sem grãos (VCH) e número de vagens totais (NVT) avaliadas em cultivares de soja nos ambientes sob deficit hídrico e controle.....55

Tabela 6 Índice de seleção de soma de “rank” Mulamba e Mock nos diferentes cenários: deficit hídrico e controle, considerando os GEBVs e os fenótipos das sete características avaliadas..56

SUMÁRIO

| | |
|---|----|
| 1 INTRODUÇÃO GERAL | 11 |
| 2 REFERÊNCIAS | 15 |
| RESUMO | 18 |
| ABSTRACT | 19 |
| 1 INTRODUÇÃO | 20 |
| 2 MATERIAL E MÉTODOS | 22 |
| 2.1 OBTENÇÃO DOS DADOS | 22 |
| 2.4 ANÁLISES ESTATÍSTICAS | 24 |
| 3 RESULTADOS | 26 |
| 3.1 ACURÁCIA | 26 |
| 3.2 EFICIÊNCIA RELATIVA DA PREDIÇÃO GENÔMICA | 27 |
| 3.3 COEFICIENTE DE COINCIDÊNCIA | 28 |
| 4 DISCUSSÃO | 30 |
| 5 REFERÊNCIAS | 35 |
| RESUMO | 41 |
| ABSTRACT | 42 |
| 1 INTRODUÇÃO | 43 |
| 2 MATERIAL E MÉTODOS | 46 |
| 2.1 ANÁLISES ESTATÍSTICAS DOS FENÓTIPOS | 49 |
| 2.2 GENOTIPAGEM | 50 |
| 2.3 PREDIÇÃO GENÔMICA | 51 |
| 2.4 SELEÇÃO DE CULTIVARES VIA ÍNDICE DE SELEÇÃO | 52 |
| 3 RESULTADOS | 52 |
| 3.1 ANÁLISE DOS DADOS FENOTÍPICOS | 52 |
| 3.2 CAPACIDADE PREDITIVA E COINCIDÊNCIA DE SELEÇÃO | 55 |
| 3.3 ÍNDICE DE SELEÇÃO | 55 |
| 4 DISCUSSÃO | 59 |
| 5 REFERÊNCIAS | 64 |
| CONCLUSÕES GERAIS | 70 |

1 INTRODUÇÃO GERAL

A soja é uma planta originária de clima temperado, com uma ampla adaptação agronômica aos climas subtropicais e tropicais. Considerada uma das mais importantes leguminosas cultivadas no mundo. A soja e seus subprodutos são utilizados para consumo humano e animal e para fins industriais (HARTMAN et al., 2011), apresentando grande importância social e econômica. Essa oleaginosa atingiu esse patamar por apresentar características desejáveis, como o elevado teor de óleo (20%) e proteína (40%), junto com inúmeros nutrientes favoráveis (SINGH et al., 2008).

Atualmente, os Estados Unidos é o maior produtor mundial da oleaginosa, com uma área plantada na safra de 2018 de 35,657 milhões de hectares, com uma produção de 123,664 milhões de toneladas de soja (USDA, 2019). O Brasil ocupa o segundo lugar no ranque, com área plantada na safra 2018/2019 de 35,874 milhões de hectares e produção de 115 milhões de toneladas de soja (CONAB, 2019). O Brasil teve um aumento de área cultivada (2,1%), no entanto, apresentou um decréscimo de produção de aproximadamente 3,6% em relação à safra 2017/2018, esse decréscimo pode ser decorrente da falta de chuva nos meses de dezembro e janeiro nos maiores estados produtores, prejudicando sobretudo a produtividade das cultivares precoces (CONAB, 2019).

No Brasil, na maioria das safras são registrados quebras de produtividade relacionadas à estiagem, na safra de 2017/2018 houve frustração em relação as expectativas de safra devido aos veranicos em praticamente todas as regiões produtoras, no Paraná houve queda na produtividade de 14,8% associada aos baixos índices pluviométricos e altas temperaturas (CONAB, 2019), na safra 2014/2015 o estado de Minas Gerais e o Distrito Federal tiveram respectivamente 5,1 e 12,1% de redução na produtividade devido ao longo período de estiagem (CONAB, 2015).

Na safra de 2015/2016 foram identificados problemas na maioria das regiões produtoras do Centro-Oeste, causando uma queda de produtividade de 1,3% no Mato Grosso em relação à safra anterior. Nas regiões Norte e Nordeste, a escassez de chuvas ocasionada pelo fenômeno “El Niño” também acarretou queda de 23,2% na produção, principalmente nos estados do Maranhão, Tocantins e Bahia (CONAB, 2016). Podendo também, apresentar problemas a nível nacional, como ocorreu na safra 2011/2012 onde a seca nas regiões produtoras, reduziu a produção em mais de nove milhões de toneladas (CONAB, 2013).

Na cultura da soja para um desempenho adequado, necessita-se de um volume de água de 450 a 800 mm, e uma boa distribuição de chuvas ao longo do seu ciclo, atendendo as

necessidades, principalmente, durante as fases críticas de desenvolvimento, sendo a fase de floração e enchimento de grãos, os mais exigentes em termos de necessidade hídrica, com exigência de 7 a 8 mm.dia⁻¹ (GAVA, 2014). No entanto, a extensão da perda de rendimento depende da duração e intensidade do estresse. Grandes períodos de veranicos ou um curto período de 15 dias, principalmente durante o período reprodutivo, afetam drasticamente a cultura da soja, causando alterações morfológicas no crescimento vegetativo da planta e reduzindo a quantidade e a qualidade das sementes, bem como o número de vagens e o peso seco.

O deficit hídrico é uma grande ameaça e a restrição mais imprevisível, com efeitos adversos na produção agrícola em todo o mundo (GOLLDACK et al., 2014; ANJUM et al., 2017; HUSSAIN et al., 2018). O deficit hídrico induz vários efeitos devastadores nas plantas, perturbando várias atividades da planta, como a taxa de assimilação de carbono, diminuição do turgor, aumento do dano oxidativo e alterações nas trocas gasosas, levando a uma redução no rendimento (HUSSAIN et al., 2018; CHOWDHURY et al., 2016). A sensibilidade da planta à falta de água é um fenômeno complexo e depende de vários fatores, incluindo o estágio de crescimento da planta, potencial genético, duração e severidade do estresse (ZHU et al., 2002). A falta de água também afeta o desenvolvimento foliar, a atividade de enzimas, o equilíbrio iônico e, por fim, leva à redução do rendimento (ANJUM et al., 2017; TOKADA et al., 2017).

O deficit hídrico na fase vegetativa diminui a fotossíntese, a área foliar e a biomassa e, uma vez terminado o deficit, a planta pode se recuperar em certa medida. No entanto, sabe-se que o deficit no estágio reprodutivo pode prejudicar os processos reprodutivos, como aborto de flores, eficiência reprodutiva, desenvolvimento de sementes e vagens jovens. Além disso, o deficit nesta fase geralmente não tem chance de recuperação e, portanto, resulta em severa perda de produtividade da soja (BHATIA e JUMRANI, 2016; JUMRANI et al., 2017).

Baseado nesse cenário, são fundamentais estudos relacionados ao melhoramento genético, auxiliando no desenvolvimento de cultivares tolerantes e adaptadas a regiões com menor disponibilidade hídrica. No entanto, para selecionar genótipos tolerante à seca é necessária uma avaliação criteriosa independentemente do método que será utilizado. Isto ocorre porque, os diferentes genótipos de soja tolerantes à seca exibem um espectro de mudanças morfológicas e fisiológicas diferenciais sob o estresse deficit hídrico, devido a diferença entre os genótipos (KU et al., 2013). Genótipos que apresentam diversidade na resposta ao deficit hídrico constituem excelentes materiais para serem utilizados em programas de melhoramento genético (NOGUEIRA et al., 2001).

Diversos esforços vêm sendo realizados por empresas de pesquisa no intuito de aumentar a eficiência de genótipos de soja quando submetidos a situações de deficit hídrico. Para tanto, várias metodologias têm sido propostas para mensurar o estresse e agrupar genótipos e características de interesse relacionadas à tolerância ao deficit hídrico em diferentes estádios fenológicos da planta, como o uso de marcadores moleculares para construção de mapas de ligação (WANG et al., 2012) ou por meio de estratégias com abordagens fisiológicas (RIES et al., 2011; DEVI et al., 2009) e genômicas (MANAVALAN et al., 2009) até ao uso de engenharia genética (BARBOSA et al., 2013).

Contudo, identificar esses genótipos tolerantes, por meio de metodologias de simulação de deficit hídrico confiáveis e com resultados reprodutíveis a nível de campo é o grande desafio em um programa de melhoramento (MITRA, 2001). Não existe um consenso de qual a melhor maneira de aplicar o deficit hídrico, bem como quanto de deficit deve ser aplicado, além da dificuldade em avaliar um grande número de genótipos.

Outro fator dificultador é devido as características relacionadas com o deficit hídrico apresentarem controle genético quantitativo, com ação aditiva e forte interação entre genótipos e ambientes, e essas características apresentam forte interação com genes que estão diretamente envolvidos com a produtividade (ASHRAF, 2010). Por este fator, a predição genômica aparece ultimamente com um grande potencial para o estudo e entendimento de características complexas em que a interação genótipos x ambientes exerce grande efeito, tal como a tolerância ao deficit hídrico (KUJUR et al., 2013).

A GWS (*Genome Wide Selection*) ou predição genômica (Meuwissen et al., 2001) preconiza o uso do polimorfismo de única base (*Single Nucleotide Polymorphisms* – SNP) como marcadores moleculares. Estes marcadores estão amplamente dispersos e, aproximadamente, equidistantemente distribuídos no genoma (SMARAGDOV, 2009). Sendo assim, os milhares ou milhões de marcadores que cobrem o genoma maximizam o número de QTLs (*Quantitative trait loci*) em desequilíbrio de ligação com pelo menos uma parte dos marcadores. Consequentemente, um maior número de QTLs com efeito no fenótipo serão capturados, inclusive aqueles de pequeno efeito (HEFFNER et al., 2009). O processo central da seleção genômica consiste na estimação de valores genéticos genômicos para todos os marcadores de cada genótipo para posteriormente prever os valores genético genômico estimado (do inglês, *Genomic Estimated Breeding Value* – GEBV) de progênie futuras utilizando somente suas informações genéticas (HEFFNER et al., 2009). Esta predição é realizada utilizando os efeitos dos marcadores genéticos estimados em um modelo treinado a partir de dados fenotípicos e genotípicos de outros indivíduos (MEUWISSEN et al., 2001).

No entanto, é difícil determinar antecipadamente a precisão da predição para uma população específica e uma característica específica, pois a predição é influenciada por muitos fatores. A acurácia da predição depende de fatores como a densidade do marcador em relação ao tamanho efetivo da população, o desequilíbrio de ligação entre marcadores e QTLs, a adequação do modelo de predição em relação à genética da população e a arquitetura da característica de interesse. Segundo Xavier et al. (2016) o tamanho da população de treinamento teve o maior impacto na precisão da predição genômica. Devido à dificuldade de fenotipagem de uma população de grande tamanho e com precisão, para características relacionadas ao déficit hídrico, foram verificadas as respostas de diferentes tamanhos populacionais na acurácia da GWS através de populações simuladas.

Além disso, é necessária a realização de experimentos de predição baseados em dados reais, principalmente relacionados ao déficit hídrico, e, com isso, identificar indivíduos que possuam maiores valores genéticos genômicos (GEBVs) e auxiliar nas estratégias de seleção de um programa de melhoramento de soja visando à seleção de cultivares tolerantes ao déficit hídrico.

2 REFERÊNCIAS

- Anjum SA, Ashraf U, Zohaib A, Tanveer M, Naeem M, Ali I, Tabassum T, Nazir U (2017) Growth and developmental responses of crop plants under drought stress: A review. *Zemdirbyste-Agriculture* 104:267–276.
- Ashraf M (2010) Inducing drought tolerance in plants: recent advances. *Biotechnology Advances* 28:169–183.
- Barbosa EGG, Leite JP, Marin SRR, Marinho JP, Carvalho JFC, Farias JRB, Neumaier N, Marcelino-Guimaraes FC, Oliveira MCN, Yamaguchi-Shinozaki K, Nakashima K, Maruyama K, Kanamori N, Fujita Y, Yoshida T, Nepomuceno AL (2013) Overexpression of the ABA-Dependent AREB1 Transcription Factor from *Arabidopsis thaliana* Improves Soybean Tolerance to Water Deficit. *Plant Molecular Biology Reporter* 31:719-730.
- Bhatia VS, Jumrani K (2016) A maximin-minimax approach for classifying soybean genotypes for drought tolerance based on yield potential and loss. *Plant Breed.* 136:691–700.
- Chowdhury JA, Karim MA, Khaliq QA, Ahmed AU, Khan MSA (2016) Effect of drought stress on gas exchange characteristics of four soybean genotypes. *Bangladesh J. Agr. Res* 41:195–205.
- Conab - Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos- Safra 2019/2020. (2019) 7:1-110. Disponível em: < <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>>. Acesso em: 15/11/2019.
- Conab - Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos- Safra 2015/2016. (2016) 3:1-115. Disponível em: < https://www.conab.gov.br/.../safras/graos/...safra...graos/.../1305_92934803e5c32864b0>. Acesso em: 01/07/2018.
- Conab - Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira. 6º levantamento. Grãos safra 2012-2013. (2013) Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/13_03_07_10_39_19_levantamento_safras_graos_6.pdf>. Acesso em 03/07/2018.
- Conab - Companhia Nacional de Abastecimento. Levantamentos de safras (2015) Disponível em: <<https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos?start=50>>. Acesso em: 18/06/2018.
- Devi MJ, Sinclair TR, Vadez V, Krishnamurthy L (2009) Peanut genotypic variation in transpiration efficiency and decreased transpiration during progressive soil drying. *Field Crops Res.* 114:280-285.
- Gava R (2014) Os efeitos do estresse hídrico na cultura da soja (*Glicine Max*, (L.) Merrill.). Tese (Doutor em Ciências). Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba- SP. 124.

- Golldack D, Li C, Mohan H, Probst N (2014) Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front. Plant Sci.* 5:151.
- Hartman GL, West ED, Herman TK (2011) Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Security* 3:5-17.
- Heffner EL, Sorrells ME, Jannink J (2009) Genomic selection for crop improvement. *Crop Science* 49:1-12.
- Hussain M, Farooq S, Hasan W, Ul-allah S, Tanveer M (2018) Drought stress in sunflower: Physiological effects and its management through breeding and agronomic alternatives. *Agric. Water Manag. J.* 201:152–166.
- Jumrani K, Bhatia VS, Pandey GP (2017) Impact of elevated temperatures on specific leaf weight, stomatal density, photosynthesis and chlorophyll fluorescence in soybean. *Photosynth Res.* 131:333–350.
- Ku YS, Yeung WKA, Yung YL, Li MW, Wen CQ, Liu X, Lam HM (2013) Drought stress and tolerance in soybean. In: James E. Board Ed., *A comprehensive survey of international soybean research - genetics, physiology, agronomy and nitrogen relationships* 209–238.
- Kujur A, Saxena MS, Bajaj D, Laxmi N, Parida SK (2013) Integrated genomics and molecular breeding approaches for dissecting the complex quantitative traits in crop plants. *Journal of Biosciences* 38:971–987.
- Manavalan LP, Guttikonda SK, Tran LS, Nguyen HT (2009) Physiological and molecular approaches to improve drought resistance in soybean. *Plant Cell Physiology* 50:1260–1276.
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Mitra J (2001) Genetics and genetic improvement of drought resistance in crop plants. *Current science* 758–763.
- Nogueira RJMC, Moraes JAPV, Burity HA (2001) Alterações na resistência à difusão de vapor das folhas e relações hídricas em aceroleira submetidas a deficit de água. *Revista Brasileira de Fisiologia Vegetal* 13:75-87.
- Ries LL, Purcell LC, Carter TE, Edwards JT, King CA (2012) Physiological traits contributing to differential canopy wilting in soybean under drought. *Crop Sci* 52:272–281.
- Singh P, Kumar R, Sabapathy SN, Bawa AS (2008) Functional and edible uses of soy protein products. *Comprehensive reviews in food science and food safety* 7:14-28.
- Smaragdov MG (2009) Genomic selection as a possible accelerator of traditional selection. *Russian journal of genetics* 45:633-636.

Todaka D, Zhao Y, Yoshida T, Kudo M, Kidokoro S, Mizoi J, Kodaira K (2017) Temporal and spatial changes in gene expression, metabolite accumulation and phytohormone content in rice seedlings grown under drought stress conditions. *Plant J.* 90:61–78.

USDA – United States Department Agriculture: Economics, Statistics and Market Information System (2019) Disponível em: < <https://usda.library.cornell.edu/>>. Acesso em: 15/11/2019.

Xavier A, Muir WM, Rainey KM (2016) Impact of imputation methods on the amount of genetic variation captured by a single-nucleotide polymorphism panel in soybeans. *BMC bioinformatics* 17:55.

Wang M, Yang WM, Du WJ (2012) Construction of a molecular marker linkage map and its use for quantitative trait locus (QTLs) underlying drought tolerance at germination stage in soybean. *African Journal of Biotechnology* 11:12830-12838.

Zhu JK (2002) Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53:247–273

RESUMO

CAVALLIN, Isabella Cristina, M.Sc., Universidade Federal de Viçosa, março de 2020. Capítulo I. **Predição genômica em diferentes tamanhos populacionais na cultura da soja para características relacionadas ao deficit hídrico.** Orientador: Felipe Lopes da Silva. Coorientadores: Moysés Nascimento e Leonardo Lopes Bhering.

A predição genômica ampla tem sido utilizada nos programas de melhoramento para prever o desempenho de genótipos para características complexas, como, por exemplo, deficit hídrico. No entanto, a limitação na fenotipagem para um grande tamanho populacional tem comprometido a utilização desta técnica em larga escala. Diante do exposto, este trabalho objetivou comparar a acurácia de duas populações de linhagens endogâmicas recombinantes simuladas de 100 e 500 indivíduos através da predição genômica, tomando como referência a cultura da soja. As herdabilidades das quatro características fisiológicas relacionadas ao deficit hídrico foram estabelecidas como sendo igual a 0,20, 0,40, 0,60 e 0,80. Os indivíduos foram genotipados com 4000 marcadores. Realizaram-se as análises por meio do método *Genomic Best Linear Unbiased Predictor*. Para comparar a eficiência da seleção genômica ampla para os dois tamanhos populacionais foram estimadas as acurácias, a eficiência relativa e os coeficientes de coincidência de seleção. A acurácia obtida em todos os cenários avaliados foi considerada de boa a alta, com erro padrão variando de 0,097 a 0,052 para a população de 100 indivíduos e de 0,035 a 0,013 para a população de 500 indivíduos. Os coeficientes de coincidência entre o valor genético genômico e os valores genéticos verdadeiros diferem pouco para os dois tamanhos populacionais quando se seleciona a partir dos 20 melhores indivíduos. Sendo assim, não há necessidade de fenotipagem de uma população cinco vezes maior, visto que eleva o gasto de tempo e mão-de-obra.

