

LIDIANE APARECIDA SILVA

**ESTRATÉGIAS MULTIVARIADAS APLICADAS À SELEÇÃO
GENÔMICA AMPLA**

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*.

VIÇOSA
MINAS GERAIS- BRASIL
2018

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

S586e
2018 Silva, Lidiane Aparecida, 1986-
Estratégias multivariadas aplicadas à seleção genômica
ampla / Lidiane Aparecida Silva. – Viçosa, MG, 2018.
xv, 75 f. : il. ; 29 cm.

Orientador: Leonardo Lopes Bhering.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Plantas -Seleção. 2. Genética quantitativa. 3. Biometria.
4. Análise multivariada. I. Universidade Federal de Viçosa.
Departamento de Biologia Geral. Programa de Pós-Graduação
em Genética e Melhoramento. II. Título.

CDD 22. ed. 581.35

LIDIANE APARECIDA SILVA

**ESTRATÉGIAS MULTIVARIADAS APLICADAS À SELEÇÃO
GENÔMICA AMPLA**

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: 04 de julho de 2018.


Cosme Damião Cruz


Camila Ferreira Azevedo


Felipe Lopes da Silva


Lisandra Magha Moura


Leonardo Lopes Bhering
(Orientador)

*A Deus, que sempre esteve comigo ao longo
de todo esse caminho!*

Ofereço

*Aos meus pais, Sebastião Pacheco da Silva e
Maria Aparecida Silvestre Pacheco, pelo amor
incondicional e apoio constante!*

Dedico

*“O saber a gente aprende com os mestres e os livros. A
sabedoria, se aprende é com a vida e com os humildes.”*

Cora Coralina

*“Nas grandes batalhas da vida, o primeiro passo
para a vitória é o desejo de vencer.”*

Mahatma Gandhi

AGRADECIMENTOS

Agradeço primeiramente a Deus, por iluminar meus passos e acalmar meu coração diante às dificuldades. “Tudo posso naquele que me fortalece.” Filipenses 4:13.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento por todos esses anos de estudos, pela oportunidade de realização desse curso de alto nível em excelência, contribuindo com toda a estrutura necessária para meu crescimento pessoal e profissional.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa e apoio financeiro durante a condução dessa pesquisa.

Ao Prof. Leonardo Lopes Bhering pela orientação, incentivo, confiança, amizade, paciência e por estar sempre disposto a dividir seus conhecimentos. Você é um grande exemplo de profissionalismo e dedicação que levarei comigo para toda a vida!

Aos professores do Programa de Pós-Graduação em Genética e Melhoramento, pelos ensinamentos e pela amizade ao longo desses anos de estudo.

Aos secretários do Programa de Pós-Graduação em Genética e Melhoramento, Marco Túlio e Odilon, por sempre nos receberem com um sorriso no rosto, nos mostrando os caminhos mais simples para resolução de qualquer problema acadêmico.

Ao Prof. Cosme Damião Cruz, Prof. Felipe Lopes da Silva, Prof. Camila Azevedo e D.Sc. Lisandra Moura, que se prontificaram em participar da banca avaliadora com disposição para ensinar e contribuir com a minha formação, bem como com a pesquisa desenvolvida.

Aos meus pais, simplesmente pelo amor incondicional, pelas orações, pela preocupação ao longo desses anos de distância, por acreditarem em meus sonhos oferecendo o incentivo e o apoio necessário para torná-los reais!

Aos meus irmãos, Flávio e Pablo, por serem meus melhores amigos, por toda a confiança e companheirismo, por sempre estarem dispostos para me acolher em qualquer situação. Que assim seja para toda a vida!

À minha vovó Maria, que sempre trouxe alegria para os meus dias através de longas ligações, pela simplicidade, pelo grande exemplo de vida e por sempre confortar meu coração com tanta bondade e carinho.

Ao Wallisson, que sempre esteve comigo nessa caminhada, me fortalecendo e incentivando a acreditar nos meus sonhos. Obrigada pelo amor, paciência e compreensão ao longo de todos esses anos!

À minha segunda família, José Leal, Maria, Vô Orlando, Vó Eva, Luiz, Samuel, Isabela e Tatiele, pelo carinho, orações, torcida e por me proporcionarem ótimos momentos!

Aos meus futuros afilhados Leonardo e Tassia, pela amizade que levarei para toda a vida, por todo carinho, incentivo e por sempre acreditarem na minha capacidade. Vocês foram fundamentais para a conclusão dessa etapa!

Ao amigo Juan, pela grande colaboração no desenvolvimento dessa pesquisa, pela paciência e pelo otimismo na busca por soluções.

À minha irmãzinha de coração, Thais, por todo cuidado, carinho e confiança. Deus te colocou em meu caminho no momento exato, para me mostrar que no mundo existem anjos capazes de tornar tudo mais simples em nossas vidas. Não importa para onde iremos a partir de agora, o importante é que a amizade cultivada aqui existirá para sempre!

Ao grande amigo Rodrigo Pena, por todos os momentos de risadas, pelos conselhos valiosos, sermões honestos e por compartilhar comigo tanta experiência de vida.

Às amigas Sara e Rebeca, que cada uma com seu jeitinho rude ou delicado, me proporcionaram momentos de muita felicidade. Obrigada por se fazerem presentes ao longo dessa caminhada, vocês são muito importantes em minha vida!

Ao amigo e Professor Rodrigo, pelas aulas de inglês mais divertidas, pelos melhores conselhos e por nunca me deixar perder as esperanças na luta pelos meus objetivos. Obrigada por me fazer acreditar que com o esforço e a dedicação necessária, tudo é possível!

Aos amigos do Laboratório de Biometria, Nadson, Gabriela, Bruno, Jâneo, Larissa, Paulo, Arthur, Natalia, Rafael, Vinicius, Michele Jorge, Michele Brandão, Jeniffer, Igor, Andrea, Guilherme, Adriel, Felipe, Rodrigo e João, pela amizade, por todos os momentos de alegria e pelo ótimo ambiente de trabalho.

Aos amigos da agronomia 2007, especialmente Andrea, Paulo, Geverson e Ítalo, que continuaram comigo nesse caminho de pós-graduação, dividindo todas as experiências, vitórias e frustrações. Obrigada pela amizade de uma vida inteira, pelo companheirismo, pelas gargalhadas infinitas e pelo apoio na busca por esse título.

À minha prima irmã Luciana, que com seu jeitinho ciumento e carinhoso sempre esteve presente em minha vida! Um grande exemplo de que para verdadeiras amigas não existem distâncias!

À minha afilhada Alessandra, que sempre tão meiga e doce, soube entender com carinho a minha ausência em momentos tão importantes de sua vida. Obrigada por me motivar na luta pelos meus objetivos e por me considerar com tanto orgulho e admiração!

Ao amigo irmão Gabriel, por tantos momentos especiais, por todos os conselhos e por me fazer sentir uma pessoa tão importante em sua vida!

Às grandes amigas Valquíria, Sabrina e Mariele, que me acompanham com todo carinho e apoio desde o ensino médio! “Amigas para sempre é o que nós iremos ser...”

A todos os primos, tios e tias que, às vezes, mesmo sem entender muito bem meus objetivos, sempre me apoiaram e acreditaram nessa realização! Amo muito todos vocês!

Enfim, a todos aqueles que contribuíram direta ou indiretamente para a conclusão de mais esta etapa, meus sinceros agradecimentos!

BIOGRAFIA

LIDIANE APARECIDA SILVA, filha de Sebastião Pacheco da Silva e Maria Aparecida Silvestre Pacheco, nasceu na cidade de Patos de Minas, estado de Minas Gerais, em 21 de outubro de 1986.

Iniciou o curso de graduação em Agronomia em 2007 pela Universidade Federal de Viçosa (UFV), em Viçosa, MG, recebendo o título de Engenheira Agrônoma em janeiro de 2012.

Em agosto de 2012 iniciou seus estudos no Programa de Pós-Graduação em Genética e Melhoramento pela Universidade Federal de Viçosa para obtenção do título *Magister Scientiae*, apresentou sua dissertação no dia 25 de julho de 2014.

Em agosto de 2014, iniciou seu doutorado no Programa de Pós-Graduação em Genética e Melhoramento da UFV, para obtenção do título de *Doctor Scientiae* em Genética e Melhoramento, apresentou sua tese no dia 04 de julho de 2018.

SUMÁRIO

LISTA DE TABELAS	ix
RESUMO	xii
ABSTRACT	xiv
INTRODUÇÃO GERAL	1
REVISÃO DE LITERATURA	3
ÍNDICES DE SELEÇÃO	3
SELEÇÃO GENÔMICA AMPLA	5
ÍNDICES DE SELEÇÃO APLICADOS A DADOS GENÔMICOS	9
REFERÊNCIAS BIBLIOGRÁFICAS	11
COMPARAÇÃO ENTRE MÉTODOS UNI E MULTIVARIADOS APLICADOS À SELEÇÃO GENÔMICA AMPLA	17
RESUMO	17
INTRODUÇÃO	19
MATERIAL E MÉTODOS	21
RESULTADOS	26
DISCUSSÃO	34
CONCLUSÕES	38
REFERÊNCIAS BIBLIOGRÁFICAS	39
CAPÍTULO 2	45
COMPARAÇÃO ENTRE ÍNDICES APLICADOS À SELEÇÃO GENÔMICA AMPLA	45
RESUMO	45
INTRODUÇÃO	47
MATERIAL E MÉTODOS	48
RESULTADOS	55
DISCUSSÃO	66
CONCLUSÕES	70
REFERÊNCIAS BIBLIOGRÁFICAS	71

LISTA DE TABELAS

- Tabela 1.** Comparação entre os métodos de seleção genômica Random Regression Best Linear Unbiased Predictor (RRblup), Bayesian Lasso (Blasso), Univariate Partial Least Square (Upls) e Multivariate Partial Least Square (Mpls), avaliando o tempo de processamento de análise para as quatro características com diferentes herdabilidades. 26
- Tabela 2.** Comparação entre os métodos de seleção genômica Random Regression Best Linear Unbiased Predictor (RRblup), Bayesian Lasso (Blasso), Univariate Partial Least Square (Upls) e Multivariate Partial Least Square (Mpls), avaliando as acurácias de seleção genotípica e fenotípica para as quatro características avaliadas. 27
- Tabela 3.** Ganhos de seleção diretos e indiretos, com base nos indivíduos selecionados a partir dos valores fenotípicos e valores genéticos verdadeiros, para as quatro características avaliadas com diferentes herdabilidades. 28
- Tabela 4.** Ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs obtidos pelos métodos de seleção genômica ampla, para as características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente. 29
- Tabela 5.** Ganhos de seleção estimados nos valores fenotípicos e valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs obtidos pelos índices de seleção genômica, para as características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente. 31
- Tabela 6.** Comparação entre os diferentes métodos e índices de seleção genômica, via análise de coincidência, para as características simuladas: C1 ($h^2 = 20\%$), acima da diagonal principal e C2 ($h^2 = 40\%$), abaixo da diagonal principal. 32

Tabela 7. Comparação entre os diferentes métodos e índices de seleção genômica, via análise de coincidência, para as características simuladas: C3 ($h^2 = 60\%$), acima da diagonal principal e C4 ($h^2 = 80\%$), abaixo da diagonal principal.....	33
Tabela 8. Ganhos de seleção diretos e indiretos estimados com base nos indivíduos selecionados a partir dos maiores valores fenotípicos e valores genéticos verdadeiros, para as quatro características com diferentes herdabilidades.	55
Tabela 9. Ganhos de seleção diretos e indiretos estimados com base nos indivíduos selecionados a partir dos maiores valores fenotípicos e valores genéticos verdadeiros para as características C1, C3 e C4 e menores valores para a característica C2.	56
Tabela 10. Ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs obtidos pelos métodos Random Regression Best Linear Unbiased Predictor (RRblup) e Multivariate Partial Least Square (Mpls), para as quatro características avaliadas com diferentes herdabilidades.....	57
Tabela 11. Ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs para as características C1, C3 e C4 e em ordem crescente para a característica C2, considerando os métodos Random Regression Best Linear Unbiased Predictor (RRblup) e Multivariate Partial Least Square (Mpls).....	58
Tabela 12. Ganhos de seleção estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs obtidos pelos índices de seleção, para as características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.	59
Tabela 13. Ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs para as características C1, C3 e C4 (herdabilidades 20%, 60% e 80%, respectivamente) e em ordem crescente para a	

característica C2 (herdabilidade igual a 40%), considerando os diferentes índices de seleção.
..... 60

Tabela 14. Análise de coincidência de seleção dos indivíduos selecionados em ordem decrescente de GEBVs estimados por cada índice de seleção e os indivíduos selecionados pelos maiores valores genéticos verdadeiros, para as características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente. 62

Tabela 15. Análise de coincidência dos indivíduos selecionados em ordem decrescente de GEBVs para as características C1, C3 e C4 (herdabilidades 20%, 60% e 80%, respectivamente) e em ordem crescente para característica C2 (herdabilidade igual a 40%), estimados por cada índice de seleção e os indivíduos selecionados, seguindo o mesmo critério, pelos valores genéticos verdadeiros. 63

Tabela 16. Comparação entre as diferentes estratégias multivariadas, via análise de coincidência, para dois cenários de seleção: i) Cenário 1 (acima da diagonal principal): seleção em ordem decrescente dos indivíduos para todas as características avaliadas; ii) Cenário 2 (abaixo da diagonal principal): seleção em ordem decrescente para as características de herdabilidades 20%, 60% e 80% e em ordem crescente para herdabilidade 40%. 65

RESUMO

SILVA, Lidiane Aparecida, D.Sc., Universidade Federal de Viçosa, julho de 2018. **Estratégias multivariadas aplicadas à seleção genômica ampla**. Orientador: Leonardo Lopes Bhering.

A seleção simultânea de caracteres, integrada à seleção genômica ampla (GWS), tem se tornado uma estratégia de grande interesse para os programas de melhoramento de plantas. Neste sentido, os objetivos desse estudo foram: i) comparar a acurácia e eficiência de seleção do método Multivariate Partial Least Square (Mpls) em relação aos métodos univariados: Random Regression Best Linear Unbiased Predictor (RRblup), Bayesian Lasso (Blasso) e Univariate Partial Least Square (Upls); ii) verificar a eficiência da seleção direta e indireta na GWS; iii) elaborar e comparar diferentes estratégias multivariadas, via índices de seleção integrados à GWS, eficientes na identificação e seleção precoce de indivíduos geneticamente superiores, em diferentes características simultaneamente. Dez populações F_2 com 800 indivíduos foram simuladas considerando quatro características com diferentes herdabilidades. Na primeira etapa da pesquisa, os dados simulados foram submetidos às análises de GWS via RRblup, Blasso, Upls e Mpls. Quatro índices de seleção genômica foram elaborados pelo somatório dos efeitos dos marcadores obtidos para cada característica, ponderados pela sua respectiva variância residual, sendo elaborado um índice para cada metodologia avaliada. Na segunda etapa da pesquisa, os dados simulados foram submetidos às análises de GWS via RRblup e Mpls. Foram elaboradas e comparadas diferentes estratégias de índices de seleção aplicados à GWS: i) ponderar os efeitos dos marcadores pela variância residual (elaborado na primeira etapa); ii) codificar e padronizar os efeitos dos marcadores; iii) aplicar média nos efeitos dos marcadores; iv) aplicar o índice de Mulamba e Mock (1978) nos valores genéticos genômicos; v) codificar e padronizar os valores fenotípicos, antes das análises de GWS. Além disso, na segunda etapa dessa pesquisa, foram considerados dois cenários de seleção. No primeiro cenário, foram selecionados os indivíduos com maiores valores fenotípicos, valores genéticos verdadeiros e valores genéticos genômicos para as quatro características avaliadas. Já no segundo cenário, foi considerado diferente sentido de seleção para uma das características simuladas. As comparações entre os métodos e os índices de seleção foram realizadas considerando o tempo de processamento, as acurácias de predição, os ganhos de seleção e os coeficientes de coincidência de seleção. Foi aplicado o índice de Mulamba e Mock (1978) nos valores fenotípicos e valores genéticos verdadeiros. Os métodos de seleção genômica foram mais eficientes que a seleção fenotípica. O método Mpls foi similar ao método Upls para as

características de menores herdabilidades e foi menos eficiente quando comparado aos métodos RRblup e Blasso. A seleção direta e indireta baseada nos valores genéticos genômicos foi mais eficiente que a seleção fenotípica. Nenhuma das estratégias avaliadas foi eficiente considerando diferente sentido de seleção para uma das características simuladas. Os índices ponderados pela variância residual apresentaram alta eficiência para aplicação na GWS, no entanto tenderam a maximizar os ganhos para as características de maiores herdabilidades. As estratégias de aplicar índices, via RRblup, a partir da média dos efeitos dos marcadores, dos valores fenotípicos codificados e padronizados e da aplicação do índice de Mulamba e Mock nos valores genéticos genômicos, resultaram em altos ganhos de seleção e mais se aproximaram aos ganhos obtidos pelo índice de Mulamba e Mock aplicado aos valores genéticos verdadeiros. A estratégia de codificar e padronizar os efeitos dos marcadores proporcionou os menores ganhos genéticos totais. De modo geral, os índices de seleção genômica propostos, proporcionaram maior eficiência de seleção, quando comparados ao índice de seleção de Mulamba e Mock fenotípico. Esses resultados sugerem que as estratégias multivariadas, via índices de seleção integrados à GWS, são promissoras para aplicação em programas de melhoramento genético de plantas, visando a seleção precoce direta de várias características simultaneamente.

ABSTRACT

SILVA, Lidiane Aparecida, D.Sc., Universidade Federal de Viçosa, July, 2018. **Multivariate strategies applied to genome-wide selection.** Adviser: Leonardo Lopes Bhering.

Simultaneous traits selection, integrated with genome wide selection (GWS), has become a strategy of great interest for plant breeding programs. In this sense, the objectives of this study were: i) to compare the accuracy and efficiency of the Multivariate Partial Least Square (Mpls) method in relation to univariate methods: Random Regression Best Linear Unbiased Predictor (RRblup), Bayesian Lasso (Blasso) and Univariate Partial Least Square (Upls); ii) verify the efficiency of direct and indirect selection in GWS; iii) to elaborate and compare different multivariate strategies, through selection indexes integrated to GWS, efficient in the identification and early selection of genetically superior individuals, in several traits simultaneously. Ten F₂ populations with 800 individuals were simulated considering four traits with different heritabilities. In the first research step, the simulated data were submitted to the GWS analysis via RRblup, Blasso, Upls and Mpls. Four GWS indexes were elaborated by the sum of the markers effects obtained for each trait, weighted by their respective residual variance, and was elaborated an index for each methodology. In the second research step, the simulated data were submitted to GWS analysis via RRblup and MPLs. Different selection index strategies applied to GWS were elaborated and compared: i) to weigh the effects of the markers by the residual variance (elaborated in the first step); ii) to encode and standardize the effects of markers; iii) to apply average markers effects; iv) to apply the Mulamba and Mock index (1978) to genomic breeding values; v) to encode and standardize phenotypic values, before to GWS analysis. In addition, in the second research step, two scenarios selection were considered. In the first scenario, individuals with higher phenotypic values, genetic values and genomic breeding values were selected for the four traits evaluated. In the second scenario, a different sense of selection was considered for one of the simulated traits. The comparisons among the methods and the selection indexes were performed considering the processing time, the prediction accuracy, the selection gains and the selection coincidence coefficients. The Mulamba and Mock index was applied to the phenotypic values and genetic values. The GWS methods were more efficient than phenotypic selection. The Mpls method was similar to the Upls method for the smaller heritabilities traits and was less efficient when compared to the RRblup and Blasso methods. The direct and indirect selection based on genomic breeding values was more efficient than phenotypic selection. None of the evaluated strategies was

efficient considering a different sense of selection for one of the simulated traits. The residual variance weighted indexes showed high application efficiency to GWS, but tended to maximize gains for the traits of higher heritabilities. The strategies of applying indexes, via RRblup, from the markers effects mean, the coded and standardized phenotypic values and the application of the Mulamba and Mock index on the genomic breeding values, resulted in high selection gains and more approached to the gains obtained by the Mulamba and Mock index applied to the genetic values. The coding and standardizing strategy of the effects of markers provided the lowest total genetic gains. In general, the genomic selection indexes, provided greater selection efficiency than Mulamba and Mock selection index phenotypic. These results suggest that multivariate strategies, via selection indexes integrated to the GWS, are promising for application in plant genetic improvement programs, aiming at the direct early selection of several traits simultaneously.

