

**MÁGNO SÁVIO FERREIRA VALENTE**

**SELEÇÃO GENÔMICA EM DIFERENTES ESTRUTURAS POPULACIONAIS  
NO MELHORAMENTO VEGETAL**

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  
2014

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

T

V154s  
2014

Valente, Mágnio Sávio Ferreira, 1983-  
Seleção genômica em diferentes estruturas  
populacionais no melhoramento vegetal / Mágnio Sávio  
Ferreira Valente. - Viçosa, MG, 2014.  
ix, 29f. : il. ; 29 cm.

Orientador : José Marcelo Soriano Viana.  
Tese (doutorado) - Universidade Federal de Viçosa.  
Referências bibliográficas: f. 20-23.

1. Genômica. 2. Genética quantitativa. 3. Marcadores  
genéticos. 4. Modelos matemáticos. 5. Plantas -  
Melhoramento genético. I. Universidade Federal de Viçosa.  
Departamento de Biologia Geral. Programa de  
Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 576.5

**MÁGNO SÁVIO FERREIRA VALENTE**

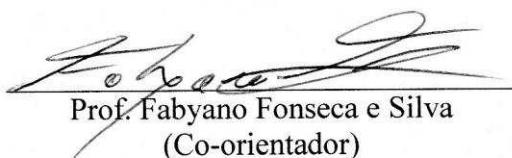
**SELEÇÃO GENÔMICA EM DIFERENTES ESTRUTURAS POPULACIONAIS  
NO MELHORAMENTO VEGETAL**

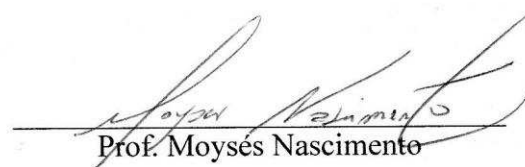
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.


Aprovada: 13 de março de 2014

  
Prof. Antonio Augusto Franco Garcia

  
Prof. Cosme Damião Cruz

  
Prof. Fabyano Fonseca e Silva  
(Co-orientador)

  
Prof. Moyses Nascimento

  
Prof. José Marcelo Soriano Viana  
(Orientador)

A DEUS.

Aos meus pais Geraldo e Dalva.

Ao meu irmão Halan.

À minha esposa Ana Paula e à minha filha Manuela.

## AGRADECIMENTOS

A Deus, pelas oportunidades recebidas e pela linda família que tenho.

À Universidade Federal de Viçosa, pela oportunidade de realizar este curso.

Ao CNPq, pelo auxílio financeiro.

Ao meu orientador, professor José Marcelo Soriano Viana, pela orientação, pela confiança e pela amizade.

Aos professores Marcos Deon Vilela de Resende e Fabyano Fonseca e Silva pelas valiosas sugestões, pela atenção e pelo apoio.

Aos meus pais, por serem um exemplo de vida e terem dedicado parte de suas vidas a mim e ao meu irmão.

Ao meu irmão Halan, pelo apoio e pela amizade.

À Ana Paula, pelo companheirismo e pelos cuidados comigo e com nossa filha Manuela.

Aos funcionários Vicente, Márcio e Antônio, pelos esforços despendidos nos trabalhos de campo.

Aos amigos de repúblicas, de graduação e pós-graduação, pela excelente convivência durante estes longos anos.

A todos que colaboraram para a realização deste trabalho, meu sincero agradecimento.

## **BIOGRAFIA**

MÁGNO SÁVIO FERREIRA VALENTE, filho de Geraldo Gandra Valente e Dalva Lúcia Ferreira A. Valente, nasceu na cidade de Viçosa - MG, em 10 de junho de 1983.

Concluiu o segundo grau na Escola Estadual Professor Samuel João de Deus, Paula Cândido – MG, em dezembro de 2001.

Em março de 2003, iniciou o curso de Agronomia na Universidade Federal de Viçosa, obtendo o diploma de Engenheiro Agrônomo em janeiro de 2008.

Magister Scientiae em Genética e Melhoramento de Plantas, no ano de 2010, pela Universidade Federal de Viçosa. Em março de 2010 iniciou o Programa de Doutorado em Genética e Melhoramento, pela mesma instituição, submetendo-se à defesa de tese em março de 2014.

## SUMÁRIO

<b>Resumo</b>	<b>vi</b>
<b>Abstract</b>	<b>viii</b>
<b>1. Introdução</b>	<b>1</b>
<b>2. Material e Métodos</b>	<b>4</b>
2.1. Simulação	4
2.2. Análises Estatísticas	6
<b>3. Resultados</b>	<b>8</b>
3.1. Estimação e avaliação dos efeitos dos SNPs na própria população.	8
3.2. Estimação e avaliação dos efeitos dos SNPs em diferentes populações.	10
3.3. Estimação e avaliação dos efeitos dos SNPs em diferentes amostras da população de referência	10
3.4. Avaliação do efeito das relações de parentesco sobre a eficiência da seleção genômica.	11
<b>4. Discussão</b>	<b>11</b>
4.1. Estimação e avaliação dos efeitos dos SNPs na própria população.	11
4.2. Estimação e avaliação dos efeitos dos SNPs em diferentes populações.	14
4.3. Estimação e avaliação dos efeitos dos SNPs em diferentes amostras da população de referência.	16
4.4. Avaliação do efeito das relações de parentesco sobre a eficiência da seleção genômica.	17
<b>5. Conclusão</b>	<b>18</b>
<b>6. Referências</b>	<b>20</b>
<b>Tabelas</b>	<b>24</b>

## RESUMO

VALENTE, Mágno Sávio Ferreira, D.Sc., Universidade Federal de Viçosa, março de 2014. **Seleção genômica em diferentes estruturas populacionais no melhoramento vegetal.** Orientador: José Marcelo Soriano Viana. Co-orientadores: Fabyano Fonseca e Silva e Marcos Deon Vilela de Resende.

A seleção genômica pode modificar significativamente a forma como é feita a seleção nos programas de melhoramento, aumentando a acurácia do processo seletivo e os ganhos por unidade de tempo. A eficiência da seleção genômica pode ser avaliada por meio da confiabilidade do valor genético (GBV) estimado por modelos de predição em comparação ao valor genético real do indivíduo. Embora a maioria dos estudos sobre a eficiência da seleção genômica considerem em seus modelos de análise apenas uma população e suas gerações, o pesquisador pode estar interessado em selecionar indivíduos de outras subestruturas da população de referência ou até mesmo de outras populações em que o desequilíbrio de ligação (LD) entre marcador e QTL pode ser diferente. No entanto, a obtenção dos GBVs em um contexto de utilização de múltiplas populações ou amostras pode não ser vantajosa, sendo necessário avaliar a perda de acurácia com a estimação e avaliação dos efeitos dos marcadores em diferentes cenários. Embora o conceito de seleção genômica dependa do LD entre QTL e marcadores, a confiabilidade do valor genético pode ser fortemente influenciada por diversos outros fatores e assim, não apresentar reais vantagens, por exemplo, em relação à seleção fenotípica. Para investigar isso, foram simuladas populações de milho pipoca com diferentes padrões de LD e variância genética. Foram considerados dois caracteres (produção de grãos e capacidade de expansão) determinados por 100 QTLs de mesmo efeito (10 QTLs por grupo de ligação), três densidades de SNPs (um SNP a cada 0,1, 1 e 10cM), duas herdabilidades (0,3 e 0,7) e seis populações representadas na geração 0 e considerando 5 gerações de acasalamento ao acaso, totalizando 144 cenários. Para cada cenário, 30 simulações foram realizadas e 500 plantas foram genotipadas. Em cenário adicional, a fim de avaliar o efeito das relações de parentesco na eficiência da seleção genômica, foram simuladas populações estruturadas em família S1 (FS1), famílias de meios irmãos (FMI), famílias de irmãos completos (FMI) e

em população de polinização aberta (OP). A acurácia de predição dos GBVs foi obtida pela correlação entre os valores genéticos paramétricos e os valores genéticos estimados por RR-BLUP. Como resultados, verificamos que a eficiência da seleção genômica em relação à seleção fenotípica é inversamente proporcional a herdabilidade e que para populações com maior LD, menores densidades de marcadores (1SNP/cM) podem ser usadas sem afetar drasticamente a acurácia de predição. No entanto, ao considerar populações de baixo LD e menor variância genética, o uso de maiores densidades de SNP (1SNP/0,1cM) seria necessária a fim de obter acurácia do valor genético maior que 0,55 e 0,70 em herdabilidade de 0,3 e 0,7, respectivamente. Nossos resultados também evidenciaram a necessidade de grande parte dos alelos presentes na população de seleção estejam representados na população de referência. Assim, os modelos de predição do GBV não podem ser usados em populações que apenas possuem estruturas genéticas similares à população onde os efeitos dos marcadores foram estimados, pois a acurácia do GBV tendeu a zero nestes casos. Além disso, quando diferentes amostras de uma mesma população foram usadas como população de referência e de seleção, houve redução de pelo menos 8 % de acurácia ao considerar a população de seleção em mesma geração da população de referência e redução de no mínimo 22% considerando distanciamento de 5 gerações de acasalamento ao acaso. A estruturação dos indivíduos em famílias de maior relacionamento resultou em maior eficiência da seleção genômica, assim como, um menor tamanho efetivo entre as populações teve impacto positivo sobre os valores de acurácia. Neste caso, a acurácia dos valores genéticos estimados em FS1, para capacidade de expansão em herdabilidade 0,3, foi superior em aproximadamente 15, 30 e 50% em relação à acurácia obtida para FIC, FMI e OP, respectivamente.