Palavras-chave: Fenotipagem. Características fisiológicas. Acurácia. Valor genético genômico (VGG). *Glycine max*.

ABSTRACT

CAVALLIN, Isabella Cristina, M.Sc., Universidade Federal de Viçosa, march of 2020. Chapter I. **Genomic Prediction in different population sizes of soybean for traits related to drought stress.** Adviser: Felipe Lopes da Silva. Co-advisers: Moysés Nascimento and Leonardo Lopes Bhering.

Whole-genome prediction has been used in breeding programs to predict the performance of genotypes for complex traits such as, drought stress. However, the limitation in phenotyping large population sizes has hampered the use of this technique on a large scale. Given the above, this work aimed to compare the accuracy of genomic prediction in two recombinant inbred lines populations simulated with 100 and 500 individuals, taking soybean culture as reference. The heritability of four physiologic traits related to drought stress was established as being equal to 0,20, 0,40, 0,60, and 0,80. The individuals were genotyped with 4000 markers. The analyses were performed by the method Genomic Best Linear Unbiased Prediction. To compare the efficiency of genome-wide selection for the two population sizes the accuracy, relative efficiency, and selection coincidence coefficients were estimated. The accuracy obtained in all evaluated scenarios was considered from moderate to high, with standard errors ranging from 0,097 to 0,052 for the population of 100 individuals and from 0,035 to 0,013 for the population of 500 individuals. The coincidence coefficient between the genomic estimated breeding values and true genetic values differed little for the two population sizes when selecting the top 20 individuals. Thus, there is no need for phenotyping a five times larger population as it would increase both costs and time.

Keywords: Phenotyping. Physiological traits. Accuracy. Genomic estimated breeding values (GEBV). *Glycine max*.

1 INTRODUÇÃO

O deficit hídrico é considerado um dos principais fatores limitantes do desenvolvimento e do crescimento de inúmeras culturas, acarretando na redução da produtividade e em alterações na qualidade do produto final. Na ocorrência de escassez hídrica, as plantas sofrem alterações em processos fisiológicos e bioquímicos, afetando a fotossíntese e alterando a atividade enzimática (CASAGRANDE et al., 2001; FLEXAS, 2006; WANG et al., 2006).

Diante disso, alguns genótipos de soja (*Glycine max* L.) possuem características morfofisiológicas diferenciais quando submetidas ao deficit hídrico, o que constitui uma grande possibilidade para a utilização em programas de melhoramento (NOGUEIRA et al., 2001; KU et al., 2013). Devido a essas respostas diferenciais buscam-se a identificação de características de interesse relacionadas a tolerância ao deficit hídrico em diversos estádios fenológicos da planta e a seleção criteriosa de cultivares que apresentem maior tolerância a seca (WANG et al., 2012; RIES et al., 2011; DEVI et al., 2009; MANAVALAN et al., 2009; BARBOSA et al., 2013).

Uma metodologia que demonstra ser promissora para auxiliar na detecção e escolha de genótipos mais tolerantes as condições adversas do ambiente é a seleção genômica ampla (*Genome Wide Selection - GWS*) (MEUWISSEN et al., 2001).

O processo central da seleção genômica consiste na estimação de efeitos genéticos para todos os marcadores para posteriormente predizer os valores genético genômicos (GEBVs) de progênie futuras utilizando somente suas informações genéticas (HEFFNER et al., 2009). Esta predição é realizada utilizando os efeitos dos marcadores genéticos estimados em um modelo treinado a partir de dados fenotípicos e genotípicos de outros indivíduos (MEUWISSEN et al., 2001).

Espera-se que a GWS preveja desempenhos para características complexas, características de baixa herdabilidade, com mais precisão por não descartar marcadores SNPs (*Single Nucleotide Polymorphisms*) com efeitos pequenos (HEFFNER et al., 2009). Esta abordagem tem auxiliado os melhoristas na seleção, particularmente para características com baixa repetibilidade ou características para as quais a fenotipagem é dispendiosa (HEFFNER et al., 2009; ENDELMAN et al., 2014).

Para soja, estudos relataram altas acurácia na seleção genômica para características de produção, características de desenvolvimento, resistência a patógenos, teor de óleo, proteínas e ácidos graxos (JARQUÍN et al., 2014; BAO et al., 2015; ZHANG et al., 2016; SMALLWOOD

et al., 2019). Jarquín et al. (2014) relataram as acurácias de predição para rendimento e outras características agronômicas obtidas usando dados de genotipagem por sequenciamento. Zhang et al. (2016) relataram os resultados da análise de associação e avaliaram a precisão da predição genômica para o peso de sementes, que é um componente da produtividade de grãos. Smallwood et al. (2019) utilizaram três métodos de GWS para predizer os ganhos com a seleção de linhagens endogâmicas recombinantes. Em todos os estudos, os autores concluíram que a seleção genômica tem um bom potencial para acelerar o ganho genético, preconizando pela redução do tempo gasto com a seleção de genótipos superiores nos programas de melhoramento de soja.

No entanto, não há relato de estudos com o uso de GWS para selecionar genótipos para o deficit hídrico em soja. Esse fato pode ser devido à dificuldade de fenotipagem para características fisiológicas. Além disso, para o uso na GWS, a fenotipagem deve ser precisa quando se avalia uma população de treinamento porque esse conjunto de dados fornecerá a base para o desenvolvimento do modelo estatístico usado para prever o desempenho fenotípico de indivíduos em um programa de melhoramento. Isso indica que abordagens mais rigorosas para a fenotipagem em GWS são essenciais (COBB, 2013).

Segundo Singh e Singh (2015) a obtenção de dados fenotípicos ainda é um desafio, especialmente para características quantitativas como tolerância a estresses abióticos. Os autores concluíram que os métodos tradicionais para a estimação de processos fisiológicos e compostos químicos são geralmente lentos e não passíveis de medição de alto rendimento, não garantindo uma fenotipagem rápida e efetiva financeiramente para essas características.

A fenotipagem a um nível fisiológico refinado pode ser efetivamente aplicada apenas a um número limitado de genótipos, sendo recomendado de 15 a 25 genótipos e quase nunca mais de 50 genótipos (SINCLAIR, 2011), por isso, com número mais elevado de genótipos necessita-se da avaliação em multi experimentos. No entanto, para a maximização da eficiência da GWS um número suficiente de indivíduos devem ser fenotipados e genotipados de forma a se obter acurácias satisfatórias, evidenciado a dificuldade em realizar a GWS para o deficit hídrico em soja.

Diante do exposto, o presente estudo tem como objetivo comparar a acurácia da predição genômica para populações simuladas de 100 e 500 indivíduos que mimetizam quatro características fisiológicas relacionadas ao deficit hídrico.

2 MATERIAL E MÉTODOS

2.1 OBTENÇÃO DOS DADOS

Neste estudo foi avaliada a acurácia da seleção genômica ampla para ranquear e selecionar indivíduos em populações com 100 e 500 indivíduos. Os dados moleculares e fenotípicos empregados neste estudo foram gerados por simulação através do software GENES (CRUZ, 2013).

Neste estudo, o genoma da soja foi tomado como referência, sendo a espécie diploide com $2n=2x=40$ cromossomos. As informações do genoma foram obtidas junto ao banco de dados online da USDA-ARS, o SoyBase disponível em (<http://soybase.org>) (GRANT et al., 2010).

O genoma foi gerado com nível de saturação de 200 marcas moleculares espaçadas de forma aproximadamente equidistante (0,5 cM) em cada um dos 20 grupos de ligação, resultando em 4000 marcadores moleculares no genoma. As marcas moleculares geradas eram bialélicas e codominantes, permitindo a identificação de indivíduos heterozigotos (marcadores SNPs).

Em seguida, foram obtidos os genitores homozigotos contrastantes em fase de aproximação, de modo que para cada loco um genitor foi designado como homozigoto do tipo A_1A_1 e o outro como homozigoto contrastante do tipo A_2A_2 . Desta forma, o cruzamento entre pai 1 e pai 2 gerou a população F1 com todas as marcas em heterozigose (A_1A_2). Com as informações de genoma e genitores, a partir de sucessivas autofecundações dos indivíduos da F1, foram simuladas duas populações de linhagens endogâmicas recombinantes (*Recombinant Inbred Line* – RILs) com 100 e 500 indivíduos. As populações de RILs simuladas foram codificadas com 0 e 2, sendo que 0 correspondeu aos indivíduos homozigotos (A_2A_2) e 2 aos indivíduos homozigotos (A_1A_1), para um determinado loco.

A partir dos dados genotípicos das duas populações de RILs, foram simuladas quatro características fenotípicas com 10 repetições, ação gênica aditiva, ou seja, relação intraalélica de ausência de dominância ($gmd=d/a=0$) para os *loci* controladores das mesmas.

A relação intraalélica de ausência de dominância foi considerada a partir de trabalhos apresentados por Kamikoga (1989) e Brim, Cockerham (1961), cujo valores apresentados para a característica produtividade foram inferiores a unidade e próximos a zero. Desta forma, para fins de padronização, a relação de ausência de dominância foi aplicada também para as demais características.

As herdabilidades das características simuladas foram estabelecidas como sendo igual a 0,20, 0,40, 0,60 e 0,80, fazendo alusão a produtividade, comprimento radicular, murchamento de dossel e discriminação isotópica de carbono, respectivamente. A determinação de tais valores de herdabilidade e o número de *loci* (Tabela 1) foi determinado com base nas informações contidas no banco de dados da SoyBase (<http://soybase.org>) e em artigos científicos e a posição relativa dos *loci* controladores de cada característica foram obtidos aleatoriamente.

Tabela 1. Número de *loci* (L) controladores das características produtividade (Prod.), comprimento do sistema radicular (CSR), murcha de dossel (MD) e discriminação isotópica de carbono (CID) e suas respectivas herdabilidades (h^2), valores máximos (max), valores mínimos (min), média (\bar{x}), variância (σ^2).

| Característica | h^2 | L | Min | Max | \bar{x} | σ^2 | Referência |
|------------------------------|-------|----|--------|--------|-----------|------------|--|
| Prod. (kg.ha ⁻¹) | 0,2 | 82 | 2331,5 | 4236,7 | 3250 | 159661,08 | Zhang et al. (2015); Fang et al. (2017) |
| CSR (cm) | 0,4 | 60 | 10 | 29,99 | 20 | 41,87 | Prince et al. (2015); Abdel-Haleem et al. (2011) |
| MD (nota) | 0,6 | 51 | 6,52 | 48,51 | 25 | 81,49 | Kaler et al. (2017a); Gusso (2018); Charlson et al. (2009) |
| CID (%) | 0,8 | 21 | 15,5 | 24,63 | 20 | 3,58 | Kaler et al. (2017b) |

O posicionamento de cada loco em relação ao grupo de ligação para cada característica pode ser visualizado na Figura 1.

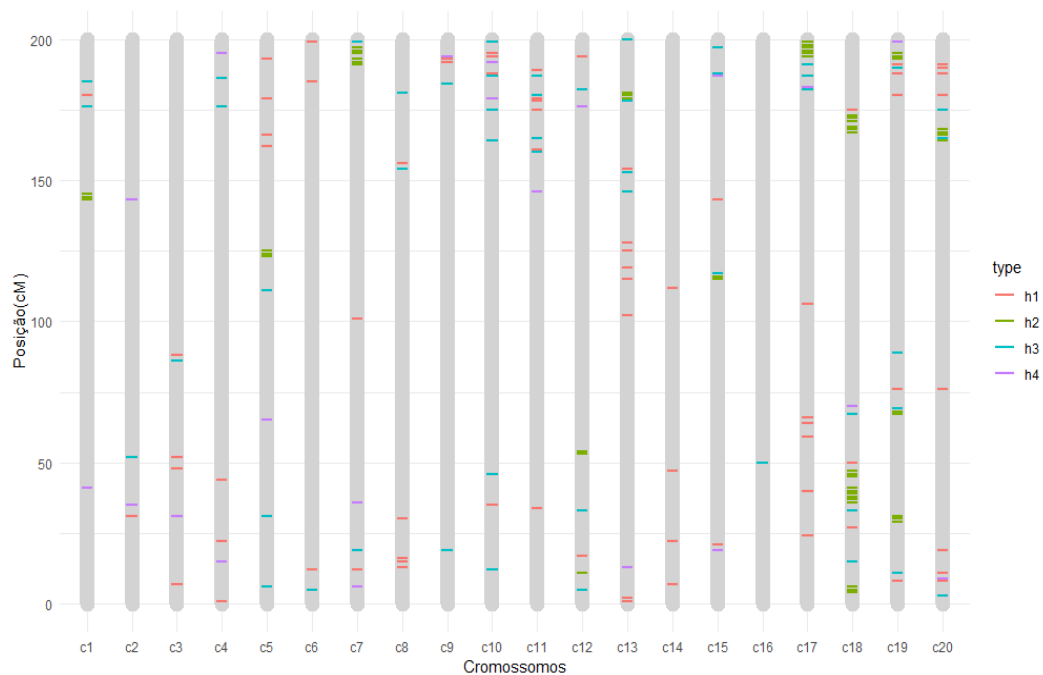


Figura 1. Cariograma: Posição dos *loci* controladores das características produtividade (h1), comprimento do sistema radicular (h2), murcha de dossel (h3) e discriminação isotópica de carbono (h4) em relação a cada grupo de ligação.

Os fenótipos dos indivíduos (i) foram gerados segundo o modelo:

$$Y_i = \mu + \alpha_j + e_i \text{ em que:}$$

Y_i é o fenótipo do indivíduo i ,

μ é a média geral da característica,

α_j é o efeito genético dado pelo somatório dos efeitos genéticos de cada loco e ,

e_i o efeito ambiental, gerado segundo uma distribuição normal com média zero e variância compatível com a herdabilidade da característica simulada.

2.4 ANÁLISES ESTATÍSTICAS

O método estatístico aplicado a predição genômica (GWS) foi o método *Genomic Best Linear Unbiased Predictor* (G-BLUP), através de uma interface do programa Rbio® (BHERING, 2017) com o software R, utilizando o pacote rr-BLUP (ENDELMAN et al., 2011; RODRIGUEZ, DE LOS CAMPOS, 2012). O método G-BLUP foi analisado utilizando as funções `mixed.solve` e `kin.BLUP` (parte do pacote `rrBLUP`) (ENDELMAN, 2011). Todo o processo de simulação foi repetido 10 vezes.

O G-BLUP assume que todos os efeitos dos marcadores são normalmente distribuídos e têm variância igual (MEUWISSEN et al., 2001). Este método utiliza informações de marcadores genéticos para calcular associações entre indivíduos com base na matriz de relação genômica (HABIER et al., 2007). A equação de modelos mistos lineares utilizado foi:

$$Y = Xb + Z\mu_a + \varepsilon$$

Onde: y é o vetor de valores fenotípicos ($N \times 1$, em que N é o número de indivíduos); b é o vetor de efeitos fixos ($p \times 1$, onde p é o número de efeitos fixos considerados); μ_a é o vetor de efeitos genéticos aditivos dos indivíduos ($N \times 1$); ε é o vetor de resíduos do modelo, onde $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$ e σ_ε^2 é a variância residual. X ($N \times p$) e Z ($N \times N$) são as matrizes de incidência para b e μ_a , respectivamente. A estrutura de variância dada por $\mu_a \sim N(0, G_a\sigma_a)$ em que σ_a é a variância aditiva e G_a ($N \times N$) é a matriz de parentesco genômica para efeitos aditivos.

Sob essas configurações, a equação de modelos mistos genômicos para a predição de μa através do BLUP genômico (G-BLUP) é equivalente a:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I G - 1_a / \left(\frac{\sigma^2_\varepsilon}{\sigma^2_a} \right) \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mu a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

onde σ^2_a se refere à variância genética total da característica e σ^2_ε é a variância residual.

Sem padronização a variação genética de cada *loci* é dada por σ^2_a/n onde n está relacionada com o número de marcadores utilizados e é dada por $n = 2 \sum_i^n p_i (1 - p_i)$ (RESENDE et al., 2008; GIANOLA et al., 2009), onde p_i é a frequência dos alelos dos loci i .

Para acessar a eficiência da GWS foi necessário definir as populações de treinamento e validação. Para a metodologia de GWS avaliada, foi realizada a validação cruzada considerando o procedimento 10-folds, com 10 repetições. Desta forma, cada população RIL com 100 e 500 indivíduos foram divididas em 10 grupos iguais contendo 10 e 50 indivíduos cada, respectivamente. Assim, as populações de treinamento foram compostas por 90 e 450 indivíduos e utilizadas para estimar os efeitos dos marcadores, enquanto a concordância entre os valores genéticos preditos via estimativas provenientes da população de treinamento foi validada em cada grupo de 10 e 50 indivíduos, considerados como população de validação.

Uma vez que as estimativas dos efeitos dos marcadores estão disponíveis, o valor genético genômico predito (GEBV) é definido da seguinte maneira:

$$GEBV = \mu a$$

A acurácia realizada ($r_{g\hat{g}}$) foi estimada sendo a correlação de Pearson entre o valor genético genômico (GEBV) estimado pelos modelos e o valor genético verdadeiro, respectivamente. É importante ressaltar que o valor genético verdadeiro g somente é obtido por se tratarem de dados simulados, possibilitando a obtenção dos *True Breeding Value* - TBV.

$$(r_{g\hat{g}}) = cor(GEBV, g)$$

Os erros padrão, associados as acurácias, foram calculados através do desvio padrão dividido por \sqrt{k} , sendo $k=10$.

A eficiência relativa (ER) da utilização de GWS foi calculada através da comparação da acurácia ($r_{g\hat{g}}$) de cada característica com a raiz da herdabilidade das mesmas, adaptado do modelo de Hoffstetter et al. (2016). A herdabilidade foi calculada através das variâncias fenotípicas dos dados simulados.

$$ER = r_{g\hat{g}} / \sqrt{\hat{h}^2}$$

Para comparar a eficiência na seleção dos dois tamanhos populacionais (100 e 500) através das duas metodologias avaliadas na identificação e seleção de indivíduos geneticamente superiores, foram estimados os coeficientes de coincidência de seleção (CS) entre os 10, 20, 30, 40 e 50 indivíduos selecionados a partir dos GEBVs estimados por cada método, em relação aos indivíduos selecionados a partir do valor genético para as quatro características.