INTRODUÇÃO GERAL

A correta definição dos objetivos de seleção é essencial para o sucesso dos programas de melhoramento genético de plantas, consistindo em grande responsabilidade para os melhoristas. Estratégias de melhoramento sofisticadas e bem planejadas são irrelevantes se for baseada em objetivos inadequados. Nesse sentido, é necessário a escolha dos critérios de seleção que representam o caráter ou o conjunto de caracteres em que a seleção se baseia, visando avaliar e ordenar os genótipos candidatos à seleção (Pires et al., 2011).

Segundo Resende e Rosa-Perez (1999), considerando os objetivos e os critérios de seleção, algumas estratégias podem ser definidas: i) seleção direta: quando a seleção é baseada em um único caráter e o critério de seleção é o mesmo caráter; ii) seleção indireta: quando a seleção é baseada em um único caráter e o critério de seleção é outro (s) caráter (s); iii) seleção empregando caracteres auxiliares no melhoramento: quando a finalidade da seleção é um único caráter e o critério de seleção é baseado em uma combinação de vários caracteres; iv) e índices de seleção: quando a seleção é fundamentada em um agregado de caracteres de importância e o critério de seleção é baseado em uma combinação desses vários caracteres.

A seleção simultânea de caracteres consiste em uma estratégia multivariada de soma importância nos programas de melhoramento de plantas, visto que é desejado o desenvolvimento de cultivares que apresentem ótimo desempenho em várias características de interesse econômico simultaneamente (Cruz et al., 2012). Para esta finalidade, a teoria de índices de seleção se destaca por permitir a seleção baseada em uma combinação linear dos caracteres de interesse, cujos coeficientes de ponderação são estimados de modo a maximizar a correlação entre o índice e o agregado genotípico (Cruz et al., 2012; Pires et al., 2011).

O processo de seleção de genótipos superiores em programas de melhoramento demanda tempo e recursos e, por essa razão, deve ser o mais eficiente possível. Desse modo,

com o intuito de reduzir o tempo requerido para o processo seletivo, a seleção precoce é uma estratégia tradicional que busca maximizar os ganhos genéticos por unidade de tempo. A seleção precoce fenotípica é uma forma de seleção indireta, via correlação entre os caracteres avaliados em fase precoce com caracteres, geralmente de importância econômica, na fase de maturação fisiológica desejada. Assim, os caracteres avaliados em idades prévias são utilizados como preditores de caracteres economicamente importantes (Pires et al., 2011).

Entretanto, como os genes podem se expressar de forma diferenciada ao longo do ciclo de desenvolvimento das plantas, a seleção precoce fenotípica pode ser inadequada. Neste contexto, a Seleção Genômica Ampla (GWS), proposta por Meuwissen et al. (2001), tem se mostrado como uma ferramenta promissora. Segundo Pires et al. (2011), a grande vantagem da GWS é que, enquanto os métodos tradicionais utilizam informações fenotípicas para inferir sobre os efeitos dos genótipos nos indivíduos, a GWS utiliza informações de milhares de marcadores moleculares distribuídos no genoma dos indivíduos, juntamente com as informações fenotípicas, para inferir sobre os valores genéticos dos indivíduos candidatos à seleção. Todavia, geralmente a GWS tem sido aplicada em abordagem univariada, visando obtenção de ganhos com a seleção de uma característica de interesse por vez.

Verifica-se que, a maioria dos programas de melhoramento buscam, além da seleção precoce, a melhoria de diversos caracteres de importância econômica simultaneamente, visando o desenvolvimento de cultivares com maior aceitação no mercado consumidor (Cruz et al., 2012). Portanto, para essa finalidade, a aplicação de estratégias multivariadas na GWS pode ser promissora na busca por maiores ganhos com a seleção em vários caracteres de interesse simultaneamente, a partir da seleção direta baseada em informações genotípicas correlacionadas aos fenótipos dos indivíduos.

REVISÃO DE LITERATURA

ÍNDICES DE SELEÇÃO

Na seleção direta o melhorista está interessado em obter ganho em uma única característica sobre a qual irá praticar a seleção. No entanto, em programas de melhoramento, a seleção com base em uma, ou em poucas características, tem se mostrado inadequada, por resultar em um produto final que não seja favorável em várias outras características de importância, levando a não aceitação pelos produtores. Assim, uma contribuição importante para o melhoramento de plantas, é a avaliação da resposta indireta pela seleção de determinadas características (Cruz et al., 2012).

Os índices de seleção consideram, simultaneamente, todos os caracteres de interesse econômico, gerando uma variável adicional que resulta na ponderação dos caracteres por meio de coeficientes calculados, geralmente, com base nas herdabilidades, valores econômicos relativos e correlações entre as características. A construção dos índices se baseia em equações de regressão múltipla que permitem prever o valor genético dos indivíduos, que são funções lineares dos valores genéticos associados a cada caráter (Cruz et al., 2012; Pires et al., 2011).

Desta forma, o índice de seleção (I) refere-se a qualquer variável utilizada para fins de seleção, obtida por meio do somatório de n variáveis, ponderadas por seus respectivos coeficientes, e pode ser descrita da seguinte forma:

$$I = \sum_{i=1}^n b_i y_i$$

em que b_i é o peso da i-ésima característica e y_i é o valor genético ou fenotípico da i-ésima característica.

Smith (1936) foi o primeiro a propor o uso de índice de seleção nos programas de melhoramento de plantas como critério de seleção simultânea de duas ou mais características correlacionadas. Esse procedimento foi adaptado ao melhoramento genético animal por Hazel (1943). O índice de seleção clássico proposto por esses autores consiste numa combinação

linear dos valores fenotípicos dos vários caracteres de importância econômica, cujos coeficientes de ponderação são estimados de modo a maximizar a correlação entre o índice de seleção e o agregado genotípico. Esse agregado é estabelecido por outra combinação linear, envolvendo valores genéticos, os quais são ponderados pelos respectivos valores econômicos (Cruz et al., 2012).

Existem diversas propostas de índices de seleção descritas na literatura proporcionando resultados promissores no melhoramento de plantas (Cruz et al., 2012). Como exemplo, podem ser citados os índices de Kempthorne e Nordskog (1959), que propuseram uma solução matemática para maximizar o ganho num conjunto de caracteres, impondo a restrição de que o ganho em outro conjunto de caracteres fosse nulo; Pesek e Baker (1969), que propuseram um índice com base nos ganhos desejados de acordo com a importância estabelecida pelo melhorista; índice proposto por Tai (1977) que divide os caracteres em primários e secundários; índice base (Williams, 1962) que estabelece pesos e pondera diretamente os valores fenotípicos médios dos caracteres, entre vários outros.

Existem ainda os índices que não necessitam das estimativas de variâncias e covariâncias fenotípicas e genotípicas e do estabelecimento dos pesos econômicos. Podem ser citados como exemplos o índice com base na soma de postos ou ranks (Mulamba e Mock, 1978) e o índice livre de pesos e livre parâmetros (Elston, 1963).

O índice de Mulamba e Mock (1978) consiste em classificar, ou ordenar os materiais genotípicos, em relação a cada um dos caracteres, em ordem favorável ao melhoramento. Uma vez classificados, são somadas as ordens de cada genótipo referente ao caráter, resultando em uma média adicional tomada como índice de seleção (Cruz et al., 2012). O cálculo desse índice depende apenas das médias fenotípicas, no entanto caso seja de interesse do melhorista, pode-se atribuir pesos econômicos aos caracteres. Em outras palavras, o princípio do índice de Mulamba e Mock (1978) é a transformação das médias fenotípicas ajustadas dos genótipos para

cada caráter, em postos de acordo com o interesse do melhorista, ou seja, no sentido de aumentar ou diminuir a expressão fenotípica (Ramalho et al., 2012).

De modo geral, diferentes índices de seleção representam diferentes alternativas de seleção e podem resultar em diferentes ganhos com a seleção. Essa metodologia é promissora para obtenção de maiores ganhos, proporcionando a identificação de maneira rápida e eficiente de genótipos mais adequados com os objetivos do melhoramento (Cruz et al., 2012). Portanto, diante das inúmeras opções de índices de seleção descritas na literatura, é de grande responsabilidade do melhorista a escolha da proposta mais adequada para a seleção e desenvolvimento de cultivares com alto desempenho em vários caracteres simultaneamente.

SELEÇÃO GENÔMICA AMPLA

A seleção assistida por marcadores (SAM) foi a primeira metodologia proposta para utilização de marcadores moleculares no melhoramento genético (Lande e Thompson, 1990). A SAM utiliza simultaneamente dados fenotípicos e de marcadores moleculares em desequilíbrio de ligação com alguns QTLs (*Quantitative trait locus*). Todavia, esta metodologia é pouco eficiente para detectar os QTLs de pequenos e raros efeitos, dificultando sua aplicação para caracteres de natureza poligênica (Goddard e Hayes, 2007; Lorenz et al., 2011).

No entanto, a maioria dos caracteres de interesse em programas de melhoramento são classificados como quantitativos, sendo controlados por inúmeros genes, apresentando baixa herdabilidade e são fortemente influenciados pelo ambiente (Falconer e Mackay, 1996). Assim, a Seleção Genômica Ampla (*Genome Wide Selection - GWS*), proposta por Meuwissen et al. (2001), tem se destacado como ferramenta promissora nos programas de melhoramento, visto que propõe a análise de genomas completos e não limita o número de QTLs considerados na expressão do caráter quantitativo (Bernardo e Yu, 2007).

A GWS tem como princípio a predição do valor genético genômico (GEBV) a partir da utilização de grande número de marcadores amplamente distribuídos no genoma, aumentando a probabilidade de que cada QTL de interesse esteja em desequilíbrio de ligação com pelo menos parte dos marcadores, capturando todos os genes que afetam a característica quantitativa (Desta e Ortiz, 2014; Goddard e Hayes, 2007; Lorenz et al., 2011; Meuwissen et al., 2001). Diferente da SAM, a GWS não necessita da prévia identificação dos marcadores com efeitos significativos e do mapeamento dos QTLs, além de ser eficiente para caracteres de baixa herdabilidade, resultando em superioridade em termos de ganhos por unidade de tempo (Heffner et al., 2011; Lorenzana e Bernardo, 2009; Pires et al., 2011).

Com o desenvolvimento de marcadores de alta eficiência e baixo custo, a GWS tem se tornado cada vez mais atrativa. Os marcadores do tipo SNP (*Single Nucleotide Polymorphism*), caracterizados pela detecção de polimorfismos resultantes da alteração de uma única base no genoma, são os preferidos para a aplicação na GWS, em relação a outros marcadores, devido à baixa taxa de mutação, facilidade de genotipagem e codominância. Além desses, outros tipos de marcadores, como os microssatélites e DArTs (*Diversity Array Technology*) podem ser utilizados com sucesso na GWS por serem abundantes e permitirem amostragem ampla no genoma (Pires et al., 2011).

A GWS requer o uso de uma população de estimação (treinamento), para estimar os efeitos dos marcadores nos fenótipos, e uma população de validação, para estimar o GEBV e analisar a eficiência da estimação destes efeitos na recuperação do valor genômico em uma população independente (Meuwissen et al., 2001). A acurácia de predição é medida pela correlação entre o GEBV estimado e o valor fenotípico dos indivíduos genotipados. Assim, após a estimação e validação de um conjunto de marcadores, estes poderão ser utilizados para a seleção precoce em uma população que se dispõe apenas de genotipagem.

De acordo com Resende et al. (2010a), para aplicação da GWS, podem ser definidas as seguintes populações: população de treinamento, validação e seleção. Essas populações podem ser: i) fisicamente distintas, sendo três populações diferentes ii) exercer duas funções ao mesmo tempo, sendo uma só população destinada tanto para treinamento, como para validação; iii) exercer três funções ao mesmo tempo, uma população utilizada para estimação e validação.

Os principais problemas da GWS são devido a estimação de um grande número de efeitos de marcadores a partir de um limitado número de observações fenotípicas e as colinearidades advindas do desequilíbrio de ligação entre os marcadores. Assim, um método ideal para GWS deve acomodar a arquitetura genética da característica em termos de genes de pequenos e grandes efeitos e suas distribuições e realizar o processo de estimação em presença de multicolinearidade e grande número de marcadores (Resende et al., 2011).

Segundo Heffner et al. (2009) o número maior de marcadores em relação ao número de fenótipos faz com que os métodos de quadrados mínimos não possam ser utilizados na GWS devido à falta de graus de liberdade para o resíduo. Atualmente, existem diversos métodos estatísticos aplicados à GWS, dentre eles os que mais se destacam são baseados em modelos mistos tais como o RRblup (Meuwissen et al., 2001) e o GBlup (Hayes et al., 2009), e os bayesianos, como o Bayes A e Bayes B (Meuwissen et al., 2001), Bayes C π (Jannink et al., 2010) e Bayesian Lasso (de los Campos et al., 2009).

As metodologias baseadas em modelos mistos assumem que todos os marcadores apresentam a mesma variância e por isso são mais indicadas para características governadas por maior número de QTLs de menor efeito, considerando que cada loco contribui igualmente para a variação genética da característica (Resende et al., 2013). Estes métodos apresentam baixa demanda computacional quando comparados aos métodos de abordagem bayesiana (Meuwissen et al., 2001; Resende et al., 2011; Resende et al., 2010b).

Os métodos bayesianos estão associados ao sistema de equações não lineares e por isso são mais adequados para características governadas por poucos genes de efeitos maiores, ou seja, quando os efeitos dos QTLs não são normalmente distribuídos (Daetwyler et al., 2010; Meuwissen et al., 2001; Zhong et al., 2009). Estes métodos podem apresentar predições genômicas mais acuradas, entretanto requerem uma análise estatística mais elaborada e são mais morosos. Os Lasso bayesianos podem ser mais vantajosos em relação aos métodos bayesianos de Meuwissen et al. (2001) por serem assintoticamente livres de informação a priori (Resende et al., 2011).

Além dessas abordagens, existem também os métodos baseados em regressão com redução de dimensionalidade que podem ser aplicados na GWS, tais como regressão via quadrados mínimos parciais (*Partial Least Square Regression - PLS*) e regressão via componentes principais (*Principal Component Regression - PCR*) (Solberg et al., 2009). Estes métodos, são eficientes para resolver os problemas da multicolinearidade e alta dimensionalidade ocasionada pelo maior número de marcadores em relação ao número de indivíduos genotipados (Azevedo et al., 2013; Resende et al., 2011).

Dentre os métodos de redução de dimensionalidade, destaca-se o método PLS, que apresenta, como ideia geral, a formação de componentes que capturem a maior quantidade de informações possíveis dispostas nas variáveis explicativas para predizer a variável dependente principal (Boulesteix e Strimmer, 2007). O PLS pode ser aplicado à seleção de marcadores com efeitos significativos em um caráter. É um método utilizado com frequência na quimiometria com o objetivo de construção de bons modelos preditivos para a variável resposta, na presença de grande número de variáveis com relações desconhecidas (Wold et al., 2001). Além disso, o método PLS apresenta como vantagem a aplicação sobre abordagem multivariada à GWS (Azevedo et al., 2013).

Neste sentido, a GWS é uma estratégia inovadora que vem demonstrado alta eficiência na predição dos GEBVs e proporcionando resultados promissores em programas de melhoramento genético vegetal, para caracteres qualitativos e principalmente para os caracteres quantitativos, proporcionando maior acurácia e aumento nos ganhos com a seleção (Bernardo e Yu, 2007; Crossa et al., 2010; Crossa et al., 2014; de Oliveira et al., 2012; Grattapaglia e Resende, 2011; Heffner et al., 2011). Portanto, a escolha do método de GWS deve ser adequada aos dados e aos objetivos do programa de melhoramento para obtenção de maiores ganhos com a seleção por unidade de tempo.

ÍNDICES DE SELEÇÃO APLICADOS A DADOS GENÔMICOS

No contexto da SAM, Lande e Thompson (1990) propuseram um índice de seleção que combina informações de marcadores com informações fenotípicas, enquanto Dekkers (2007) propôs um índice de seleção que combina GEBVs com informações fenotípicas. Ambos os índices de seleção foram avaliados usando dados simulados e 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, demonstrando que essa estratégia pode ser promissora.

Goddard (2009) desenvolveu um índice para seleção de indivíduos diretamente nos alelos de QTL. Este índice varia o peso dado a um marcador de acordo com a sua frequência, de forma que os marcadores em que o alelo favorável tem baixa frequência recebem maior peso no índice. O autor concluiu que o índice pode maximizar o diferencial de seleção ao longo dos ciclos de seleção.

Por meio da teoria do índice de seleção, os GEBVs podem ser utilizados para calcular um único GEBV (Goddard e Hayes, 2007). Esta abordagem demonstrou aumentar a precisão

de GEBVs em camundongos, trigo e tem sido usada em vários programas de melhoramento de gado leiteiro (de los Campos et al., 2009; Heffner et al., 2011; König e Swalve, 2009; Lee et al., 2008; Lorenz et al., 2011).

Heffner et al. (2011) compararam a eficiência de aplicação de índices de seleção fenotípicos e genômicos na avaliação de 11 caracteres agronômicos em uma população de melhoramento de trigo. Esses autores aplicaram o índice clássico (Hazel, 1943; Smith, 1936) e o índice base (Williams, 1962) aos valores fenotípicos e GEBVs, utilizando três diferentes pesos econômicos. Os autores verificaram maior acurácia de ambos os índices ao se utilizar informações genômicas do que apenas informações fenotípicas.

De acordo com Lorenz et al. (2011), considerando a herdabilidade e a arquitetura genética da característica, é esperado que os índices de seleção fenotípica sejam mais precisos e tenha uma maior resposta de seleção por ciclo de seleção do que os índices aplicados na GWS. No entanto, de acordo com esses autores, em termos de ganho genético por unidade de tempo, os índices de seleção genômica precisam de um terço, ou menos, do tempo exigido pelo índice de seleção fenotípica.

Massman et al. (2013) verificaram como características correlacionadas ao rendimento de grãos e qualidade no milho respondem a múltiplos ciclos por meio da aplicação de índices de seleção na GWS e SAM. Os autores concluíram que os ganhos obtidos pelos índices de seleção, através de três ciclos, foram maiores com a GWS quando comparada com a SAM. Como a GWS usa todos os marcadores disponíveis, foi possível calcular os efeitos dos marcadores para cada característica independentemente e, em seguida, calcular o índice com base no desempenho previsto para cada característica. Esta abordagem não foi possível para a SAM, visto que diferentes características identificam marcadores diferentes como significativos.

Ceron-Rojas et al. (2015) aplicaram índices de seleção genômica em dados simulados e dados reais de milho, avaliando quatro características de herdabilidades variando de 40% a 80%. Os autores compararam o índice de seleção genômica com o índice fenotípico baseado na teoria proposta por Smith (1936) e 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.