## ABSTRACT

VALENTE, Mágnio Sávio Ferreira, D.Sc., Universidade Federal de Viçosa, march, 2014. **Genomic selection in different population structures in plant breeding.** Advisor: José Marcelo Soriano Viana. Co-advisors: Fabyano Fonseca e Silva and Marcos Deon Vilela de Resende.

Genomic selection can change significantly the way in which the selection is made in breeding programs, by increasing the selective accuracy and earns per unit time. The efficiency of genomic selection can be evaluated through the reliability of the genomic breeding values (GBVs) estimated by prediction models compared to the true breeding values of the individuals. Although most studies about the efficiency of genomic selection consider in the analyses only models with one population and its generations, the researcher may be often interested in individuals from other substructures of the reference population, or even from other populations in which the linkage disequilibrium (LD) between marker and QTL may be different. However, the GBVs prediction in a context of using multiple populations or samples may not be advantageous, being necessary to calculate the decrease in accuracy with estimation and evaluation of the marker effects in different scenarios. Although the concept of genomic selection depends on LD between markers and QTL, the reliability of the breeding value can be strongly influenced by many other factors and thus, do not present real advantage in relation to phenotypic selection, for example. Aiming to investigate these facts, popcorn populations with different patterns of LD and genetic variances were simulated. The dataset refers to grain yield and expansion volume, both controlled by 100 QTLs with the same effect (10 QTLs on each linkage group), three SNP densities (one SNP each 0.1, 1, and 10 cM), two heritabilities (0.3 and 0.7) and six populations represented in generation 0 and considering 5 generations of random mating, totaling 144 scenarios. For each scenario, 30 simulations were carried out and 500 plants were genotyped. To evaluate the effect of family relationships on the efficiency of genomic selection, populations structured in S1, half-sib (HSF), full-sib families (FSF) and open-pollinated (OP) populations were simulated as an additional scenario. The prediction accuracy of the GBVs was obtained through the correlation between the true breeding values and breeding values predicted by RR-BLUP. The results showed that the efficiency of genomic selection in relation to

phenotypic selection is inversely proportional to the heritability and low marker densities (1SNP/cM) can be used without affect drastically the prediction accuracy for populations with high LD. However, high SNP density (1SNP/0.1cM) would be required to obtain prediction accuracy greater than 0.55 and 0.70, for the heritabilities of 0.3 and 0.7, respectively, in populations with low LD and genetic variance. Our results also showed that, most of the alleles present in the population under selection must be represented in the reference population. Thus, the prediction model of the GBVs cannot be used in populations who have only similar genetic structures in relation to the population where the marker effects were estimated, because the GBV accuracy tends to zero in these cases. Furthermore, when different samples of the same population were used both as reference and selection populations, there was at least 8% reduction in accuracy when compared to considering the selection population in the same generation of the reference population and, a minimum 22% reduction when considering an interval of five generations of random mating. Individuals structured in families with greater relationship resulted in greater efficiency of genomic selection, as well as, a lower effective size, and had a positive impact on the accuracy. In this case, the prediction accuracy using S1 families was approximately 15, 30 and 50% higher for expansion volume in heritability of 0.3, when compared to the accuracy obtained for FSF, HSP and OP populations, respectively.

## 1. Introdução

A seleção genômica ampla (Genome Wide Selection – GWS) é o processo de identificação dos indivíduos superiores com base nas predições de seus valores genéticos, obtidos pela análise de milhares de dados de polimorfismo de DNA e de um reduzido conjunto de dados fenotípicos (MEUWISSEN et al., 2001). Em razão da enorme quantidade de polimorfismo possível e da propriedade de codominância, os marcadores SNP (Single Nucleotide Polymorphism) estão entre os mais adequados a esse tipo de análise. Seu fundamento está na expectativa de desequilíbrio de ligação (LD) entre marcadores moleculares e regiões controladoras de características quantitativas (Quantitative trait loci – QTL). Devida à alta densidade de marcadores utilizada, espera-se que a seleção genômica seja capaz de capturar os efeitos de todos os QTLs determinantes do caráter e assim melhorar a predição dos valores genéticos.

Vários trabalhos têm abordado aspectos teóricos e práticos da GWS, indicando que o método é altamente acurado (MEUWISSEN et al., 2001; GRATTAPAGLIA e RESENDE, 2011) e pode ser efetivamente implementado como diferencial em programas de melhoramento animal e de espécies vegetais (HAYES et al., 2009; VANRADEN et al., 2009; RESENDE JR et al., 2012). Contudo, JONAS e DE KONING (2013) defendem que a implementação da seleção genômica em espécies vegetais deve ser associada a conceitos e métodos diferentes daqueles utilizados em experimentação com animais, principalmente devido à complexidade e variedade dos ciclos de melhoramento de plantas.

BERNARDO e YU (2007) foram um dos primeiros a propor a aplicabilidade da GWS no melhoramento vegetal. Estes autores obtiveram, via simulação de características quantitativas de milho, ganhos de seleção de 18 a 43% maiores com o uso da GWS em comparação ao uso da seleção recorrente assistida por marcadores moleculares. Avaliando a seleção genômica no melhoramento vegetal, JANNINK et al. (2010) concluíram que a seleção genômica, em comparação com a seleção com base em poucos QTLs de efeitos maiores, deverá proporcionar ganhos genéticos mais rápidos e com

menor custo, em razão da acentuada redução dos custos da genotipagem em relação aos custos da fenotipagem.

A GWS pode modificar significativamente a forma como é feita a seleção nos programas de melhoramento, aumentando a acurácia seletiva e os ganhos por unidade de tempo, além de reduzir os custos despendidos na experimentação (FRITSCHÉ-NETO et al., 2012). No entanto, diversos trabalhos têm demonstrado que a acurácia dos valores genéticos estimados pela GWS é influenciada por vários fatores, tais como, o LD, o tamanho efetivo populacional, a densidade dos marcadores, o número de QTLs envolvidos no controle da característica de interesse e também pela herdabilidade do caráter em análise (GODDARD, 2009; GRATTAPAGLIA e RESENDE, 2011).

Os modelos de predição, criados a partir da GWS utilizando a informação de dados genotípicos e fenotípicos, permitem que sejam usadas somente informações genotípicas para prever fenótipos futuros. Contudo, vale resaltar que os efeitos dos marcadores não serão necessariamente os mesmos em diferentes populações e ambientes, de modo que os modelos preditivos são, em sua maioria, para populações e ambientes específicos (RESENDE JR et al., 2012). A escolha da população de referência é importante aspecto a ser tratado, devida exercer grande influência na acurácia de predição dos modelos obtidos (PSZCZOLA et al., 2012), assim como limitar o tipo de população de seleção a ser utilizada (MASSMAN et al., 2013).

Embora a maioria dos estudos sobre a eficiência da GWS terem considerado apenas uma população e suas gerações, o pesquisador pode estar interessado em selecionar indivíduos de outras subestruturas da população de referência ou até mesmo de outras populações em que o LD entre SNP e QTL pode ser diferente. É difícil definir até quando o LD entre marcadores genéticos e QTL na população de referência persistirá para outros indivíduos ao longo das gerações e é ainda mais difícil se considerarmos candidatos à seleção não relacionados com a população onde os efeitos dos marcadores foram estimados. Tentando minimizar este problema, alguns autores tem sugerido combinar diferentes populações em um conjunto de estimação e obter os efeitos dos marcadores a partir de um maior número de fenótipos (DE ROOS et al., 2009; TOOSI et al., 2010). BERRY e KEARNEY (2011) consideram o aumento no tamanho da população de referência para o

cálculo dos efeitos dos marcadores uma forma viável de aumentar a acurácia de predição, sobretudo utilizando estratégias que possibilitem a redução dos custos de genotipagem e fenotipagem.

A estimação dos valores genéticos em um contexto de utilização de múltiplas populações ou amostras pode não apresentar reais vantagens, por exemplo, em relação à seleção fenotípica, sendo necessário avaliar a perda de acurácia com a estimação e avaliação dos efeitos dos marcadores em diferentes cenários. Maiores valores de acurácia têm sido frequentemente relatados quando a população de seleção é derivada, ou fortemente relacionada, com a população de referência (HABIER et al., 2010). O estreitamento genético entre as populações é tão importante que a reestimação dos efeitos dos marcadores deve ser realizada em gerações avançadas, a fim de evitar grande queda na acurácia de predição ao longo das gerações (MUIR, 2007). A queda do LD entre QTL e SNP é muitas vezes citada como principal fator na redução da acurácia ao longo das gerações, mas outros fatores, tais como, variância genética e parentesco entre indivíduos podem exercer grande influência e devem ser melhores estudados. São inegáveis os benefícios obtidos ao se utilizar populações de referência e de seleção fortemente relacionadas, mas até agora, pouco se sabe sobre a utilização destas populações na ausência de parentesco ou relacionamento, mas que possuem, por exemplo, valores de variância genética e LD similares.

Embora seja inquestionável a eficiência da seleção genômica no melhoramento genético, há escassez de informações na literatura sobre os efeitos na acurácia com a estimação dos valores genéticos em um contexto de utilização de múltiplas populações. Então, o nosso objetivo foi avaliar a eficiência da seleção genômica em diferentes cenários, com estimação e utilização dos efeitos dos marcadores em mesma e em diferentes populações, assim como, avaliar sua eficiência em relação à seleção fenotípica de acordo com o caráter, a herdabilidade, a densidade de marcadores, a geração, o LD e a variância aditiva. Outro objetivo deste estudo foi examinar o efeito das relações de parentesco entre os indivíduos na população de referência, assim como, a influência do tamanho efetivo desta sobre a eficiência da seleção genômica.