O coeficiente de coincidência de seleção (CS) foi calculado da seguinte forma:

$$CS = \left(\frac{NIS}{NT} \right) * 100$$

em que NIS é o número de indivíduos selecionados com base no valor genético que foram os mesmos selecionados com base no GEBV, e NT é o número total de indivíduos selecionados.

Para comparar com a eficiência da seleção fenotípica, foi calculado o coeficiente de coincidência de seleção fenotípica, onde o NIS é o número de indivíduos selecionados com base no valor genético que foram os mesmos selecionados com base no valor fenotípico, e NT é o número total de indivíduos selecionados

3 RESULTADOS

3.1 ACURÁCIA

A partir das correlações entre os GEBVs e os valores genéticos simulados (*True Breeding Value* - TBV) foram obtidas as acurácias para cada população (100 e 500 indivíduos) e para cada uma das quatro características avaliadas ($h^2=0,2$; $h^2=0,4$; $h^2=0,6$; $h^2=0,8$), conforme pode ser visto na Figuras 2. Os erros padrão associados as acurácias foram apresentados nos respectivos gráficos.

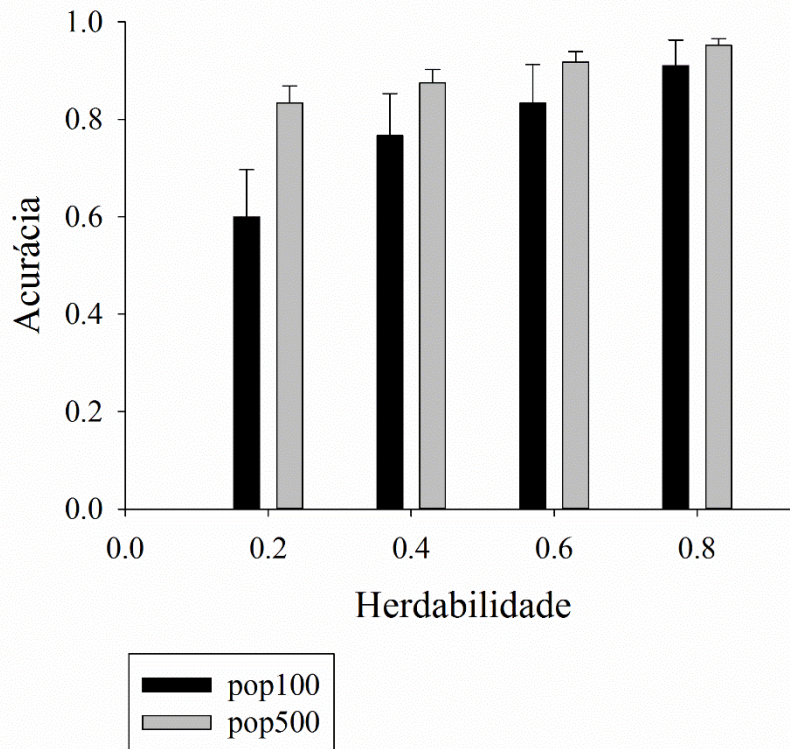


Figura 2. Acurácia realizada considerando o ajuste do G-BLUP para os dois tamanhos populacionais e as quatro características avaliadas ($h^2=0,2$, $h^2=0,4$, $h^2=0,6$ e $h^2=0,8$).

As acurácias realizadas variaram de 0,6 a 0,91 para a população com 100 indivíduos e de 0,83 a 0,95 para a população de 500 indivíduos utilizando o método do G-BLUP. Os erros padrão associados as acurácias foram de 0,097 a 0,052 para a população de 100 indivíduos e de 0,035 a 0,013 para a população de 500 indivíduos, para ambos os métodos de seleção genômica.

Ao compararmos os dois tamanhos populacionais, a população de 500 indivíduos apresenta acurácia 23,4% superior do que na população de 100 indivíduos para a característica com herdabilidade igual a 0,2, essa superioridade cai para 10,8% para a características com herdabilidade de 0,4, e assim por diante, apresentando 8,4% e 4,2% de diferença nas acurácias entre as populações de 500 e 100 indivíduos para as características com herdabilidades de 0,6 e 0,8, respectivamente.

3.2 EFICIÊNCIA RELATIVA DA PREDIÇÃO GENÔMICA

A eficiência relativa da predição genômica calculada através da comparação da acurácia de cada característica com a raiz da herdabilidade estimada das mesmas demonstrou

que a eficiência do uso de GWS diminui conforme aumenta a herdabilidade da característica, esse fato pode ser observado na Tabela 2.

Sendo assim, a GWS é mais eficiente para as características de baixa herdabilidade, como as herdabilidades próximas a 0,2 e 0,4. A ER foi de 17,4%, com os dados obtidos pelos métodos de G-BLUP para a característica com herdabilidade próxima a 0,2 na população com 100 indivíduos. Na população de 500 indivíduos, para características com herdabilidade de 0,2, a ER foi de 53% para os métodos de G-BLUP. Para herdabilidades próximas a 0,4, a ER caiu para 8,4% e 32% para os métodos G-BLUP nas populações de 100 e 500 indivíduos, respectivamente.

Para herdabilidades superiores a 0.6 a ER atingiu valores próximos a zero ou mostrou-se que a utilização de GWS apresentaria a mesma eficiência da utilização da seleção fenotípica para a população de 100 indivíduos. Na população de 500 indivíduos, os valores de ER foram de 17%, para herdabilidades próximas a 0,6 e 5,5% para herdabilidades próximas a 0,8.

Apesar da ER não apresentar valores tão elevados na população de 100 indivíduos, como na população de 500 indivíduos, devemos considerar a utilização de um tamanho populacional menor para características de baixa herdabilidade, visto a dificuldade de fenotipagem de uma população 5 vezes maior para características mais complexas.

Tabela 2. Eficiência relativa (ER) do uso de seleção genômica ampla.

| Tamanho populacional | Herdabilidade | Acurácia | Eficiência relativa |
|----------------------|---------------|----------|---------------------|
| 100 | 0,26073 | 0,59958 | 1,17423 |
| 100 | 0,49929 | 0,76645 | 1,08469 |
| 100 | 0,68220 | 0,83328 | 1,00887 |
| 100 | 0,86499 | 0,91001 | 0,97846 |
| 500 | 0,29590 | 0,83365 | 1,53253 |
| 500 | 0,43563 | 0,87473 | 1,32530 |
| 500 | 0,61316 | 0,91712 | 1,17122 |
| 500 | 0,81251 | 0,95180 | 1,05593 |

3.3 COEFICIENTE DE COINCIDÊNCIA

Os coeficientes de coincidência na seleção dos 10, 20, 30, 40 e 50 melhores indivíduos entre os GEBVs e os valores genéticos mostraram resultados muito semelhantes entre os dois métodos de seleção genômica, apresentados nas Tabelas 3.

Para o método G-BLUP a coincidência estimada entre os indivíduos selecionados para a população de 100 indivíduos variou de 66,6 e 100% e para a população de 500 indivíduos

variou de 30% a 100%. O coeficiente de coincidência demonstrou-se crescente conforme elevava o valor da herdabilidade da característica em praticamente todos os cenários avaliados (Tabela 3).

Pode-se observar que os coeficientes de coincidência diferem pouco para os dois tamanhos populacionais quando se seleciona a partir dos 20 melhores indivíduos, apresentando melhores resultados com a utilização de população com 100 indivíduos, esse fato pode ser observado na Tabela 3.

Tabela 3. Comparação entre os diferentes tamanhos populacionais, via análise de coincidência*, para a características simuladas: $h^2=0,20$, $h^2=0,40$, $h^2=0,60$ e $h^2=0,80$, com a seleção dos 10, 20, 30, 40 e 50 melhores indivíduos, os valores no corpo da tabela, fora do parênteses, utilizando os GEBVs obtidos pelo método G-BLUP e os valores genéticos obtidos pela simulação. Entre parênteses estão os valores da análise de coincidência*, utilizando os valores fenotípicos (vfen) e genéticos (vgen) obtidos pela simulação.

| h ² | Pop. Total | Selecionados | | | | |
|----------------|------------|--------------|------------|----------------|--------------|------------|
| | | GBLUP | | | | |
| | | 10 | 20 | 30 | 40 | 50 |
| 20 | 100 | 70 (40) | 75 (65) | 66,6 (66,7) | 80 (80) | 80 (80) |
| | 500 | 70 (50) | 80 (70) | 90 (73,3) | 80 (65) | 82 (64) |
| 40 | 100 | 70 (50) | 80 (70) | 90 (80) | 90 (95) | 92 (94) |
| | 500 | 30 (20) | 55 (45) | 73,3 (56,7) | 65 (62,5) | 70 (64) |
| 60 | 100 | 90 (70) | 85 (95) | 86,6 (90) | 95 (95) | 94 (98) |
| | 500 | 80 (90) | 75 (85) | 80 (80) | 80 (80) | 84 (84) |
| 80 | 100 | 100 (100) | 90 (95) | 90 (93,3) | 95 (95) | 98 (98) |
| | 500 | 90 (90) | 85 (80) | 90 (83,3) | 80 (90) | 86 (92) |

*Os valores no corpo da tabela estão em porcentagem.

De acordo com os coeficientes de coincidência entre os valores fenotípicos e os valores genéticos apresentados na Tabela 3, as coincidências de seleção variaram entre 40 e 100% para a população com 100 indivíduos. Na população de 500 indivíduos, os valores de coincidência variaram entre 20 a 92%, apresentando os menores valores para a herdabilidade de 0,4, semelhante ao ocorrido na comparação entre os GEBVs e os valores genéticos (Tabela 3).

Pode-se observar que as melhores coincidências de seleção entre os valores genéticos e fenotípicos são obtidas para as herdabilidades iguais ou superiores a 0,6, demonstrando que a

seleção fenotípica é eficiente para características de herdabilidade altas, não justificando o custo para se realizar a GWS, esses resultados confirmam os resultados obtidos pela ER (Tabela 2).

4 DISCUSSÃO

A soja tem mostrado uma redução no potencial de ganhos genéticos ao longo dos anos (ROGERS et al., 2015). Por isso, o uso de informações genômicas através de modelos de predição é uma solução possível para um melhoramento genético mais eficaz.

Através do uso de informações genômicas, a acurácia obtida em todos os cenários avaliados foi considerada de boa a ótima, apresentando resultados surpreendentes por se tratar de um estudo que utilizou uma população considerada, por muitos autores, de tamanho insuficiente e uma população de tamanho razoável, para a utilização de GWS (RESENDE et al., 2014). Os valores de acurácia obtidos foram semelhantes e superiores aos encontrados por SmallWood et al. (2019), que obtiveram acurácias que variaram de 0,4862 a 0,6715 para o G-BLUP em uma população de treinamento com 860 RILs. Estudos de simulação computacional encontraram acurácias entre 0,62 e 0,85 usando dados simulados de marcadores, valores genéticos e fenótipos (HABIER et al., 2007; MEUWISSEN et al., 2001; ZHONG et al., 2009)

Para Resende et al. (2014), a acurácia da GWS depende de alguns fatores, como: proporção (r^2_{mq}) da variação genética explicada pelos marcadores e da acurácia da predição dos efeitos dos marcadores em desequilíbrio de ligação com os QTLs. O primeiro parâmetro está relacionado com a densidade de marcadores e da extensão e padrão do desequilíbrio de ligação existente na população. Ou seja, depende do número e espaçamento dos marcadores e do tamanho do genoma.

Já a acurácia da predição dos efeitos dos marcadores em desequilíbrio de ligação com os QTLs depende do número de indivíduos na população de treinamento e do tamanho efetivo populacional. Xavier et al. (2016) mostraram que o tamanho da população de treinamento é o principal fator limitante para a precisão na soja. Jannink et al. (2010) relataram que um bom conjunto de treinamento deve ser representativo do germoplasma sob avaliação para poder capturar a estrutura populacional e ter um tamanho suficiente para uma acurada estimação dos efeitos dos alelos. Todavia, uma população de treinamento suficientemente grande deve ser empregada quando objetivo final é a seleção de materiais não fenotipados (HEFFNER et al., 2009).

Em características complexas, como a produtividade e características fisiológicas, que possuem baixa herdabilidade, a acurácia da GWS é reduzida, pois a variação genotípica explica

uma reduzida fração da variação fenotípica, devido a influência ambiental. Goddard, Hayes (2007) afirmaram que com a redução da herdabilidade, uma menor proporção da variação aditiva é explicada pelos marcadores genéticos e, portanto, menores acurácias são observados. Para atenuar essa diminuição da acurácia em características complexas deve-se garantir uma boa coleta dos dados fenotípicos e controle ambiental, realizar estudos repetidos em vários ambientes ao longo de várias safras e, com isso, obter valores genético genômicos robustos (HEFFNER et al., 2009; ENDELMAN et al., 2014). Assim, reduzindo os efeitos ambientais, maior parte da variação fenotípica recai sobre a variação genotípica.

Todavia, o tamanho da população de treinamento deve permitir uma fenotipagem eficiente e acurada em vários ambientes, a fim de estimar com precisão os efeitos da interação GxE e do marcador, esse fato torna-se uma entrave ao uso de GWS, pois para avaliar um número maior de genótipos seria necessário realizar experimentos em multi ambientes (MET) devido a dificuldade de se realizar a fenotipagem. Com isso haveria um aumento na interação GxE, podendo levar a uma redução da capacidade preditiva pois os modelos preditivos são, em sua maioria, população e ambientes específicos (RESENDE et al., 2012).

Portanto, no melhoramento de plantas é difícil adicionar mais linhagens a uma população de treinamento para aumentar a precisão do GWS, pois isso aumenta muito os recursos de fenotipagem, visto que, ferramentas de fenotipagem disponíveis exigem, em sua grande maioria, colheitas destrutivas em horários fixos ou em estádios fenológicos específicos além de serem lentas e apresentarem custo elevado. Devido a este alto custo de mão de obra e tempo, a fenotipagem se torna um fator crucial que limita os ganhos genéticos no melhoramento de plantas. Com isso, esforços substanciais estão sendo dedicados ao desenvolvimento de plataformas de fenotipagem de alto rendimento (HTP) em muitas culturas, a fim de gerar fenotipagem em larga escala e em profundidade a baixo custo e intensidade de trabalho (ARAUS, CAIRNS 2014; YANG et al., 2014; ARAUS, 2018). No entanto, a fenotipagem de alto rendimento tornou-se mais comum em laboratórios no setor comercial, mas essa informação raramente entra no domínio público não solucionando o problema com a fenotipagem para características complexas.

Segundo Mujibi et al. (2011), em populações com maior diversidade genética, especialmente aquelas de polinização cruzada, o tamanho da população de treinamento requerido é maior para que se obtenha resultado satisfatório. Por isso, para a cultura da soja, por se tratar de uma cultura autógama e apresentar menor diversidade genética, os resultados obtidos com uma população de 100 indivíduos apresentou acurácia satisfatória. Os resultados desse trabalho demonstram que o incremento nos valores de acurácia obtidos nas populações

de 100 para 500 indivíduos não justificam as dificuldades em se realizar a fenotipagem de uma população cinco vezes superior.

Os altos valores de acurácia obtidos nesse estudo podem estar relacionados ao tamanho efetivo populacional, por se tratar de RILs o tamanho efetivo é reduzido. Segundo Sved (1971), o tamanho efetivo (N_e) da população possui um impacto muito maior que o tamanho absoluto da população de treinamento. Isso porque o tamanho efetivo determina, ao menos parcialmente, a extensão do desequilíbrio de ligação (DL) na população (SVED, 1971). Em populações com N_e reduzido a extensão do DL é consideravelmente maior e, de maneira geral, o número de alelos presentes na população para cada loco é menor. Isso implica que população de baixo N_e , um menor número de marcadores é requerido, bem como um menor número de indivíduos genotipados na população de treinamento, a fim de garantir que ao menos um deles esteja em DL com cada um dos QTLs que controlam a característica de interesse (CRUZ et al., 2013b).

Resende (2002) afirmou que manter o N_e em níveis intermediários pode-se restringir o número de genitores que são inter cruzados para gerar as famílias segregantes, no entanto, essa redução pode ter impactos no programa de melhoramento pois, ao restringir o N_e , reduz a variabilidade genética da população e conseqüentemente, reduzir os ganhos esperados com a seleção em gerações futuras. Tamanhos efetivos na faixa de 10 a 50 são suficientes para o melhoramento da população-elite por várias gerações (WHITE et al., 2007).

Outro fator que afeta a acurácia na GWS é a densidade de marcadores e a extensão do DL existente na população. O presente trabalho utilizou 4000 marcadores espaçados equidistantes a uma distância de aproximadamente 0,5 cM e os *loci* controladores das características fenotípicas foram previamente escolhidos. Moser et al. (2010) mostraram que um chip com 3000 marcadores uniformemente espaçados poderia fornecer aproximadamente 90% de precisão alcançada por meio de alta densidade. Sousa et al. (2019) observaram que o uso de 2500 SNPs distribuídos uniformemente pelo genoma foram capazes de obter boa precisão na predição genômica para características de produtividade e altura de planta de arroz e milho, demonstrando que a eliminação de marcadores que apresentam pequeno efeito é uma boa estratégia e que o construir um chip de baixa densidade e específico para cada característica pode reduzir os custos de genotipagem em programas de melhoramento.

SmallWood et al. (2019), relataram que para a cultura da soja ao usar matrizes de marcadores muito densas, muitos dos marcadores poderiam mapear a mesma localização genética com base na recombinação limitada. Estes mesmos autores, observaram redundância em seus estudos, onde o número inicial de 11633 SNPs foi reduzido para 4867 após a remoção daqueles em locais duplicados. Além disso, relataram que o método G-BLUP não foi

amplamente afetado pela queda de 4867 SNPs para 867 SNPs. Da mesma forma, VanRaden et al. (2011) reportaram que uma maior densidade de marcadores não necessariamente aumenta a acurácia da GWS e o uso de subconjuntos de marcadores às vezes superam o conjunto de dados inteiro (ZHANG et al., 2010; MA et al., 2016).