Schulthess et al. (2016) combinaram diferentes índices de seleção com a GWS para melhoria de rendimento de grãos e teor de proteína em centeio. Eles utilizaram o índice clássico (Hazel, 1943; Smith, 1936) e o índice de Kempthorne e Nordskog (1959) 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.

Portanto, alguns estudos têm apontado perspectivas favoráveis para aplicação de índices de seleção a dados genômicos, desta forma a aplicação de índices na GWS visando maximizar os ganhos com a seleção simultânea de caracteres em tempo reduzido pode ser promissora para programas de melhoramento de plantas. 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).

REFERÊNCIAS BIBLIOGRÁFICAS

AZEVEDO, C. F., SILVA, F. F., REZENDE, M. D. V. D., PETERNELLI, L. A., GUIMARÃES, S. E. F., LOPES, P. S., Uni and multivariate partial least squares applied to genomic selection for carcass traits in pigs. **Ciência Rural**, v. 43, p. 1642-1649, 2013.

BERNARDO, R., YU, J., Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v. 47, p. 1082-1090, 2007.

BOULESTEIX, A.-L., STRIMMER, K., Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. **Briefings in bioinformatics**, v. 8, p. 32-44, 2007.

CERON-ROJAS, J. J., CROSSA, J., ARIEF, V. N., BASFORD, K., RUTKOSKI, J., JARQUÍN, D., ALVARADO, G., BEYENE, Y., SEMAGN, K., DELACY, I., A genomic selection index applied to simulated and real data. **G3: Genes, Genomes, Genetics**, v. 5, p. 2155-2164, 2015.

CROSSA, J., DE LOS CAMPOS, G., PÉREZ, P., GIANOLA, D., BURGUEÑO, J., ARAUS, J. L., MAKUMBI, D., SINGH, R. P., DREISIGACKER, S., YAN, J., Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, v. 186, p. 713-724, 2010.

CROSSA, J., PÉREZ, P., HICKEY, J., BURGUEÑO, J., ORNELLA, L., CERÓN-ROJAS, J., ZHANG, X., DREISIGACKER, S., BABU, R., LI, Y., Genomic prediction in CIMMYT maize and wheat breeding programs. **Heredity**, v. 112, p. 48-60, 2014.

CRUZ, C. D., REGAZZI, A. J., CARNEIRO, P. C. S., 2012. Modelos biométricos aplicados ao melhoramento genético. UFV, Viçosa.

DAETWYLER, H. D., PONG-WONG, R., VILLANUEVA, B., WOOLLIAMS, J. A., The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, p. 1021-1031, 2010.

DE LOS CAMPOS, G., NAYA, H., GIANOLA, D., CROSSA, J., LEGARRA, A., MANFREDI, E., WEIGEL, K., COTES, J. M., Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, v. 182, p. 375-385, 2009.

DE OLIVEIRA, E. J., DE RESENDE, M. D. V., DA SILVA SANTOS, V., FERREIRA, C. F., OLIVEIRA, G. A. F., DA SILVA, M. S., DE OLIVEIRA, L. A., AGUILAR-VILDOSO, C. I., Genome-wide selection in cassava. **Euphytica**, v. 187, p. 263-276, 2012.

DEKKERS, J., Prediction of response to marker- assisted and genomic selection using selection index theory. **Journal of Animal Breeding and Genetics**, v. 124, p. 331-341, 2007.

DESTA, Z. A., ORTIZ, R., Genomic selection: genome-wide prediction in plant improvement. **Trends in plant science**, v. 19, p. 592-601, 2014.

ELSTON, R., A weight-free index for the purpose of ranking or selection with respect to several traits at a time. **Biometrics**, v., p. 85-97, 1963.

FALCONER, D., MACKAY, T., 1996. Introduction to Quantitative Genetics. Longman Scientific & Technical, Harlow, UK.

GODDARD, M., Genomic selection: prediction of accuracy and maximisation of long term response. **Genetica**, v. 136, p. 245-257, 2009.

GODDARD, M. E., HAYES, B. J., Genomic selection. **Journal of Animal breeding and Genetics**, v. 124, p. 323-330, 2007.

GRATTAPAGLIA, D., RESENDE, M. D. V., Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, p. 241-255, 2011.

HAYES, B. J., BOWMAN, P. J., CHAMBERLAIN, A. J., GODDARD, M. E., Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of dairy science**, v. 92, p. 433-443, 2009.

HAZEL, L. N., The genetic basis for constructing selection indexes. **Genetics**, v. 28, p. 476-490, 1943.

HEFFNER, E. L., JANNINK, J.-L., IWATA, H., SOUZA, E., SORRELLS, M. E., Genomic selection accuracy for grain quality traits in biparental wheat populations. **Crop Science**, v. 51, p. 2597-2606, 2011.

HEFFNER, E. L., SORRELLS, M. E., JANNINK, J.-L., Genomic selection for crop improvement. **Crop Science**, v. 49, p. 1-12, 2009.

JANNINK, J. L., LORENZ, A. J., IWATA, H., Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**, v. 9, p. 166-177, 2010.

KEMPTHORNE, O., NORDSKOG, A. W., Restricted selection indices. **Biometrics**, v. 15, p. 10-19, 1959.

KONIG, S., SWALVE, H., Application of selection index calculations to determine selection strategies in genomic breeding programs. **Journal of Dairy Science**, v. 92, p. 5292-5303, 2009.

LANDE, R., THOMPSON, R., Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, p. 743-756, 1990.

LEE, S. H., VAN DER WERF, J. H., HAYES, B. J., GODDARD, M. E., VISSCHER, P. M., Predicting unobserved phenotypes for complex traits from whole-genome SNP data. **PLoS genetics**, v. 4, p. e1000231, 2008.

LORENZ, A. J., CHAO, S., ASORO, F. G., HEFFNER, E. L., HAYASHI, T., IWATA, H., SMITH, K. P., SORRELLS, M. E., JANNINK, J.-L., 2 Genomic Selection in Plant Breeding: Knowledge and Prospects. **Advances in agronomy**, v. 110, p. 77, 2011.

LORENZANA, R. E., BERNARDO, R., Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and applied genetics**, v. 120, p. 151-161, 2009.

MASSMAN, J. M., JUNG, H.-J. G., BERNARDO, R., Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. **Crop Science**, v. 53, p. 58-66, 2013.

MEUWISSEN, T. H. E., HAYES, B. J., GODDARD, M. E., Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.

PESEK, J., BAKER, R., Desired improvement in relation to selection indices. **Canadian journal of plant science**, v. 49, p. 803-804, 1969.

PIRES, I. E., RESENDE, M. D. V., SILVA, R. L., RESENDE JÚNIOR, M., 2011. Genética florestal. Arka, Viçosa, MG.

RAMALHO, M., ABREU, A. D. F., SANTOS, J. D., NUNES, J., Aplicações da genética quantitativa no melhoramento de plantas autógamas. **UFPA, Lavras**, v., p. 522, 2012.

RESENDE, M. D., ROSA-PEREZ, J., Genética quantitativa e estatística no melhoramento animal. **Curitiba: UFPR**, v., p., 1999.

RESENDE, M. D. V., FONSECA, F., VIANA, S. J. M. S., PETERNELLI, L. A., RESENDE JR, M. F. R., DEL VALLE, P. M., Métodos estatísticos na seleção genômica ampla. **Embrapa Florestas-Documents (INFOTECA-E)**, v., p., 2011.

RESENDE, M. D. V., JÚNIOR, M. F. R. R., AGUIAR, A. M., ABAD, J. I. M., MISSIAGGIA, A. A., SANSALONI, C., PETROLI, C., GRATTAPAGLIA, D., Computação da seleção genômica ampla (GWS). **Embrapa Florestas. Documents**, v., p., 2010a.

RESENDE, M. D. V., RESENDE JUNIOR, M., AGUIAR, A. M., ABAD, J. I. M., MISSIAGGIA, A. A., SANSALONI, C., PETROLI, C., GRATTAPAGLIA, D., Computação da seleção genômica ampla (GWS). **Embrapa Florestas-Documents (INFOTECA-E)**, v., p., 2010b.

RESENDE, M. F. R., ALVES, A. A., SÁNCHEZ, C. F. B., RESENDE, M. D. V., CRUZ, C. D., 2013. Seleção genômica ampla, in: CRUZ, C. D., SALGADO, C. C., BHERING, L. L. (Eds.), *Genômica aplicada*. Suprema, Visconde do Rio Branco, MG, pp. 375-424.

SCHULTHESS, A. W., WANG, Y., MIEDANER, T., WILDE, P., REIF, J. C., ZHAO, Y., Multiple-trait-and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. **Theoretical and applied genetics**, v. 129, p. 273-287, 2016.

SMITH, H. F., A discriminant function for plant selection. **Annals of Human Genetics**, v. 7, p. 240-250, 1936.

SOLBERG, T. R., SONESSON, A. K., WOOLLIAMS, J. A., MEUWISSEN, T. H., Reducing dimensionality for prediction of genome-wide breeding values. **Genetics Selection Evolution**, v. 41, p. 29, 2009.

TAI, G., Index selection with desired gains [Potatoes]. **Crop Science**, v., p., 1977.

WILLIAMS, J., The evaluation of a selection index. **Biometrics**, v. 18, p. 375-393, 1962.

WOLD, S., SJÖSTRÖM, M., ERIKSSON, L., PLS-regression: a basic tool of chemometrics. **Chemometrics and intelligent laboratory systems**, v. 58, p. 109-130, 2001.

ZHONG, S., DEKKERS, J. C., FERNANDO, R. L., JANNINK, J.-L., Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, p. 355-364, 2009.

CAPÍTULO 1

COMPARAÇÃO ENTRE MÉTODOS UNI E MULTIVARIADOS APLICADOS À SELEÇÃO GENÔMICA AMPLA

RESUMO

A maioria dos estudos de seleção genômica ampla aplicados ao melhoramento de plantas, realiza a comparação entre diferentes métodos estatísticos com enfoque univariado. Todavia, é interessante a utilização de metodologias multivariadas, que possibilitam a seleção de genótipos superiores em vários caracteres de importância econômica simultaneamente. Nesse sentido, os objetivos deste estudo foram: i) comparar a acurácia e eficiência de seleção do método Multivariate Partial Least Square (Mpls) em relação aos métodos univariados de seleção genômica: Random Regression Best Linear Unbiased Predictor (RRblup), Bayesian Lasso (Blasso) e Univariate Partial Least Square (Upls); ii) elaborar índices de seleção genômica eficientes na identificação e seleção precoce de indivíduos geneticamente superiores. Dez populações F_2 com 800 indivíduos foram simuladas considerando quatro características com diferentes herdabilidades. Os dados simulados foram submetidos às análises de seleção genômica via RRblup, Blasso, Upls e Mpls. Quatro índices de seleção genômica foram elaborados pelo somatório dos efeitos dos marcadores obtidos para cada característica, ponderados pela respectiva variância residual. Foi elaborado um índice para cada metodologia avaliada, designados como: IRRblup, IBlasso, IUpls e IMpls. As comparações entre os métodos e os índices de seleção foram realizadas considerando o tempo de processamento, as acurácias de predição, os ganhos de seleção e os coeficientes de coincidência de seleção. O método Mpls foi similar ao método Upls para as características de menores herdabilidades, e foi menos eficiente quando comparado aos métodos RRblup e Blasso. Os índices de seleção genômica proporcionaram os maiores ganhos genéticos totais. O índice IRRblup e IMpls se destacaram devido à alta eficiência na seleção de indivíduos geneticamente superiores em menor tempo de

processamento. Portanto, os resultados desse estudo sugerem que os índices de seleção genômica propostos podem ser promissores para aplicação em programas de melhoramento genético de plantas.

Palavras-chave: Genética quantitativa; Biometria; Análise multivariada; Índice de seleção.

INTRODUÇÃO

A seleção genômica ampla (*Genome Wide Selection - GWS*), proposta por Meuwissen et al. (2001), tem como princípio a predição do valor genético genômico (GEBV) a partir da relação entre o fenótipo dos indivíduos e milhares de marcadores moleculares amplamente distribuídos no genoma. Essa alta densidade de marcadores aumenta a probabilidade de que cada QTL (*Quantitative Trait Loci*) de interesse esteja em desequilíbrio de ligação com pelo menos um marcador, possibilitando maior acurácia de predição e a seleção direta precoce, sem limitar o número de QTLs que governam a expressão de um caráter quantitativo (Bernardo e Yu, 2007; Meuwissen et al., 2001).

Atualmente a GWS vem sendo descrita por vários autores e proporcionando resultados promissores em diversas áreas da genética e melhoramento vegetal (Arruda et al., 2016; Bhering et al., 2015; de Azevedo Peixoto et al., 2017; Fernandes et al., 2017; Spindel et al., 2015). Existem diferentes métodos estatísticos aplicados à GWS (Resende et al., 2011), visto que a escolha e a eficiência de aplicação destes métodos depende principalmente da arquitetura genética da característica quantitativa em estudo (Daetwyler et al., 2013; de los Campos et al., 2013; Resende et al., 2013).

No melhoramento genético de plantas é interessante a utilização de metodologias multivariadas que possibilitam a seleção de genótipos superiores em vários caracteres de importância econômica simultaneamente (Cruz et al., 2012). Todavia, de modo geral, os métodos de GWS são aplicados em abordagem univariada, ou seja, avaliando uma característica de interesse por vez. Desta forma, é importante o desenvolvimento de estudos que verifiquem a eficiência da GWS multivariada em programas de melhoramento de plantas. Neste contexto, o método de regressão via Quadrados Mínimos Parciais (*Partial Least Square – PLS*) se apresenta como uma alternativa eficiente, visto que este permite a seleção multivariada (Azevedo et al., 2013).

O PLS apresenta como ideia geral a formação de componentes que capturem a maior quantidade de informações possíveis dispostas nas variáveis explicativas para prever a variável dependente principal (Boulesteix e Strimmer, 2007). Além disso, este método é eficiente para resolver os problemas da multicolinearidade e alta dimensionalidade ocasionada pelo maior número de marcadores em relação ao número de indivíduos genotipados (Azevedo et al., 2013).

Uma estratégia multivariada preconizada pelos melhoristas e biometristas é a utilização dos índices de seleção, que consistem em estabelecer um novo caráter a partir de uma combinação linear de todos os caracteres de importância, considerando que os coeficientes de ponderação são estimados de modo a maximizar a correlação entre o índice e os valores genéticos verdadeiros dos genótipos a serem selecionados (Cruz et al., 2012). Existem diversos estudos sobre a aplicação de índices de seleção ao melhoramento de plantas evidenciando a obtenção de ganhos em um conjunto de caracteres simultaneamente (França et al., 2016; Junqueira et al., 2016; Kumar et al., 2016; Silva et al., 2016; Silva et al., 2017; Vieira et al., 2016). Assim, atualmente 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).

Diante do exposto, a partir de dados simulados, este estudo foi realizado com os seguintes objetivos: i) comparar a acurácia e eficiência de seleção do método Multivariate Partial Least Square (Mpls) em relação aos métodos univariados: Random Regression Best Linear Unbiased Predictor (RRblup), Bayesian Lasso (Blasso) e Univariate Partial Least Square (Upls); ii) elaborar índices de seleção genômica eficientes na identificação e seleção precoce de indivíduos geneticamente superiores em diferentes características simultaneamente, para aplicação em programas de melhoramento vegetal.

MATERIAL E MÉTODOS

Simulação dos dados fenotípicos e genotípicos

Os dados fenotípicos e genotípicos para aplicação dos métodos estatísticos de GWS, foram simulados a partir do *software* Genes (Cruz, 2013). A simulação foi realizada conforme descrito por Peixoto et al. (2016).

Inicialmente foi simulado o genoma considerando 10 grupos de ligação, similar a uma espécie diploide $2n = 2x = 20$. Cada grupo de ligação foi simulado com tamanho de 100 cM, com 100 marcas moleculares codominantes espaçadas de forma equidistante (1 cM) ao longo do genoma, totalizando 1000 marcadores.

Após a simulação do genoma, foi realizada a simulação dos genitores considerando pais homocigotos contrastantes, de modo que para cada loco um pai foi designado como homocigoto do tipo A_1A_1 e o outro como homocigoto contrastante do tipo A_2A_2 . Desta forma, o cruzamento entre pai 1 e pai 2 gerou a população F_1 com todas as marcas em heterocigose (A_1A_2).

De posse do genoma e dos genitores, a partir da autofecundação dos indivíduos da população F_1 foram simuladas 10 populações de mapeamento do tipo F_2 , com 800 indivíduos cada. A formação de cada indivíduo das populações F_2 foi realizada de forma que cada indivíduo da população F_1 produzisse 5,000 gametas que se recombinaram ao acaso. As populações F_2 simuladas foram codificadas com 0, 1 e 2, sendo que 0 e 2 corresponderam aos indivíduos homocigotos (A_1A_1 ou A_2A_2) e 1 aos indivíduos heterocigotos (A_1A_2), para um determinado loco.

Em seguida foram simulados os fenótipos, ou seja, as características quantitativas, considerando distribuição binomial, ação gênica aditiva e ausência de dominância entre os alelos (modelo aditivo). Foram simulados quatro caracteres quantitativos controlados cada um por 200 locos, distribuídos de modo aleatório para cada característica. As características foram simuladas com herdabilidades 20%, 40%, 60% e 80% e designadas como C1, C2, C3 e C4,

respectivamente. Foram considerados dois alelos por loco, sem presença de QTLs de efeitos maiores. Assim o efeito de cada QTL foi definido por: $A_1A_1=\mu + a$; $A_1A_2=\mu$; $A_2A_2=\mu - a$, em que a expressa, na população F_2 , o efeito aditivo de cada gene. Desta forma, os fenótipos dos indivíduos (Y_i) foram gerados segundo o modelo:

$$Y_i = \mu + \sum_{j=1}^{200} \alpha_j + \varepsilon_i$$

em que μ é a média geral da característica; α_j é o efeito genético em cada loco; e ε_i o efeito ambiental.

Validação Cruzada

Para proceder com as análises de GWS foi necessário definir as populações de treinamento e validação. Para todas as metodologias de GWS avaliadas, foi realizada a validação cruzada considerando 5-folds com 50 repetições. Desta forma, cada população F_2 com 800 indivíduos foram divididas em 5 grupos iguais contendo 160 indivíduos cada. Assim, as populações de treinamento foram compostas por 640 indivíduos e utilizadas para estimar os efeitos dos marcadores, enquanto a concordância entre os valores genéticos preditos via estimativas provenientes da população de treinamento foram validadas em cada grupo de 160 indivíduos, considerados como população de validação.

Métodos de GWS utilizados

As análises de seleção genômica RRblup e Blasso foram realizadas no *software* Rbio (Bhering, 2017).

O método RRblup, baseado em modelos mistos, foi descrito por Meuwissen et al. (2001) e utiliza preditores do tipo BLUP, considerando que todos os marcadores possuem a mesma

variância (ausência de genes de efeitos maiores). O RRblup foi analisado utilizando a função `mixed.solve` do pacote RRblup (Endelman, 2011).

Já o método Blasso (Park e Casella, 2008) considera uma variância para cada marcador. Este método foi analisado utilizando a função BGLR, parte do pacote BGLR (Pérez & de los Campos, 2012). Foi definido burn-in de 20,000, thin de 10 e 100,000 iterações da Markov-Chain Monte Carlo (MCMC).

As análises dos métodos Upls e Mpls foram realizados no *software* R (R Core Team, 2015) utilizando o pacote pls e a função `plsr`. Para esses métodos, optou-se pelo número ótimo de componentes para predição dos GEBVs, que explicassem 80% da variação total dos marcadores (variável X). Desta forma, em média 30 componentes foram considerados na análise para as diferentes populações simuladas.