## 2. Material e Métodos

### 2.1. Simulação

Neste trabalho, o termo eficiência da seleção genômica é empregado para expressar a confiabilidade do valor genético obtido via RR-BLUP em comparação ao valor genético paramétrico. A eficiência da seleção genômica com utilização de diferentes populações de seleção foi estimada a partir da análise de dados simulados. Foi usado o programa de simulação – REALbreeding – desenvolvido pelo Prof. J. M. S. Viana, da Universidade Federal de Viçosa, Brasil, usando o software REALbasic 5.5. Foram considerados dois caracteres determinados por 100 QTLs, três densidades de SNPs, duas herdabilidades, seis populações representadas em duas gerações, totalizando 144 cenários. Para cada cenário foram realizadas 30 simulações.

Os 100 QTLs foram distribuídos em 10 grupos de ligação, com 10 QTLs por grupo e todos possuíam o mesmo efeito. Em cada grupo de ligação, foram inseridos 500, 51 ou 6 SNPs e a distância média entre SNPs adjacentes foi de 0,1, 1 e 10 cM, respectivamente. Então, em cada grupo de ligação, os QTLs e SNPs foram distribuídos em média em 50 cM. Duas características de milho pipoca foram simuladas com valores genotípicos mínimos e máximos dos homozigotos de 20 e 200 g planta<sup>-1</sup> e 5 e 50 mL g<sup>-1</sup> para produção de grãos e capacidade de expansão (CE), respectivamente. Dominância unidirecional positiva ( $0 < d/a \leq 1,2$ ) foi assumida para produção de grãos e dominância bidirecional ( $-1,2 \leq d/a \leq 1,2$ ) foi assumida para CE. As estimativas de herdabilidade ( $h^2$ ) foram de 0,3 e 0,7, obtendo assim, pela  $\sqrt{h^2}$ , acurácias dos valores fenotípicos de 0,55 e 0,84, respectivamente. O tamanho amostral considerado foi de 500 indivíduos genotipados.

Para este estudo, foi tomado, como população de referência (população 0), um composto obtido de duas linhagens com diferença mínima de frequência gênica de 0,6 para QTLs e SNPs. Populações adicionais foram consideradas com o intuito de observar o impacto na acurácia de predição com a utilização de diferentes populações de seleção. Cabe destacar que o tipo de população utilizada tem impacto relevante sobre os padrões de LD, assim como, sobre o

número de marcadores necessário para identificar genes que controlam características de interesse ao melhoramento. As populações adicionais incluíram populações com similar LD (população 1, 2, 3), menor LD (P4) e maior LD (P5) em relação à população de referência (Tabela 1). As estatísticas usadas para medir o LD foram a diferença entre as frequências gaméticas observada e esperada sob equilíbrio de ligação gênica ( $D$ ) e o quadrado da correlação entre alelos de dois locos ( $r^2$ ). A média dos valores de  $D$  nos grupos de ligação foram próximas a 0,09 (populações 0, 1, 2 e 3), 0,04 (P4) e 0,15 (P5), considerando a geração 0. Na geração 5, valores de  $D$  correspondentes tiveram redução e ficaram próximos a 0,05, 0,02 e 0,09. O uso de diferentes densidades de SNP não provocou grande variação do LD na maioria das situações. Outros parâmetros também foram considerados, como similar variância aditiva e de dominância (população 1, 4 e 5), menor variância aditiva (P2), e maiores valores de variância aditiva (P3) em relação à população de referência (Tabela 2).

Geração extra de cada população também foi avaliada. De cada população foi derivada uma nova população obtida após cinco gerações de acasalamentos ao acaso, sem seleção, mutação e migração. De modo geral, as simulações foram destinadas a representar populações de milho-pipoca, com diferentes padrões de LD e variância genética. No entanto, os nossos estudos são relevantes para todas as espécies onde a seleção genômica é viável e os fatores em estudos podem ser considerados.

Novo conjunto de dados foi tomado a fim de estudar o efeito das relações de parentesco entre indivíduos sobre a eficiência da seleção genômica. Estruturas de populações comumente usadas no melhoramento vegetal foram simuladas. Para tal, 500 indivíduos foram estruturados em famílias de meios irmãos (FMI), famílias de irmãos completos (FIC), famílias S1 (FS1) e população de polinização aberta (OP). Foram realizadas análises considerando 10 famílias com 50 indivíduos cada, gerando populações com tamanho efetivo ( $N_e$ ) de 38, 20, 10 para FMI, FIC e FS1, respectivamente. Análises de 20 famílias com 25 indivíduos também foram consideradas ( $N_e = 72, 38, 20$  foram obtidas para mesmas condições). As estimativas de tamanho efetivo foram obtidas pelas seguintes expressões (RESENDE e BARBOSA, 2005):

$$N_e = \frac{4N_f n}{n+3} : \text{para meios irmãos};$$

$$N_e = \frac{2N_f n}{n+1} : \text{para irmãos completos};$$

$$N_e = \frac{N_f n}{n+0,5} : \text{para irmãos S1};$$

onde  $n$  é o número de indivíduos por família e  $N_f$  o número de genitores femininos. Para a população de polinização aberta o tamanho efetivo considerado foi de 1000 indivíduos. Foram tomados dois caracteres (capacidade de expansão e produção de grãos) determinados por 100 QTLs distribuídos em 10 grupos de ligação (10 QTLs por grupo), duas herdabilidades (0,3 e 0,7) e 2000 marcadores SNPs com distância média entre marcadores adjacentes de 0,1cM. Para cada cenário foram realizadas 10 simulações. As simulações foram conduzidas visando à obtenção de magnitude similar de LD entre as populações. Deste modo, os valores de LD variaram aproximadamente de 0,03 a 0.04.

## 2.2. Análises Estatísticas

Vários métodos de predição de valores genéticos por meio de valores genômicos têm sido empregados em diferentes estudos e, de modo geral, dependendo da estrutura da população e da natureza do caráter, eles têm apresentado diferentes resultados. Contudo, devido à sua eficiência, fácil implementação, e ampla utilização em aplicações em vegetais, o método RR-BLUP foi o método adotado neste trabalho. Para a obtenção dos valores genômicos por RR-BLUP, foi utilizado o pacote rrBLUP (ENDELMAN, 2011) do software estatístico R (R Development Core Team, 2013). Foi assumindo o modelo aditivo e a acurácia de predição dos valores genéticos foi obtida pela correlação entre os valores genéticos paramétricos, gerados pelo REALbreeding, e os valores genômicos obtidos por RR-BLUP.

O modelo linear misto usado equivale a:

$$y = 1u + Zg + e$$

em que  $y$  é um vetor coluna de fenótipos,  $1$  é um vetor coluna que contém o número 1 representado  $N$  (número de indivíduos) vezes,  $Z$  é a matriz de incidência que aloca o genótipo de cada loco marcador a cada indivíduo,  $u$  é um escalar contendo o efeito fixo da média geral,  $g$  é um vetor dos efeitos aleatórios de marcadores e  $e$  é um vetor de erros aleatórios.

As equações de modelo misto equivalem a

$$\begin{bmatrix} 1'1 & 1'Z \\ Z'1 & Z'Z + I \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1'y \\ Z'y \end{bmatrix}$$

A variação genética contribuída por um loco marcador,  $\sigma_g^2$ , é dada por

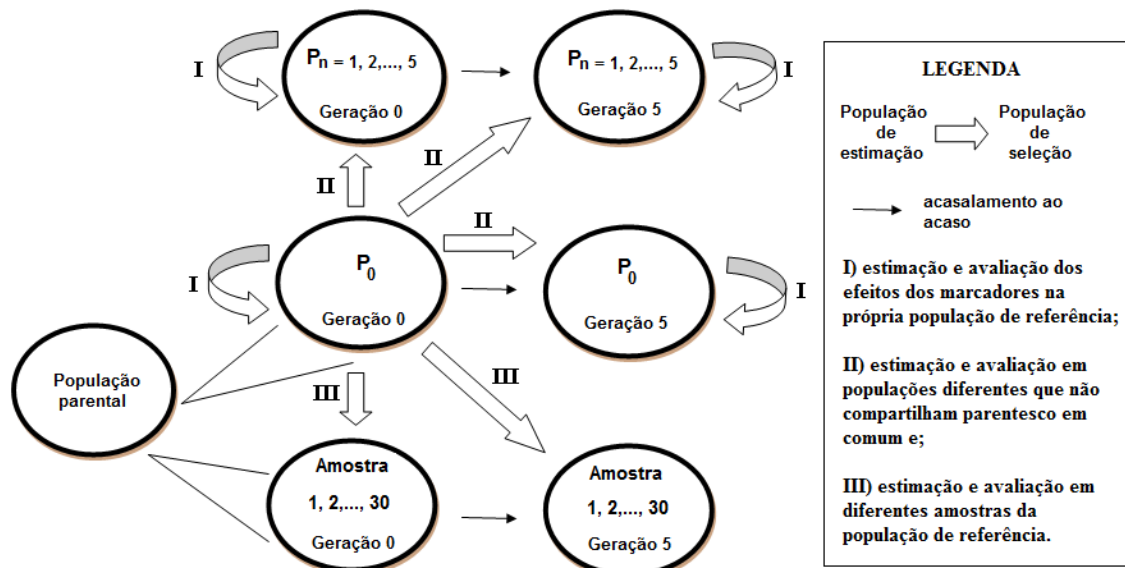
$\sigma_g^2 = \sigma_a^2 / [2 \sum_{i=1}^n p_i (1 - p_i)]$ , em que  $\sigma_a^2$  é a variância genética aditiva do caráter e  $p_i$  é a frequência alélica no loco marcador  $i$ .