Xavier et al. (2016) avaliaram as propriedades preditivas da GWS na cultura da soja e testaram subconjuntos dos dados genotípicos, utilizando o painel completo, meio painel e um quarto de painel de marcadores, correspondendo a 4077, 2039 e 1020 SNPs, respectivamente. Os autores relataram que 1020 marcadores foram suficientes para prover uma predição consistente, enquanto que o aumento da densidade de marcadores promoveu apenas ganhos marginais na capacidade de predição. Possivelmente, este resultado está associado às propriedades genômicas da soja, como o grande desequilíbrio de ligação existente (HYTEN et al., 2007) e a distribuição desproporcional dos marcadores SNPs (LI et al., 2014). Com isso, podemos afirmar que com a utilização de 4000 marcadores e espaçados equidistantes foram suficientes para uma boa cobertura do genoma e para auxiliar na obtenção de boas acurácias.

Além disso, o uso de SNPs previamente selecionados de estudos de associação genômica ampla (GWAS) na matriz de relação genômica seria uma forma de aumentar a precisão genômicas (SU et al., 2014; SPINDEL et al., 2016). Zhang et al. (2015) mostraram que a matriz G ponderada de acordo com SNPs individuais que apresentam sinais fortes e robustos de associação pode efetivamente melhorar as predições genômicas, ou seja, a escolha dos *loci* controladores das características através de trabalhos descritos na literatura e na base da plataforma (soybase.com) garantiu uma melhor acurácia para todas as características avaliadas.

O emprego de maiores densidades de SNPs é necessário em populações com baixo desequilíbrio de ligação entre os marcadores e QTL, com maior variância genética ou de estrutura populacional desconhecida (BASTIAANSEN et al., 2012; DAETWYLER et al., 2013; VALENTE et al., 2015). No entanto Valente et al. (2015) destacaram que a seleção de um caráter com elevada herdabilidade requer uma alta densidade de marcadores no genoma para que a acurácia da GWS fosse superior à seleção fenotípica, nesse caso, possivelmente a eficiência relativa do uso de GWS iria melhorar para as características de herdabilidade de 0,6 e 0,8 na população de 100 indivíduos com a utilização de um maior número de marcadores.

Pode-se observar que a eficiência relativa do uso da GWS quando comparada a seleção fenotípica é inversamente proporcional aos valores de herdabilidade. Valente et al. (2015), reportaram que, quanto menor a herdabilidade, maior a superioridade da GWS frente a seleção fenotípica, visto que a herdabilidade proporciona menor correlação com o fenótipo observado

em relação ao valor genético-genômico estimado. Muir (2007) apontou que características com baixa herdabilidade têm maior potencial para serem exploradas pela GWS, para incrementar a acurácia de seleção e, por conseguinte, os ganhos genéticos.

Se analisarmos a porcentagem de eficiência do uso de seleção genômica e as coincidências de seleção obtidas pela utilização dos métodos de GWS frente a seleção fenotípica, podemos garantir o uso de dados genômicos para características com menor herdabilidade podem fornecer resultados comparáveis a não utilização de dados genômicos na seleção de características com valores de herdabilidade altos mesmo com populações de 100 indivíduos. Desta forma, resultados indicam que até as características com baixa herdabilidade podem prover acurácia razoável e seleção eficiente com o uso de GWS, conclusão semelhante a obtida por Xavier et al. (2016).

Dessa forma, características complexas que apresentam baixa herdabilidade e efeitos pequenos de marcadores apresentam maior eficiência na GWS. Grattapaglia, Resende (2011) demonstraram que a acurácia aumentou apenas 10-20% à medida que a herdabilidade aumentou de 0,2 para 0,6, independentemente do tamanho da população.

Ao se analisar os resultados obtidos pelas coincidências de seleção para os dois tamanhos populacionais, confirmam mais uma vez que a GWS é eficiente em selecionar os mesmos indivíduos geneticamente superiores para ambos os tamanhos populacionais, não justificando o gasto com a fenotipagem de uma população cinco vezes maior para nenhuma das características avaliadas.

Os métodos de seleção genômica apresentaram resultados bem semelhantes para todas as variáveis avaliadas nas duas populações, em todos os níveis de herdabilidade. Esses achados são concordantes com os achados de Duhnen et al. (2017), Bhering et al. (2015); Azevedo et al. (2015), Habier et al. (2011) e Jiménez-Montero et al. (2013). Esses resultados indicam que o método G-BLUP mostra-se eficiente na predição dos GEBVs por apresentar altos valores de acurácias preditivas e coincidência na seleção de indivíduos geneticamente superiores em menor tempo de processamento. Sendo assim, nesse caso, recomenda-se o uso de G-BLUP por se tratar de um método que demanda de menor tempo de processamento bem como menor demanda computacional.

Por fim, conclui-se que as acurácias obtidas nos dois tamanhos populacionais foram consideradas de boas a altas, além de apresentarem ótimas porcentagens de coincidências na seleção de indivíduos geneticamente superiores, não justificando a fenotipagem de um número de indivíduos cinco vezes superior, visto que eleva tempo e mão-de-obra e torna-se mais dispendioso para características da natureza das utilizadas nesse experimento.

5 REFERÊNCIAS

- Abdel-Haleem H, Lee G-J, Boerma RH (2011) Identification of QTL for increased fibrous roots in soybean. *Theoretical and applied genetics*. 122:935-946
- Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends in plant science* 19:52-61.
- Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS, Cairns JE (2018). Translating high-throughput phenotyping into genetic gain. *Trends in plant science*, 23:451-466.
- Azevedo CF, de Resende MDV, e Silva FF, Viana JMS, Valente MSF, Resende MFR, Muñoz P (2015) Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC genetics*, 16:105.
- Barbosa EGG, Leite JP, Marin SRR, Marinho JP, Carvalho JDFC, Fuganti-Pagliarini R, Yamaguchi-Shinozaki K (2013) Overexpression of the ABA-dependent AREB1 transcription factor from *Arabidopsis thaliana* improves soybean tolerance to water deficit. *Plant molecular biology reporter*, 31:719-730.
- Bao Y, Kurlle JE, Anderson G, Young ND (2015) Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. *Molecular Breeding*, 35:128.
- Bastiaansen JW, Coster A, Calus MP, Van Arendonk JA, Bovenhuis H (2012) Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. *Genetics Selection Evolution* 44:3.
- Bernardo R, Yu JM (2007) Prospects for genomewide selection for quantitative traits in maize. *Crop Sci* 47:1082-1090.
- Bhering LL (2017) Rbio: A Tool For Biometric And Statistical Analysis Using The R Platform. *Crop Breeding and Applied Biotechnology* 17:187-190.
- Bhering LL, Junqueira VS, Peixoto LA, Cruz CD, Laviola BG (2015) Comparison of methods used to identify superior individuals in genomic selection in plant breeding. *Genet. Mol. Res* 14:10888-10896.
- Brim CA, Cockerham CC (1961) Inheritance of Quantitative Characters in Soybeans 1. *Crop Science* 3:187-190.
- Casagrande EC, Farias JRB, Neumaier N, Oya T, Pedroso J, Martins PK, Breton M, Nepomucano AL (2001) Expressão gênica durante deficit hídrico em soja. *Revista Brasileira de Fisiologia Vegetal* 13:168-184.
- Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics* 126:867-887.

- Charlson DV, Bhatnagar S, King CA, Ray JD, Sneller CH, Carter TE, Purcell LC (2009) Polygenic inheritance of canopy wilting in soybean [*Glycine max* (L.) Merr.]. *Theoretical and Applied Genetics*, 119:587-594.
- CRUZ CD (2013a) Genes: a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum* 35:271-276.
- Cruz CD, Salgado CC, Bhering LL (2013) *Genômica aplicada*. Suprema, Visconde do Rio Branco.
- Daetwyler HD, Calus MP, Pong-Wong R, De Los Campos G, Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193:347–365
- Devi MJ, Sinclair TR, Vadez V, Krishnamurthy L (2009) Peanut genotypic variation in transpiration efficiency and decreased transpiration during progressive soil drying. *Field Crops Res* 114:280-285.
- Duhnen A, Gras A, Teyssèdre S, Romestant M, Claustres B, Daydé J, Mangin B (2017) Genomic selection for yield and seed protein content in soybean: a study of breeding program data and assessment of prediction accuracy. *Crop Science* 57:1325-1337.
- Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells ME, Jannink JL (2014) Optimal design of preliminary yield trials with genome-wide markers. *Crop Science* 54:48-59.
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* 4:250-255.
- Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, Pan Y et al. (2017) Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome biology* 18:161.
- Flexas J, Bota J, Galmes J, Medrano H, Ribas-Carbó M (2006) Keeping a positive carbon balance under adverse conditions: responses of photosynthesis and respiration to water stress. *Physiologia Plantarum* 127:343-352.
- Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R (2009) Additive genetic variability and the Bayesian alphabet. *Genetics* 183:347-363.
- Goddard ME, Hayes BJ (2007) Genomic selection. *Journal of Animal breeding and Genetics* 124:323-330.
- Grant D, Nelson RT, Cannon SB, Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database.
- Grattapaglia D, Resende MD (2011) Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7:241-255.
- Gusso A (2018) Canopy temperatures distribution over soybean crop fields using satellite data in the Amazon biome frontier. *European Journal of Remote Sensing* 51: 901-910.

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397.

Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. *BMC bioinformatics*, 12:186.

Hayes BJ, Daetwyler HD, Bowman P, Moser G, Tier B, Crump R, Goddard ME et al (2009) Accuracy of genomic selection: comparing theory and results. In *Proc Assoc Advmt Anim Breed Genet* 18:34-37.

Heffner EL, Sorrells ME, Jannink, JL (2009) Genomic selection for crop improvement. *Crop Science* 49:1-12.

Hoffstetter A, Cabrera A, Huang M, Sneller C (2016) Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat. *G3: Genes, Genomes, Genetics* 6:2919-2928.

Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, Cregan PB et al (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175:1937-1944.

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* 9:166-177.

Jarquín D, Kocak K, Posadas L, Hyma K, Jedlicka J, Graef G, Lorenz A (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC genomics* 15:740.

Jiménez-Montero JA, González-Recio O, Alenda R (2013) Comparison of methods for the implementation of genome-assisted evaluation of Spanish dairy cattle. *Journal of dairy science* 96:625-634.

Kaler AS, Ray JD, Schapaugh WT, King CA, Purcell LC (2017a) Genome-wide association mapping of canopy wilting in diverse soybean genotypes. *Theoretical and Applied Genetics* 130:2203-2217.

Kaler AS, Dhanapal AP, Ray JD, King CA, Fritschi FB, Purcell LC (2017) Genome-wide association mapping of carbon isotope and oxygen isotope ratios in diverse soybean genotypes. *Crop Science*, 57(6), 3085-3100.

Kamikoga MK (1989) Herança do acamamento e associações com outros caracteres em soja (*Glycine max* (L.) Merrill). Dissertação, Escola Superior de Agronomia “Luiz de Queiroz”, Universidade de São Paulo.

Ku YS, Yeung WKA, Yung YL, Li MW, Wen CQ, Liu X, Lam HM (2013) Drought stress and tolerance in soybean. In: James E. Board Ed, *A comprehensive survey of international soybean research - genetics, physiology, agronomy and nitrogen relationships*. InTech 209–238.

Li YH, Liu YL, Reif JC, Liu ZX, Liu B, Mette MF, Qiu LJ et al (2014) Biparental resequencing coupled with SNP genotyping of a segregating population offers insights into the landscape of

recombination and fixed genomic regions in elite soybean. *G3: Genes, Genomes, Genetics* 4:553-560.

Ma Y, Reif JC, Jiang Y, Wen Z, Wang D, Liu Z, Wang H (2016) Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.). *Molecular Breeding* 36:113.

Manavalan LP, Guttikonda SK, Phan Tran LS, Nguyen HT (2009) Physiological and molecular approaches to improve drought resistance in soybean. *Plant and Cell Physiology* 50:1260-1276.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.

Moser G, Khatkar MS, Hayes BJ, Raadsma HW (2010) Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genet Sel Evol* 42:37.

Muir WM (2007) Comparison of genomic and traditional BLUP estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet* 124:342-355.

Mujibi FDN, Nkrumah JD, Durunna, ON, Stothard P, Mah J, Wang Z, Moore SS et al (2011) Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. *Journal of animal science* 89:3353-3361.

Nogueira RJMC, Moraes JAPV, BURITY HA, Bezerra Neto E (2001) Alterações na resistência à difusão de vapor das folhas e relações hídricas em aceroleiras submetidas a deficit de água. *Revista Brasileira de Fisiologia Vegetal* 13:75-87.

Prince SJ, Song L, Qiu D, dos Santos JVM, Chai C, Joshi T, Krampis K (2015) Genetic variants in root architecture-related genes in a *Glycine soja* accession, a potential resource to improve cultivated soybean. *BMC genomics* 16:132.

Resende MDV (2002) *Genômica biométrica e estatística no melhoramento de plantas perenes*. Brasília.

Resende Jr, MFR, Munoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Kirst M et al (2012) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193:617-624.

Resende MDV, Lopes PS, da Silva RL, Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesquisa Florestal Brasileira*, 56:63.

Resende MDV, Silva FF, Azevedo CF (2014) *Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência*. Suprema, Viçosa.

Ries LL, Purcell LC, Carter TE, Edwards JT, King CA (2012) Physiological traits contributing to differential canopy wilting in soybean under drought. *Crop Science* 52:272-281.

- Rogers J, Che P, Shi A, Zhang B, Scaboo A, Smith SF, Zeng A (2015) Agronomic performance and genetic progress of selected historical soybean varieties in the southern USA. *Plant breeding* 134:85-93.
- Rodriguez, PP, de Los Campos G (2012) BGLR: A Statistical Package for Whole-Genome Regression.
- Sinclair TR (2011) Challenges in breeding for yield increase for drought. *Trends in plant science* 16:289-293.
- Singh BD, Singh AK (2015) *Marker-assisted plant breeding: principles and practices*. Springer, New Delhi.
- Smallwood CJ, Saxton AM, Gillman JD, Bhandari HS, Wadl PA, Fallen BD, Pantalone VR et al (2019) Context-Specific Genomic Selection Strategies Outperform Phenotypic Selection for Soybean Quantitative Traits in the Progeny Row Stage. *Crop Science* 59:54-67.
- Smaragdov MG (2009) Genomic selection as a possible accelerator of traditional selection. *Russian journal of genetics* 45:633-636.
- Sousa MB, Galli G, Lyra DH, Granato ÍSC, Matias FI, Alves FC, Fritsche-Neto R (2019). Increasing accuracy and reducing costs of genomic prediction by marker selection. *Euphytica* 215:18.
- Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redona E, McCouch SR et al (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS genetics* 11:e1004982.
- Su G, Madsen P, Nielsen US, Mäntysaari EA, Aamand GP, Christensen OF, Lund MS (2012) Genomic prediction for Nordic Red Cattle using one-step and selection index blending. *Journal of Dairy Science* 95:909-917.
- Su G, Christensen OF, Janss L, Lund MS (2014) Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *Journal of dairy science* 97:6547-6559.
- Sved JA (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology* 2:125-141.
- Valente BD, Morota G, Peñagaricano F, Gianola D, Weigel K, Rosa GJ (2015) The causal meaning of genomic predictors and how it affects construction and comparison of genome-enabled selection models. *Genetics* 200:483-494.
- VanRaden PM, O'Connell JR, Wiggans GR, Weigel KA (2011) Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10-1186.
- Wang L, Zhang T, Ding S (2006) Effect of drought and rewatering on photosynthetic physiological characteristics of soybean. *Acta Ecologica Sinica* 26:2073-2078.

Wang M, Yang WM, Du WJ (2012) Construction of a molecular marker linkage map and its use for quantitative trait locus (QTLs) underlying drought tolerance at germination stage in soybean. *African Journal of Biotechnology* 11:12830-12838.

White TL, Adams WT, Neale DB (2007) *Forest Genetics*. CABI publisher, Wallingford.

Xavier A, Muir WM, Rainey KM (2016) Impact of imputation methods on the amount of genetic variation captured by a single-nucleotide polymorphism panel in soybeans. *BMC bioinformatics* 17:55.

Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Wang G et al (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature communications* 5:5087.

Zhang Z, Liu J, Ding X, Bijma P, de Koning DJ, Zhang Q (2010) Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PloS one* 5:e12648.

Zhang Z, Erbe M, He J, Ober U, Gao N, Zhang H, Li, J et al (2015) Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. *G3: Genes, Genomes, Genetics* 5:615-627.

Zhang J, Song Q, Cregan PB, Jiang GL (2016) Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theoretical and Applied Genetics* 129:117-130.

Zhong S, Dekkers JC, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182:355-364.

RESUMO

CAVALLIN, Isabella Cristina, M.Sc., Universidade Federal de Viçosa, março de 2020. Capítulo II. **Predição genômica em soja sob condição adequada de disponibilidade hídrica e sob deficit hídrico induzido em casa de vegetação.** Orientador: Felipe Lopes da Silva. Coorientadores: Moysés Nascimento e Leonardo Lopes Bhering.

O deficit hídrico afeta seriamente a produtividade da soja. Uma maneira de aumentar a produção da cultura é através do desenvolvimento de cultivares com tolerância à seca. No entanto, as características relacionadas ao déficit hídrico são de natureza complexa. Diante disto, espera-se que a predição genômica e índice de seleção com a utilização de valores genéticos genômicos forneçam uma ferramenta para a seleção destas características. O experimento consistiu na avaliação de 91 cultivares de soja divididas em dois grupos de 47 cultivares, cada grupo foi avaliado em dois experimentos. Três cultivares pertenceram aos dois grupos como testemunhas comuns. O delineamento experimental utilizado para cada ensaio, foi em blocos casualizados (DBC) com três repetições e em esquema fatorial com dois fatores. O primeiro fator composto pelas cultivares e o segundo pela disponibilidade de água, sendo: o tratamento controle e o tratamento sob deficit hídrico. A imposição do deficit hídrico se baseia na capacidade de retenção do solo e ocorreu no estágio reprodutivo (R5) durante quinze dias. Ao final do ciclo foram avaliadas sete características relacionadas aos componentes de produtividade. As mesmas cultivares foram genotipadas com 3566 marcadores. Os valores fenotípicos corrigidos e o conjunto de marcadores foram utilizados para a realização das análises de predição genômica por meio do método *Genomic Best Linear Unbiased Predictor*. Posteriormente, foram estimadas as capacidades preditivas, coeficientes de coincidência e, utilizou-se o índice de seleção de soma de “ranks” de Mulamba e Mock. Os valores de capacidade preditiva no tratamento sob deficit hídrico variaram entre 0.4404 a 0.8599 e no tratamento controle as capacidades preditivas variaram de 0.5766 a 0.8414. A maioria dos resultados de coincidência entre a seleção praticada com os dados fenotípicos e valores genéticos genômicos (GEBVs) foram iguais ou superiores a 50%. Com o uso do índice de Mulamba e Mock aplicados aos GEBVs, selecionou-se as cultivares 5G770 RR, P98Y70, CD 226 RR, BRS 282, CD 224, CD 208, 5D711RR, CD253, BRS MT Pintado, CD 242 RR em ambos os tratamentos. Os cultivares selecionados podem ser utilizados em programas de melhoramento genético de soja, como fonte de germoplasma, visando a tolerância ao deficit hídrico.