Para rodar as análises de seleção genômica foram utilizados dois computadores de alta performance com as seguintes especificações: Intel Xeon com processador E5-26 12ª geração 3.30 GHz, com memória RAM de 64 e 96 GB respectivamente, e hard drive de 1024 GB.

Elaboração dos índices de seleção genômica

O índice de seleção genômica proposto neste estudo, tem como princípio a estimativa do efeito final para cada marcador, a partir do somatório dos efeitos deste marcador para cada característica ponderado pela sua respectiva variância residual, como apresentado a seguir:

$$Ef_m = \sum_{i=1}^4 \frac{M_{mi}}{\sigma_i^2} = \frac{M_{m1}}{\sigma_1^2} + \frac{M_{m2}}{\sigma_2^2} + \frac{M_{m3}}{\sigma_3^2} + \frac{M_{m4}}{\sigma_4^2}$$

em que Ef_m é o efeito final do marcador m ponderado pelas quatro características avaliadas, considerando $m= 1, 2, \dots, 1000$ marcadores; M_{mi} é o efeito estimado para o marcador m para a característica i ($i= 1, 2, 3, 4$); e σ_i^2 é a variância residual obtida para a característica i .

Seguindo este princípio, foram elaborados quatro índices de seleção genômica, sendo um para cada método avaliado: Índice de seleção RRblup (IRRblup); Índice de seleção Blasso (IBlasso); Índice de seleção Upls (IUpls); e Índice de seleção Mpls (IMpls). Desta forma, os efeitos finais dos marcadores ($Ef_1, Ef_2, \dots, Ef_{1000}$) estimados pelo índice de seleção genômica proposto para cada método foram considerados para estimar os GEBVs de cada indivíduo.

Comparação entre os métodos e índices de seleção genômica

Com a finalidade de comparar as metodologias avaliadas neste estudo, foram determinadas as acurácias de predição fenotípica (ac_f) e genotípica (ac_g), como descrito nas equações abaixo:

$$ac_f = \frac{cor(VF, GEBV)}{\sqrt{h^2}}$$

$$ac_g = cor(VG, GEBV)$$

em que h^2 é a herdabilidade da característica; VF é o valor fenotípico; VG é o valor genético verdadeiro; e $GEBV$ é o valor genético genômico.

O tempo de processamento de cada método foi computado em segundos. Foi realizada uma análise de fatores AxB em delineamento inteiramente casualizado (DIC) e utilizado o teste de média de Tukey a 5% de probabilidade para comparação das acurácias e tempo de cada método de GWS, considerando cada característica avaliada.

Em seguida foram selecionados 80 (10%) indivíduos com maiores valores fenotípicos e valores genéticos verdadeiros. De posse dos indivíduos selecionados, foram estimados os ganhos de seleção (GS) pelo método de seleção direta e indireta, utilizando a seguinte equação:

$$GS \% = \frac{DS \times h^2 \times 100}{\bar{X}_o}$$

em que DS é o diferencial de seleção que foi estimado por:

$$DS = \bar{X}_s - \bar{X}_o$$

em que \bar{X}_s é a média dos indivíduos selecionados e \bar{X}_o é a média da população inicial.

Da mesma forma, foram selecionados 80 indivíduos com os maiores valores de GEBVs estimados a partir de cada método de GWS e índice de seleção. De posse dos indivíduos ranqueados a partir dos maiores GEBVs, foram calculados os respectivos GS utilizando os valores fenotípicos e genéticos verdadeiros.

É importante ressaltar que, para os cálculos de GS utilizando os valores genéticos verdadeiros simulados para cada característica, foi considerada herdabilidade igual a 100%, ou seja, $h^2 = 1$. Já para o cálculo dos GS utilizando os valores fenotípicos foram consideradas as herdabilidades simuladas 20%, 40%, 60% e 80% para as características C1, C2, C3 e C4, respectivamente.

Para comparar a eficiência de todas as metodologias avaliadas na identificação e seleção de indivíduos geneticamente superiores, foram estimados os coeficientes de coincidência de seleção (CS) entre os 80 indivíduos selecionados a partir dos GEBVs estimados por cada método e índices de seleção, bem como em relação aos indivíduos selecionados a partir do valor fenotípico e valor genético verdadeiro para as quatro características. Assim os CS foram estimados da seguinte forma:

$$CS \% = \frac{NS}{NT} \times 100$$

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

Todas as medidas utilizadas para comparação entre as diferentes metodologias avaliadas foram estimadas considerando a média das 10 populações simuladas para as quatro características com diferentes herdabilidades.

RESULTADOS

Estimativas do tempo de processamento e acurácia de seleção

É possível verificar na Tabela 1, que o tempo de processamento de análise do método Blasso foi significativamente maior que o tempo de análise dos demais métodos de GWS avaliados, pelo teste de Tukey ao nível de 5% de probabilidade, independente da herdabilidade da característica. Desta forma, considerando a menor demanda computacional (menor tempo de processamento), esses resultados sugerem que os métodos RRblup, Upls e Mpls foram mais vantajosos.

Tabela 1. Comparação entre os métodos de seleção genômica Random Regression Best Linear Unbiased Predictor (RRblup), Bayesian Lasso (Blasso), Univariate Partial Least Square (Upls) e Multivariate Partial Least Square (Mpls), avaliando o tempo de processamento de análise para as quatro características com diferentes herdabilidades.

Métodos	C1	C2	C3	C4
Blasso	304,27 ^a	295,89 ^a	273,24 ^a	253,57 ^a
RRblup	2,31 ^b	2,62 ^b	2,63 ^b	2,30 ^b
Upls	0,47 ^b	0,46 ^c	0,53 ^c	0,53 ^b
Mpls	0,72 ^b	0,98 ^{bc}	0,86 ^{bc}	0,94 ^b

*Médias seguidas pela mesma letra na coluna não diferem estatisticamente a nível de 5% de probabilidade pelo teste de Tukey. Características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

De modo geral, pelo teste de comparação de médias de Tukey, os métodos de seleção genômica apresentaram diferenças significativas para acurácia de seleção, considerando as quatro características com diferentes herdabilidades, a partir dos dados fenotípicos e genéticos verdadeiros (Tabela 2). Não houve diferença significativa entre os resultados obtidos pelos métodos RRblup e Blasso, em nenhuma das avaliações de acurácias de seleção. O método Upls proporcionou resultados inferiores aos demais métodos de GWS, para as acurácias de seleção, considerando as características de menores herdabilidades. Todavia, considerando a

característica de maior herdabilidade, as acurácias dos resultados obtidos pelo método Upls se igualam estatisticamente aos métodos RRblup e Blasso (Tabelas 2).

Tabela 2. Comparação entre os métodos de seleção genômica Random Regression Best Linear Unbiased Predictor (RRblup), Bayesian Lasso (Blasso), Univariate Partial Least Square (Upls) e Multivariate Partial Least Square (Mpls), avaliando as acurácias de seleção genotípica e fenotípica para as quatro características avaliadas.

Métodos	C1	C2	C3	C4
Acurácia Genotípica				
Blasso	0,91 ^a	0,94 ^a	0,96 ^a	0,98 ^a
RRblup	0,91 ^a	0,94 ^a	0,96 ^a	0,98 ^a
Upls	0,46 ^b	0,65 ^b	0,79 ^b	0,90 ^{ab}
Mpls	0,79 ^a	0,84 ^a	0,78 ^b	0,84 ^b
Acurácia Fenotípica				
Blasso	0,89 ^a	0,94 ^a	0,96 ^a	0,98 ^a
RRblup	0,89 ^a	0,94 ^a	0,96 ^a	0,97 ^a
Upls	0,46 ^b	0,64 ^b	0,79 ^b	0,90 ^{ab}
Mpls	0,80 ^a	0,84 ^a	0,79 ^b	0,82 ^b

*Médias seguidas pela mesma letra na coluna não diferem estatisticamente a nível de 5% de probabilidade pelo teste de Tukey. Características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

Diferente dos demais métodos, os resultados de acurácias para o método Mpls não foram crescentes de acordo com o aumento da herdabilidade da característica (Tabela 2). O método Mpls proporcionou resultados de acurácia estatisticamente iguais aos métodos RRblup e Blasso para os caracteres de menores herdabilidades. Contudo, para as características de maiores herdabilidades esses resultados foram inferiores aos obtidos pelos métodos RRblup e Blasso.

Ganhos de seleção via diferentes métodos e índices de seleção genômica

Na Tabela 3 estão apresentadas as estimativas de ganhos de seleção direta e indireta, considerando os 80 indivíduos selecionados a partir dos maiores valores fenotípicos e valores genéticos verdadeiros, sobre as quatro características avaliadas. Os maiores ganhos foram

obtidos via seleção direta. Como esperado, os ganhos diretos em cada característica obtidos a partir dos valores genéticos verdadeiros foram maiores que os ganhos com base nos valores fenotípicos.

Tabela 3. Ganhos de seleção diretos e indiretos, com base nos indivíduos selecionados a partir dos valores fenotípicos e valores genéticos verdadeiros, para as quatro características avaliadas com diferentes herdabilidades.

Seleção direta	Resposta indireta				Total
	C1	C2	C3	C4	
	Valor Fenotípico				
C1	1,68	0,47	0,74	1,07	3,96
C2	0,30	2,36	0,85	1,04	4,55
C3	0,35	0,68	2,89	1,60	5,52
C4	0,49	0,69	1,36	3,38	5,92
	Valor Genético Verdadeiro				
C1	3,79	2,25	2,33	2,38	10,75
C2	2,26	3,70	2,34	2,04	10,34
C3	2,34	2,36	3,73	2,53	10,95
C4	2,35	2,00	2,47	3,78	10,59

*Características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

Observando os resultados na Tabela 3, os maiores ganhos totais a partir da seleção considerando os valores fenotípicos foram para as características de maiores herdabilidades (C3 e C4), que ocasionaram ganhos totais 5,52% e 5,92%, atingindo apenas 50,41% e 54,31% dos ganhos totais máximos possíveis de serem alcançado com base nos valores genéticos verdadeiros para esses caracteres (10,95% e 10,59%).

Além disso, é possível notar na Tabela 3 que os ganhos totais obtidos para os caracteres de menores herdabilidades, considerando a seleção com base nos valores fenotípicos (3,96% e 4,55%), se distanciaram do máximo possível de ser alcançado a partir dos valores genéticos

verdadeiros (10,75% e 10,34%) à medida que a herdabilidade diminuiu, confirmando que a seleção fenotípica é menos eficiente para os caracteres de menores herdabilidades.

Na Tabela 4 estão apresentados os ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos 80 indivíduos ranqueados pelos GEBVs obtidos em cada método de GWS.

Tabela 4. Ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs obtidos pelos métodos de seleção genômica ampla, para as características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

	Características				Total
	C1	C2	C3	C4	
	RRblup				
C1	3,46	2,27	2,39	2,53	10,65
C2	2,30	3,51	2,26	2,07	10,15
C3	2,47	2,44	3,59	2,59	11,09
C4	2,43	2,06	2,45	3,70	10,63
	Blasso				
C1	3,46	2,29	2,40	2,54	10,69
C2	2,30	3,51	2,26	2,08	10,15
C3	2,45	2,43	3,60	2,58	11,05
C4	2,45	2,06	2,46	3,70	10,67
	Upls				
C1	1,94	1,17	1,26	1,44	5,82
C2	1,51	2,51	1,52	1,44	6,99
C3	1,94	1,91	3,04	2,10	8,99
C4	2,17	1,85	2,26	3,45	9,73
	Mpls				
C1	1,98	1,24	1,31	1,35	5,88
C2	1,97	3,18	1,94	1,77	8,87
C3	2,38	2,37	3,56	2,52	10,82
C4	2,37	2,05	2,47	3,67	10,57

*RRblup: Random Regression Best Linear Unbiased Predictor; Blasso: Bayesian Lasso; e Upls: Univariate Partial Least Square; Mpls: Multivariate Partial Least Square.

Com esses resultados (Tabela 4), é possível verificar que os métodos RRblup e Blasso proporcionaram ganhos diretos e indiretos, bem como ganhos totais, muito similares entre eles e dentro do esperado ao observar os ganhos máximos possíveis de serem obtidos via valores genéticos verdadeiros (Tabela 3). Estes métodos se destacaram pelos maiores ganhos totais obtidos na seleção para as características de menores herdabilidades (C1 e C2), demonstrando superioridade quando comparados com a seleção baseada apenas nos fenótipos dos indivíduos (Tabela 3).

É possível notar que os métodos Upls e Mpls, apresentaram os menores ganhos diretos para as quatro características avaliadas ao se comparar com as metodologias RRblup e Blasso (Tabela 4). Além disso, esses métodos foram eficientes apenas para as características de maiores herdabilidades, visto que os ganhos totais com a seleção tenderam a se aproximarem aos ganhos obtidos com a seleção fenotípica para os caracteres de menores herdabilidades (Tabela 3).

Os ganhos de seleção estimados nos valores fenotípicos e valores genéticos verdadeiros, com base nos 80 indivíduos selecionados com maiores GEBVs considerando os índices de seleção genômica estão expostos na Tabela 5. É possível observar que os ganhos totais obtidos a partir dos valores genéticos verdadeiros, ao se aplicar qualquer índice de seleção genômica proposto neste estudo (Tabela 5) são maiores que os ganhos totais obtidos pela seleção direta e indireta em qualquer metodologia apresentada anteriormente (Tabelas 3 e 4). Estes resultados sugerem superioridade dos índices de seleção em relação tanto ao método multivariado Mpls, como aos métodos univariados RRblup, Blasso e Upls.

Como esperado, apesar dos índices de seleção alcançarem maiores ganhos totais, os ganhos obtidos para cada característica são menores que os ganhos máximos possíveis de serem obtidos na seleção direta de cada característica com base nos valores genéticos verdadeiros (Tabela 3). Dentre os índices de seleção, o IUpls apresentou o menor ganho total ao selecionar

os indivíduos em relação aos valores genéticos verdadeiros, enquanto apresentou o maior ganho total com a seleção a partir dos valores fenotípicos (Tabela 5).

Tabela 5. Ganhos de seleção estimados nos valores fenotípicos e valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs obtidos pelos índices de seleção genômica, para as características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

Índices*	C1	C2	C3	C4	Total
Valor Fenotípico					
IBlasso	0,60	1,05	1,90	2,88	6,43
IRRblup	0,60	1,05	1,90	2,88	6,43
IUpls	0,65	1,09	1,94	2,95	6,63
IMpls	0,66	1,09	1,94	2,82	6,51
Valor genético verdadeiro					
IBlasso	2,73	2,55	3,06	3,57	11,90
IRRblup	2,73	2,55	3,05	3,57	11,90
IUpls	2,52	2,38	2,88	3,36	11,15
IMpls	2,72	2,60	3,10	3,50	11,92

* IRRblup: Índice de seleção genômica via RRblup; IBlasso: Índice de seleção genômica via Blasso; IUpls: Índice de seleção genômica via Upls; e IMpls: Índice de seleção genômica via Mpls.

Coincidências na seleção de indivíduos superiores via métodos de GWS e índices de seleção genômica

Os coeficientes de coincidência na seleção de 80 indivíduos entre os métodos de GWS, índices de seleção genômica, valores fenotípicos e valores genéticos verdadeiros para as características de menores (C1 e C2) e maiores herdabilidades (C3 e C4) estão apresentados nas Tabelas 6 e 7, respectivamente.

A coincidência estimada entre os indivíduos selecionados a partir dos valores fenotípicos e valores genéticos verdadeiros variou de 30,75% a 65,13% da menor para maior herdabilidade (Tabelas 6 e 7). Os métodos RRblup e Blasso foram mais eficientes na seleção dos indivíduos geneticamente superiores que a seleção baseada apenas nos fenótipos dos

indivíduos, principalmente para as características de menores herdabilidades. Este fato pode ser evidenciado na Tabela 6, ao verificar que as coincidências dos indivíduos selecionados pelos GEBVs obtidos pelos métodos RRblup e Blasso com o valor genético verdadeiro para a característica C1 são maiores que o dobro (70% para ambos os métodos) quando comparado a coincidência entre o valor fenotípico e o valor genético verdadeiro (30,75 %).

Tabela 6. Comparação entre os diferentes métodos e índices de seleção genômica, via análise de coincidência, para as características simuladas: C1 ($h^2 = 20\%$), acima da diagonal principal e C2 ($h^2 = 40\%$), abaixo da diagonal principal.

	VG	VF	Blasso	RRblup	Upls	Mpls	IBlasso	IRRblup	IUpls	IMpls
VG		30,75	70,00	70,00	32,25	33,75	48,00	47,88	43,50	47,63
VF	42,25		34,00	34,38	56,50	54,50	26,00	26,25	27,38	29,00
Blasso	76,38	44,25		99,13	36,88	38,13	50,88	50,75	46,25	50,00
RRblup	76,50	44,38	99,50		36,63	37,63	50,75	50,75	46,38	49,88
Upls	45,50	62,13	50,13	50,00		58,25	26,13	26,25	28,88	28,75
Mpls	63,00	46,88	69,50	69,63	50,38		26,88	26,88	27,88	29,63
IBlasso	45,75	29,25	48,13	47,88	32,50	40,25		99,25	76,25	88,00
IRRblup	45,88	29,38	48,25	48,00	32,75	40,25	99,25		76,38	88,25
IUpls	42,25	31,13	43,50	43,25	35,38	37,13	76,25	76,38		72,38
IMpls	47,00	30,50	48,88	48,88	34,00	42,75	88,00	88,25	72,38	

*VG: Valor genético verdadeiro; VF: Valor fenotípico; RRblup: Random Regression Best Linear Unbiased Predictor; Blasso: Bayesian Lasso; Upls: Univariate Partial Least Square; Mpls: Multivariate Partial Least Square; IRRblup: Índice de seleção genômica via RRblup; IBlasso: Índice de seleção genômica via Blasso; IUpls: Índice de seleção genômica via Upls; e IMpls: Índice de seleção genômica via Mpls.

Além disso, é possível notar que as coincidências obtidas entre os métodos RRblup e Blasso na seleção de 80 indivíduos, atingiram coeficientes maiores que 98%, independente da herdabilidade da característica (Tabelas 6 e 7). Estes resultados, confirmam mais uma vez que estes métodos, apesar de diferentes abordagens, apresentaram comportamento e resultados similares, sendo igualmente eficientes na identificação e seleção de indivíduos geneticamente superiores.

Tabela 7. Comparação entre os diferentes métodos e índices de seleção genômica, via análise de coincidência, para as características simuladas: C3 ($h^2 = 60\%$), acima da diagonal principal e C4 ($h^2 = 80\%$), abaixo da diagonal principal.

	VG	VF	Blasso	RRblup	Upls	Mpls	IBlasso	IRRblup	IUpls	IMpls
VG		54,38	81,25	80,63	58,63	78,88	58,00	57,88	53,75	58,38
VF	65,13		56,00	56,13	69,38	53,13	42,63	42,75	44,88	43,88
Blasso	85,13	68,50		98,75	61,63	85,00	60,75	60,50	56,13	60,63
RRblup	85,00	68,50	98,63		61,75	84,75	61,00	60,75	56,38	60,88
Upls	70,00	78,75	74,75	75,63		57,13	45,75	45,88	51,63	45,75
Mpls	82,38	63,88	87,75	88,13	70,13		57,75	57,88	53,13	59,50
IBlasso	75,50	61,50	80,38	80,13	67,63	77,63		99,25	76,25	88,00
IRRblup	75,50	61,38	80,63	80,38	67,75	77,75	99,25		76,38	88,25
IUpls	66,38	66,50	70,25	70,25	73,75	66,38	76,25	76,38		72,38
IMpls	73,00	58,50	74,63	74,38	62,00	76,25	88,00	88,25	72,38	

*VG: Valor genético verdadeiro; VF: Valor fenotípico; RRblup: Random Regression Best Linear Unbiased Predictor; Blasso: Bayesian Lasso; Upls: Univariate Partial Least Square; Mpls: Multivariate Partial Least Square; IRRblup: Índice de seleção genômica via RRblup; IBlasso: Índice de seleção genômica via Blasso; IUpls: Índice de seleção genômica via Upls; e IMpls: Índice de seleção genômica via Mpls.