Depois de estimados os valores genéticos dos marcadores em  $g$ , os valores aditivos genômicos dos indivíduos foram obtidos por  $\hat{a} = Z\hat{g}$ . A matriz de incidência  $Z$  contém os valores  $Z_{ij}$  iguais a 0, 1 e 2 para os marcadores dos tipos  $mm$ ,  $Mm$  e  $MM$ , respectivamente (RESENDE et al., 2012). Verifica-se que não há necessidade de uso da matriz de parentesco. A matriz de parentesco baseada em pedigree usada no BLUP tradicional é substituída por uma matriz de parentesco estimada pelos marcadores, sendo esta, função da própria matriz  $Z'Z$  presente nas equações de modelo misto.

A eficiência da seleção genômica foi obtida sob diferentes cenários (Figura 1): **I**) estimação e avaliação dos efeitos dos marcadores na própria população de referência; **II**) estimação e avaliação em populações que não compartilham parentesco em comum; **III**) estimação e avaliação em diferentes amostras da população de referência.

De acordo com o cenário de análise considerado, as populações foram usadas como população de referência sendo genotipadas e fenotipadas ou usada como população de seleção, sendo então, apenas genotipadas.

A avaliação do efeito das relações de parentesco sobre a estimação do valor genético (GBV) foi obtida assumindo estimação e avaliação dos efeitos dos marcadores na própria população.



**Figura 1** – Esquema representativo da avaliação da eficiência da seleção genômica em diferentes cenários.

O procedimento de estimação e avaliação de marcadores genéticos na mesma população é um procedimento pouco interessante em termos práticos devido ao alto custo de genotipagem e fenotipagem e por muitas vezes gerarem valores superestimados de acurácia quando o objetivo é a utilização do modelo em outra amostra. No entanto, este se mostra como uma ferramenta interessante em estudos de simulação, uma vez que, é possível isolar fatores discrepantes em comparação com a utilização de múltiplas populações. Os valores de acurácia obtidos neste tipo de análise devem ser entendidos como acurácia máxima de predição para o conjunto de dados em estudo.

### 3. Resultados

#### 3.1. Estimação e avaliação dos efeitos dos SNPs na própria população.

O aumento da herdabilidade proporcionou maiores valores de acurácia independentemente de outros fatores analisados (Tabela 3). O aumento da acurácia variou de 8 a 37%, valores estes abaixo do aumento da acurácia do valor fenotípico (53%), que passou de 0,548 para 0,837. A acurácia de predição para CE e produção de grãos teve acréscimo médio de 21% ao aumentar a herdabilidade de 0,3 para 0,7. Para mesma herdabilidade, tanto

menor densidade de SNP, como menor LD devido a cinco gerações de acasalamento ao acaso, proporcionaram menores acurácias. Podemos destacar que CE e produção de grãos tenderam a possuir valores similares de acurácia quando submetidos à mesma condição de análise.

Comparando populações com LD equivalentes (população 0, 1, 2 e 3), o aumento da variância aditiva observada na população 3 em contraste com as populações 0 e 1 (Tabela 2), proporcionou um aumento médio de acurácia de 19 e 14% nas gerações 0 e 5, respectivamente (Tabela 3). Comportamento inverso, contudo consistente, foi observado na população 2 ao apresentar menor variância aditiva comparada a aquelas das populações 0 e 1. Neste caso, houve uma perda média de acurácia de 36 e 7% nas gerações 0 e 5, respectivamente. Devido a maior diferença apresentada entre as variâncias aditivas nas populações 2 e 3, observamos diferença média de acurácia de 62 e 23% nas gerações 0 e 5, respectivamente.

A redução da distância entre marcadores adjacentes de 1 para 0,1 cM, não causou aumento relevante na acurácia de predição (8%) e, na maioria dos casos, os valores de acurácia estavam contemplados em ambos intervalos de confiança, independentemente da geração (Tabela 3). No entanto, o aumento da distância de 0,1 para 10 cM resultou em uma perda média da acurácia de 39% (variou de 10 a 87%) para CE e 35% (variou de 8 a 79%) para produção na geração 0 e de 64% para CE e produção na geração 5. Estes valores sugerem que aumentos na densidade de SNPs é vantajoso até 1 SNP/cM e densidades maiores podem não resultar em ganhos significativos de acurácia para mesmas condições de experimentação simuladas neste estudo.

Assim como observado em outros trabalhos, em populações com menores LD (população 4) e/ou menor variância genética (população 2), a acurácia de predição tende a ser mais afetada em baixa densidade de SNPs do que em condições opostas. Neste caso, a redução da distância entre SNPs adjacentes de 10 para 1 cM causaram acréscimo de acurácia de aproximadamente 10 e 15% (geração 0 e 5, respectivamente) na população 5 e valores correspondentes de 80 e 105% na população 4. O mesmo pode ser observado em relação a populações de similar LD com diferentes valores de variância aditiva, onde o aumento da densidade de SNP provocou acréscimos na acurácia de aproximadamente 18 e 35% na população 3 (geração 0 e 5,

respectivamente) e de 55 e 74% na população 2. Embora o acréscimo na acurácia com o aumento na densidade de SNP ter sido inversamente proporcional ao aumento de LD e variância aditiva, estes maiores acréscimos estão associados a menores valores absolutos de acurácia. Independente da densidade de marcadores utilizada, assim como outros fatores em estudo, a população de maior LD (P5) obteve as maiores acurácias em contraste com a de menor LD (P4), que obteve as menores acurácias. Para as populações com magnitude de LD similar, maiores acurácias foram obtidas para a população de maior variância aditiva (P3).

### 3.2. Estimação e avaliação dos efeitos dos SNPs em diferentes populações.

Ao obter os efeitos dos SNPs estimados na população de referência (P0, geração 0) e usá-los na estimação dos GBVs das demais populações e gerações, não houve correlação positiva significativa entre os valores estimados e os verdadeiros (Tabela 4). A exceção foi quando a população de seleção e a de referência eram a mesma população ou apenas distanciadas por 5 gerações de acasalamento ao acaso assumindo ausência de seleção, mutação, migração e deriva genética. No entanto, a acurácia de predição na geração 5 teve redução média de 23% (em  $h^2=0,3$ ) e 30% (em  $h^2=0,7$ ) em contraste com aqueles obtidos na própria geração.

Os valores de herdabilidade, densidade de SNP, LD, assim como a variância genética encontrada nas populações, não causaram efeitos sobre a acurácia de predição para os cenários onde a população de referência e a de seleção não eram derivadas de mesma amostra. Neste caso, na maioria das situações a acurácia foi próxima de zero.

### 3.3. Estimação e avaliação dos efeitos dos SNPs em diferentes amostras da população de referência.

É esperado perda de acurácia com a estimação e avaliação dos efeitos dos marcadores em diferentes amostras da população. Considerando a geração 0, tanto para CE como para produção de grãos, houve perda média de acurácia de 10 e 19% considerando herdabilidades de 0,3 e 0,7,

respectivamente (Tabela 5). Para a geração 5, as perdas foram de 26 e 34% em mesmas condições. No entanto, em 11 dos 24 cenários simulados, o intervalo de confiança continha mesmos valores de acurácia quando candidatos à seleção pertenciam à amostra de estimação ou apenas tiveram origem de mesma população.

#### 3.4. Avaliação do efeito das relações de parentesco sobre a eficiência da seleção genômica.

Os valores de acurácia obtidos foram inversamente proporcionais ao tamanho efetivo encontrado nas diferentes famílias. No entanto, não houve grande diferença em termos percentuais ao quase duplicar o tamanho efetivo na mesma população. Para todos os cenários simulados a estruturação dos indivíduos em FS1 obteve os maiores valores de acurácia, seguida de FIC, FMI e por último a OP (Tabela 6).

Para capacidade de expansão na herdabilidade 0,3, a acurácia dos GBVs estimados em FS1 foi superior em aproximadamente 15, 30 e 50% em relação à acurácia obtida para FIC, FMI e OP, respectivamente.

## 4. Discussão

### 4.1. Estimação e avaliação dos efeitos dos SNPs na própria população.

Independentemente do caráter em estudo, o aumento da acurácia de predição foi proporcional ao aumento da herdabilidade, embora nem sempre associada a altos incrementos. Com a redução da herdabilidade, menor proporção da variância genética aditiva é explicada por marcadores genéticos, assim menores acurácias são obtidas (GODDARD e HAYES, 2007). No entanto, em diversos estudos, a herdabilidade das características analisadas assim como o número de QTLs controladores destas, têm-se mostrado de menor influência sobre a acurácia da GWS. Trabalhos como os de CALUS et al. (2008) e de GRATTAPAGLIA e RESENDE (2011), demonstraram que a GWS é eficiente em selecionar indivíduos superiores para características de baixa herdabilidade obtendo inclusive, pequenos aumentos na acurácia (<

10%) em detrimento de grandes aumentos de herbabilidade. Estes últimos autores também relataram que o número de QTLs envolvidos no controle da característica de interesse tem impacto significativo na acurácia da GWS somente quando a densidade de marcadores utilizada não é suficiente para amostrar adequadamente o genoma. Neste caso, nem todos os QTLs ficam em desequilíbrio de ligação com pelo menos um marcador.