Palavras-chave: Fenotipagem. Deficit hídrico. Predição genômica. Índice de seleção genômico.

ABSTRACT

CAVALLIN, Isabella Cristina, M.Sc., Universidade Federal de Viçosa, march of 2020. Chapter II. **Genomic Prediction in soybean under adequate water condition and under induced drought stress in a greenhouse.** Adviser: Felipe Lopes da Silva. Co-advisers: Moyses Nascimento and Leonardo Lopes Bhering.

Drought stress affects severely soybean production. One way to increase crop production is through the development of drought stress cultivars. However, the characteristics related to the water deficit are complex. Given this, it is expected that the genomic prediction and selection index with the use of genomic genetic values will provide a tool for the selection of these characteristics. The experiment consisted of the evaluation of 91 soybean cultivars divided into two groups of 47 cultivars, each group was evaluated in two experiments. Three cultivars were used as common cultivars between the two groups. The experimental design used for each trial was in randomized blocks with three replications and in a factorial scheme with two factors. The first factor composed of cultivars and the second by the availability of water, being: the control treatment and the treatment under drought stress. The imposition of drought stress is based on the capacity of soil retention and occurred in the reproductive stage (R5) for fifteen days. At the end of the cycle, seven characteristics related to the productivity components were evaluated. The same cultivars were genotyped with 3566 markers. The phenotypic values and the set of markers were used to perform the genomic prediction analysis using the Genomic Best Linear Unbiased Predictor method. Subsequently, predictive capacities, coincidence coefficients were estimated, and the Mulamba and Mock index was used. The predictive capacity values in the treatment with drought stress ranged from 0.4404 to 0.8599 and in the control treatment, the predictive capacities ranged from 0.5766 to 0.8414. Most of the results of coincidence between the selection practiced with the phenotypic data and GEBVs were equal to or greater than 50%. Using the Mulamba and Mock index applied to GEBVs, the cultivars 5G770 RR, P98Y70, CD 226 RR, BRS 282, CD 224, CD 208, 5D711RR, CD253, BRS MT Pintado, CD 242 RR were selected in both treatments. The selected cultivars can be used in soybean breeding programs, as a source of germplasm, aiming at tolerance to drought stress.

Keywords: Phenotyping. Drought stress. Genomic prediction. Genomic selection index.

1 INTRODUÇÃO

A soja (*Glycine max*) é a leguminosa de semente mais importante do mundo e contribui significativamente para o óleo comestível, o concentrado de proteínas para alimentação animal e vários produtos industriais (LIMA et al. 2014). A soja contém 40 a 42% de proteína de boa qualidade e 18 a 22% de óleo, o que a torna ideal para a dieta animal e humana (ARSHAD et al. 2006). Devido à sua importância econômica, a cultura tem sido o foco de pesquisas, especialmente na área de melhoramento genético, para obter cultivares melhoradas, portando genes capazes de expressar ampla adaptação e tolerância a fatores bióticos e abióticos, representando contribuições significativas para o setor produtivo (SOARES et al., 2015).

Dentre estes fatores abióticos, o déficit hídrico pode afetar seriamente a produção de soja, sendo que a fase reprodutiva é a mais sensível (BHATIA e JUMRANI, 2016; JUMRANI et al., 2017). Mesmo um episódio de seca de, aproximadamente 15 dias, pode resultar em uma redução substancial do rendimento (JUMRANI et al., 2017). Por isso, o desenvolvimento de genótipos de soja com tolerância à seca é uma das melhores maneiras de aumentar a produção de soja (PENNISI 2008).

No entanto, a seleção de genótipos superiores de soja é um processo complexo, porque as características agrônomicas de importância econômica são de natureza quantitativa, sendo algumas correlacionadas entre si e de baixa herdabilidade (NOGUEIRA et al., 2012). Além disso, os genótipos promissores devem unir simultaneamente alguns atributos desejáveis, visando a altos rendimentos para atender a demanda do setor produtivo (CRUZ, 2013).

Devido à esta complexidade, espera-se que a predição genômica forneça uma ferramenta para a seleção de características relacionadas ao déficit hídrico. A predição genômica (MEUWISSEN et al., 2001) é um método estatístico que prevê valores genéticos de indivíduos com base em dados genômicos e na modelagem genótipo-fenótipo. Um modelo é ajustado usando uma população de treinamento para a qual existem dados genômicos e fenotípicos e posteriormente, o modelo pode ser aplicado em dados genômicos de indivíduos que não foram fenotipados. A predição genômica estima simultaneamente os efeitos de todos os marcadores disponíveis, sem a necessidade de seleção prévia de marcadores, assim, evita a superestimação dos efeitos de marcadores selecionados e pode explicar polimorfismos de base única com pequenos efeitos no fenótipo (HEFFNER et al., 2009). Essa abordagem tem potencial para ajudar os melhoristas em seu trabalho de seleção, particularmente para características com baixa herdabilidade ou características para as quais a fenotipagem é dispendiosa (HEFFNER et al., 2009; ENDELMAN et al., 2014; HESLOT et al., 2015).

Os valores genéticos genômicos estimados (GEBVs) permitem que os programas de melhoramento acelerem o processo de desenvolvimento de cultivares (HEFFNER et al. 2009; ENDELMAN et al. 2014). A seleção baseada em GEBVs é mais confiável do que a baseada apenas em fenótipos (NAKAYA e ISOBE 2012), e fornece maior ganho genético a longo prazo quando comparados aos valores genéticos baseados em pedigree (MUIR 2007). Os GEBVs são usados para selecionar genótipos não fenotipados (HEFFNER et al. 2009), para realizar uma seleção mais precisa de linhagens (ENDELMAN et al. 2014), incorporar germoplasma útil (CHUNG et al. 2014), monitorar a perda de diversidade genética (HENRYON et al. 2014) e selecionar combinações parentais para cruzamentos (MOHAMMADI et al. 2015). No entanto, o melhor modelo de predição varia entre características e de cultura para cultura (DE LOS CAMPOS et al. 2013). Além disso, a contribuição da densidade de genotipagem, tamanho da população e fenotipagem para a predição genômica não é clara quando aplicada a dados reais (WIMMER et al. 2013). Os estudos de predição geralmente fornecem resultados conflitantes que variam de acordo com a base genética da população em avaliação (DE LOS CAMPOS et al. 2013).

Outra maneira de melhorar a eficiência de um programa de melhoramento é selecionar um conjunto de caracteres de importância agrônômica simultaneamente (REZENDE et al., 2014; VASCONCELOS et al., 2010). Para isso, os índices de seleção associam as informações de vários caracteres e permitem fazer a seleção com eficiência, aumentando a possibilidade de sucesso de um programa (VASCONCELOS et al., 2010; ROSADO et al. al., 2012; CRUZ, 2013; VIANNA et al., 2013; LEITE et al., 2016).

O uso de índices para selecionar genótipos de soja pode promover melhores ganhos totais, distribuídos entre todas as características disponíveis, o que é adequado para programas de melhoramento genético (COSTA et al., 2004; BIZARI et al., 2017; LEITE et al., 2016). Como exemplo de índice de seleção, o proposto por Mulamba e Mock (1978) consiste em classificar os genótipos de acordo com cada característica em ordem favorável de melhoria e tem a vantagem de não precisar estabelecer importância econômica e estimativas de variância e covariância (REZENDE et al., 2014).

A aplicação de índices de seleção evidenciando a obtenção de ganhos em um conjunto de caracteres simultaneamente tem sido bastante utilizada no melhoramento de plantas (FRANÇA et al., 2016; JUNQUEIRA et al., 2016; KUMAR et al., 2016; SILVA et al., 2016; SILVA et al., 2017). Com isso, tem aumentado o interesse em estudos sobre a eficiência de aplicação desta metodologia na GWS (CERON-ROJAS et al., 2015; FERNANDES et al., 2017; HEFFNER et al., 2011).

Neste sentido, acredita-se que a identificação de estratégias de seleção, aliando a GWS e o teoria de índices de seleção, é de grande importância para os programas de melhoramento de soja visando à seleção de genótipos tolerantes ao deficit hídrico.

O presente estudo terá como objetivo identificar e selecionar genótipos tolerantes de soja, sob condição de deficit hídrico, via utilização de predição genômica e de índice de seleção genômico.

2 MATERIAL E MÉTODOS

Foram utilizados 91 cultivares de soja oriundas do Banco de Germoplasma do Programa Soja da Universidade Federal de Viçosa (Tabela 1). As cultivares foram divididas em dois grupos de 47 cultivares. O primeiro grupo de cultivares foi avaliado em dois experimentos, sendo um realizado em 2016, denominado neste trabalho de 2016-1, e um em 2017. O segundo foi avaliado nos anos de 2016 e 2018, um em cada ano, denominados de 2016-2 e 2018, respectivamente. As cultivares BMX Potência RR, CD 235 RR e MG BR Conquista pertenciam aos dois grupos avaliados, sendo consideradas como testemunhas comuns (de efeito fixo no modelo estatístico adotado).

Os experimentos foram desenvolvidos em casa de vegetação e foram utilizados como unidade experimental vasos com capacidade de 7,5 litros. O delineamento experimental utilizado para cada ensaio, foi em blocos casualizados (DBC) em esquema fatorial com dois fatores (cultivares e disponibilidade hídrica), com três repetições. O primeiro fator foi composto pelas cultivares e o segundo fator foi composto pela disponibilidade de água, sendo: o tratamento controle, solo irrigado até a capacidade de campo, e o tratamento sob déficit hídrico. A imposição das disponibilidades de água será detalhada posteriormente.

O substrato foi composto por solo e areia na proporção de 3:2, respectivamente. Após a confecção do substrato foram realizadas as análises físicas dos mesmos (Tabela 2) necessárias para a imposição da metodologia de estresse hídrico utilizada no presente trabalho. Para a correção da fertilidade do substrato utilizou-se análises químicas dos mesmos e, posteriormente, as exigências nutricionais da cultura da soja seguindo os valores de extração e exportação da cultura para produção de três toneladas por hectare, conforme demonstrado por Câmara (2015).

Antes do plantio, as sementes foram tratadas com fungicida Carbedazin + Thiram, na dose de 200 mL por 100 kg de sementes. Na semeadura foram utilizadas cinco sementes por vaso e, após a emergência das plantas, foi efetuado o desbaste sendo deixado somente uma planta por vaso.

As plantas receberam suprimento hídrico, necessário para atingir a capacidade de campo do solo, até o estágio fenológico R5.3, correspondente ao período de enchimento de grãos de acordo com a escala de Fehr a Caviness (1977). A partir desse estágio, os tratamentos com aplicação do déficit hídrico, foram induzidos durante o período de quinze dias, pela metodologia desenvolvida pelo Programa de Melhoramento da Soja da Universidade Federal de Viçosa, conforme descrito por Dantas (2018).

Quadro 1. Lista das cultivares de soja (*Glycine max* (L.) Merrill) utilizadas em cada experimento.

| 2016 -1 | 2016-2 | 2017 | 2018 |
|----------------|------------------|----------------|------------------|
| 5D660 RR | 5G770 RR | 5D660 RR | 5G770 RR |
| 5D690 RR | BMX Energia RR | 5D690 RR | BMX Energia RR |
| 5D711 RR | BMX Titan RR | 5D711 RR | BMX Titan RR |
| BMX Turbo RR | Bragg | BMX Turbo RR | Bragg |
| BRS 133 | BRS 184 | BRS 133 | BRS 184 |
| BRS 185 | BRS 213 | BRS 185 | BRS 213 |
| BRS 232 | BRS 257 | BRS 232 | BRS 257 |
| BRS 243 RR | BRS 283 | BRS 243 RR | BRS 283 |
| BRS 245 RR | BRS 284 | BRS 245 RR | BRS 284 |
| BRS 246 RR | BRS MT Pintado | BRS 246 RR | BRS MT Pintado |
| BRS 268 | CD 201 | BRS 268 | CD 201 |
| BRS 282 | CD 202 | BRS 282 | CD 202 |
| CD 206 | CD 205 | CD 206 | CD 205 |
| CD 208 | CD 215 | CD 208 | CD 215 |
| CD 213 RR | CD 217 | CD 213 RR | CD 217 |
| CD 216 | CD 225 RR | CD 216 | CD 225 RR |
| CD 224 | CD 226 RR | CD 224 | CD 226 RR |
| CD 232 | CD 229 RR | CD 232 | CD 229 RR |
| CD 233 RR | CD 237 RR | CD 233 RR | CD 237 RR |
| CD 234 RR | CD 242 RR | CD 234 RR | CD 242 RR |
| CD 236 RR | CD 246 | CD 236 RR | CD 246 |
| CD 240 RR | CD 250 RR | CD 240 RR | CD 250 RR |
| CD 243 RR | CD 253 | CD 243 RR | CD 253 |
| CD 244 RR | CD 254 RR | CD 244 RR | CD 254 RR |
| CD 248 RR | CD 2630 RR | CD 248 RR | CD 2630 RR |
| CD 2737 RR | CD 2721 RR | CD 2737 RR | CD 2721 RR |
| CD 2800 | CD 2792 RR | CD 2800 | CD 2792 RR |
| CD 2840 | CD 288 | CD 2840 | CD 288 |
| CD/FAPA 220 | CD 5807 | CD/FAPA 220 | CD 5807 |
| DMARIO 70i | CD251 RR | DMARIO 70i | CD251 RR |
| Embrapa 48 | Embrapa 59 | Embrapa 48 | Embrapa 59 |
| FT Guaíra | Emgopa 304 | FT Guaíra | Emgopa 304 |
| Fundacep 56 RR | Fundacep 33 | Fundacep 56 RR | Fundacep 33 |
| Fundacep 57 RR | Fundacep 58 | Fundacep 57 RR | Fundacep 58 |
| M 7211 RR | M 6707 RR | M 7211 RR | M 6707 RR |
| M-SOY 6101 | MG/BR 48 Garimpo | M-SOY 6101 | MG/BR 48 Garimpo |
| NA 4990 RG | MSOY 7901 | NA 4990 RG | MSOY 7901 |
| NA 5909 RG | MSOY 8001 | NA 5909 RG | MSOY 8001 |
| NK 7059 RR | Ocepar 3 | NK 7059 RR | Ocepar 3 |
| OC 14 | P98Y70 | OC 14 | P98Y70 |

| | | | |
|-----------------|-----------------|-----------------|-----------------|
| P98Y11 | TMG 1066 RR | P98Y11 | TMG 1066 RR |
| R7 | TMG 115 RR | R7 | TMG 115 RR |
| RA 518 | Tropical RR | RA 518 | Tropical RR |
| TMG 1067 RR | RA 628 | TMG 1067 RR | RA 628 |
| BMX Potência RR | BMX Potência RR | BMX Potência RR | BMX Potência RR |
| CD 235 RR | CD 235 RR | CD 235 RR | CD 235 RR |
| MG BR Conquista | MG BR Conquista | MG BR Conquista | MG BR Conquista |

Essa metodologia se baseia na determinação da capacidade de vaso pelo método da curva de retenção de água no solo. Para aplicação dessa metodologia de imposição do deficit hídrico, foi necessário a realização da análise física do solo dos experimentos conforme apresentado na Tabela 2. Em seguida, elaborou-se uma curva de retenção com base no teor de argila do solo, nas tensões -6, -10, -40, -80, -100, -400 e -1500 kPa, baseado no conteúdo de água no solo para cada tensão. Por meio desta, ajustou-se uma equação de regressão do tipo potencial, sendo utilizado um solo de textura argilosa, pelo modelo de Van Genuchten (1980) para cada experimento. Com essas informações, calculou-se o conteúdo de água no solo em capacidade de campo estabelecendo para isto os potenciais de -10 kPa, para os experimentos de 2016-1, 2017 e 2018, e -33kpa para o experimento de 2016-2. Para o conteúdo de água aplicado no deficit hídrico adotou-se para todos os experimentos o potencial de -900 kPa.

No momento do enchimento dos vasos, a massa do sistema vaso + solo foi padronizada para 6,50 kg. Dez amostras de solo foram coletadas para determinação da umidade e obtenção da massa úmida do solo (M_{su}). Para a determinação da massa do solo seco (M_{ss}) as amostras foram levadas para a estufa de circulação de ar forçada a 105 °C por 48 horas e posteriormente foram pesadas em balança.

Após a instalação dos ensaios, foi elaborado um sistema, composto pela massa de água ($M_{\text{água}}$), do solo seco (M_{ss}), do vaso (M_{vas}) e da planta (M_{planta}), sendo definida a equação da massa do sistema (M_{sist}) por meio do estimador:

$$M_{\text{sist}} = M_{\text{vas}} + M_{\text{ss}} + M_{\text{água}} + M_{\text{planta}}$$

A massa da planta (M_{planta}) foi determinada a partir da mensuração da matéria fresca de plantas no respectivo estágio fenológico. Para isso, acrescentou-se plantas adicionais às parcelas, sendo cultivadas paralelamente com as plantas que foram avaliadas.

Tabela 2. Resumo da análise física do solo dos 4 experimentos.

| Ano/Experimento | Areia | Silte | Argila | Classe textural |
|-----------------|-------------|-------|--------|-------------------------|
| | -----%----- | | | |
| 2016.1 | 50 | 15 | 35 | Franco argiloso arenoso |
| 2016.2 | 56 | 13 | 31 | Franco argiloso arenoso |
| 2017 | 66 | 4 | 30 | Franco-argiloso arenoso |
| 2018 | 58 | 8 | 34 | Franco argiloso arenoso |

(Classificação de acordo com Lemos e Santos, 1984)

A partir desses cálculos, foi confeccionada uma planilha com o peso do sistema (massa do vaso, massa da planta, massa do solo seco e massa de água), em seus respectivos tratamentos. Com base nas informações da planilha foi realizada a manutenção destes potenciais e monitorado pessoa massa a ser alcançada todos os dias através do gerenciamento de irrigação. Para isso, durante o período de imposição do deficit hídrico, uma amostra de 10 vasos (cinco vasos sem estresse e cinco com estresse) de cada bloco foram pesadas às 10 e 16 horas durante os dias de imposição do deficit hídrico. Com base na média da massa desses vasos foi aplicada uma lâmina de água necessária para a manutenção dos potenciais pré-fixados, correspondente a cada tratamento e a cada bloco. Passado o período de imposição do deficit hídrico as plantas voltaram a ser irrigadas normalmente até o final do ciclo.