Por outro lado, para a característica C1 (Tabela 6) as coincidências obtidas pelos indivíduos selecionados com base nos GEBVs, estimados pelos métodos Upls e Mpls, e os valores genéticos verdadeiros foram 32,25% e 33,75% respectivamente, sendo muito similares à coincidência obtida pelos indivíduos selecionados via valor fenotípico e valor genético verdadeiro (30,75%). Desta forma, os métodos Mpls e Upls apresentaram a mesma eficiência que a seleção fenotípica para caracteres de baixas herdabilidades, não justificando a aplicação dessas metodologias nesses casos.

Na maioria dos cenários é possível observar que as coincidências entre os indivíduos selecionados pelos índices de seleção genômica e os valores genéticos verdadeiros, independente da herdabilidade da característica, apresentaram coeficientes maiores em comparação com a coincidência de seleção entre os valores fenotípicos e os valores genéticos verdadeiros (Tabelas 6 e 7).

Dentre os índices propostos, a maior coincidência na seleção dos indivíduos superiores foi obtida entre os índices IBlasso e IRRblup (99,25%), que selecionaram praticamente os mesmos indivíduos. Apesar da baixa eficiência apresentada pelo método Mpls na seleção de indivíduos geneticamente superiores para as características de menores herdabilidades, o IMpls foi eficiente nesses cenários (Tabela 6). Já os indivíduos selecionados pelo IUpls resultaram nos mais baixos coeficientes de coincidência com os indivíduos selecionados via valores genéticos verdadeiros, comparado aos demais índices, sendo que para algumas características foi menor ou igual ao coeficiente de coincidência entre o valor fenotípico e valor genético verdadeiro (Tabelas 6 e 7).

DISCUSSÃO

Comparação da eficiência de aplicação dos métodos de GWS

Segundo Falconer e Mackay (1996) a seleção fenotípica no melhoramento de plantas é dificultada pois, a maioria dos caracteres de interesse são quantitativos, ou seja, são controlados por inúmeros genes, apresentam baixa herdabilidade e por consequência são fortemente influenciados pelo ambiente. Este fato pode ser evidenciado pelos resultados do presente estudo, visto que a seleção fenotípica demonstrou baixa eficiência na coincidência de seleção com os indivíduos geneticamente superiores, principalmente para caracteres de menores herdabilidades (Tabelas 6 e 7). Além disso, os ganhos diretos e indiretos de seleção via valores fenotípicos foram muito inferiores comparado aos ganhos genéticos máximos possíveis de serem obtidos com a seleção (Tabela 3), resultando em ganhos totais reduzidos em relação ao que poderia ser alcançado com a utilização de métodos de seleção mais acurados, como a GWS (Spindel et al., 2015; Zhang et al., 2016).

Os métodos Upls e Mpls permitiram alta redução da dimensionalidade (97%), que resultou em baixo tempo de processamento de análise, sendo mais vantajosos que o Blasso

(Tabela 1). Estas metodologias baseadas em regressão via quadrados mínimos parciais permitem que grandes quantidades de dados sejam rapidamente analisadas, apresentando alta eficiência estatística e velocidade computacional, para obter estimativas dos GEBVs dos indivíduos (Boulesteix e Strimmer, 2006; Solberg et al., 2009). Todavia, esses métodos não apresentaram maiores vantagens computacionais sobre o método RRblup, considerando a avaliação de 800 indivíduos e 1000 marcadores (Tabela 1).

De modo geral, o método Upls proporcionou baixas acurácias de seleção para as características de menores herdabilidades (Tabela 2). Solberg et al. (2009), a partir de dados simulados, compararam diferentes metodologias de GWS para aplicação no melhoramento animal, e verificaram superioridade da acurácia de seleção de um método bayesiano (Bayes B) em relação ao método Upls. Estes autores demonstraram que a acurácia do método Upls diminuiu para características de menores herdabilidades. Azevedo et al (2014) relataram baixas capacidades preditivas para o método Upls, demonstrando comportamento inferior deste método em relação ao RRblup.

Neste sentido, o método Upls foi muito similar a seleção fenotípica dos indivíduos, principalmente para as características de baixas herdabilidades, resultando em baixos ganhos totais (Tabelas 3 e 4). Por conseguinte, houve alta coincidência na seleção dos indivíduos superiores pelo método Upls e a seleção fenotípica, sendo esta coincidência maior que em relação aos valores genéticos verdadeiros dos indivíduos (Tabelas 6 e 7). Desta forma, esses resultados sugerem que o método Upls foi pouco eficiente na identificação e seleção acurada dos indivíduos geneticamente superiores. A ineficiência de predição deste método também foi reportada por Azevedo et al. (2014), em uma análise comparativa de seis metodologias de GWS aplicadas ao melhoramento animal.

No presente estudo, as acurácias de predição obtidas pelo método Mpls foram maiores ou iguais ao método Upls (Tabela 2). Azevedo et al. (2013), relataram maiores acurácias para

o método Mpls e inferiram que esta metodologia é mais apropriada para GWS que o método Upls, visto que o Mpls captura de forma realística a natureza das características, considerando as correlações entre elas.

Contudo, foi possível verificar neste estudo, que para as características de menores herdabilidades, ao considerar os ganhos de seleção (Tabela 4) e as coincidências na seleção dos indivíduos geneticamente superiores (Tabelas 6 e 7), o método Mpls teve comportamento similar ao Upls, resultando em baixos ganhos totais e baixa eficiência na identificação e seleção de indivíduos geneticamente superiores. Estes resultados indicam que para caracteres de baixa herdabilidade o método Mpls se aproxima da seleção fenotípica, sendo recomendado para aplicação na GWS apenas para caracteres de herdabilidade maior ou igual a 60%.

Os métodos RRblup e Blasso foram mais eficientes na seleção de indivíduos geneticamente superiores que os métodos Mpls e Upls, independente da herdabilidade da característica (Tabelas 6 e 7). Existem diversos estudos na literatura relatando a eficiência dos métodos de GWS com abordagens tradicionais e bayesianas no melhoramento de plantas (Bernardo e Yu, 2007; Crossa et al., 2010; Crossa et al., 2014; de Oliveira et al., 2012; Grattapaglia e Resende, 2011).

Neste estudo, foram simulados 200 QTLs para cada característica quantitativa, distribuídos aleatoriamente no genoma e sem presença de QTLs de efeitos maiores, que resultaram na similaridade entre os métodos Blasso e RRblup. De acordo com (Resende et al., 2013), considerando distribuição normal dos QTLs ao longo do genoma, os métodos baseados em modelos mistos são igualmente eficientes aos métodos bayesianos. Bhering et al. (2015), comparando os métodos RRblup e Blasso, verificaram que não houve diferença significativa entre os métodos para computo da acurácia de predição. Portanto, os resultados do presente estudo indicam que o método RRblup foi o mais eficiente por apresentar altos valores de

acurácias preditivas, altos ganhos e coincidência na seleção de indivíduos geneticamente superiores em menor tempo de processamento.

Viabilidade e eficiência de aplicação de índices de seleção genômica

Os índices de seleção propostos neste estudo foram mais eficientes e proporcionaram os maiores ganhos totais com a seleção quando comparados aos demais métodos univariados (Blasso, RRblup e Upls) e multivariado (Mpls) (Tabelas 4 e 5). Segundo Dekkers (2007), a aplicação da teoria de índices na GWS é promissora e pode aumentar significativamente os ganhos com a seleção, especialmente para características de baixas herdabilidades.

Todavia, os ganhos diretos estimados via índices de seleção para cada característica foram menores que aqueles obtidos pelos demais métodos de GWS (Tabelas 4 e 5). Estes resultados estão de acordo com a teoria de índices de seleção, que proporcionam ganhos totais relativamente superiores, ocasionando ganhos reduzidos em cada caráter individualmente, compensando essa redução pela melhor distribuição de ganhos favoráveis nos demais caracteres (Cruz et al., 2012). Neste sentido, em relação aos GEBVs, as coincidências obtidas para os índices de seleção foram sempre menores que as coincidências obtidas via métodos Blasso e RRblup (Tabelas 6 e 7). Entretanto é importante ressaltar que os indivíduos selecionados pelos índices apresentam maior desempenho em todas as características simultaneamente, diferente dos métodos de GWS que consideram melhoria em uma característica por vez.

O IMpls resultou em altos ganhos genéticos totais e alta eficiência na seleção e identificação dos indivíduos superiores, apresentando resultados similares ao IBlasso e IRRblup (Tabela 5). Segundo Junqueira et al. (2016), a utilização de modelos multivariados integrados a seleção genômica podem aumentar a acurácia de predição e conseqüentemente podem resultar em índices de seleção mais poderosos e confiáveis.

É importante ressaltar que para obtenção dos índices de seleção genômica propostos é necessário analisar os dados inicialmente via métodos de GWS para obtenção dos efeitos dos marcadores. Desta forma, considerando a baixa eficiência demonstrada pelo método univariado Upls, as menores coincidências entre os índices na seleção de indivíduos superiores são em relação ao IUpls (Tabelas 6 e 7). Assim, este índice apresentou a menor eficiência na identificação e seleção de indivíduos geneticamente superiores, resultando nos menores ganhos de seleção e nas menores coincidências ao se considerar o valor genético verdadeiro, sendo o menos indicado para aplicação em programas de melhoramento genético.

A aplicação de índices de seleção genômica é uma ferramenta promissora para ser aplicada em programas de melhoramento de plantas, uma vez que possibilita a seleção precoce direta, por meio de informações de marcadores moleculares, bem como a seleção de indivíduos superiores em um conjunto de caracteres de interesse econômico, em tempo reduzido (Ceron-Rojas et al., 2015; Dekkers, 2007). Portanto, de acordo com os resultados obtidos neste estudo, os índices IRRblup e o IMpls são os mais indicados, visto que estes índices apresentaram alta eficiência na identificação e seleção de indivíduos geneticamente superiores em menor tempo e demanda computacional.

CONCLUSÕES

O método Mpls apresentou menor eficiência que os métodos RRblup e Blasso na seleção de indivíduos geneticamente superiores em uma população F_2 . O RRblup é o método mais recomendado por apresentar altos valores de acurácias preditivas, altos ganhos e coincidência na seleção de indivíduos geneticamente superiores em menor tempo de processamento.

Os índices de seleção genômica elaborados neste estudo foram eficientes na identificação e seleção precoce de indivíduos geneticamente superiores em diferentes características simultaneamente, sendo adequados para aplicação nos programas de

melhoramento plantas. Os índices IRRblup e IMpls foram os mais vantajosos, devido a maior eficiência de seleção de indivíduos geneticamente superiores em menor tempo de processamento.

REFERÊNCIAS BIBLIOGRÁFICAS

ARRUDA, M. P., LIPKA, A. E., BROWN, P. J., KRILL, A. M., THURBER, C., BROWN-GUEDIRA, G., DONG, Y., FORESMAN, B. J., KOLB, F. L., Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum*). **Molecular Breeding**, v. 36, p. 1-11, 2016.

AZEVEDO, C., SILVA, F., RESENDE, M. D., LOPES, M., DUIJVESTIJN, N., GUIMARÃES, S., LOPES, P., KELLY, M., VIANA, J., KNOL, E., Supervised independent component analysis as an alternative method for genomic selection in pigs. **Journal of Animal Breeding and Genetics**, v. 131, p. 452-461, 2014.

AZEVEDO, C. F., SILVA, F. F., REZENDE, M. D. V. D., PETERNELLI, L. A., GUIMARÃES, S. E. F., LOPES, P. S., Uni and multivariate partial least squares applied to genomic selection for carcass traits in pigs. **Ciência Rural**, v. 43, p. 1642-1649, 2013.

BERNARDO, R., YU, J., Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v. 47, p. 1082-1090, 2007.

BHERING, L. L., Rbio: A tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology**, v. 17, p. 187-190, 2017.

BHERING, L. L., JUNQUEIRA, V. S., PEIXOTO, L. A., CRUZ, C. D., LAVIOLA, B. G., Comparison of methods used to identify superior individuals in genomic selection in plant breeding. **Genetics and molecular research: GMR**, v. 14, p. 10888, 2015.

BOULESTEIX, A.-L., STRIMMER, K., Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. **Briefings in bioinformatics**, v. 8, p. 32-44, 2006.

BOULESTEIX, A.-L., STRIMMER, K., Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. **Briefings in bioinformatics**, v. 8, p. 32-44, 2007.

CERON-ROJAS, J. J., CROSSA, J., ARIEF, V. N., BASFORD, K., RUTKOSKI, J., JARQUÍN, D., ALVARADO, G., BEYENE, Y., SEMAGN, K., DELACY, I., A genomic selection index applied to simulated and real data. **G3: Genes, Genomes, Genetics**, v. 5, p. 2155-2164, 2015.

CROSSA, J., DE LOS CAMPOS, G., PÉREZ, P., GIANOLA, D., BURGUEÑO, J., ARAUS, J. L., MAKUMBI, D., SINGH, R. P., DREISIGACKER, S., YAN, J., Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, v. 186, p. 713-724, 2010.

CROSSA, J., PÉREZ, P., HICKEY, J., BURGUEÑO, J., ORNELLA, L., CERÓN-ROJAS, J., ZHANG, X., DREISIGACKER, S., BABU, R., LI, Y., Genomic prediction in CIMMYT maize and wheat breeding programs. **Heredity**, v. 112, p. 48-60, 2014.

CRUZ, C. D., GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v. 35, p. 271-276, 2013.

CRUZ, C. D., REGAZZI, A. J., CARNEIRO, P. C. S., 2012. Modelos biométricos aplicados ao melhoramento genético. UFV, Viçosa.

DAETWYLER, H. D., CALUS, M. P., PONG-WONG, R., DE LOS CAMPOS, G., HICKEY, J. M., Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, v. 193, p. 347-365, 2013.

DE AZEVEDO PEIXOTO, L., LAVIOLA, B. G., ALVES, A. A., ROSADO, T. B., BHERING, L. L., Breeding *Jatropha curcas* by genomic selection: A pilot assessment of the accuracy of predictive models. **PloS one**, v. 12, p. e0173368, 2017.

DE LOS CAMPOS, G., VAZQUEZ, A. I., FERNANDO, R., KLIMENTIDIS, Y. C., SORENSEN, D., Prediction of complex human traits using the genomic best linear unbiased predictor. **PLoS Genet**, v. 9, p. e1003608, 2013.

DE OLIVEIRA, E. J., DE RESENDE, M. D. V., DA SILVA SANTOS, V., FERREIRA, C. F., OLIVEIRA, G. A. F., DA SILVA, M. S., DE OLIVEIRA, L. A., AGUILAR-VILDOSO, C. I., Genome-wide selection in cassava. **Euphytica**, v. 187, p. 263-276, 2012.

DEKKERS, J., Prediction of response to marker- assisted and genomic selection using selection index theory. **Journal of Animal Breeding and Genetics**, v. 124, p. 331-341, 2007.

ENDELMAN, J. B., Ridge regression and other kernels for genomic selection with R package rrBLUP. **The Plant Genome**, v. 4, p. 250-255, 2011.

FALCONER, D., MACKAY, T., 1996. Introduction to Quantitative Genetics. Longman Scientific & Technical, Harlow, UK.

FERNANDES, S. B., DIAS, K. O., FERREIRA, D. F., BROWN, P. J., Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. **Theoretical and Applied Genetics**, v., p. 1-9, 2017.

FRANÇA, Á. E. D. D., PARRELLA, R. A. D. C., SOUZA, V. F. D., BASTOS, G. Q., NUNES, J. A. R., SCHAFFERT, R. E., Simultaneous selection in sweet sorghum progenies by selection indices. **Pesquisa Agropecuária Brasileira**, v. 51, p. 1737-1743, 2016.

GRATTAPAGLIA, D., RESENDE, M. D. V., Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, p. 241-255, 2011.

HEFFNER, E. L., JANNINK, J.-L., IWATA, H., SOUZA, E., SORRELLS, M. E., Genomic selection accuracy for grain quality traits in biparental wheat populations. **Crop Science**, v. 51, p. 2597-2606, 2011.

JUNQUEIRA, V. S., PEIXOTO, L. D. A., LAVIOLA, B. G., BHERING, L. L., MENDONÇA, S., COSTA, T. D. S. A., ANTONIASSI, R., Bayesian Multi-Trait Analysis Reveals a Useful Tool to Increase Oil Concentration and to Decrease Toxicity in *Jatropha curcas* L. **PloS one**, v. 11, p. e0157038, 2016.

KUMAR, B., GULERIA, S. K., KHANORKAR, S. M., DUBEY, R. B., PATEL, J., KUMAR, V., PARIHAR, C. M., JAT, S. L., SINGH, V., YATISH, K., 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**, v. 67, p. 1087-1095, 2016.

MEUWISSEN, T. H. E., HAYES, B. J., GODDARD, M. E., Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.

PARK, T., CASELLA, G., The bayesian lasso. **Journal of the American Statistical Association**, v. 103, p. 681-686, 2008.

PEIXOTO, L., BHERING, L., CRUZ, C., Determination of the optimal number of markers and individuals in a training population necessary for maximum prediction accuracy in F2 populations by using genomic selection models. **Genetics and molecular research: GMR**, v. 15, p., 2016.

R CORE TEAM, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

RESENDE, M. D. V., FONSECA, F., VIANA, S. J. M. S., PETERNELLI, L. A., RESENDE JR, M. F. R., DEL VALLE, P. M., Métodos estatísticos na seleção genômica ampla. v., p., 2011.

RESENDE, M. F. R., ALVES, A. A., SÁNCHEZ, C. F. B., RESENDE, M. D. V., CRUZ, C. D., 2013. Seleção genômica ampla, in: CRUZ, C. D., SALGADO, C. C., BHERING, L. L. (Eds.), Genômica aplicada. Suprema, Visconde do Rio Branco, MG, pp. 375-424.

SILVA, L., RESENDE, R., FERREIRA, R., SILVA, G., KIST, V., BARBOSA, M., NASCIMENTO, M., BHERING, L., Selection index using the graphical area applied to sugarcane breeding. **Genetics and molecular research: GMR**, v. 15, p., 2016.

SILVA, L., TEODORO, P., PEIXOTO, L., ASSIS, C., GASPARINI, K., BARBOSA, M., BHERING, L., Selecting sugarcane genotypes by the selection index reveals high gain for technological quality traits. **Genetics and molecular research: GMR**, v. 16, p., 2017.

SOLBERG, T. R., SONESSON, A. K., WOOLLIAMS, J. A., MEUWISSEN, T. H., Reducing dimensionality for prediction of genome-wide breeding values. **Genetics Selection Evolution**, v. 41, p. 29, 2009.

SPINDEL, J., BEGUM, H., AKDEMIR, D., VIRK, P., COLLARD, B., REDOÑA, E., ATLIN, G., JANNINK, J.-L., MCCOUCH, S. R., 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 Genet**, v. 11, p. e1004982, 2015.

VIEIRA, R., ROCHA, R., SCAPIM, C., AMARAL JR, A., VIVAS, M., Selection index based on the relative importance of traits and possibilities in breeding popcorn. **Genet. Mol. Res**, v. 15, p. 1-10, 2016.