Em termos da eficiência em relação à seleção fenotípica, a seleção genômica tende a ser mais eficiente quando empregada em caracteres de baixa herdabilidade em contraste com aqueles de maior herdabilidade. Para valores similares de LD, variância aditiva e densidade de marcadores, podemos dizer que a eficiência da seleção genômica em relação à seleção fenotípica é inversamente proporcional a herdabilidade. Dependendo da magnitude da variância genética e do valor de LD apresentado na população, a GWS conseguiu ser superior em até 48,5% a seleção fenotípica ao considerar herdabilidade 0,3. A superioridade da seleção fenotípica se deu nos casos onde baixa densidade de marcadores foi utilizada conciliada ao uso de populações de menor LD ou de menor variância genética. Ao considerar herdabilidade de 0,7, a superioridade da GWS não ficou tão evidente, uma vez que, alta densidade de marcadores associada a populações de alto LD e/ou alta variância aditiva foi requerida a fim de obter acurácias similares à seleção fenotípica.

Conforme ao já exposto, parece evidente que a eficiência da GWS em relação à seleção fenotípica decai com o aumento da herdabilidade. De modo geral, a GWS tem sido superior a seleção fenotípica e seleção assistida por marcadores, apresentando resultados acurados com economia de tempo e recursos durante o ciclo de melhoramento das culturas (HEFFNER et al., 2011; DE LOS CAMPOS et al., 2013). É importante reconhecer que diferentes acurácias podem ser obtidas de acordo com os métodos de predição dos GBVs utilizado, sendo estes afetados por diversos fatores. ZHONG et al. (2009) observaram que o método de predição da GWS de maior acurácia mudava de acordo com diferentes níveis de LD entre marcador e QTL, efeitos de QTLs, e gerações de indivíduos. Em geral, semelhante acurácia é obtida em vários métodos de predição quando grande número de QTLs é assumido (DAETWYLER et al., 2013).

A densidade de marcadores é um dos fatores que mais afetam a acurácia dos modelos preditivos, uma vez que, o desequilíbrio de ligação entre dois locos é afetado principalmente pela taxa de recombinação entre eles (FLINT-GARCIA et al., 2003). Assim, empregando maior densidade de marcadores espera-se que maior número destes esteja associado de forma mais próxima à QTLs de interesse. Para este conjunto de dados simulados, não foi observado aumentos relevantes na acurácia de predição ao reduzir a distância entre SNPs adjacentes de 1 para 0,1 cM. Enquanto que a redução da distância de 10 para 0,1 cM causou expressivos aumentos na acurácia de predição, sobretudo nas populações de baixo LD e menor variância genética. BERNARDO e YU (2007) demonstraram que a acurácia do valor genético estimado melhora quando diminuído os intervalos entre marcadores adjacentes. Outros autores relataram aumentos lineares da acurácia da seleção genômica com aumentos na densidade dos marcadores sendo esta estabilizada quando alta densidade de marcadores é praticada (MEUWISSEN et al., 2009; COMBS e BERNARDO, 2013).

Visando obter maior cobertura de marcadores no genoma associado a menores custos de genotipagem, o uso de 1 SNP a cada cM pareceu ser a decisão mais acertada para a maioria dos cenários aqui simulados. Contudo, ao considerar as populações de baixo LD e menor variância genética, o uso de maiores densidades de SNP seria necessária a fim de obter satisfatória acurácia de predição. NAKAYA e ISOBE (2012) comentam que a disparidade na quantidade de marcadores utilizadas em diferentes trabalhos é causada pela diferença do intervalo entre marcadores adjacentes com o LD encontrado na população de mapeamento. Os autores ainda citam que empregar uma população originada a partir de algumas linhagens parentais é eficaz na redução do número de marcadores necessários, especialmente para espécies cujo LD decai rapidamente entre indivíduos não aparentados. No caso de populações biparentais, comuns no melhoramento vegetal, maiores densidades de marcadores serão necessárias se o LD for baixo (ASORO et al., 2011).

O LD pode ser muito diferente entre populações de melhoramento tendo grande impacto sobre a genotipagem e eficiência da GWS e, a perda de alelos favoráveis tem sido em muitas vezes relatada como inevitável quando em presença de fraca associação com os marcadores (JANNINK et al., 2010).

Neste trabalho verificamos que, para populações de maior LD, menores densidades de marcadores podem ser utilizadas sem afetar drasticamente a acurácia de predição. A utilização de 1 SNP a cada 10 cM na população 5 (de maior LD) resultou em acurácia média de 0,742 para CE na geração 0 e herdabilidade 0,3 e acurácia de 0,317 em mesmas condições na população 4 (de menor LD). No entanto, a redução da distância entre SNPs adjacentes de 10 para 0,1 cM proporcionaram acurácias de 0,814 (acréscimo de 10%) e 0,567 (acréscimo de 80%) nas populações 5 e 4, respectivamente. Assim, podemos concluir que é possível obter alta acurácia sob o uso de baixas densidades de marcadores se a população de genotipagem estiver associada a alto LD. De forma semelhante, alta densidade de SNP será necessária em populações de menor LD, como geralmente verificado em linhagens de gerações avançadas.

Ao considerarmos outros fatores constantes e somente o valor da variância aditiva variando entre as populações, observamos que o aumento da acurácia é proporcional ao aumento da variância genética. BASTIAANSEN et al. (2012) observaram que, sob mesma população de referência, ao avançar a população de seleção em 10 gerações de acasalamento ao acaso, a acurácia de predição tinha significativa redução, sendo esta causada principalmente pela redução da variância genética. Trabalhos de simulação como os apresentados por HABIÉR et al. (2007), tem relatado o declínio da acurácia ao longo das gerações, sugerindo como populações de genotipagem a utilização de populações mestiças com maior variabilidade genética ao invés de populações homogêneas ou de raça pura.

#### 4.2. Estimação e avaliação dos efeitos dos SNPs em diferentes populações.

Mesmo utilizando populações de referência e de seleção com similares valores de LD, variância aditiva e herdabilidade, o uso da GWS resultou em acurácia próxima de zero. Estes resultados demonstram que os modelos preditivos não podem ser usados para populações de melhoramento que apenas possuem estruturas genéticas similares à população de referência. NAKAYA e ISOBE (2012) argumentam que os marcadores em LD significativo com os QTLs de interesse, em geral, não são mantidos em diferentes

cruzamentos dificultando o uso da GWS nestes casos. Como exemplo, se efeitos de marcadores forem estimados a partir de híbridos simples de milho, os valores genéticos genômicos estimados em populações heterogêneas não irão fornecer informações acuradas (MASSMAN et al., 2013). HARRIS et al. (2009) utilizando duas raças de bovinos como populações de referência verificaram que o uso de uma raça como população de referência e sua validação em outra, resultava em baixa acurácia (-0,10 a 0,3).

Conforme relatado por JONAS e DE KONING (2013), a população de referência e a de seleção devem ser geneticamente mais similares a fim de atingir alta acurácia. No entanto, verificamos que valores similares de LD, variância aditiva e herdabilidade entre as populações não se traduzem em alta acurácia na GWS, sendo necessário que o candidato à seleção seja relacionado com a população da qual os efeitos de marcadores foram estimados. De acordo com DE ROOS et al. (2009), a dificuldade na utilização da seleção genômica entre várias populações é o cálculo dos GBVs, porque na maioria dos casos, os efeitos de marcadores serão diferentes entre elas. JONAS e DE KONING (2013) ainda citam que na maioria dos estudos em espécies vegetais, a validação dos marcadores é realizada dentro da mesma geração, sugerindo nestes casos a superestimação da acurácia dos modelos de predição da GWS. As influências do LD e da variância genética sobre a precisão de GWS ao longo das gerações tem sido relatada na literatura (SOLBERG et al., 2009), diminuindo a acurácia de predição conforme a população de seleção se distancia geneticamente da população de referência (BASTIAANSEN et al., 2012). Este problema é agravado se alguns QTLs estiverem fixados em qualquer uma das populações, sendo ineficientes aumentos expressivos na densidade de marcadores.

Os resultados sugerem a necessidade de grande parte dos alelos presentes na população de seleção estejam representados na população de referência. Estes resultados concordam com os obtidos por TOOSI et al. (2010), os autores observaram que a acurácia é reduzida drasticamente quando os alelos específicos de uma amostra ou de um subgrupo não estão incluídos na população de referência.

#### 4.3. Estimaco e avaliao dos efeitos dos SNPs em diferentes amostras da populao de referncia.

Quando mesma populao  usada para estimaco e validao dos efeitos dos marcadores, estratgias de validao devem ser utilizadas a fim de evitar a superestimaco destes efeitos (USAI et al., 2009). Neste estudo, por se tratar de populaes simuladas, a acurcia da GWS foi obtida pela simples correlao de Spearman entre o GBV estimado pelo mtodo RR-BLUP e o seu valor gentico paramtrico. Com base nos resultados discutidos anteriormente, chegamos  concluso de que para a GWS ser eficiente  necessria que as populaes de referncia e seleo compartilhassem genes em comum, sendo originadas, por exemplo, de uma mesma populao ancestral. Contudo, apesar de originarem de mesma populao, o simples fato das populaes de referncia e seleo serem tomadas de amostras diferentes conduz a diferentes perdas de acurcia em funo dos parmetros considerados.