No final do ciclo, estádio R8, foram avaliados as características produção de grãos por planta (g.planta^{-1})(PG), número de grãos (NG), número de vagens (NV), número de grãos por vagens (NGV), peso de cem grãos (P100G), número de vagens sem grãos (VCH) e número de vagens totais (NVT).

2.1 ANÁLISES ESTATÍSTICAS DOS FENÓTIPOS

As análises foram realizadas considerando os modelos lineares mistos (procedimento REML/BLUP), por meio do software Echidna® (GILMOUR, 2018), cujos componentes de variância são estimados pelo método da máxima verossimilhança restrita (REML) e os valores genotípicos preditos pela melhor predição linear não viesada (BLUP). Após análise dos *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC) dos modelos testados, considerou-se a utilização do modelo abaixo. Considerou-se ainda cada ambiente em separado, sendo o tratamento controle e o sob deficit hídrico, considerando o seguinte modelo:

$$y = Xb + Zg + Wge + e, \text{ em que:}$$

y , b , g , ge , e = vetores de dados, de efeitos fixos (médias de blocos através dos locais, testemunhas e interação testemunhas x experimentos), de efeitos genotípicos (aleatório), de efeitos da interação genótipos x experimentos (aleatório) e de erros aleatórios, respectivamente. X , Z e W = matrizes de incidência para b , g e ge , respectivamente.

Nesse modelo, os valores genotípicos preditos livres da interação, considerando todos os experimentos são dados por $\mu + g$, em que μ é a média de todos os locais. E os valores fenotípicos corrigidos por meio da adição da média dos erros por parcela (y^*), dado pelo estimador:

$$y^* = \mu + g + e$$

A significância dos efeitos aleatórios foi obtida através da análise de deviance, por meio do teste da razão da máxima verossimilhança (LRT) (RESENDE, 2007). As deviances foram obtidas conforme descrito por Resende (2007), utilizando-se o modelo com e sem os respectivos efeitos, subtraindo a deviance obtida no modelo completo do modelo sem o efeito e comparada com o valor do Qui-quadrado (χ^2) com um grau de liberdade.

O coeficiente de variação genética (CVg %) foi calculado por meio do seguinte estimador:

$$CVg (\%) = (\sqrt{\sigma_g^2} / \mu) * 100.$$

em que: $\sqrt{\sigma_g^2}$ é a variância genética; e , μ é média geral de cada característica.

O coeficiente de variação residual é dado por:

$$(CVe \%) \text{ por: } CVe (\%) = (\sqrt{\sigma_e^2} / \mu) * 100$$

em que: $\sqrt{\sigma_e^2}$ é variância residual.

Foram obtidas estimativas dos coeficientes de correlação pelo método de Pearson, descrito em Steel & Torrie (1960). As análises foram realizadas com o auxílio do aplicativo computacional GENES (CRUZ, 2013).

2.2 GENOTIPAGEM

A genotipagem foi realizada na Deoxi Biotecnologia Ltda, em Araçatuba/SP, utilizando-se a plataforma Illumina iScan e o painel BARCSoySNP6K BeadChip Song et al. (2014), um subconjunto da plataforma SoySNP50K BeadChip (SONG et al., 2013). A genotipagem foi realizada de acordo com protocolo descrito pela fabricante (MISEQ-SUPPORT, 2018).

Inicialmente os dados genotípicos foram filtrados considerando valores mínimos de frequência alélica no genoma (MAF) e proporção no genoma (call rate) para que os SNPs

possam ser identificados de forma confiável. Foram adotadas $MAF \geq 0,05$ e $call\ rate \leq 80$, restando 3566 SNPs após a limpeza, realizada utilizando o software R.

2.3 PREDIÇÃO GENÔMICA

Os valores fenotípicos corrigidos e o conjunto de marcadores foram utilizados para a realização das análises de predição genômica com o software RBio (BHERING, 2017). O método estatístico aplicado à predição genômica (GWS) foi o *Genomic Best Linear Unbiased Prediction* (G-BLUP).

O G-BLUP assume que todos os efeitos dos marcadores são normalmente distribuídos e têm variância igual (MEUWISSEN et al., 2001). Este método utiliza informações de marcadores genéticos para calcular associações entre indivíduos com base na matriz de relação genômica (HABIER et al., 2007). A equação de modelos mistos lineares utilizado foi:

$$Y = Xb + Z\mu a + \varepsilon$$

em que: y é o vetor de valores fenotípicos ($N \times 1$, em que N é o número de indivíduos); b é o vetor de efeitos fixos ($p \times 1$, onde p é o número de efeitos fixos considerados); μa é o vetor de efeitos genéticos aditivos dos indivíduos ($N \times 1$); ε é o vetor de resíduos do modelo, onde $\varepsilon \sim N(0, I\sigma^2_\varepsilon)$ e σ^2_ε é a variância residual. X ($N \times p$) e Z ($N \times N$) são as matrizes de incidência para b e μa , respectivamente. A estrutura de variância dada por $\mu a \sim N(0, G_a\sigma_a)$ em que σ_a é a variância aditiva e G_a ($N \times N$) é a matriz de parentesco genômica para efeitos aditivos. Sob essas configurações, a equação de modelos mistos genômicos para a predição de μa através do BLUP genômico (G-BLUP) é equivalente a:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I G - 1_a / (\frac{\sigma^2_\varepsilon}{\sigma^2_a}) \end{bmatrix} \begin{bmatrix} b \\ \widehat{\mu a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

em que σ^2_a se refere à variância genética total da característica e σ^2_ε é a variância residual.

Sem padronização a variação genética de cada locus é dada por σ^2_a/n onde n está relacionada com o número de marcadores utilizados e é dada por $n = 2 \sum_i^n p_i (1 - p_i)$ (RESENDE et al., 2008 ; GIANOLA et al., 2009), onde p_i é a frequência dos alelos dos loci i .

Para todas as metodologias de GWS avaliadas, foi realizada a validação cruzada considerando o procedimento 5-folds com 5 repetições. Desta forma, para cada característica,

a população de 88 cultivares foi dividida em cinco grupos. Assim, a populações de treinamento foi composta por 71 indivíduos e utilizadas para estimar os efeitos dos marcadores, enquanto a concordância entre os valores genéticos preditos via estimativas provenientes da população de treinamento foi validada em cada grupo de 17 indivíduos, considerados como população de validação.

Uma vez que as estimativas dos efeitos dos marcadores foram realizadas, o valor genético genômico predito (GEBV) foi definido da seguinte maneira:

$$GEBV = \hat{\mu}a$$

em que: μ é média geral; e, a é o efeito genético aditivo dos indivíduos.

A capacidade preditiva (r_{gy}) foi estimada como sendo a correlação de Pearson entre o valor genético genômico (GEBV) estimado pelos modelos e o valor fenotípico corrigido, respectivamente (y^*).

$$r_{gy} = cor(GEBV, y^*)$$

Os erros padrão associados às capacidades preditivas foram calculados através do desvio padrão dividido por \sqrt{k} , sendo $k=5$.

2.4 SELEÇÃO DE CULTIVARES VIA ÍNDICE DE SELEÇÃO

Utilizou-se a metodologia do índice soma de “ranks” (MULAMBA e MOCK 1978) via os valores fenotípicos e os GEBVs preditos para cada uma das características avaliadas, considerando os ambientes controle e o sob deficit hídrico.

O índice foi realizado no sentido de aumentar os valores de todas as características avaliadas, exceto para o número de vagens sem grãos, em que a seleção foi no sentido de reduzir a mesma.

O coeficiente de coincidência foi calculado entre as seleções praticadas via os valores fenotípicos e os GEBVs.

3 RESULTADOS

3.1 ANÁLISE DOS DADOS FENOTÍPICOS

Os efeitos de variância genotípica e variância da interação cultivares x experimentos do número de grãos (NG), número de vagens (NV), número de grãos por vagens (NGV), peso de cem grãos (P100G), número de vagens sem grãos (VCH) e número total de vagens (NVT)

foram significativos ($P < 0,01$) quando as cultivares foram avaliadas nos dois ambientes, controle e sob deficit hídrico. Para a característica produção de grãos por planta (PG) o efeito de variância genotípica não foi significativo no ambiente estresse.

Praticamente todas as características apresentaram redução nos valores das médias (μ) no ambiente com indução do deficit hídrico em relação ao controle, exceto para a variável número de vagens sem grãos, que foi relativamente o dobro no ambiente sob deficit hídrico (Tabela 3).

Os valores de herdabilidade (h^2) variaram de 0,049 a 0,668 no ambiente de deficit hídrico e variaram de 0,153 a 0,659 no controle. A característica com menor herdabilidade foi a produção de grãos por planta ($g.planta^{-1}$) e maior herdabilidade foi o número de vagens total, em ambos os ambientes. A acurácia seletiva (r_{yy}) apresentou valores de 0,223 a 0,8174 no ambiente sob deficit hídrico e variou de 0,3919 a 0,8122 no controle, para as características produção de grãos por planta ($g.planta^{-1}$) e número de vagens total, respectivamente.

Quanto a razão entre o coeficiente de variação genético e ambiental (CVg/Cve), os valores mais próximos a unidade indicam que a seleção para as características número de vagens totais (0,8581), número de vagens (0,8435) e número de grãos por vagem (0,7516) apresentam condições mais favoráveis em termos de ganhos genéticos imediatos para o ambiente sob deficit hídrico.

Para o ambiente controle, as características que apresentaram valores de razão (CVg/Cve) mais próxima a unidade foram semelhantes, número de vagens totais (0,8404), número de vagens (0,8341) e número de grãos por vagem (0,8188).

Tabela 3. Média (μ), herdabilidade (h^2), acurácia seletiva (r_{yy}), razão entre o coeficiente de variação genético e coeficiente de variação ambiental (CVg/CVe), variância genética entre cultivares (σ_g^2), variância da interação cultivar por ambiente (σ_{ga}^2) e variância residual (σ_e^2) dos componentes de produção: produção de grãos por planta ($g.planta^{-1}$)(PG), número de grãos (NG), número de vagens (NV), número de grãos por vagens (NGV), peso de cem grãos (P100G), número de vagens sem grãos (VCH) e número de vagens totais (NVT) nos ambientes deficit hídrico e controle.

| | DEFICIT HÍDRICO | | | | | | |
|-------|-----------------|--------|----------|---------------|--------------|--------------|-----------------|
| | μ | h^2 | r_{yy} | CVg/CVe (%) | σ_e^2 | σ_g^2 | σ_{ga}^2 |
| PG | 19,59 | 0,0498 | 0,2232 | 0,1405 | 5,7923 | 0,1400 | 3,0016 |
| P100G | 13,30 | 0,2008 | 0,4481 | 0,2961 | 21,182 | 1,8574 | 3,9735 |
| NG | 142,7 | 0,6048 | 0,7777 | 0,7516 | 454,94 | 257,03 | 195,57 |
| NV | 68,85 | 0,6607 | 0,8128 | 0,8435 | 94,277 | 67,071 | 36,292 |
| NGV | 2,068 | 0,4680 | 0,6841 | 0,5478 | 0,0993 | 0,0298 | 0,0092 |
| VCH | 6,270 | 0,2653 | 0,5151 | 0,3545 | 28,358 | 3,5643 | 4,9914 |
| NVT | 74,38 | 0,6681 | 0,8174 | 0,8581 | 44,373 | 83,940 | 44,373 |

| CONTROLE | | | | | | | |
|----------|-------|--------|----------|-------------|--------------|--------------|-----------------|
| | μ | h^2 | r_{yy} | CVg/CVe (%) | σ_e^2 | σ_g^2 | σ_{ga}^2 |
| PG | 26,38 | 0,1535 | 0,3918 | 0,2506 | 10,777 | 0,6768 | 1,6707 |
| P100G | 13,93 | 0,4794 | 0,6924 | 0,5906 | 5,1969 | 1,8129 | 2,8402 |
| NG | 183,9 | 0,6429 | 0,8019 | 0,8188 | 627,92 | 421,08 | 293,99 |
| NV | 85,25 | 0,6557 | 0,8097 | 0,8341 | 148,83 | 103,56 | 57,298 |
| NGV | 2,155 | 0,4732 | 0,6879 | 0,5660 | 0,1189 | 0,0381 | 0,0333 |
| VCH | 3,100 | 0,2092 | 0,4573 | 0,6994 | 11,048 | 1,0199 | 2,0849 |
| NVT | 88,35 | 0,6598 | 0,8122 | 0,8407 | 153,20 | 108,29 | 57,392 |

Os valores das correlações entre as características avaliadas em condição de deficit hídrico encontram-se na Tabela 4. As correlações foram altas e positivas entre o número de vagens e número total de vagens (0,898), número de vagens e número de grãos (0,749), número total de vagens e número de grãos (0,709), número de vagens sem grãos e número total de vagens (0,507) e número de grãos por vagem e número de grãos (0,422). Já para os caracteres número de grãos e peso de 100 grãos (-0,819) e número de vagens e peso de 100 grãos (-0,627), os valores foram altos e negativos.

Considerando as características avaliadas no ambiente controle, observou-se correlações altas e positivas entre o número de vagens e número total de vagens (0,881), número de vagens e número de grãos (0,757), número total de vagens e número de grãos (0,556), número de grãos por vagem e número de grãos (0,416), além de produção de grãos por planta (g.planta^{-1}) e peso de 100 grãos (0,426) e produção de grãos por planta e número de grãos por vagem (0,327) (Tabela 4). Apresentou correlação alta e negativa entre os caracteres número de grãos e peso de 100 grãos (-0,843), número de vagens e peso de 100 grãos (-0,661) e número total de vagens e peso de 100 grãos (-0,437).

Tabela 4. Correlação entre produção de grãos por planta (g.planta^{-1}) (PG), número de grãos (NG), número de vagens (NV), número de grãos por vagens (NGV), peso de 100 grãos (P100G), número de vagens sem grãos (VCH) e número de vagens totais (NVT), avaliados em cultivares de soja nos ambientes sob deficit hídrico (acima da diagonal) e controle (abaixo da diagonal).

| | PG | NG | NV | NGV | P100G | VCH | NVT |
|-------|---------|----------|----------|----------|----------|---------|----------|
| PG | 1 | 0,296** | 0,062 | 0,158 | 0,1129 | -0,011 | 0,109 |
| NG | 0,313** | 1 | 0,749** | 0,422** | -0,819** | 0,146 | 0,709** |
| NV | 0,211* | 0,757** | 1 | -0,270* | -0,627** | 0,249* | 0,898** |
| NGV | 0,327** | 0,416** | -0,111 | 1 | -0,323** | -0,129 | -0,249* |
| P100G | 0,426** | -0,843** | -0,661** | -0,253* | 1 | -0,218* | -0,332** |
| VCH | -0,074 | 0,066 | 0,123 | -0,123 | -0,202 | 1 | 0,507** |
| NVT | 0,134 | 0,556** | 0,881** | -0,305** | -0,437** | 0,245* | 1 |

*5% de significância, ** 1% de significância

3.2 CAPACIDADE PREDITIVA E COINCIDÊNCIA DE SELEÇÃO

Os valores de capacidade preditiva no ambiente sob deficit hídrico variaram entre 0,4404 para a característica peso de 100 grãos a 0,8599 para número de vagens totais, conforme apresentado na Tabela 5. Já para o ambiente controle, as capacidades preditivas variaram de 0,5766 a 0,8414, para as características peso de grãos por planta e número de vagens, respectivamente.

Os erros padrão, no ambiente sob deficit hídrico, apresentaram variação de 0,0352 a 0,0516 para as características número de vagens sem grãos e peso de grãos, respectivamente. No controle, os erros foram de 0,0287 para o peso de 100 grãos até 0,0483 no peso de grãos por planta.

No ambiente sob deficit hídrico, as variáveis que apresentaram valores de viés mais próximos à unidade foram o peso de 100 grãos, seguido do peso de grãos por planta. Da mesma forma, no tratamento controle, os menores valores de viés foram para o peso de 100 grãos, seguido do peso de grãos, sendo de 1,48 e 1,59, respectivamente. Em relação aos valores de viés que mais se distanciaram da unidade foram para o número de grãos por vagem no estresse e número de vagens sem grãos no controle.

Tabela 5. Capacidade preditiva (r_{gy}), viés e erro padrão (sd) das características peso de grãos por planta (PG), número de grãos por planta (NG), número de grãos por vagem (NGV), número de vagens com grãos (NV), peso de 100 grãos (P100G), número de vagens sem grãos (VCH) e número de vagens totais (NVT) avaliadas em cultivares de soja nos ambientes sob deficit hídrico e controle.

| Característica | ESTRESSE | | | CONTROLE | | |
|----------------|----------|------|--------|----------|------|--------|
| | r_{gy} | Viés | sd | r_{gy} | viés | Sd |
| PG | 0,5366 | 1,25 | 0,0516 | 0,5766 | 1,59 | 0,0483 |
| NG | 0,7287 | 2,56 | 0,0422 | 0,7900 | 1,90 | 0,0456 |
| NGV | 0,6027 | 4,24 | 0,0483 | 0,6572 | 3,08 | 0,0466 |
| NV | 0,8554 | 1,63 | 0,0413 | 0,8414 | 1,64 | 0,0319 |
| P100G | 0,4404 | 1,19 | 0,0447 | 0,6137 | 1,48 | 0,0287 |
| VCH | 0,5431 | 2,38 | 0,0352 | 0,6049 | 3,21 | 0,0445 |
| NVT | 0,8599 | 1,57 | 0,0374 | 0,8239 | 2,07 | 0,0367 |

3.3 INDÍCE DE SELEÇÃO

Através dos resultados obtidos pela soma de ranks de Mulamba e Mock (Tabela 6), podemos observar que as dez melhores cultivares selecionadas com base no GEBV, quando avaliadas no ambiente sob estresse hídrico, foram: 5G770 RR, CD 233 RR, CD 224, BRS 282,

CD 208, 5D690 RR, CD 226 RR, CD 244 RR, P98Y70 e 5D660 RR. Considerando a seleção com base nos valores fenotípicos as cultivares selecionadas foram: 5G770 RR, CD 233 RR, CD 224, 5D690 RR, CD 244 RR, 5D660 RR, BRS 213, CD 242 RR, MSOY 7901 e FUNDACEP 58. O índice de coincidência de seleção foi de 60%. Quando aplicou-se a seleção de 20 melhores cultivares o índice de coincidência aumentou para 65%.