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

CAPÍTULO 2

COMPARAÇÃO ENTRE ÍNDICES APLICADOS À SELEÇÃO GENÔMICA AMPLA

RESUMO

A seleção simultânea de caracteres, integrada a seleção genômica ampla (GWS), é uma estratégia promissora em programas de melhoramento genético de plantas. Diante deste fato, os objetivos deste estudo foram: i) verificar a eficiência da seleção direta e indireta na GWS; ii) e comparar estratégias multivariadas aplicadas à GWS via índices de seleção. Dez populações F_2 com 800 indivíduos foram simuladas considerando quatro características com diferentes herdabilidades. Os dados simulados foram submetidos as análises de seleção genômica via métodos Random Regression Best Linear Unbiased Predictor (RRblup) e Multivariate Partial Least Square (Mpls). Cinco diferentes estratégias de índices de seleção aplicados à GWS foram elaboradas: i) ponderar os efeitos dos marcadores pela variância residual; ii) codificar e padronizar os efeitos dos marcadores; iii) aplicar média nos efeitos dos marcadores; iv) aplicar índice de Mulamba e Mock (1978) nos valores genéticos genômicos; v) codificar e padronizar os valores fenotípicos antes das análises de GWS. Foram considerados dois cenários de seleção. No primeiro cenário, foram selecionados os indivíduos com maiores valores fenotípicos, valores genéticos verdadeiros e valores genéticos genômicos para as quatro características avaliadas. Já no segundo cenário foi considerado diferente sentido de seleção para uma das características simuladas. As comparações entre os métodos e os índices de seleção foram realizadas considerando os ganhos de seleção e os coeficientes de coincidência de seleção. Foi aplicado o índice de Mulamba e Mock (1978) nos valores fenotípicos e valores genéticos verdadeiros. Os métodos de GWS foram mais eficientes que a seleção fenotípica, destacando-se o RRblup. A seleção direta e indireta baseada nos valores genéticos genômicos foi mais eficiente que a seleção fenotípica. Nenhuma das estratégias avaliadas foi eficiente para

o segundo cenário de seleção. Os índices ponderados pela variância residual apresentaram alta eficiência para aplicação na GWS, no entanto tenderam a maximizar os ganhos para as características de maiores herdabilidades. As estratégias de aplicar índices, via RRblup, a partir da média dos efeitos dos marcadores, dos valores fenotípicos codificados e padronizados e da aplicação do índice Mulamba e Mock nos valores genéticos genômicos, resultaram em altos ganhos de seleção e mais se aproximaram aos ganhos obtidos pelo índice de Mulamba e Mock aplicado aos valores genéticos verdadeiros. A estratégia de codificar e padronizar os efeitos dos marcadores proporcionou os menores ganhos genéticos totais. De modo geral, as estratégias multivariadas avaliadas, proporcionaram maior eficiência de seleção quando comparadas à seleção direta e indireta fenotípica, à seleção direta e indireta baseada nos valores genéticos genômicos e ao índice de Mulamba e Mock fenotípico. Portanto, esses resultados sugerem que as estratégias multivariadas, via índices de seleção integrados à GWS, são promissoras para aplicação em programas de melhoramento genético de plantas, visando a seleção precoce direta de várias características simultaneamente.

Palavras-chave: Genética quantitativa; Análise multivariada; Biometria; Índice de seleção.

INTRODUÇÃO

A busca por ferramentas que permitam a seleção precoce de genótipos superiores possibilitando maximizar os ganhos com a seleção em tempo reduzido é um dos principais propósitos nos programas de melhoramento genético de plantas. Com esse objetivo, a seleção genômica ampla (*Genome Wide Selection – GWS*), proposta por Meuwissen et al. (2001), vem despertando cada vez mais o interesse entre os melhoristas. A GWS apresenta como grande vantagem a seleção precoce direta baseada em informações genotípicas correlacionadas ao fenótipo desejado (Bernardo e Yu, 2007; Heffner et al., 2009; Lorenz et al., 2011; Meuwissen et al., 2001).

Além disso, em programas de melhoramento de plantas é de suma importância a aplicação de estratégias multivariadas para obtenção de ganhos com a seleção em várias características desejáveis simultaneamente. Neste contexto, a teoria de índices de seleção se destaca por permitir a formação de um caráter adicional estabelecido pela combinação linear ótima de várias características de interesse econômico (Cruz et al., 2012). Existem diversos estudos evidenciando a eficiência de aplicação de índices de seleção no melhoramento de plantas (França et al., 2016; Junqueira et al., 2016; Kumar et al., 2016; Silva et al., 2016; Silva et al., 2017; Vieira et al., 2016).

Neste sentido, é esperado que a implementação da teoria de índices na GWS seja uma proposta promissora para os programas de melhoramento de plantas para possibilitar a seleção precoce direta de indivíduos superiores em um conjunto de caracteres de interesse econômico, em tempo reduzido (Dekkers, 2007). Atualmente, existem alguns estudos descritos na literatura sobre a eficiência de aplicação da teoria de índices de seleção na GWS para obtenção de ganhos genéticos no melhoramento de plantas (Ceron-Rojas et al., 2015; Fernandes et al., 2017; Heffner et al., 2011).

Todavia, é importante destacar que diferentes índices de seleção resultam em diferentes alternativas de seleção, e ao se considerar um mesmo índice, existem diferentes cenários de seleção e diversas maneiras de tratar estatisticamente os dados (Cruz et al., 2012). Assim a aplicação de índices na GWS carece de maiores investigações para o ajuste de melhores estratégias que permitam ao melhorista a rápida e eficiente identificação de genótipos promissores e mais adequados aos propósitos dos programas de melhoramento de plantas (Ceron-Rojas et al., 2015).

Diante do exposto, a partir de dados simulados, utilizando os métodos Random Regression Best Linear Unbiased Predictor (RRblup) e Multivariate Partial Least Square (Mpls) em diferentes cenários de seleção, os objetivos deste estudo foram: i) verificar a eficiência da seleção direta e indireta na GWS; ii) e comparar estratégias multivariadas aplicadas à GWS via índices de seleção.

MATERIAL E MÉTODOS

Simulação dos dados fenotípicos e genéticos verdadeiros

Os dados fenotípicos e genotípicos para aplicação dos métodos estatísticos de GWS, foram simulados a partir do *software* Genes (Cruz, 2013). A simulação foi realizada conforme descrito por Peixoto et al. (2016).

Inicialmente foi simulado o genoma considerando 10 grupos de ligação, similar a uma espécie diploide $2n = 2x = 20$. Cada grupo de ligação foi simulado com tamanho de 100 cM, com 100 marcas moleculares codominantes espaçadas de forma equidistante (1 cM) ao longo do genoma, totalizando 1000 marcadores.

Após a simulação do genoma, foi realizada a simulação dos genitores considerando pais homocigotos contrastantes, de modo que para cada loco um pai foi designado como homocigoto

dominante (A_1A_1) e o outro como homozigoto recessivo (A_2A_2). Desta forma, o cruzamento entre pai 1 e pai 2 gerou a população F_1 com todas as marcas em heterozigose (A_1A_2).

A partir da autofecundação dos indivíduos da população F_1 foram simuladas 10 populações de mapeamento do tipo F_2 , com 800 indivíduos cada. A formação de cada indivíduo das populações F_2 foi realizada de forma que cada indivíduo da população F_1 produzisse 5.000 gametas que se recombinaram ao acaso. As populações F_2 simuladas foram codificadas com 0, 1 e 2, sendo que 0 e 2 corresponderam aos indivíduos homozigotos (A_1A_1 ou A_2A_2) e 1 aos indivíduos heterozigotos (A_1A_2), para um determinado loco.

Em seguida foram simulados os fenótipos, ou seja, as características quantitativas, considerando distribuição binomial, ação gênica aditiva e ausência de dominância entre os alelos (modelo aditivo). Foram simulados quatro caracteres quantitativos controlados cada um por 100 locos, distribuídos de modo aleatório para cada característica. As características foram simuladas com herdabilidades 20%, 40%, 60% e 80% e designadas como C1, C2, C3 e C4, respectivamente. Foram considerados dois alelos por loco, sem presença de QTLs de efeitos maiores. Assim o efeito de cada QTL foi definido por: $A_1A_1 = \mu + a$; $A_1A_2 = \mu$; $A_2A_2 = \mu - a$, em que a expressa, na população F_2 , o efeito aditivo de cada gene. Desta forma, os fenótipos dos indivíduos (Y_i) foram gerados segundo o modelo:

$$Y_i = \mu + \sum_{j=1}^{100} \alpha_j + \varepsilon_i$$

em que μ é a média geral da característica; α_j é o efeito genético em cada loco; e ε_i o efeito ambiental.

Validação Cruzada

Para proceder com as análises de GWS foi necessário definir as populações de treinamento e validação. Para todas as metodologias de GWS avaliadas, foi realizada a

validação cruzada considerando 5-folds com 10 repetições. Desta forma, cada população F_2 com 800 indivíduos foram divididas em 5 grupos iguais contendo 160 indivíduos cada. Assim, as populações de treinamento foram compostas por 640 indivíduos e utilizadas para estimar os efeitos dos marcadores, enquanto a concordância entre os valores genéticos preditos via estimativas provenientes da população de treinamento foram validadas em cada grupo de 160 indivíduos, considerados como população de validação.

Métodos de GWS utilizados

As análises de seleção genômica via método RRblup foram realizadas no *software* Rbio (Bhering, 2017). O método RRblup, baseado em modelos mistos, foi descrito por Meuwissen et al. (2001) e utiliza preditores do tipo BLUP, considerando que todos os marcadores possuem a mesma variância (ausência de genes de efeitos maiores). O RRblup foi analisado utilizando a função `mixed.solve` do pacote RRblup (Endelman, 2011).

As análises do método Mpls foram realizadas no *software* R (R Core Team, 2015), utilizando o pacote `pls` e a função `plsr`. Para esse método, optou-se pelo número ótimo de componentes para predição dos GEBVs, que explicassem 80% da variação total dos marcadores (variável X). Desta forma, em média 30 componentes foram considerados na análise para as diferentes populações simuladas.

Para a realização das análises de seleção genômica foram utilizados dois computadores de alta performance com as seguintes especificações: Intel Xeon com processador E5-26 12^o geração 3.30 GHz, memória RAM de 64 e 96 GB respectivamente, e hard drive de 1024 GB.

Elaboração dos índices aplicados a GWS de acordo com diferentes estratégias

Com o intuito de investigar a eficiência de aplicação de índices na GWS para identificação e seleção de indivíduos geneticamente superiores, foram consideradas cinco diferentes estratégias.

Estratégia 1: Efeitos dos marcadores ponderados pela variância residual

O índice de seleção foi baseado na estimativa do efeito final para cada marcador, a partir do somatório dos efeitos deste marcador para cada característica ponderado pela sua respectiva variância residual, como apresentado a seguir:

$$Ef_m = \sum_{i=1}^4 \frac{M_{mi}}{\sigma_i^2} = \frac{M_{m1}}{\sigma_1^2} + \frac{M_{m2}}{\sigma_2^2} + \frac{M_{m3}}{\sigma_3^2} + \frac{M_{m4}}{\sigma_4^2}$$

em que Ef_m é o efeito final do marcador m ponderado pelas quatro características avaliadas, considerando $m= 1, 2, \dots, 1000$ marcadores; M_{mi} é o efeito estimado para o marcador m para a característica i ($i= 1, 2, 3, 4$); σ_i^2 é a variância residual obtida para a característica i .

Seguindo este princípio, foram criados dois índices de seleção, sendo um para cada método avaliado: Índice de seleção ponderado pela variância residual RRblup (IRRblup) e Índice de seleção ponderado pela variância residual Mpls (IMpls). Desta forma, os efeitos finais dos marcadores ($Ef_1, Ef_2, \dots, Ef_{1000}$) estimados pelo índice de seleção proposto para cada método foram considerados para estimar os GEBVs de cada indivíduo.

Estratégia 2: Efeitos dos marcadores codificados e padronizados

O índice de seleção foi baseado na estimativa do efeito final para cada marcador, a partir do somatório dos efeitos dos marcadores codificados (ou centrados na média) e padronizados para cada característica, como apresentado a seguir:

$$Ef_m = \sum_{i=1}^4 \frac{(M_{mi} - \overline{M_{mi}})}{\sigma_i} = \frac{(M_{m1} - \overline{M_{m1}})}{\sigma_1} + \frac{(M_{m2} - \overline{M_{m2}})}{\sigma_2} + \frac{(M_{m3} - \overline{M_{m3}})}{\sigma_3} + \frac{(M_{m4} - \overline{M_{m4}})}{\sigma_4}$$

em que Ef_m é o efeito final do marcador m , considerando $m= 1, 2, \dots, 1000$ marcadores; M_{mi} é o efeito estimado para o marcador m para a característica i ($i= 1, 2, 3, 4$); $\overline{M_{mi}}$ é a média dos efeitos dos 1000 marcadores para a característica i ; σ_i é o desvio padrão obtido para os efeitos dos marcadores para a característica i .

Seguindo este princípio, foram criados dois índices de seleção, sendo um para cada método avaliado: Índice de seleção codificado e padronizado RRblup (IRbcp) e Índice de seleção codificado e padronizado Mpls (Iplcp). Desta forma, os efeitos finais dos marcadores ($Ef_1, Ef_2, \dots, Ef_{1000}$) estimados pelo índice de seleção proposto para cada método foram considerados para estimar os GEBVs de cada indivíduo.

Estratégia 3: Efeito médio dos marcadores

O índice de seleção foi estimado a partir da média dos efeitos dos marcadores para as quatro características avaliadas, como apresentado a seguir:

$$Ef_m = \frac{\sum_{i=1}^4 M_{mi}}{t} = \frac{(M_{m1} + M_{m2} + M_{m3} + M_{m4})}{4}$$

em que Ef_m é o efeito final do marcador m , considerando $m= 1, 2, \dots, 1000$ marcadores; M_{mi} é o efeito estimado para o marcador m para a característica i ($i= 1, 2, 3, 4$); t é o número total de características avaliadas ($t = 4$).

Seguindo este princípio, foram criados dois índices de GWS, sendo um para cada método avaliado: Índice de seleção efeito médio dos marcadores RRblup (IRbmed) e Índice de seleção efeito médio dos marcadores Mpls (IPlmed). Desta forma, os efeitos finais dos marcadores ($Ef_1, Ef_2, \dots, Ef_{1000}$) estimados pelo índice de seleção proposto para cada método foram considerados para estimar os GEBVs de cada indivíduo.

Estratégia 4: Índice de Mulamba e Mock aplicado aos GEBVs

O índice de seleção baseado em soma de ranks descrito por Mulamba e Mock (1978) foi aplicado na matriz dos GEBVs das quatro características, estimados pelos métodos de seleção genômica RRblup e Mpls.

Seguindo este princípio, foram criados dois índices de GWS, sendo um para cada método avaliado: Índice de Mulamba e Mock aplicado ao RRblup (MMRb) e Índice de Mulamba e Mock aplicado ao Mpls (MMPI).

Estratégia 5: Valores fenotípicos codificados e padronizados

O índice de seleção foi determinado pelo somatório dos valores fenotípicos codificados (ou centrados na média) e padronizados, antes da aplicação dos métodos de GWS. Desta forma, após aplicação do índice nos valores fenotípicos, os dados foram submetidos a análise de GWS pelos métodos RRblup e Mpls.

Seguindo este princípio, foram criados dois índices de GWS, sendo um para cada método avaliado: Índice RRblup a partir de valores fenotípicos codificados e padronizados (IndRb) e Índice Mpls a partir de valores fenotípicos codificados e padronizados (IndPI).

Comparação entre as estratégias multivariadas aplicadas à GWS

Com o propósito de comparar as estratégias avaliadas neste estudo, foram determinados dois cenários considerando diferentes sentidos de seleção. No primeiro cenário foram selecionados os indivíduos com maiores valores fenotípicos, genéticos verdadeiros e GEBVs para as quatro características avaliadas (C1, C2, C3 e C4). Já no segundo cenário foram selecionados os indivíduos com maiores valores para as características C1, C3 e C4, de herdabilidades iguais a 20%, 60% e 80%, respectivamente e selecionados os indivíduos com os menores valores para a característica C2, de herdabilidade igual a 40%.

Foram selecionados 80 indivíduos (10%) com melhor desempenho de acordo com cada cenário, considerando os valores fenotípicos e valores genéticos verdadeiros. De posse dos indivíduos selecionados foram estimados os ganhos de seleção (*GS*) pelo método de seleção direta e indireta, utilizando a seguinte equação:

$$GS \% = \frac{DS \times h^2 \times 100}{\bar{X}_o}$$

em que DS é o diferencial de seleção que foi estimado por:

$$DS = \bar{X}_s - \bar{X}_o$$

em que \bar{X}_s é a média dos indivíduos selecionados e \bar{X}_o é a média da população inicial.

Da mesma forma, foram selecionados 80 indivíduos com os melhores desempenhos para GEBVs obtidos pelos métodos de GWS e pelas estratégias multivariadas utilizadas, considerando os diferentes cenários. De posse dos indivíduos ranqueados a partir dos GEBVs, foram calculados os respectivos GS utilizando os valores fenotípicos e genéticos verdadeiros.

É importante ressaltar que, para os cálculos de GS utilizando os valores genéticos verdadeiros simulados para cada característica, foi considerada herdabilidade igual a 100%, ou seja, $h^2 = 1$. Já para o cálculo dos GS utilizando os valores fenotípicos foram consideradas as herdabilidades simuladas 20%, 40%, 60% e 80% para as características C1, C2, C3 e C4, respectivamente.

Para comparar a eficiência dos diferentes métodos e estratégias multivariadas na identificação e seleção de indivíduos geneticamente superiores, foram estimados os coeficientes de coincidência de seleção (CS) entre os 80 indivíduos selecionados a partir dos valores fenotípicos, genéticos verdadeiros e GEBVs, da seguinte forma:

$$CS \% = \frac{NS}{NT} \times 100$$

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

Além disso, com o intuito de comparar as diferentes estratégias multivariadas, o índice de soma de ranks (Mulamba e Mock, 1978) foi aplicado nas matrizes de valores fenotípicos (MMf) e valores genéticos verdadeiros (MMg) considerando as quatro características. Do

mesmo modo, como descrito acima, foram estimados os ganhos de seleção e as coincidências de seleção obtidas pelos índices MMf e MMg. Todas as medidas utilizadas para comparação entre as diferentes metodologias avaliadas foram estimadas considerando a média das 10 populações simuladas para as quatro características com diferentes herdabilidades.

RESULTADOS

Estimativas de ganhos de seleção via diferentes métodos e estratégias multivariadas aplicados à GWS

Na Tabela 8 estão apresentados os resultados de ganhos com a seleção direta e indireta a partir dos valores fenotípicos e genéticos verdadeiros, considerando o mesmo sentido de seleção para as quatro características avaliadas. É possível notar que os ganhos totais ao se realizar a seleção fenotípica são inferiores aos ganhos máximos possíveis de serem alcançados considerando os valores genéticos verdadeiros, sendo ainda menores para as características de menores herdabilidades.

Tabela 8. Ganhos de seleção diretos e indiretos estimados com base nos indivíduos selecionados a partir dos maiores valores fenotípicos e valores genéticos verdadeiros, para as quatro características com diferentes herdabilidades.