H uma reduo de pelo menos 8 % de acurcia ao considerar a populao de seleo em mesma gerao da populao de referncia e uma reduo de no mnimo 22% considerando distanciamento de 5 geraes de acasalamento ao acaso. Este resultado no  surpreendente, pois alguns marcadores podem estar em elevado LD com um QTL em uma amostra e no na outra, especialmente quando as populaes esto distanciadas por muitas geraes ou quando menor densidade de marcadores  utilizada (ANDREESCU et al., 2007; DE ROOS et al., 2008). Neste caso, verificamos que apesar de sempre haver reduo da acurcia de predio com a utilizao de amostras diferentes para estimaco e seleo, a diferena de gerao entre as populaes teve maior importncia do que a herdabilidade e a densidade de marcadores considerada, pois alguns marcadores no permanecem em fase de ligao com o QTL ao longo as geraes. MUIR (2007) observou que a acurcia dos GBVs diminu rapidamente quando o nmero de geraes era aumentado entre a populao de referncia e a populao de seleo, e concluram que devido  baixa quantidade de QTLs simulados, a diferena de variabilidade gentica entre as geraes foi a principal causa para este resultado.

#### 4.4. Avaliação do efeito das relações de parentesco sobre a eficiência da seleção genômica.

É uma tarefa difícil separar os efeitos do LD e relações de parentesco na acurácia de predição da seleção genômica, pois estes efeitos estão interligados. A estruturação dos indivíduos em famílias de maior relacionamento resultou em maior eficiência da seleção genômica. A diminuição das relações de parentesco entre populações tem sido reportada como um empecilho no uso da seleção genômica (GIANOLA et al., 2009; HABIÉR et al., 2010) e podendo ter maior efeito sobre a acurácia dos GBVs do que o valor *per se* de LD (WIENTJES et al., 2013).

A diferença do tamanho efetivo entre as populações teve impacto relevante sobre os valores de acurácia obtidos, uma vez que, o grau de parentesco entre os indivíduos é reduzido com o aumento do tamanho efetivo da população de referência (CLARK et al., 2012). No entanto, o aumento do  $N_e$  dentro de cada população não conduziu a reduções expressivas de acurácia, apresentando inclusive, em alguns cenários para produção de grãos, valores maiores de acurácias com o aumento do  $N_e$ . Uma explicação para tal fato é ter sido usado alta densidade de marcadores (1SNP/0,1cM) e assim o tamanho efetivo ter exercido menor influência nestes casos. Em GRATTAPAGLIA e RESENDE (2011) podemos observar que valores de acurácia obtidos para diferentes tamanhos efetivos vão se estreitando com o aumento da densidade de marcadores, ou seja, o tamanho efetivo só teria maior influência se a densidade de marcadores utilizada fosse insuficiente para amostrar o genoma. Segundo os autores, para obter boa acurácia na GWS, um marcador a cada 0,1cM devem ser usados na presença de maiores tamanhos efetivos populacionais ( $N_e > 60$ ) e comparativamente, valores similares de acurácia podem ser obtidos quando até um marcadores a cada 0,5cM são usados considerando menores tamanhos efetivos ( $N_e \leq 30$ ). Outro aspecto importante é que quando o tamanho efetivo populacional é pequeno e os valores associados aos genótipos são grandes, isso também pode provocar uma superestimação do efeito do genótipo (NAKAYA e ISOBE, 2012).

Geralmente menores tamanhos da população de referência são requeridas em populações de autopolinização, pois populações com maior

diversidade genética requerem tamanhos populacionais maiores visando obter alta acurácia na estimação dos GBVs (MUJIBI et al., 2011). Neste contexto, é de se esperar que para condições similares de seleção, maiores acurácias sejam obtidas para FS1, FIC, FMI e OP, respectivamente. No entanto, devemos ser cuidados ao extrapolar estes resultados para qualquer cenário de análise, pois o valor de LD e a magnitude da variância genética das populações aqui simuladas podem não refletirem a magnitude destes parâmetros encontrada nas diferentes populações de melhoramento. De acordo com WIENTJES et al. (2013), em relação à eficiência da GWS, haverá maior importância das relações de parentesco em relação à magnitude do LD quanto menor for o tamanho efetivo na população de referência.

## **5. Conclusões**

Verificamos que a eficiência da seleção genômica em relação à seleção fenotípica é inversamente proporcional a herdabilidade e que para populações com maior LD, menores densidades de marcadores (1SNP/cM) podem ser usadas sem afetar drasticamente a acurácia de predição. No entanto, ao considerar populações de menor LD e menor variância genética, o uso de maiores densidades de SNP (1SNP/0,1cM) seria necessária a fim de obter acurácia do valor genético maior que 0,55 e 0,70 em herdabilidade de 0,3 e 0,7, respectivamente.

Nossos resultados também evidenciaram a necessidade de grande parte dos alelos presentes na população de seleção estejam representados na população de referência. Assim, os modelos de predição do GBV não podem ser usados em populações que apenas possuem estruturas genéticas similares à população onde os efeitos dos marcadores foram estimados, pois a acurácia do valor genético tendeu a zero nestes casos.

Além disso, quando diferentes amostras de uma mesma população foram usadas como população de referência e de seleção, houve redução de pelo menos 8 % de acurácia ao considerar a população de seleção em mesma geração da população de referência e redução de no mínimo 22% considerando distanciamento de 5 gerações de acasalamento ao acaso.

Pôde se concluir também, que estruturação dos indivíduos em famílias de maior relacionamento resulta em maior eficiência da seleção genômica, assim como, um menor tamanho efetivo entre as populações tem impacto positivo sobre os valores de acurácia.

## 6. Referências

ANDREESCU, C.; AVENDANO, S.; BROWN, S. R.; HASSEN, A.; LAMONT, S. J.; DEKKERS, J. C. Linkage disequilibrium in related breeding lines of chickens. **Genetics**, v. 177, n. 4, p. 2161-2169, 2007.

ASORO, F. G.; NEWELL, M. A.; BEAVIS, W. D.; SCOTT, M. P.; JANNINK, J. L. Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. **Plant Genome**, v. 4, n. 2, p. 132-144, 2011.

BASTIAANSEN, J. W.; COSTER, A.; CALUS, M. P.; VAN ARENDONK, J. A.; BOVENHUIS, H. Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. **Genet Sel Evol**, v. 44, p. 3, 2012.

BERNARDO, R.; YU, J. Prospects for Genomewide Selection for Quantitative Traits in Maize. **Crop Sci.**, v. 47, n. 3, p. 1082-1090, 2007.

BERRY, D. P.; KEARNEY, J. F. Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. **Animal**, v. 5, n. 8, p. 1162-1169, 2011.

CALUS, M. P. L.; MEUWISSEN, T. H. E.; DE ROOS, A. P. W.; VEERKAMP, R. F. Accuracy of genomic selection using different methods to define haplotypes. **Genetics**, v. 178, n. 1, p. 553-561, 2008.

CLARK, S. A.; HICKEY, J. M.; DAETWYLER, H. D.; VAN DER WERF, J. H. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. **Genet Sel Evol**, v. 44, p. 4, 2012.

COMBS, E.; BERNARDO, R. Accuracy of Genomewide Selection for Different Traits with Constant Population Size, Heritability, and Number of Markers. **Plant Genome**, v. 6, n. 1, 2013.

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, n. 2, p. 347-365, 2013.

DE LOS CAMPOS, G.; HICKEY, J. M.; PONG-WONG, R.; DAETWYLER, H. D.; CALUS, M. P. L. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. **Genetics**, v. 193, n. 2, p. 327-+, 2013.

DE ROOS, A. P.; HAYES, B. J.; GODDARD, M. E. Reliability of genomic predictions across multiple populations. **Genetics**, v. 183, n. 4, p. 1545-1553, 2009.

DE ROOS, A. P.; HAYES, B. J.; SPELMAN, R. J.; GODDARD, M. E. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. **Genetics**, v. 179, n. 3, p. 1503-1512, 2008.

ENDELMAN, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Gen.*, v. 4, n. 3, p. 250-255, 2011.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, v. 54, p. 357-374, 2003.

FRITSCHÉ-NETO, R.; RESENDE, M. D. V.; MIRANDA, G. V.; DOVALE, J. C. Seleção genômica ampla e novos métodos de melhoramento do milho. **Rev. Ceres**, v. 59, n.6, p. 794-802, 2012.

GIANOLA, D.; DE LOS CAMPOS, G.; HILL, W. G.; MANFREDI, E.; FERNANDO, R. Additive genetic variability and the Bayesian alphabet. *Genetics*, v. 183, n. 1, p. 347-363, 2009.

GODDARD, M. Genomic selection: prediction of accuracy and maximisation of long term response. **Genetica**, v. 136, n. 2, p. 245-257, 2009.

GODDARD, M. E.; HAYES, B. J. Genomic selection. **J Anim Breed Genet**, v. 124, n. 6, p. 323-330, 2007.

GRATTAPAGLIA, D.; RESENDE, M. D. V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, n. 2, p. 241-255, 2011.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, n. 4, p. 2389-2397, 2007.

HABIER, D.; TETENS, J.; SEEFRIED, F. R.; LICHTNER, P.; THALLER, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genet Sel Evol**, v. 42, p. 5, 2010.

HARRIS, B. L.; JOHNSON, D. L.; SPELMAN, R. J. Genomic selection in New Zealand and the implications for national genetic evaluation. 13. 2009, Roma. **Anais**. Roma: International Committee for Animal Recording (ICAR), 2009. p. 325-330.

HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Invited review: Genomic selection in dairy cattle: progress and challenges. **J Dairy Sci**, v. 92, n. 2, p. 433-443, 2009.

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, n. 6, p. 2597-2606, 2011.