As dez melhores cultivares selecionadas com base no GEBV, quando avaliadas no ambiente controle, foram: CD 251 RR, CD 254 RR, 5G770 RR, CD 242 RR, CD 2792 RR, BRS MT Pintado, P98Y70, CD 2840, CD 2800 e CD 226 RR. Considerando a seleção com base nos valores fenotípicos as cultivares selecionadas foram: 5G770 RR, CD 252 RR, CD 242 RR, BRS 282, CD 288, P98Y11, 5D690 RR, TMG 1067 RR, CD 226 RR, CD 213 RR. O índice de coincidência de seleção foi de apenas 40%. Quando aplicou-se a seleção de 20 melhores cultivares o índice de coincidência aumentou para 65%.

Tabela 6. Índice de seleção de soma de rank Mulamba e Mock nos diferentes cenários: déficit hídrico e controle, considerando os GEBVs e os fenótipos das sete características avaliadas.

| Cultivares | GEBV | | FENÓTIPO | |
|--------------|--------------------|------------------|--------------------|--------------------|
| | Deficit hídrico | Controle | Deficit hídrico | Controle |
| 5D660 RR | 210 ^{1,2} | 329 | 228 ^{1,2} | 287 |
| 5D690 RR | 148 ^{1,2} | 259 | 200 ^{1,2} | 202 ^{1,2} |
| 5D711 RR | 226 ² | 249 ² | 245 ² | 248 |
| BMX TURBO RR | 311 | 357 | 263 | 349 |
| BRS 133 | 395 | 324 | 345 | 256 |
| BRS 185 | 481 | 465 | 421 | 397 |
| BRS 232 | 396 | 354 | 394 | 424 |
| BRS 243 RR | 480 | 318 | 361 | 285 |
| BRS 245 RR | 426 | 369 | 383 | 332 |
| BRS 246 RR | 439 | 388 | 247 ² | 376 |
| BRS 268 | 413 | 382 | 384 | 348 |
| BRS 282 | 190 ^{1,2} | 217 ² | 251 ² | 184 ^{1,2} |
| CD 206 | 428 | 407 | 310 | 434 |
| CD 208 | 190 ^{1,2} | 237 ² | 263 | 274 |
| CD 213 RR | 277 | 276 | 282 | 223 ^{1,2} |
| CD 216 | 371 | 417 | 386 | 386 |
| CD 224 | 181 ^{1,2} | 237 ² | 176 ^{1,2} | 229 ² |
| CD 232 | 453 | 332 | 394 | 299 |
| CD 233 RR | 132 ^{1,2} | 279 | 162 ^{1,2} | 242 ² |
| CD 234 RR | 413 | 416 | 353 | 355 |
| CD 236 RR | 241 ² | 394 | 365 | 367 |
| CD 240 RR | 266 | 269 | 396 | 282 |
| CD 243 RR | 296 | 347 | 291 | 385 |

continua...

continuação...

| Cultivares | GEBV | | FENÓTIPO | |
|----------------|--------------------|--------------------|--------------------|--------------------|
| | Deficit hídrico | Controle | Deficit hídrico | Controle |
| CD 244 RR | 175 ^{1,2} | 300 | 169 ^{1,2} | 337 |
| CD 248 RR | 335 | 325 | 369 | 301 |
| CD 2737 RR | 277 | 301 | 258 ² | 300 |
| CD 2800 | 263 | 195 ^{1,2} | 264 | 238 ² |
| CD 2840 | 250 | 185 ^{1,2} | 293 | 284 |
| CD/FAPA 220 | 312 | 297 | 234 ² | 294 |
| DMARIO 70i | 374 | 346 | 433 | 363 |
| EMBRAPA 48 | 253 | 271 | 276 | 335 |
| FT GUAÍRA | 442 | 471 | 426 | 443 |
| FUNDACEP 56 RR | 290 | 354 | 389 | 380 |
| FUNDACEP 57 RR | 399 | 334 | 322 | 277 |
| M 7211 RR | 332 | 363 | 293 | 360 |
| MA 4990 RG | 333 | 389 | 349 | 410 |
| M-SOY 6101 | 402 | 322 | 258 ² | 235 ² |
| NA 5909 RG | 319 | 352 | 326 | 343 |
| NK 7059 RR | 351 | 321 | 421 | 358 |
| OC 14 | 318 | 320 | 322 | 304 |
| P98Y11 | 266 | 226 ² | 290 | 198 ^{1,2} |
| R7 | 293 | 261 | 272 | 260 |
| RA 518 | 358 | 270 | 350 | 253 |
| RA 628 | 310 | 359 | 379 | 438 |
| TMG 1067 RR | 301 | 364 | 308 | 219 ^{1,2} |
| 5G770 RR | 178 ^{1,2} | 153 ^{1,2} | 87 ^{1,2} | 149 ^{1,2} |
| BMX Energia RR | 375 | 389 | 381 | 381 |
| BMX Titan RR | 397 | 370 | 387 | 378 |
| BRAGG | 299 | 253 ² | 286 | 322 |
| BRS 184 | 458 | 427 | 394 | 382 |
| BRS 213 | 217 ² | 349 | 149 ^{1,2} | 335 |
| BRS 257 | 222 ² | 292 | 327 | 370 |
| BRS 283 | 412 | 508 | 447 | 489 |
| BRS 284 | 381 | 472 | 437 | 487 |
| BRS MT Pintado | 243 ² | 179 ^{1,2} | 282 | 225 ² |
| CD 201 | 220 ² | 278 | 282 | 328 |
| CD 202 | 248 | 383 | 331 | 494 |
| CD 205 | 273 | 270 | 326 | 319 |
| CD 215 | 271 | 452 | 275 | 440 |
| CD 217 | 395 | 281 | 457 | 248 |
| CD 225 RR | 307 | 427 | 299 | 415 |
| CD 226 RR | 205 ^{1,2} | 195 ^{1,2} | 344 | 219 ^{1,2} |
| CD 229 RR | 303 | 333 | 301 | 358 |
| CD 237 RR | 378 | 285 | 269 | 340 |
| CD 242 RR | 246 ² | 153 ^{1,2} | 231 ^{1,2} | 170 ^{1,2} |

continua....

continuação...

| Cultivares | GEBV | | FENÓTIPO | |
|------------------|--------------------|--------------------|--------------------|--------------------|
| | Deficit hídrico | Controle | Deficit hídrico | Controle |
| CD 246 | 312 | 209 ² | 304 | 235 ² |
| CD 250 RR | 345 | 418 | 363 | 335 |
| CD 253 | 238 ² | 221 ² | 247 ² | 235 ² |
| CD 254 RR | 277 | 153 ^{1,2} | 315 | 238 ² |
| CD 2630 RR | 243 ² | 353 | 262 ² | 288 |
| CD 2721 RR | 292 | 319 | 266 | 278 |
| CD 2792 RR | 297 | 175 ^{1,2} | 380 | 329 |
| CD 288 | 321 | 255 | 277 | 187 ^{1,2} |
| CD 5807 | 349 | 318 | 388 | 282 |
| CD251 RR | 311 | 150 ^{1,2} | 317 | 151 ^{1,2} |
| EMBRAPA 59 | 477 | 355 | 296 | 360 |
| EMGOPA 304 | 285 | 305 | 307 | 278 |
| FUNDACEP 33 | 281 | 259 | 372 | 265 |
| FUNDACEP 58 | 260 | 288 | 212 ^{1,2} | 319 |
| M 6707 RR | 329 | 257 | 375 | 232 ² |
| MG/BR 48 GARIMPO | 280 | 277 | 325 | 295 |
| MSOY 7901 | 241 ² | 272 | 218 ^{1,2} | 318 |
| MSOY 8001 | 312 | 225 ² | 262 ² | 253 |
| OCEPAR 3 | 445 | 416 | 367 | 423 |
| P98Y70 | 203 ^{1,2} | 180 ^{1,2} | 231 ² | 227 ² |
| TMG 1066 RR | 309 | 369 | 351 | 290 |
| TMG 115 RR | 254 | 250 ² | 264 | 305 |
| Tropical RR | 407 | 391 | 320 | 346 |

¹ selecionada entre as 10 melhores cultivares e ² selecionada entre a décima primeira e vigésima posição do rank para cada ambiente.

Considerando a seleção via GEBV das 20 melhores cultivares, verificou-se que as cultivares 5G770 RR, P98Y70, CD 226 RR, BRS 282, CD 224, CD 208, 5D711RR, CD253, BRS MT Pintado, CD 242 RR e foram selecionados em ambos os ambientes, fazendo com que estas sejam consideradas tolerantes ao deficit hídrico e responsivas ao aumento da disponibilidade de água. Com destaque para as cultivares 5G770 RR, P98Y70, CD 226 RR ranqueadas entre as 10 melhores em ambos os ambientes (controle e deficit hidrico).

Da mesma forma, considerando a seleção via valores fenotípicos, as cultivares selecionadas em ambos os ambientes foram: 5G770 RR, 5D690 RR, CD 242 RR, BRS 282, CD 224, CD 233 RR, CD 253, M-SOY 6101 e P98Y70. Sendo as cultivares 5G770 RR, 5D690 RR, CD 242 RR ranqueadas entre as 10 melhores nos ambientes controle e deficit hídrico simultaneamente.

4 DISCUSSÃO

A maioria dos caracteres avaliados teve uma redução na média quando submetidos a condições de baixa disponibilidade de água (deficit hídrico) quando comparados as condições de boa disponibilidade hídrica (controle), atingindo queda de 25,73% na produção de grãos por planta, 4,5% no peso de 100 grãos, 22,40% no número de grãos por planta, 19,24% no número de vagens por planta, 4,03% no número de grãos por vagem, 15,81% no número total de vagens e um aumento de 50,55% no número de vagens sem grãos. Estes resultados corroboraram com os resultados obtidos por Anjum et al. (2011), estes autores observaram diminuições do número de vagens por planta (18,22%), número de grãos por planta (14,81%), número de grãos por vagem (3,41%), peso de 100 sementes (13,16%) e produtividade de grãos por planta (21,88%) sob deficit hídrico comparado ao controle e atribuíram a redução de rendimento ao aumento do agente oxidante malondialdeído (MDA) e menor nível de água nos tecidos causando a inativação de enzimas envolvidas na carboxilação.

Jumrani e Bhatia (2018) observaram perdas significativas de rendimento entre os tratamentos de estresse hídrico quando impostos nos estádios vegetativo (V4) e reprodutivo (R5) em comparação às plantas não estressadas. O rendimento médio em plantas não estressadas foi de 11,3 gramas por planta e foi reduzido em 28 e 74% quando as plantas foram submetidas a estresse hídrico no estágio vegetativo e reprodutivo, respectivamente. Os mesmos autores relataram redução de 39% do número de vagens/planta e 62% no número de sementes/planta no estágio reprodutivo.

Muitas pesquisas relataram que as vagens por planta foram reduzidas com deficit hídrico (DEMIRTA et al., 2010; MAK et al., 2014; DESCLAUX, ROUMET, 2000; LI et al., 2013). Rotundo e Westgate, 2009; Demirta et al., 2010; Li et al., 2013; He et al., 2016 registraram um menor número de sementes por planta sob deficit hídrico durante o período de produção de sementes e, finalmente, reduz o rendimento final (KOBRAAE et al., 2011). Um estudo mais recente observou uma diminuição significativa no rendimento de sementes (73-82%) por planta na soja sob deficit hídrico (WEI et al., 2018). Redução de 42% no rendimento na fase de enchimento de grãos (MALEKI et al., 2013) e 46-71% na fase reprodutiva (SAMARAH et al., 2006).

As correlações positivas que ocorreram no tratamento controle e não ocorreram no tratamento estresse, peso de 100 grãos com produção de grãos por planta e produção de grãos por planta com número de grãos por vagem, demonstram que no controle os grãos tendem a ser maiores, ou seja, há um aumento do peso de 100 grãos quando ocorre um aumento na produção

de grãos por planta, bem como ocorre o aumento da produção de grãos por planta quando ocorre o aumento do número de grãos por vagem.

Em condições de deficit hídrico, com o aumento do número de grãos por vagem há uma redução no peso dos grãos. Dornbos e Mullen (1991) relataram que sob seca severa, a proporção de sementes de diâmetro maior que 4,8 mm foi reduzida em 30% a 40%, enquanto a proporção de sementes de diâmetro menor que 3,2 mm foi aumentada de 3% a 15%. Segundo Jumrani e Bhatia (2018) sob seca, a rápida senescência das folhas e a reduzida fotossíntese obstruem o suprimento de assimilados disponíveis para a semente, afetando o tamanho da semente. Ghassemi-Golezani et al. (2012) demonstraram que a qualidade das sementes das cultivares de soja foi significativamente reduzida pelo estresse hídrico durante as fases reprodutivas, principalmente durante a fase de enchimento de sementes.

Em relação à correlação entre o número total de vagens e o número de vagens sem sementes em ambiente sob deficit hídrico, pode-se inferir que condições estressantes de cultivo podem afetar o desenvolvimento das sementes e reduzir o número de sementes em uma vagem. Jumrani e Bhatia (2018) observaram um maior número de vagens sem sementes quando submetidas ao estresse hídrico no estágio reprodutivo. Da mesma forma, Wei et al. (2018) observaram um aumento de até 23,8% no número de vagens vazias.

Visto que as características relacionadas com o deficit hídrico apresentam controle genético complexo, a predição genômica aparece ultimamente com um grande potencial para o estudo e entendimento destas características, (KUJUR et al., 2013), bem como para auxiliar na seleção de genótipos tolerantes.

As capacidades preditivas obtidas pela correlação entre os dados fenotípicos e os GEBVs foram de magnitude moderada a alta (RESENDE e DUARTE 2007), para todas as características avaliadas em ambos os ambientes. Os resultados corroboram com os obtidos por Heffner et al. (2011), demonstrando uma acurácia média de 0,52 com uma população de treinamento de 96 indivíduos de trigo através do método de predição genômica RR-BLUP. Ou seja, os resultados das análises de GWS foram promissores, indicando que mesmo características com herdabilidades mais baixas, como é o caso do peso de grãos por planta e peso de 100 grãos, avaliadas em uma população de treinamento pequena (88 cultivares), ainda fornecem precisão razoável, acima dos valores de herdabilidades das mesmas. Muir (2007) apontou que características com baixa herdabilidade têm maior potencial para serem exploradas pela GWS.

Quanto à densidade de marcadores, considerou-se que os 3566 marcadores promoveram boa cobertura do genoma da soja. Maior densidade de genotipagem nem sempre

umentam a precisão (VANRADEN et al. 2011). De acordo com Xavier et al (2016) 1020 marcadores são suficientes para fornecer uma predição consistente. Esse resultado provavelmente está associado às propriedades genômicas da soja, como a existência de grandes blocos de desequilíbrio de ligação e (HYTEN et al. 2007) e a distribuição desigual de SNPs no genoma da soja (LI et al. 2014). A limitação imposta pelo pequeno tamanho das populações de treinamento pode ter sido reduzida pela boa quantidade de marcadores disponíveis, como demonstrado em estudos com dados simulados obtidos por Resende (2008). Diante disso, demonstra-se o potencial na utilização de GWS em maximizar os ganhos com seleção por unidade de tempo e diminuir a necessidade de fenotipagem de características de difícil mensuração em seleções posteriores.

Por se tratar de seleção de cultivares tolerantes ao deficit hídrico, necessita-se da seleção simultânea de características relacionadas ao deficit hídrico juntamente com a seleção para produtividade. Para a realização da seleção simultânea com eficiência, os índices de seleção são uma alternativa que permite a seleção combinada de vários caracteres (ROSADO et al., 2012; CRUZ, 2013). Com isso, o índice de soma de “ranks” mostrou-se uma alternativa para a seleção simultânea de caracteres através das altas coincidências de ranqueamento obtidas com a utilização do índice nos dados fenotípicos e nos GEBVs. Gomes et al. (2018) observaram um coeficiente de coincidência de 62% comparando a seleção através do índice de Mulamba e Mock com a seleção através da metodologia REML (Método de Máxima Verossimilhança Restrita) / BLUP (Melhor Preditor Linear Não Viesado), para as características número de vagens e produtividade de grãos em feijão-vagem.

Segundo Leite et al. (2016) a utilização o índice baseado na soma de “ranks” em linhagens de soja proporcionou maiores valores de ganhos totais, cujas maiores estimativas foram obtidas quando todos os caracteres (produtividade, número de vagens e número de nós) foram considerados como principais, em virtudes da aquisição de ganhos mais distribuídos em todos os caracteres avaliados. Da mesma forma, Bizari et al. (2017) concluiu que o índice de Mulamba e Mock proporcionou melhores ganhos em seu estudo do que os vários outros índices aplicados à população segregante de soja utilizando produtividade e o valor agrônômico da soja como principais estimadores.

Resultados semelhantes foram obtidos por Teixeira et al. (2017) que observaram ganho total de 49,25% com a utilização do índice de soma de “ranks”, mesmo não ter apresentado os maiores ganhos com a seleção para a produtividade, os autores demonstraram que o índice forneceu uma distribuição equilibrada dos ganhos de seleção para as outras características. Val

et al. (2017) observaram ganhos expressivos quando utilizado o índice de soma de “ranks” para dez características avaliadas.

Na seleção de progênes superiores em milho pipoca alguns autores observaram maiores e mais apropriados ganhos usando o índice Mulamba e Mock (1978) para capacidade de expansão e rendimento de grãos (SANTOS et al., 2007; AMARAL JUNIOR et al., 2010; RANGEL et al., 2011). Rosado et al. (2012) verificaram que o índice de Mulamba e Mock foi mais adequado, fornecendo uma distribuição equilibrada dos ganhos de seleção por maior número de progênes de maracujazeiro-amarelo avaliadas em Minas Gerais.

Diante das citações de utilização do índice e do sucesso proporcionado, justifica-se a sua utilização no presente trabalho. Contudo, acredita-se que o emprego de valores genômicos preditos (GEBVs) favoreça a maior eficiência da seleção.