Seleção direta	Resposta indireta				Total
	C1	C2	C3	C4	
	Valor Fenotípico				
C1	6,84	2,29	2,15	3,96	15,24
C2	1,35	11,15	3,15	3,00	18,65
C3	1,14	2,72	13,50	5,04	22,39
C4	1,91	2,72	5,09	13,66	23,37
	Valor Genético Verdadeiro				
C1	15,29	11,01	7,19	10,26	43,75
C2	9,88	17,57	8,65	6,09	42,19
C3	6,39	8,86	17,32	7,39	39,96
C4	10,37	6,88	8,50	15,26	41,01

*Características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

Na Tabela 9 estão apresentados os resultados de ganhos com a seleção direta e indireta a partir dos valores fenotípicos e genéticos verdadeiros, considerando diferente sentido de seleção para a característica de herdabilidade igual a 40%. Ao selecionar os indivíduos com os menores valores genéticos verdadeiros foi possível alcançar alto ganho direto negativo com a seleção para a característica desejada (C2), no entanto a seleção indireta foi prejudicada, resultando em ganhos indiretos também negativos para as demais características.

Tabela 9. Ganhos de seleção diretos e indiretos estimados com base nos indivíduos selecionados a partir dos maiores valores fenotípicos e valores genéticos verdadeiros para as características C1, C3 e C4 e menores valores para a característica C2.

Seleção direta	Resposta indireta				Total
	C1	C2	C3	C4	
Valor Fenotípico					
C1	6,84	2,29	2,15	3,96	15,24
C2	-1,23	-10,91	-2,99	-2,53	17,65
C3	1,14	2,72	13,50	5,04	22,39
C4	1,91	2,72	5,09	13,66	23,37
Valor Genético Verdadeiro					
C1	15,29	11,01	7,19	10,26	43,75
C2	-9,54	-17,29	-8,48	-5,96	41,27
C3	6,39	8,86	17,32	7,39	39,96
C4	10,37	6,88	8,5	15,26	41,01

*Características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

Os resultados de ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos 80 indivíduos ranqueados pelos GEBVs obtidos pelos métodos RRblup e Mpls, considerando o mesmo sentido de seleção para as quatro características avaliadas, estão apresentados na Tabela 10. Esses resultados sugerem a superioridade dos métodos de GWS quando comparados à seleção fenotípica (Tabela 8), visto que os ganhos diretos e totais obtidos ao selecionar os 80 indivíduos com maiores GEBVs são superiores àqueles obtidos com a seleção fenotípica e se aproximam aos ganhos genéticos máximos

possíveis de serem alcançados com a seleção. Todavia esses resultados indicam que o método Mpls é eficiente apenas para os caracteres com maiores herdabilidades, uma vez que os ganhos diretos e totais obtidos para a característica de menor herdabilidade (C1) se aproxima da seleção baseada no valor fenotípico dos indivíduos (Tabela 8), não justificando sua aplicação nesses casos.

Tabela 10. Ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs obtidos pelos métodos Random Regression Best Linear Unbiased Predictor (RRblup) e Multivariate Partial Least Square (Mpls), para as quatro características avaliadas com diferentes herdabilidades.

Seleção Direta	Resposta indireta				Total
	C1	C2	C3	C4	
RRblup					
C1	13,85	10,48	6,82	10,46	41,61
C2	10,22	16,54	8,79	5,94	41,49
C3	6,54	8,75	16,73	7,58	39,60
C4	10,53	7,03	8,81	14,93	41,29
Mpls					
C1	8,59	6,82	3,78	5,68	24,87
C2	8,46	14,43	7,02	5,03	34,94
C3	6,50	8,83	16,45	7,61	39,39
C4	10,56	7,08	8,82	14,73	41,18

*Características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

Na Tabela 11 estão expostos os resultados de ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos 80 indivíduos ranqueados pelos GEBVs obtidos pelos métodos RRblup e Mpls, considerando diferente sentido de seleção para a característica de herdabilidade igual a 40%. Neste cenário (Tabela 11) é possível verificar que os resultados para os ganhos de seleção apresentaram o mesmo comportamento observado com a seleção direta e indireta a partir dos valores genéticos verdadeiros (Tabela 9), inviabilizando a seleção indireta, por resultar em ganhos indiretos negativos.

Tabela 11. Ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs para as características C1, C3 e C4 e em ordem crescente para a característica C2, considerando os métodos Random Regression Best Linear Unbiased Predictor (RRblup) e Multivariate Partial Least Square (Mpls).

Seleção Direta	Resposta indireta				Total
	C1	C2	C3	C4	
RRblup					
C1	13,85	10,48	6,82	10,46	41,61
C2	-9,82	-16,06	-8,45	-6,01	40,34
C3	6,54	8,75	16,73	7,58	39,60
C4	10,53	7,03	8,81	14,93	41,29
Mpls					
C1	8,59	6,82	3,78	5,68	24,87
C2	-7,86	-14,02	-6,34	-4,61	32,83
C3	6,50	8,83	16,45	7,61	39,39
C4	10,56	7,08	8,82	14,73	41,18

*Características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

O método RRblup proporcionou resultados mais próximos aos ganhos genéticos máximos possíveis de serem atingidos com a seleção dos 80 indivíduos com os menores GEBVs para a característica C2, visto que possivelmente por se tratar de uma característica de herdabilidade mediana, o método Mpls foi menos eficiente e resultou em menores ganhos totais (Tabela 11).

Na Tabela 12 estão dispostos os ganhos de seleção estimados nos valores genéticos verdadeiros, com base nos indivíduos selecionados a partir dos GEBVs obtidos pelas diferentes estratégias multivariadas, considerando o mesmo sentido de seleção para as quatro características avaliadas. As estratégias de aplicar a média nos efeitos dos marcadores (IRbmed), codificar e padronizar os valores fenotípicos antes de realizar as análises de seleção genômica (IndRb) e aplicar o índice de Mulamba e Mock nos GEBVs (MMRb), via método RRblup, proporcionaram os maiores ganhos totais (51,28%, 50,93% e 50,80%,

respectivamente), se aproximando dos resultados obtidos pelo índice de Mulamba e Mock (1978) aplicado aos valores genéticos verdadeiros (MMg), que proporcionou 51,80% de ganho genético total.

Tabela 12. Ganhos de seleção estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs obtidos pelos índices de seleção, para as características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

Índices*	C1	C2	C3	C4	Total
MMg	12,75	13,73	13,27	12,05	51,80
IRRblup	11,21	9,57	12,04	14,33	47,15
IMpls	11,34	9,89	12,24	13,96	47,43
IRbcp	5,27	6,71	5,53	2,69	20,20
Iplcp	-0,08	-1,27	7,13	6,28	14,76
IRbmed	12,58	13,65	13,29	11,76	51,28
IPlmed	11,66	12,38	11,49	10,53	46,06
MMRb	12,58	13,29	12,99	11,94	50,80
MMPI	11,56	12,23	12,59	11,05	47,43
IndRb	12,29	12,90	13,38	12,36	50,93
IndPI	10,80	11,13	11,71	11,22	44,86

*MMg: Índice Mulamba e Mock aplicado aos valores genéticos verdadeiros; IRRblup: Índice ponderado pela variância residual via RRblup; IMpls: Índice ponderado pela variância residual via Mpls; IRbcp: Índice efeitos de marcadores codificados e padronizados via RRblup; Iplcp: Índice efeitos de marcadores codificados e padronizados via Mpls; IRbmed: Índice média de efeitos de marcadores via RRblup; IPlmed: Índice média de efeitos de marcadores via Mpls; MMRb: Índice Mulamba e Mock aplicados aos GEBVs via RRblup; MMPI: Índice Mulamba e Mock aplicados aos GEBVs via Mpls; IndRb: Índice valores fenotípicos codificados e padronizados via RRblup; IndPI: Índice valores fenotípicos codificados e padronizados via Mpls.

É possível notar na Tabela 12 que a estratégia de codificar e padronizar os efeitos dos marcadores (IRbcp e Iplcp) proporcionou os menores ganhos genéticos totais (20,20 e 14,76 % respectivamente) quando comparada às demais estratégias multivariadas propostas. Além disso, essa estratégia aplicada ao método Mpls resultou em ganhos negativos para os caracteres de menores herdabilidades (-0,08% e -1,27%).

Observando os resultados proporcionados pela estratégia de ponderar os efeitos dos marcadores pela variância residual, considerando o mesmo sentido de seleção para todas as características, os índices IRRblup e IMpls pouco se diferenciaram quanto aos ganhos de seleção obtidos (Tabela 12). Além disso esses índices proporcionaram maiores resultados de ganhos para a característica de maior herdabilidade (C4)

Na Tabela 13 estão apresentados os ganhos de seleção estimados nos valores genéticos verdadeiros, com base nos indivíduos selecionados a partir dos GEBVs obtidos pelas diferentes estratégias multivariadas, considerando diferente sentido de seleção para a característica de herdabilidade igual a 40%.

Tabela 13. Ganhos de seleção diretos e indiretos estimados nos valores genéticos verdadeiros, com base nos indivíduos ranqueados em ordem decrescente de GEBVs para as características C1, C3 e C4 (herdabilidades 20%, 60% e 80%, respectivamente) e em ordem crescente para a característica C2 (herdabilidade igual a 40%), considerando os diferentes índices de seleção.

Índices*	C1	C2	C3	C4	Total
MMg	10,28	2,71	11,56	12,47	37,02
IRRblup	10,28	6,94	11,41	14,49	43,12
IMpls	10,21	6,87	11,69	14,11	42,88
IRbcp	2,42	-6,22	3,97	6,31	18,92
Iplcp	-0,77	-4,97	6,29	6,84	18,87
IRbmed	9,62	4,15	11,92	12,86	38,55
IPlmed	7,85	3,29	9,05	9,74	29,93
MMRb	10,69	2,48	11,61	12,82	37,60
MMPI	15,26	0,22	10,55	10,54	36,57
IndRb	9,76	4,91	12,27	13,26	40,20
IndPI	7,34	3,06	9,58	10,58	30,56

*MMg: Índice Mulamba e Mock aplicado aos valores genéticos verdadeiros; IRRblup: Índice ponderado pela variância residual via RRblup; IMpls: Índice ponderado pela variância residual via Mpls; IRbcp: Índice efeitos de marcadores codificados e padronizados via RRblup; Iplcp: Índice efeitos de marcadores codificados e padronizados via Mpls; IRbmed: Índice média de efeitos de marcadores via RRblup; IPlmed: Índice média de efeitos de marcadores via Mpls; MMRb: Índice Mulamba e Mock aplicados aos GEBVs via RRblup; MMPI: Índice Mulamba e Mock aplicados aos GEBVs via Mpls; IndRb: Índice valores fenotípicos codificados e padronizados via RRblup; IndPI: Índice valores fenotípicos codificados e padronizados via Mpls.

A partir dos resultados apresentados na Tabela 13 é possível verificar que apenas a estratégia de codificar e padronizar os efeitos dos marcadores proporcionou ganhos negativos na característica desejada (C2). Todavia esta estratégia proporcionou baixos ganhos para as demais características e os menores ganhos genéticos totais quando comparado as demais estratégias de seleção. De forma similar ao primeiro cenário (Tabela 12), o ganho de seleção obtido para a característica de menor herdabilidade via seleção dos indivíduos pelos GEBVs estimados pela estratégia de codificar e padronizar os efeitos dos marcadores via método Mpls foi negativo (-0,77).

De modo geral, nenhuma das estratégias de índices de seleção avaliadas neste estudo foram eficientes para maximizar os ganhos com a seleção em todas as características, ao se considerar o cenário com diferente sentido de seleção para a característica de herdabilidade igual a 40% (Tabela 13).

Coefficientes de coincidências entre as diferentes estratégias multivariadas na seleção de indivíduos geneticamente superiores

Na Tabela 14 estão apresentados os coeficientes de coincidência de seleção entre as diferentes estratégias multivariadas e os valores genéticos verdadeiros considerando o mesmo sentido de seleção para as quatro características avaliadas.

As menores coincidências de seleção com os indivíduos selecionados pelos valores genéticos verdadeiros (Tabela 14), independente da característica, foram verificadas para os métodos via estratégia de codificar e padronizar os efeitos dos marcadores (IRbcp e Iplcp). Esses resultados foram menores que as coincidências obtidas para o índice de Mulamba e Mock fenotípico (MMf) e sugerem que esta estratégia não é eficiente para seleção dos indivíduos geneticamente superiores, considerando o mesmo sentido de seleção para todas as características avaliadas.

Tabela 14. Análise de coincidência de seleção dos indivíduos selecionados em ordem decrescente de GEBVs estimados por cada índice de seleção e os indivíduos selecionados pelos maiores valores genéticos verdadeiros, para as características simuladas C1, C2, C3 e C4 com herdabilidades 20%, 40%, 60% e 80%, respectivamente.

Índices*	C1	C2	C3	C4
MMg	57,88	52,88	50,00	53,63
MMf	43,75	39,13	39,00	43,38
IRRblup	49,13	35,63	44,25	76,75
IMpls	50,00	36,75	46,13	72,13
IRbcp	22,75	23,75	23,63	16,75
Iplcp	10,63	7,00	26,25	27,63
IRbmed	58,13	54,38	51,88	53,75
IPImed	52,13	47,25	43,75	45,75
MMRb	56,75	50,63	48,75	52,75
MMPI	50,88	46,50	46,13	46,75
IndRb	55,75	49,50	52,00	57,63
IndPI	47,63	41,50	43,63	49,50

*MMg: Índice Mulamba e Mock aplicado aos valores genéticos verdadeiros; MMf: Índice Mulamba e Mock aplicado aos valores fenotípicos; IRRblup: Índice ponderado pela variância residual via RRblup; IMpls: Índice ponderado pela variância residual via Mpls; IRbcp: Índice efeitos de marcadores codificados e padronizados via RRblup; Iplcp: Índice efeitos de marcadores codificados e padronizados via Mpls; IRbmed: Índice média de efeitos de marcadores via RRblup; IPImed: Índice média de efeitos de marcadores via Mpls; MMRb: Índice Mulamba e Mock aplicados aos GEBVs via RRblup; MMPI: Índice Mulamba e Mock aplicados aos GEBVs via Mpls; IndRb: Índice valores fenotípicos codificados e padronizados via RRblup; IndPI: Índice valores fenotípicos codificados e padronizados via Mpls.

A estratégia de aplicar a média nos efeitos dos marcadores estimados via método RRblup (IRbmed) proporcionou a seleção de indivíduos com maiores coeficientes de coincidência com os indivíduos geneticamente superiores que aquelas obtidas pelo índice de Mulamba e Mock aplicado aos valores genéticos verdadeiros (MMg), para todas as características avaliadas (Tabela 14).

As estratégias de ponderar os efeitos dos marcadores pela variância residual, IRRblup e IMpls, apresentaram as maiores coincidências na seleção dos indivíduos geneticamente superiores ao se considerar a característica de maior herdabilidade (C4), proporcionando coeficientes de coincidência de 76,75% e 72,13% respectivamente. Por outro lado, a

coincidência obtida para a característica de herdabilidade igual a 40%, pelas estratégias IRRblup e IMpls, foram menores que as coincidências obtidas para o índice de Mulamba e Mock fenotípico (MMf) considerando a mesma característica (Tabela 14).

Na Tabela 15 estão apresentados os coeficientes de coincidência de seleção entre as diferentes estratégias multivariadas propostas e os valores genéticos verdadeiros considerando diferente sentido de seleção para a característica de herdabilidade igual a 40%.

Tabela 15. Análise de coincidência dos indivíduos selecionados em ordem decrescente de GEBVs para as características C1, C3 e C4 (herdabilidades 20%, 60% e 80%, respectivamente) e em ordem crescente para característica C2 (herdabilidade igual a 40%), estimados por cada índice de seleção e os indivíduos selecionados, seguindo o mesmo critério, pelos valores genéticos verdadeiros.

Índices*	C1	C2	C3	C4
	Valor Genético Verdadeiro			
MMg	43,00	3,38	41,13	56,13
MMf	26,13	3,75	32,13	36,75
IRRblup	44,13	0,63	41,75	79,00
IMpls	43,25	0,88	43,75	74,00
IRbcp	16,38	22,13	17,00	27,25
Iplcp	8,75	18,88	22,63	29,13
IRbmed	40,63	2,75	43,63	59,25
IPImed	32,75	4,63	33,13	42,13
MMRb	39,13	2,13	38,75	56,75
MMPI	30,25	3,63	35,38	43,75
IndRb	41,50	1,88	45,75	64,13
IndPI	29,38	3,88	35,00	45,50

*MMg: Índice Mulamba e Mock aplicado aos valores genéticos verdadeiros; MMf: Índice Mulamba e Mock aplicado aos valores fenotípicos; IRRblup: Índice ponderado pela variância residual via RRblup; IMpls: Índice ponderado pela variância residual via Mpls; IRbcp: Índice efeitos de marcadores codificados e padronizados via RRblup; Iplcp: Índice efeitos de marcadores codificados e padronizados via Mpls; IRbmed: Índice média de efeitos de marcadores via RRblup; IPImed: Índice média de efeitos de marcadores via Mpls; MMRb: Índice Mulamba e Mock aplicados aos GEBVs via RRblup; MMPI: Índice Mulamba e Mock aplicados aos GEBVs via Mpls; IndRb: Índice valores fenotípicos codificados e padronizados via RRblup; IndPI: Índice valores fenotípicos codificados e padronizados via Mpls.

De modo geral nenhuma das estratégias avaliadas nesse estudo foi eficiente para selecionar os indivíduos geneticamente superiores que apresentaram os menores valores genéticos verdadeiros para a característica de herdabilidade igual a 40%, evidenciado pelas baixas coincidências observadas para esta característica (Tabela 15). Neste cenário, as maiores coincidências obtidas com os indivíduos selecionados pelos menores valores genéticos verdadeiros (C2) foram obtidas pela estratégia de codificar e padronizar os efeitos dos marcadores via RRblup e Mpls, resultando em 22,13% e 18,88% para os índices IRRblup e IMpls, respectivamente.

A estratégia de ponderar os efeitos dos marcadores pela variância residual via métodos RRblup e Mpls (IRRRblup e IMpls), apresentaram as menores coincidências na seleção dos indivíduos geneticamente superiores ao se considerar a característica C2 (Tabela 15). Neste cenário de seleção, esta estratégia proporcionou as maiores coincidências ao considerar as demais características, principalmente para a característica de maior herdabilidade (C4). De modo geral, os resultados proporcionados pela estratégia de ponderar os efeitos dos marcadores pela variância residual, independentemente do método de GWS utilizado (IRRRblup e IMpls), pouco foi influenciada pela alteração no cenário de seleção (Tabelas 14 e 15).

Na Tabela 16 estão apresentados os coeficientes de coincidência entre todas as estratégias multivariadas avaliadas nos diferentes cenários quanto ao sentido de seleção das características. Independente do cenário, as coincidências de todos os índices propostos com o índice MMg foram maiores que a coincidência deste com o índice de Mulamba e Mock baseado em valores fenotípicos (MMf), exceto ao se considerar os índices baseados na codificação e padronização dos efeitos dos marcadores (IRbcp e Iplcp). De modo geral, as menores coincidências verificadas entre todas as estratégias foram para os índices IRbcp e Iplcp.

Tabela 16. Comparação entre as diferentes estratégias multivariadas, via análise de coincidência, para dois cenários de seleção: i) Cenário 1 (acima da diagonal principal): seleção em ordem decrescente dos indivíduos para todas as características avaliadas; ii) Cenário 2 (abaixo da diagonal principal): seleção em ordem decrescente para as características de herdabilidades 20%, 60% e 80% e em ordem crescente para herdabilidade 40%.