JANNINK, J. L.; LORENZ, A. J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Brief Funct Genomics**, v. 9, n. 2, p. 166-177, 2010.

JONAS, E.; DE KONING, D.-J. Does genomic selection have a future in plant breeding? **Trends in Biotechnology**, v. 31, n. 9, p. 497-504, 2013.

MASSMAN, J. M.; GORDILLO, A.; LORENZANA, R. E.; BERNARDO, R. Genomewide predictions from maize single-cross data. **Theor Appl Genet**, v. 126, n. 1, p. 13-22, 2013.

MEUWISSEN, T. H.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-1829, 2001.

MEUWISSEN, T. H.; SOLBERG, T. R.; SHEPHERD, R.; WOOLLIAMS, J. A. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. **Genet Sel Evol**, v. 41, p. 2, 2009.

MUIR, W. M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **J Anim Breed Genet**, v. 124, n. 6, p. 342-355, 2007.

MUJIBI, F. D.; NKRUMAH, J. D.; DURUNNA, O. N.; STOTHARD, P.; MAH, J.; WANG, Z.; BASARAB, J.; PLASTOW, G.; CREWS, D. H., JR.; MOORE, S. S. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. **J Anim Sci**, v. 89, n. 11, p. 3353-3361, 2011.

NAKAYA, A.; ISOBE, S. N. Will genomic selection be a practical method for plant breeding? **Ann Bot**, v. 110, n. 6, p. 1303-1316, 2012.

PSZCZOLA, M.; STRABEL, T.; VAN ARENDONK, J. A. M.; CALUS, M. P. L. The impact of genotyping different groups of animals on accuracy when moving from traditional to genomic selection. **J Dairy Sci**, v. 95, n. 9, p. 5412-5421, 2012.

RESENDE JR, M. F.; MUNOZ, P.; ACOSTA, J. J.; PETER, G. F.; DAVIS, J. M.; GRATTAPAGLIA, D.; RESENDE, M. D.; KIRST, M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. **New Phytol**, v. 193, n. 3, p. 617-624, 2012.

RESENDE, M. D. V.; BARBOSA, M. H. P. **Melhoramento genético de plantas de propagação assexuada**. Colombo: Embrapa Florestas, 2005. 130 p.

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.; PAPPAS, G. J.; KILIAN, A.; GRATTAPAGLIA, D. 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, n. 1, p. 116-128, 2012.

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

TOOSI, A.; FERNANDO, R. L.; DEKKERS, J. C. Genomic selection in admixed and crossbred populations. **J Anim Sci**, v. 88, n. 1, p. 32-46, 2010.

USAI, M. G.; GODDARD, M. E.; HAYES, B. J. LASSO with cross-validation for genomic selection. **Genet Res (Camb)**, v. 91, n. 6, p. 427-436, 2009.

VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: reliability of genomic predictions for North American Holstein bulls. **J Dairy Sci**, v. 92, n. 1, p. 16-24, 2009.

WIJNTJES, Y. C. J.; VEERKAMP, R. F.; CALUS, M. P. L. The Effect of Linkage Disequilibrium and Family Relationships on the Reliability of Genomic Prediction. **Genetics**, v. 193, n. 2, p. 621-+, 2013.

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, n. 1, p. 355-364, 2009.

**Tabela 1.** Média do valor  $D$  e do  $r^2$ , estimados em relação a 6 populações (geração 0 e geração 5) em diferentes densidades de SNP.

População	Geração	Distância entre SNPs adjacentes					
		0,1cM		1cM		10cM	
		D	$r^2$	D	$r^2$	D	$r^2$
P0	0	0,089	0,151	0,085	0,140	0,081	0,125
	5	0,046	0,059	0,045	0,058	0,041	0,047
P1	0	0,087	0,147	0,088	0,149	0,080	0,129
	5	0,045	0,059	0,045	0,059	0,044	0,050
P2	0	0,091	0,156	0,089	0,150	0,081	0,123
	5	0,051	0,067	0,048	0,064	0,049	0,059
P3	0	0,086	0,144	0,089	0,155	0,091	0,157
	5	0,049	0,064	0,050	0,067	0,047	0,062
P4	0	0,040	0,043	0,042	0,046	0,042	0,045
	5	0,023	0,021	0,024	0,022	0,023	0,019
P5	0	0,156	0,417	0,155	0,409	0,149	0,380
	5	0,087	0,182	0,082	0,172	0,073	0,141

D - diferença entre frequência gamética observada e esperada sob equilíbrio de ligação gênica;  
 $r^2$  - quadrado da correlação entre alelos de dois locos.

**Tabela 2.** Média (M), variância aditiva ( $\sigma_A^2$ ) e de dominância ( $\sigma_D^2$ ) para a capacidade de expansão (ml g<sup>-1</sup>) e produção de grãos (g planta<sup>-1</sup>) em seis populações representadas nas gerações 0 e 5.

População	Geração	Capacidade de expansão			Produção de grãos		
		M	$\sigma_A^2$	$\sigma_D^2$	M	$\sigma_A^2$	$\sigma_D^2$
P0	0	29,22	2,56	0,59	143,02	35,78	17,17
	5	29,22	2,54	0,58	143,02	35,33	15,55
P1	0	29,36	2,45	0,55	144,34	33,07	17,07
	5	29,36	2,45	0,55	144,34	33,03	15,48
P2	0	30,44	1,76	0,46	148,06	25,79	22,11
	5	30,44	1,85	0,47	148,06	27,19	19,90
P3	0	28,03	3,38	0,55	142,75	51,57	18,48
	5	28,03	3,27	0,54	142,75	49,69	16,60
P4	0	30,37	2,47	0,51	143,90	33,11	8,17
	5	30,37	2,48	0,51	143,90	33,23	7,85
P5	0	27,83	2,45	0,64	140,09	38,76	37,62
	5	27,83	2,55	0,64	140,09	40,22	32,57

**Tabela 3.** Acurácia de predição dos valores genéticos, em diferentes cenários com intervalo de confiança bootstrap de 95%, com base no método RR-BLUP para seis populações. Efeitos dos marcadores foram estimados e avaliados na própria população.

População	Ger.	Densid. de SNP	Acurácia por meio do valor fenotípico ( $\sqrt{h^2}$ )				
			0,548 ( $h^2=0,3$ )		0,837 ( $h^2=0,7$ )		
			Capacidade de expansão	Produção de grãos	Capacidade de expansão	Produção de grãos	
P0	0	0,1	0,692 (0,641:0,748)	0,666 (0,591:0,725)	0,817 (0,791:0,842)	0,778 (0,750:0,806)	
		1	0,663 (0,616:0,718)	0,643 (0,575:0,691)	0,761 (0,730:0,786)	0,733 (0,697:0,772)	
		10	0,565 (0,509:0,619)	0,548 (0,477:0,614)	0,634 (0,590:0,679)	0,612 (0,566:0,651)	
	5	0,1	0,646 (0,574:0,700)	0,621 (0,553:0,682)	0,815 (0,784:0,842)	0,787 (0,754:0,820)	
		1	0,604 (0,520:0,667)	0,578 (0,486:0,644)	0,758 (0,707:0,791)	0,724 (0,682:0,764)	
		10	0,445 (0,291:0,533)	0,430 (0,349:0,496)	0,536 (0,458:0,589)	0,525 (0,464:0,571)	
	P1	0	0,1	0,664 (0,597:0,713)	0,645 (0,558:0,705)	0,801 (0,772:0,832)	0,768 (0,730:0,804)
			1	0,627 (0,548:0,678)	0,614 (0,528:0,679)	0,737 (0,695:0,772)	0,712 (0,639:0,760)
			10	0,504 (0,420:0,574)	0,501 (0,423:0,567)	0,579 (0,533:0,629)	0,577 (0,524:0,619)
5		0,1	0,640 (0,558:0,711)	0,607 (0,535:0,682)	0,812 (0,778:0,845)	0,781 (0,736:0,814)	
		1	0,591 (0,503:0,669)	0,564 (0,482:0,644)	0,745 (0,681:0,787)	0,721 (0,669:0,761)	
		10	0,415 (0,342:0,493)	0,399 (0,297:0,532)	0,508 (0,432:0,571)	0,498 (0,428:0,569)	
P2	0	0,1	0,553 (0,450:0,614)	0,507 (0,418:0,567)	0,745 (0,662:0,788)	0,657 (0,552:0,710)	
		1	0,490 (0,392:0,567)	0,467 (0,357:0,549)	0,640 (0,565:0,690)	0,583 (0,504:0,643)	
		10	0,364 (0,267:0,445)	0,346 (0,267:0,428)	0,448 (0,375:0,513)	0,416 (0,344:0,481)	
	5	0,1	0,651 (0,600:0,699)	0,590 (0,503:0,680)	0,828 (0,794:0,858)	0,750 (0,616:0,808)	
		1	0,601 (0,546:0,660)	0,538 (0,457:0,617)	0,761 (0,710:0,800)	0,686 (0,573:0,759)	
		10	0,397 (0,321:0,462)	0,328 (0,249:0,411)	0,478 (0,388:0,534)	0,423 (0,318:0,509)	
P3	0	0,1	0,765 (0,721:0,800)	0,760 (0,720:0,799)	0,858 (0,835:0,876)	0,844 (0,817:0,863)	
		1	0,748 (0,706:0,778)	0,741 (0,700:0,788)	0,823 (0,793:0,849)	0,819 (0,788:0,836)	
		10	0,659 (0,609:0,711)	0,656 (0,606:0,710)	0,714 (0,671:0,744)	0,712 (0,652:0,743)	
	5	0,1	0,684 (0,639:0,749)	0,686 (0,645:0,727)	0,839 (0,811:0,864)	0,819 (0,788:0,850)	
		1	0,653 (0,613:0,723)	0,662 (0,636:0,706)	0,798 (0,770:0,832)	0,784 (0,752:0,817)	
		10	0,515 (0,415:0,593)	0,511 (0,427:0,566)	0,602 (0,541:0,663)	0,590 (0,531:0,648)	
P4	0	0,1	0,567 (0,514:0,629)	0,568 (0,497:0,629)	0,757 (0,726:0,793)	0,760 (0,706:0,787)	
		1	0,483 (0,425:0,550)	0,492 (0,404:0,538)	0,631 (0,580:0,677)	0,636 (0,574:0,675)	
		10	0,317 (0,229:0,385)	0,333 (0,234:0,407)	0,404 (0,335:0,442)	0,424 (0,364:0,495)	
	5	0,1	0,594 (0,535:0,650)	0,592 (0,525:0,653)	0,803 (0,769:0,836)	0,801 (0,768:0,827)	
		1	0,512 (0,444:0,577)	0,515 (0,439:0,576)	0,697 (0,644:0,747)	0,709 (0,665:0,758)	
		10	0,281 (0,182:0,381)	0,291 (0,202:0,409)	0,382 (0,306:0,444)	0,399 (0,345:0,466)	
P5	0	0,1	0,814 (0,780:0,849)	0,776 (0,706:0,820)	0,896 (0,876:0,913)	0,855 (0,825:0,885)	
		1	0,808 (0,772:0,841)	0,770 (0,709:0,819)	0,881 (0,858:0,898)	0,845 (0,816:0,874)	
		10	0,742 (0,697:0,775)	0,716 (0,655:0,768)	0,798 (0,776:0,826)	0,775 (0,748:0,808)	
	5	0,1	0,792 (0,765:0,824)	0,732 (0,670:0,813)	0,889 (0,873:0,911)	0,852 (0,823:0,878)	
		1	0,775 (0,745:0,816)	0,716 (0,638:0,803)	0,869 (0,846:0,893)	0,833 (0,799:0,862)	
		10	0,624 (0,581:0,691)	0,585 (0,511:0,679)	0,699 (0,660:0,740)	0,679 (0,627:0,729)	