A utilização dos GEBVs nos índices de seleção pode ser eficiente na busca por maiores ganhos com a seleção em vários caracteres de interesse simultaneamente, a partir da seleção direta baseada em informações genótípicas correlacionadas aos fenótipos dos indivíduos. Ceron-Rojas et al. (2015) verificaram que o índice baseado em GEBVs foi menos eficiente que o índice fenotípico quando o número de marcadores é pequeno, mas se o número de marcadores é suficiente, as respostas do índice fenotípico e índice genômico são muito semelhantes, corroborando com os nossos resultados, onde foram obtidas altas porcentagens de coincidências nos ranqueamentos entre os genótipos selecionados utilizando o índice de seleção para os GEBVs e valores fenotípicos.

Dekkers (2007) propôs um índice de seleção que combina GEBVs com informações fenotípicas, os autores verificaram que a resposta de seleção estimada foi maior em relação a situação em que apenas informações fenotípicas foram utilizadas. Dekkers (2007) ao aplicar a teoria de índices na GWS, verificou aumentos significativos nos ganhos, especialmente para características de baixas herdabilidades. Heffner et al. (2011) verificaram maior acurácia na utilização de índices de seleção quando se utiliza informações genômicas do que apenas informações fenotípicas na avaliação de onze caracteres agrônômicos em uma população de melhoramento de trigo.

Schulthess et al. (2016) combinaram diferentes índices de seleção com a GWS para melhoria de rendimento de grãos e teor de proteína em centeio e concluíram que, em geral, para informações fenotípicas é recomendado aplicar a GWS considerando uma única característica estimada com base em índices de seleção, sendo este método uma forma simples, direta e eficiente de predição. Assim, é de grande importância o incentivo para novas investigações a

partir de dados reais ou simulados que possam determinar as estratégias mais viáveis de aplicação de índices na GWS (HEFFNER et al., 2011; SCHULTHESS et al., 2016).

Assim, a seleção baseada na predição genômica para o deficit hídrico em soja permite a aquisição de genótipos com características mais favoráveis para a tolerância ao deficit hídrico e maior produtividade de grãos nessa condição e, pode permitir a seleção nos próximos ciclos sem a necessidade da fenotipagem cara e dispendiosa, com a utilização das informações dos marcadores. Com isso, os genótipos selecionados podem ser usados em novas hibridizações para formar populações segregantes com o objetivo de unir, em um indivíduo, valores maiores de componentes de rendimento quando submetidos a condições de deficit hídrico e em condições com boa disponibilidade hídrica.

A soja é uma cultura de importância mundial que vem demonstrando taxas limitadas de ganhos genéticos. O uso de informações genômicas por meio de modelos de predição é uma solução possível para o aprimoramento deste ganho genético. Neste estudo, mostramos como a predição genômica atua em características complexas da soja em cenários relacionados ao deficit hídrico e estes favoreceram a seleção de cultivares de soja com perfil de tolerância, além de responsividade ao aumento da disponibilidade hídrica. Acreditamos que este trabalho inicial com predição genômica no melhoramento para tolerância ao deficit hídrico fornecerá importante visão para alocação de recursos, planejamento e tomada de decisão nos programas que visam otimizar ganhos genéticos por meio da seleção genômica. Assim, estudos adicionais são necessários para avaliar a predição genômica nas próximas gerações com a utilização de dados genômicos e verificar a eficiência em relação ao tempo, custo e mão de obra.

5 REFERÊNCIAS

- Amaral Júnior AT, Freitas Júnior SP, Rangel R.M, Pena GF, Ribeiro RM, Morais RC, Schuelter AR (2010) Improvement of a popcorn population using selection indexes from a fourth cycle of recurrent selection program carried out in two different environments. *Genet. Mol. Res* 9:340-347.
- Anjum SA, Wang L, Farooq M, Khan I, Xue L (2011) Methyl jasmonate-induced alteration in lipid peroxidation, antioxidative defence system and yield in soybean under drought. *Journal of Agronomy and Crop Science* 197:296–301.
- Arshad M, Ali N, Ghafoor A (2006) Character correlation and path coefficient in soybean *Glycine max* (L.) Merrill. *Pakistan Journal of Botany* 38:121.
- Bhatia VS, Jumrani K (2016) A maximin-minimax approach for classifying soybean genotypes for drought tolerance based on yield potential and loss. *Plant Breed.* 136:691–700.
- Bhering LL (2017) Rbio: A Tool For Biometric And Statistical Analysis Using The R Platform. *Crop Breeding and Applied Biotechnology* 17:187-190.
- Bizari EH, Val BHP, Pereira EDM, Mauro AOD, Unêda-Trevisoli SH (2017) Selection indices for agronomic traits in segregating populations of soybean. *Revista Ciência Agronômica* 48:110-117.
- Câmara TMM, Bento DAV, Alves GF, Santos MF, Moreira JUV, De Souza CL (2007) Parâmetros genéticos de caracteres relacionados à tolerância à deficiência hídrica em milho tropical. *Bragantia* 66:595-603.
- Ceron-Rojas JJ, Crossa J, Arief VN, Basford K, Rutkoski J, Jarquín D, Alvarado G, Beyene Y, Semagn K, Delacy I (2015) A genomic selection index applied to simulated and real data. *G3: Genes, Genomes, Genetics* 5:2155-2164.
- Chung WH, Jeong N, Kim J, Lee WK, Lee YG, et al (2014) Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. *DNA Res.* 21:153–167.
- Costa MM, Di Mauro AO, Unêda-Trevisoli SH, Arriel, NHC, Bárbaro IM, Muniz FRS (2004) Ganho genético por diferentes critérios de seleção em populações segregantes de soja. *Pesquisa Agropecuária Brasileira* 39:1095-1102.
- Cruz CD (2013a) Genes: a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum* 35:271-276.
- Cruz CD, Salgado CC, Bhering LL (2013) *Genômica aplicada*. Suprema, Visconde do Rio Branco.
- Dantas SAG (2018) Estratégia de indução de deficit hídrico em soja e desenvolvimento de GWAS para germinação e vigor de sementes. Tese de Doutorado (Doutorado em Fitotecnia) – Departamento de Fitotecnia, Universidade Federal de Viçosa, Viçosa, Viçosa, MG.

- De los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327–45.
- Dekkers J (2007) Prediction of response to marker- assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics* 124:331-341.
- Demirta C, Yazgan S, Candogan BN, Sincik M (2010) Quality and yield response of soybean (*Glycine max* L. Merrill) to drought stress in sub-humid environment. *Afr. J. Biotechnol.* 9:6873–6881
- Desclaux D, Roumet P (2000) Identification of soybean plant characteristics that indicate the timing of drought stress. *Crop Sci.* 40:716–722.
- Dornbos DL, Mullen RE (1991) Influence of stress during soybean seed fill on seed weight, germination, and seedling growth rate. *Canadian Journal of Plant Science* 71:373-383.
- Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells ME, Jannink JL (2014) Optimal design of preliminary yield trials with genome-wide markers. *Crop Science* 54:48-59.
- Fehr WR, Caviness CE (1977) Stages of soybean development. Special Report, Agriculture and Home Economics Experiment Station 80:11.
- Fernandes SB, Dias KO, Ferreira DF, Brown PJ (2017) Efficiency of multitrait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theoretical and Applied Genetics* 1-9.
- França ÁEDD, Parrella RADC, Souza VFD, Bastos GQ, Nunes JAR, Schaffert RE (2016) Simultaneous selection in sweet sorghum progenies by selection indices. *Pesquisa Agropecuária Brasileira* 51:1737-1743.
- Ghassemi-Golezani K, Lotfi R, Norouzi M (2012) Seed quality of soybean cultivars affected by pod position and water stress at reproductive stages. *International Journal of Plant, Animal and Environmental Sciences* 2:119-125.
- Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R (2009) Additive genetic variability and the Bayesian alphabet. *Genetics* 183:347-363.
- Gilmour AR (2018) Echidna Mixed Models Software. Disponível em: <www.echidnamms.org>. Acesso em 22 de agosto de 2019.
- Gomes A, Oliveira TR, Cruz DP, Gravina, GA, Daher RF, Araújo LC, Araújo KC (2018) Genetic gain via REML/BLUP and selection indices in snap bean. *Horticultura Brasileira* 36:195-198.
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397.

- He J, Du Y, Wang T, Turner NC, Yang R, Jin Y, Xi Y, Zhang C, Cui T, Fang X et al (2016) Conserved water use improves the yield performance of soybean. *Agric. Water Manag.* 179:236-245.
- Heffner EL, Jannink JL, Iwata H, Souza E, Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Science* 51:2597-2606.
- Henryon M, Berg P, Sorensen AC (2014) Animal-breeding schemes using genomic information need breeding plans designed to maximise long-term genetic gains. *Livest. Sci.* 166:38–47.
- Heslot N, Rutkoski J, Poland J, Jannink JL, Sorrells ME (2013) Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS One* 8:e74612.
- Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL et al (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 4:1937–1944.
- Jumrani K, Bhatia VS, Pandey GP (2017) Impact of elevated temperatures on specific leaf weight, stomatal density, photosynthesis and chlorophyll fluorescence in soybean. *Photosynth Res.* 131:333–350.
- Junqueira VS, Peixoto LDA, Laviola BG, Bhering LL, Mendonça S, Costa TDSA, Antoniassi R (2016) Bayesian Multi-Trait Analysis Reveals a Useful Tool to Increase Oil Concentration and to Decrease Toxicity in *Jatropha curcas* L. *PloS one* 11:e0157038.
- Kobraee S, Shamsi K, Rasekhi B (2011) Soybean production under water deficit conditions. *Sch. Res. Libr.* 2:423–434.
- Kumar B, Guleria SK, Khanorkar SM, Dubey RB, Patel J, Kumar V, Parihar CM, Jat SL, Singh V, Yatish K (2016) Selection indices to identify maize (*Zea mays* L.) hybrids adapted under drought-stress and drought-free conditions in a tropical climate. *Crop and Pasture Science* 67:1087-1095.
- Leite WDS, Unêda-Trevisoli SH, Silva FMD, Silva AJD, Mauro AOD (2018) Identification of superior genotypes and soybean traits by multivariate analysis and selection index. *Revista Ciência Agronômica* 49:491-500.
- Leite WS, Pavan BE, Matos Filho CHA, de Alcantara Neto F, de Oliveira CB, Feitosa FS (2016) Estimativas de parâmetros genéticos, correlações e índices de seleção para seis caracteres agronômicos em linhagens F8 de soja. *Comunicata Scientiae* 7:302-310.
- Lemos RC, Santos RD (1984) Manual de descrição e coleta desolo no campo. Campinas: Sociedade Brasileira de Ciência do Solo 2:46.
- Li D, Liu H, Qiao Y, Wang Y, Cai Z, Dong B, Shi C, Liu Y, Li X, Liu M (2013) Effects of elevated CO₂ on the growth, seed yield, and water use efficiency of soybean (*Glycine max.* L.) under drought stress. *Agric. Water Manag.* 129:105–112

- Li YH, Liu YL, Reif JC, Liu ZX, Liu B et al (2014) Biparental resequencing coupled with SNP genotyping of a segregating population offers insights into the landscape of recombination and fixed genomic regions in elite soybean. *G3: Gene, Genomes, Genetics* 4:553–560.
- Lima FS, Kurozawa LE, Ida EI (2014) The effects of soybean soaking on grain properties and isoflavones loss. *LWT-food science and technology* 59:1274-1282.
- Mak M, Babla M, Xu S, Carrigan AO, Liu X, Gong Y, Holford P, Chen Z (2014) Leaf mesophyll K^+ , H^+ and Ca^{2+} fluxes are involved in drought-induced decrease in photosynthesis and stomatal closure in soybean. *Environ. Exp. Bot.* 98:1–12.
- Maleki A, Naderi A, Naseri R, Fathi A, Bahamin S, Maleki R (2013) Physiological performance of soybean cultivars under drought stress. *Bull. Life Sci.* 2:38–44.
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- MiSeq Support - Documentation & Literature. Disponível em: <https://support.illumina.com/sequencing/sequencing_instruments/miseq/documentation.html> Acesso em: 16 jan. 2018.
- Mohammadi M, Tiede T, Smith KP (2015) A genome-wide procedure for predicting genetic variance and correlated response in bi-parental breeding populations. *Crop Sci.* 55:2068–2077.
- Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.* 124:342–355.
- Mulamba NN, Mock JJ (1978) Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. *Journal of Genetics and Cytology* 7:40–51.
- Nakaya A, Isobe SN (2012) Will genomic selection be a practical method for plant breeding? *Ann. Bot.* 110:1303–1316.
- Nogueira APO, Sedyama T, Sousa LB, Hamawaki OT et al (2012) Análise de trilha e correlações entre caracteres em soja cultivada em duas épocas de semeadura. *Biosci. J.* 28:877-888.
- Pennisi E (2008). Plant genetics. The blue revolution, drop by drop, gene by gene. *Science* 320:171–173.
- Rangel RM, do Amaral Júnior AT, Gonçalves LS, Júnior SDPF, Candido LS (2011) Análise biométrica de ganhos por seleção em população de milho pipoca de quinto ciclo de seleção recorrente. *Rev. Ciênc. Agron* 42:473-481.
- Resende MDV, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical* 37:182–194.

Resende MDV (2007) Matemática e estatística na análise de experimentos e no melhoramento genético. Colombo: Embrapa Florestas 1:561.

Resende MDV, Lopes PS, da Silva RL, Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. Pesquisa Florestal Brasileira, 56:63.

Rezende JCD, Botelho CE, Oliveira ACBD, Silva FLD, Carvalho GR, Pereira AA (2014) Genetic progress in coffee progenies by different selection criteria. Coffee Science 9:347-353.

Rosado LDS, Santos CEMD, Bruckner CH, Nunes ES, et al. (2012) Simultaneous selection in progenies of yellow passion fruit using selection indices. Rev. Ceres 59:95-101.

Rotundo L, Westgate ME (2009) Meta-analysis of environmental effects on soybean seed composition. F. Crop. Res. 110:147–156.

Samarah NH, Mullen RE, Cianzio SR, Scott P (2006) Dehydrin-like proteins in soybean seeds in response to drought stress during seed filling. Crop Sci. 46:2141–2150.

Santos FS, Amaral Júnior AT, Freitas Júnior SP, Rangel RM, Pereira MG (2007) Predição de ganhos genéticos por índices de seleção na população de milho-pipoca unb-2u sob seleção recorrente. Bragantia 66:389-396.

Silva L, Resende R, Ferreira R, Silva G, Kist V, Barbosa M, Nascimento M, Bhering L (2016) Selection index using the graphical area applied to sugarcane breeding. Genetics and molecular research: GMR 3:15.

Silva L, Teodoro P, Peixoto L, Assis C, Gasparini K, Barbosa M, Bhering L (2017) Selecting sugarcane genotypes by the selection index reveals high gain for technological quality traits. Genetics and molecular research: GMR 2:16.

Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y (2016) Multiple-trait-and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theoretical and applied genetics 129:273-287.

Soares IO, Rezende PM, Bruzi AT, Zambiazzi EV et al (2015) Adaptability of soybean cultivars in different crop years. Genet. Mol. Res. 14:8995-9003.

Song Q, Jia G, Quigley CV, Fickus EW, Hyten D, Nelson RL, Cregan, PB (2014) Soybean BARCSoySNP6K Beadchip: A tool for soybean genetics research. In Poster presented at: plant and animal genome XXII conference, San Diego, CA, USA.

Steel RGD, Torrie JH (1960) Principles and procedures of statistics. New York: McGraw-Hill 1:418.

Teixeira FG, Hamawaki OT, Nogueira APO, Hamawaki, RL, Jorge GL, Hamawaki CL et al (2017) Genetic parameters and selection of soybean lines based on selection indexes. Genetics and Molecular Research 16:3.

- Val BHP, da Silva FM, Bizari EH, de Sousa Leite W, Groli EL, de Matos Pereira, E et al (2017) Identification of superior soybean lines by assessing genetic parameters and path analysis of grain yield components. *African Journal of Biotechnology* 8: 328-336.
- Van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1. *Soil science society of America journal* 44:892-898.
- VanRaden PM, O'Connell JR, Wiggans GR, Weigel KA (2011) Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10.
- Vasconcelos ES, Ferreira RP, Cruz CD, Moreira A et al (2010) Estimativas de ganho genético por diferentes critérios de seleção em genótipos de alfafa. *Rev. Ceres* 57:205-210.
- Vianna VF, Desideriosup JA, De Santiagosup S, Juniorsup JAF et al (2013) The multivariate approach and influence of characters in selecting superior soybean genotypes. *Afr. J. Agric. Res.* 8:4162-4169.
- Xavier A, Muir WM, Rainey KM (2016) Assessing predictive properties of genome-wide selection in soybeans. *G3: Genes, Genomes, Genetics*, 6:2611-2616.
- Wei Y, Jin J, Jiang S, Ning S, 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.
- Wimmer V, Lehermeier C, Albrecht T, Auinger HJ, Wang Y, et al (2013) Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics* 195: 573–587.

CONCLUSÕES GERAIS

No primeiro capítulo, foram obtidas acurácias consideradas de boas a altas nos dois tamanhos populacionais, além de apresentarem ótimas porcentagens de coincidências na seleção de indivíduos geneticamente superiores, não justificando a fenotipagem de um número de indivíduos cinco vezes superior, visto que eleva tempo e mão-de-obra e torna-se mais dispendioso para características da natureza das utilizadas nesse experimento.

No segundo capítulo, observou-se que as capacidades preditivas obtidas pela correlação entre os dados fenotípicos e os GEBVs foram de magnitude moderada a alta, para todos as características avaliadas. Com isso, os resultados dos coeficientes de coincidência de seleção praticada com os dados fenotípicos e GEBVs foram, em sua maioria, iguais ou superiores a 50%. Com o uso do índice de seleção de soma de ranks de Mulamba e Mock, baseado no GEBV de cada cultivar foram identificadas as cultivares melhores ranqueadas para ambos os ambientes (controle e sob déficit hídrico), sendo: 5G770 RR, P98Y70, CD 226 RR, BRS 282, CD 224, CD 208, 5D711RR, CD253, BRS MT Pintado, CD 242 RR, consideradas tolerantes ao deficit hídrico e responsivas ao aumento da disponibilidade de água.

As informações obtidas nos dois capítulos demonstram o potencial do uso da predição genômica para características de difícil mensuração, mesmo para um tamanho populacional reduzido na cultura da soja, tanto com a utilização de dados reais como simulados. Os genótipos selecionados podem ser utilizados no programa de melhoramento genético da soja, como fonte de germoplasma auxiliando no desenvolvimento de cultivares tolerantes ao deficit hídrico.