	MMg	MMf	IRRblup	IMpls	IRbcp	Iplcp	IRbmed	IPlmed	MMRb	MMPi	IndRb	IndPI
MMg		56,88	69,50	70,13	25,75	17,88	84,38	66,00	86,00	71,38	82,50	62,88
MMf	43,25		54,50	55,75	27,13	22,13	58,13	61,63	59,38	63,00	60,38	68,38
IRRblup	65,38	42,88		87,75	21,50	27,75	70,63	56,25	70,00	60,38	75,75	61,13
IMpls	64,75	45,25	85,75		22,38	30,63	71,00	60,25	70,38	65,25	75,75	61,63
IRbcp	38,50	33,63	31,88	31,38		33,25	28,75	29,50	29,00	27,75	28,13	27,38
Iplcp	32,88	30,63	33,63	36,88	53,88		18,63	21,38	18,13	22,75	21,25	22,13
IRbmed	72,25	47,50	74,00	73,13	45,25	37,63		68,13	90,13	72,13	92,38	64,63
IPlmed	46,50	51,00	47,00	53,00	35,88	34,13	54,00		67,50	85,38	68,88	68,13
MMRb	73,13	47,25	69,00	67,88	47,88	35,88	86,63	53,00		73,75	88,88	64,13
MMPi	54,25	52,00	51,13	57,63	38,50	37,75	60,63	80,88	61,38		73,63	68,13
IndRb	71,88	48,00	79,25	79,00	42,00	37,63	90,75	55,13	83,13	61,25		66,75
IndPI	49,63	61,88	50,75	52,63	34,38	32,25	54,50	53,88	53,63	54,63	54,88	

*MMg: Índice Mulamba e Mock aplicado aos valores genéticos verdadeiros; MMf: Índice Mulamba e Mock aplicado aos valores fenotípicos; IRRblup: Índice ponderado pela variância residual via RRblup; IMpls: Índice ponderado pela variância residual via Mpls; IRbcp: Índice efeitos de marcadores codificados e padronizados via RRblup; Iplcp: Índice efeitos de marcadores codificados e padronizados via Mpls; IRbmed: Índice média de efeitos de marcadores via RRblup; IPlmed: Índice média de efeitos de marcadores via Mpls; MMRb: Índice Mulamba e Mock aplicados aos GEBVs via RRblup; MMPi: Índice Mulamba e Mock aplicados aos GEBVs via Mpls; IndRb: Índice valores fenotípicos codificados e padronizados via RRblup; IndPI: Índice valores fenotípicos codificados e padronizados via Mpls.

As maiores coincidências com os indivíduos selecionados pelo índice de Mulamba e Mock genético (MMg) foram proporcionadas pelas estratégias de aplicar o índice de Mulamba e Mock nos GEBVs (MMRb), aplicar a média nos efeitos dos marcadores (IRbmed) e codificar e padronizar os valores fenotípicos antes de realizar as análises de seleção genômica (IndRb), considerando o método RRblup, para os dois cenários avaliados (Tabela 16). Além disso, os coeficientes de coincidência verificados entre essas três estratégias (MMRb, IRbmed e IndRb) foram de alta magnitude, variando de 83,13% a 92,38%.

DISCUSSÃO

Seleção direta e indireta na GWS

No presente estudo foi verificado que os ganhos obtidos com a seleção direta e indireta fenotípica foi muito inferior aos ganhos máximos possíveis de serem alcançados ao se considerar os valores genéticos verdadeiros, principalmente para os caracteres de baixas herdabilidades (Tabelas 8 e 9). Daí a importância da busca por ferramentas mais acuradas para maximizar os ganhos de seleção em programas de melhoramento. Neste sentido, a GWS se destaca pela reconhecida eficiência na seleção de caracteres de baixa herdabilidade e permite mais ciclos de seleção por unidade de tempo, resultando em maiores ganhos genéticos em comparação com a seleção fenotípica (Fernandes et al., 2017; Riedelsheimer et al., 2013; Sonesson e Meuwissen, 2009).

Na GWS os GEBVs são utilizados para classificar e selecionar indivíduos geneticamente superiores. Neste estudo, foi possível constatar superioridade dos métodos de GWS (Tabelas 10 e 11) quando comparados aos ganhos diretos e indiretos proporcionados pela seleção fenotípica (Tabelas 8 e 9). Fernandes et al. (2017) compararam a eficiência de várias estratégias de GWS para aumentar a acurácia de predição de uma característica principal, utilizando informações de características correlacionadas. Esses autores verificaram a maior eficiência da GWS indireta quando comparada à seleção fenotípica e concluíram que ao utilizar uma característica de fácil medição em fase precoce da planta, com alta herdabilidade e correlacionada a uma característica principal, é possível obter redução no tempo por ciclo e consequentemente nos custos de produção.

Os resultados deste estudo demonstraram que o método RRblup proporcionou elevados ganhos com a seleção direta e indireta (Tabelas 10 e 11), independente da herdabilidade da característica, proporcionando resultados mais próximos aos ganhos genéticos máximos possíveis de serem obtidos com a seleção (Tabelas 8 e 9). Do mesmo modo, o método RRblup

tem sido utilizado em diversas pesquisas de melhoramento de plantas e apresentado maior eficiência que a seleção fenotípica (de Azevedo Peixoto et al., 2017; de los Campos et al., 2013; Resende et al., 2012a; Resende et al., 2012b; Zhong et al., 2009).

Todavia, o método Mpls foi menos eficiente na seleção direta e indireta que o RRblup para caracteres de menores herdabilidades (Tabelas 10 e 11). Observando os resultados deste estudo, foi possível notar que os ganhos obtidos pelo método Mpls se aproximaram da seleção fenotípica ao considerar caracteres de baixas herdabilidades (Tabelas 8 e 9), sugerindo que sua aplicação na GWS é justificada apenas para caracteres de herdabilidades maiores que 60%. Azevedo et al. (2013) verificaram que o método Mpls foi 75% menos eficiente, em média, do que o RRblup na estimação dos GEBVs e dos efeitos de marcadores avaliando características de carcaça de suínos, com diferentes herdabilidades.

Ao se estimar os ganhos com a seleção dos 80 indivíduos com menores valores fenotípicos, genéticos ou genômicos para a característica de herdabilidade igual a 40%, houve inversão de sinais nos ganhos indiretos para as demais características (Tabelas 9 e 11). De acordo com Cruz et al. (2012), a seleção direta em certas características pode provocar alterações indesejáveis em outras quando houver correlações desfavoráveis, levando a baixo desempenho em características secundárias, de forma que a população melhorada poderá apresentar sérios problemas e conseqüentemente a rejeição pelos produtores.

Estratégias multivariadas via índices de seleção aplicadas à GWS

Na busca por metodologias eficientes na seleção precoce direta de vários caracteres de interesse econômico simultaneamente, cinco estratégias multivariadas via aplicação de índices na GWS, utilizando os métodos RRblup e Mpls, foram propostas. Segundo Ceron-Rojas et al. (2015) um índice de seleção genômica pode ser definido como uma combinação linear de

GEBVs, que utiliza informações de marcadores moleculares para prever o mérito genético e selecionar indivíduos em uma população não fenotipada.

De acordo com a teoria de índice de seleção genômica definida por Ceron-Rojas et al. (2015), a estratégia de codificar e padronizar os valores fenotípicos (IndRb e IndPI) antes da aplicação dos métodos de GWS, não se aplica ao conceito, visto que, de acordo com a teoria proposta por esses autores, o índice de seleção genômica é aplicado nos GEBVs. Todavia, esta estratégia considerou a combinação linear de todas as características para estimar os valores de GEBVs e proporcionou altos ganhos e coincidências na seleção de indivíduos geneticamente superiores, principalmente para o método RRblup (Tabelas 12 e 14). Segundo Fernandes et al. (2017), ao utilizar dados fenotípicos de características correlacionadas utilizadas na validação para estimar os efeitos dos marcadores e GEBVs, é possível alcançar aumentos substanciais na acurácia e ganho genéticos com a seleção.

De modo geral, os ganhos genéticos totais obtidos pelos diferentes índices de seleção propostos neste estudo tenderam a se aproximar aos ganhos obtidos pelo índice de Mulamba e Mock aplicado aos valores genéticos verdadeiros (MMg), destacando os índices IRbmed e IMMR (Tabela 12). Além disso, esses índices resultaram nas maiores coincidências com o MMg na seleção dos indivíduos, independente do cenário avaliado (Tabela 16). Ceron-Rojas et al. (2015) compararam a aplicação do índice de seleção genômica com o índice fenotípico baseado na teoria proposta por Smith (1936) e verificaram que esses índices proporcionaram ganhos de seleção semelhantes.

Os índices baseados na codificação e padronização dos efeitos dos marcadores (IRbcp e Iplcp), resultaram em baixos ganhos quando comparado ao índice MMg, se aproximando dos resultados obtidos pela seleção direta e indireta fenotípica (Tabelas 1 e 12). De acordo com Lorenz et al. (2011), é esperado que os índices de seleção fenotípica sejam mais precisos e resultem em maior resposta de seleção por ciclo de seleção do que os índices aplicados na GWS.

No entanto, independente do cenário, os índices IRbcp e Iplcp apresentaram também baixas coincidências na seleção dos indivíduos geneticamente superiores (Tabelas 14 e 15) e as menores coincidência com os demais índices de seleção (Tabela 16). Assim, esses resultados sugerem que, dentre as estratégias avaliadas, a codificação e padronização dos efeitos dos marcadores proporcionou a menor eficiência para aplicação na GWS.

Os índices baseados na ponderação dos efeitos dos marcadores pela variância residual (IRRblup e IMpls) tenderam a maximizar os ganhos em caracteres com maiores herdabilidades (Tabelas 12 e 13), fato esperado, visto que as características de maiores herdabilidades apresentam menores variâncias residuais. Além disso, as maiores coincidências ao selecionar os indivíduos geneticamente superiores, foram obtidas para esses índices considerando a característica de herdabilidade igual a 80% (Tabelas 14 e 15). De maneira geral, esses índices proporcionaram resultados eficientes ao se comparar com a seleção direta e indireta fenotípica ou com a aplicação dos métodos de GWS (Tabelas 8 e 10). König e Swalve (2009) verificaram que a aplicação de índices na GWS pode proporcionar melhores resultados para os caracteres de maiores herdabilidades e para minimizar este problema, os autores sugeriram a aplicação de índices com pesos econômicos mais elevados para as características de baixas herdabilidades.

Ao se verificar os ganhos obtidos com a seleção e as coincidências de seleção para os indivíduos geneticamente superiores, nenhuma das estratégias de índices proposta neste estudo foi eficiente ao se considerar o cenário com diferente sentido de seleção para a característica de herdabilidade igual a 40% (Tabela 13 e 15). Esses resultados sugerem que os dados simulados não foram eficientes para representar este cenário de seleção, que provavelmente podem ser explicados por altas correlações positivas entre as quatro características avaliadas. Neste sentido, é interessante a realização de novos estudos, a partir de dados simulados e reais, que possam demonstrar a eficiência dos índices propostos nesse cenário de seleção.

De modo geral, independente do cenário avaliado, as coincidências da maioria dos índices propostos foram maiores em relação ao índice MMg que em relação ao índice MMf (Tabela 16), indicando maior eficiência dos índices aplicados a GWS na seleção dos indivíduos geneticamente superiores quando comparado ao índice baseado apenas em valores fenotípicos. Heffner et al. (2011), também indicaram sucesso da aplicação de índices de seleção na GWS quando comparados aos índices de seleção aplicados a dados fenotípicos, e concluíram que mais pesquisas sobre a aplicação de estratégias multivariadas na GWS são necessárias, visto que um dos principais objetivos dos melhoristas é a maximização dos ganhos genéticos totais com a seleção em vários caracteres simultaneamente.

Portanto, os resultados apresentados nesse estudo sugerem que a aplicação de índices na GWS é uma proposta eficiente para obter ganhos em várias características simultaneamente, sendo promissora para aplicação em programas de melhoramento de plantas, visto que a maioria dos índices resultaram em ganhos superiores aos alcançados pela seleção direta e indireta fenotípica ou utilizando os métodos RRblup e Mpls.

CONCLUSÕES

A seleção direta e indireta utilizando métodos de GWS apresentou maior eficiência que a seleção fenotípica, principalmente para os caracteres de menores herdabilidades. Todavia, de modo geral, a utilização de índices de seleção genômica proporcionaram maiores ganhos genéticos totais que a seleção direta e indireta.

A estratégia de codificar e padronizar (IRbcp e Iplcp) resultou em menores ganhos totais e menores coincidências de seleção quando comparada aos demais índices propostos. Os índices baseados na estratégia de ponderar os efeitos dos marcadores pela variância residual (IRRblup e IMpls), tenderam a maximizar os ganhos em caracteres com maiores herdabilidades. Os índices IndRb, IRbmed e IMMR, apresentaram os maiores ganhos genéticos

totais com a seleção e foram os que mais se aproximaram do índice de Mulamba e Mock aplicado aos valores genéticos verdadeiros (MMg).

Nenhuma das estratégias de índices proposta neste estudo foi eficiente ao se considerar o cenário com diferente sentido de seleção para a característica C4.

REFERÊNCIAS BIBLIOGRÁFICAS

AZEVEDO, C. F., DE RESENDE, M. D. V., FONSECA, F., LOPES, P. S., GUIMARÃES, S. E. F., Regressão via componentes independentes aplicada à seleção genômica para características de carcaça em suínos. **Pesquisa Agropecuária Brasileira**, v. 48, p. 619-626, 2013.

BERNARDO, R., YU, J., Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v. 47, p. 1082-1090, 2007.

BHERING, L. L., Rbio: A tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology**, v. 17, p. 187-190, 2017.

CERON-ROJAS, J. J., CROSSA, J., ARIEF, V. N., BASFORD, K., RUTKOSKI, J., JARQUÍN, D., ALVARADO, G., BEYENE, Y., SEMAGN, K., DELACY, I., A genomic selection index applied to simulated and real data. **G3: Genes, Genomes, Genetics**, v. 5, p. 2155-2164, 2015.

CRUZ, C. D., GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v. 35, p. 271-276, 2013.

CRUZ, C. D., REGAZZI, A. J., CARNEIRO, P. C. S., 2012. Modelos biométricos aplicados ao melhoramento genético. UFV, Viçosa.

DE AZEVEDO PEIXOTO, L., LAVIOLA, B. G., ALVES, A. A., ROSADO, T. B., BHERING, L. L., Breeding *Jatropha curcas* by genomic selection: A pilot assessment of the accuracy of predictive models. **PloS one**, v. 12, p. e0173368, 2017.

DE LOS CAMPOS, G., HICKEY, J. M., PONG-WONG, R., DAETWYLER, H. D., CALUS, M. P., Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v. 193, p. 327-345, 2013.

DEKKERS, J., Prediction of response to marker- assisted and genomic selection using selection index theory. **Journal of Animal Breeding and Genetics**, v. 124, p. 331-341, 2007.

ENDELMAN, J. B., Ridge regression and other kernels for genomic selection with R package rrBLUP. **The Plant Genome**, v. 4, p. 250-255, 2011.

FERNANDES, S. B., DIAS, K. O., FERREIRA, D. F., BROWN, P. J., Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. **Theoretical and Applied Genetics**, v., p. 1-9, 2017.

FRANÇA, Á. E. D. D., PARRELLA, R. A. D. C., SOUZA, V. F. D., BASTOS, G. Q., NUNES, J. A. R., SCHAFFERT, R. E., Simultaneous selection in sweet sorghum progenies by selection indices. **Pesquisa Agropecuária Brasileira**, v. 51, p. 1737-1743, 2016.

HEFFNER, E. L., JANNINK, J.-L., IWATA, H., SOUZA, E., SORRELLS, M. E., Genomic selection accuracy for grain quality traits in biparental wheat populations. **Crop Science**, v. 51, p. 2597-2606, 2011.

HEFFNER, E. L., SORRELLS, M. E., JANNINK, J.-L., Genomic selection for crop improvement. **Crop Science**, v. 49, p. 1-12, 2009.

JUNQUEIRA, V. S., PEIXOTO, L. D. A., LAVIOLA, B. G., BHERING, L. L., MENDONÇA, S., COSTA, T. D. S. A., ANTONIASSI, R., Bayesian Multi-Trait Analysis Reveals a Useful Tool to Increase Oil Concentration and to Decrease Toxicity in *Jatropha curcas* L. **PloS one**, v. 11, p. e0157038, 2016.

KONIG, S., SWALVE, H., Application of selection index calculations to determine selection strategies in genomic breeding programs. **Journal of Dairy Science**, v. 92, p. 5292-5303, 2009.

KUMAR, B., GULERIA, S. K., KHANORKAR, S. M., DUBEY, R. B., PATEL, J., KUMAR, V., PARIHAR, C. M., JAT, S. L., SINGH, V., YATISH, K., 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**, v. 67, p. 1087-1095, 2016.

LORENZ, A. J., CHAO, S., ASORO, F. G., HEFFNER, E. L., HAYASHI, T., IWATA, H., SMITH, K. P., SORRELLS, M. E., JANNINK, J.-L., 2 Genomic Selection in Plant Breeding: Knowledge and Prospects. **Advances in agronomy**, v. 110, p. 77, 2011.

MEUWISSEN, T. H. E., HAYES, B. J., GODDARD, M. E., Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.

MULAMBA, N., MOCK, J., Improvement of yield potential of the ETO blanco maize (*Zea mays* L.) population by breeding for plant traits [Mexico]. **Egyptian Journal of Genetics and Cytology**, v., p., 1978.

PEIXOTO, L., BHERING, L., CRUZ, C., Determination of the optimal number of markers and individuals in a training population necessary for maximum prediction accuracy in F. v., p., 2016.

R CORE TEAM, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

RESENDE, M. D. V., RESENDE, M. F. R., SANSALONI, C. P., PETROLI, C. D., MISSIAGGIA, A. A., AGUIAR, A. M., ABAD, J. M., TAKAHASHI, E. K., ROSADO, A. M., FARIA, D. A., Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, v. 194, p. 116-128, 2012a.

RESENDE, M. F., MUÑOZ, P., RESENDE, M. D., GARRICK, D. J., FERNANDO, R. L., DAVIS, J. M., JOKELA, E. J., MARTIN, T. A., PETER, G. F., KIRST, M., Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). **Genetics**, v. 190, p. 1503-1510, 2012b.

RIEDELSCHEIMER, C., ENDELMAN, J. B., STANGE, M., SORRELLS, M. E., JANNINK, J.-L., MELCHINGER, A. E., Genomic predictability of interconnected biparental maize populations. **Genetics**, v. 194, p. 493-503, 2013.

SILVA, L., RESENDE, R., FERREIRA, R., SILVA, G., KIST, V., BARBOSA, M., NASCIMENTO, M., BHERING, L., Selection index using the graphical area applied to sugarcane breeding. **Genetics and molecular research: GMR**, v. 15, p., 2016.

SILVA, L., TEODORO, P., PEIXOTO, L., ASSIS, C., GASPARINI, K., BARBOSA, M., BHERING, L., Selecting sugarcane genotypes by the selection index reveals high gain for technological quality traits. **Genetics and molecular research: GMR**, v. 16, p., 2017.

SMITH, H. F., A discriminant function for plant selection. **Annals of Human Genetics**, v. 7, p. 240-250, 1936.

SONESSON, A. K., MEUWISSEN, T. H., Testing strategies for genomic selection in aquaculture breeding programs. **Genetics Selection Evolution**, v. 41, p. 37, 2009.

VIEIRA, R., ROCHA, R., SCAPIM, C., AMARAL JR, A., VIVAS, M., Selection index based on the relative importance of traits and possibilities in breeding popcorn. **Genet. Mol. Res**, v. 15, p. 1-10, 2016.

ZHONG, S., DEKKERS, J. C., FERNANDO, R. L., JANNINK, J.-L., Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, p. 355-364, 2009.