**Tabela 4.** Acurácia de predição dos valores genéticos em diferentes cenários com intervalo de confiança bootstrap de 95%, com base no método RR-BLUP para seis populações. Efeitos dos marcadores estimados na população de referência (P0, geração 0) foram os considerados nos cálculos.

População	Ger.	Densid. de SNP	Acurácia por meio do valor fenotípico ( $\sqrt{h^2}$ )			
			0,548 ( $h^2=0,3$ )		0,837 ( $h^2=0,7$ )	
			Capacidade de expansão	Produção de grãos	Capacidade de expansão	Produção de grãos
P0	0	0,1	0,692 (0,641:0,748)	0,666 (0,591:0,725)	0,817 (0,791:0,842)	0,778 (0,750:0,806)
		1	0,663 (0,616:0,718)	0,643 (0,575:0,691)	0,761 (0,730:0,786)	0,733 (0,697:0,772)
		10	0,565 (0,509:0,619)	0,548 (0,477:0,614)	0,634 (0,590:0,679)	0,612 (0,566:0,651)
	5	0,1	0,495 (0,383:0,555)	0,478 (0,408:0,548)	0,560 (0,502:0,610)	0,546 (0,486:0,606)
		1	0,451 (0,355:0,517)	0,436 (0,379:0,501)	0,507 (0,435:0,570)	0,494 (0,444:0,572)
		10	0,351 (0,225:0,407)	0,344 (0,261:0,423)	0,395 (0,296:0,459)	0,389 (0,328:0,444)
P1	0	0,1	0,040 (-0,155:0,193)	0,034 (-0,120:0,191)	0,031 (-0,110:0,204)	0,052 (-0,143:0,193)
		1	0,023 (-0,212:0,267)	-0,015 (-0,212:0,147)	0,030 (-0,140:0,198)	0,008 (-0,162:0,159)
		10	-0,067 (-0,217:0,096)	-0,080 (-0,237:0,064)	-0,080 (-0,185:0,007)	-0,080 (-0,215:0,002)
	5	0,1	0,032 (-0,128:0,139)	0,002 (-0,104:0,119)	0,006 (-0,131:0,163)	0,018 (-0,101:0,110)
		1	-0,008 (-0,183:0,103)	-0,012 (-0,115:0,092)	-0,004 (-0,133:0,123)	-0,005 (-0,123:0,113)
		10	-0,058 (-0,182:0,018)	-0,045 (-0,188:0,073)	-0,049 (-0,174:0,044)	-0,053 (-0,170:0,055)
P2	0	0,1	0,059 (-0,056:0,162)	0,02 (-0,132:0,142)	0,041 (-0,081:0,193)	0,027 (-0,119:0,114)
		1	0,098 (-0,012:0,192)	0,063 (-0,098:0,209)	0,111 (-0,017:0,197)	0,098 (0,000:0,193)
		10	0,002 (-0,159:0,098)	0,003 (-0,175:0,215)	-0,006 (-0,112:0,083)	-0,003 (-0,113:0,127)
	5	0,1	0,019 (-0,116:0,131)	-0,014 (-0,188:0,162)	0,007 (-0,121:0,120)	0,021 (-0,094:0,130)
		1	0,068 (-0,035:0,193)	0,058 (-0,100:0,179)	0,076 (0,012:0,136)	0,096 (-0,033:0,225)
		10	0,057 (-0,067:0,153)	0,038 (-0,048:0,140)	0,055 (-0,043:0,156)	0,042 (-0,065:0,114)
P3	0	0,1	0,045 (-0,158:0,217)	0,064 (-0,204:0,231)	0,090 (-0,189:0,286)	0,080 (-0,110:0,255)
		1	0,066 (-0,157:0,273)	0,016 (-0,279:0,299)	0,124 (-0,088:0,330)	0,004 (-0,182:0,213)
		10	-0,029 (-0,264:0,171)	-0,034 (-0,228:0,205)	0,017 (-0,237:0,215)	-0,010 (-0,308:0,133)
	5	0,1	0,057 (-0,080:0,174)	0,054 (-0,073:0,198)	0,075 (-0,077:0,229)	0,063 (-0,075:0,235)
		1	0,039 (-0,147:0,192)	0,006 (-0,114:0,201)	0,057 (-0,088:0,211)	-0,019 (-0,148:0,135)
		10	0,009 (-0,165:0,132)	-0,011 (-0,146:0,151)	0,038 (-0,122:0,178)	-0,006 (-0,129:0,156)
P4	0	0,1	0,000 (-0,132:0,096)	-0,007 (-0,113:0,108)	0,000 (-0,120:0,130)	-0,001 (-0,063:0,086)
		1	-0,016 (-0,101:0,127)	-0,011 (-0,139:0,122)	-0,034 (-0,130:0,054)	-0,006 (-0,120:0,104)
		10	-0,015 (-0,146:0,112)	0,005 (-0,143:0,125)	-0,028 (-0,108:0,090)	0,011 (-0,100:0,101)
	5	0,1	-0,009 (-0,14:0,096)	0,005 (-0,068:0,079)	0,002 (-0,101:0,116)	-0,001 (-0,073:0,097)
		1	-0,011 (-0,125:0,107)	-0,005 (-0,109:0,092)	-0,007 (-0,134:0,085)	0,000 (-0,094:0,098)
		10	-0,003 (-0,084:0,079)	0,024 (-0,078:0,115)	-0,005 (-0,125:0,133)	0,008 (-0,099:0,090)
P5	0	0,1	0,069 (-0,269:0,336)	0,013 (-0,330:0,353)	0,067 (-0,282:0,252)	0,024 (-0,309:0,339)
		1	0,042 (-0,254:0,293)	-0,014 (-0,276:0,267)	0,111 (-0,247:0,416)	0,025 (-0,271:0,227)
		10	-0,183 (-0,447:0,042)	-0,142 (-0,380:0,124)	-0,185 (-0,367:-0,016)	-0,196 (-0,385:0,064)
	5	0,1	0,060 (-0,177:0,325)	0,044 (-0,180:0,255)	0,067 (-0,119:0,239)	0,052 (-0,196:0,229)
		1	-0,055 (-0,258:0,127)	-0,082 (-0,249:0,144)	0,008 (-0,234:0,260)	-0,046 (-0,251:0,124)
		10	-0,118 (-0,300:0,067)	-0,110 (-0,250:0,073)	-0,118 (-0,232:-0,005)	-0,138 (-0,299:-0,011)

**Tabela 5.** Acurácia de predição dos valores genéticos em diferentes cenários com intervalo de confiança bootstrap de 95%, com base no método RR-BLUP. Efeitos dos marcadores foram estimados e avaliados em diferentes amostras da população de referência (P0).

Ger.	Densid. de SNP	Acurácia por meio do valor fenotípico ( $\sqrt{h^2}$ )			
		0,548 ( $h^2=0,3$ )		0,837 ( $h^2=0,7$ )	
		Capacidade de expansão	Produção de grãos	Capacidade de expansão	Produção de grãos
0	0,1	0,613 (0,573:0,652)	0,612 (0,570:0,666)	0,646 (0,603:0,686)	0,641 (0,597:0,699)
	1	0,581 (0,544:0,628)	0,588 (0,537:0,652)	0,612 (0,571:0,646)	0,606 (0,559:0,664)
	10	0,483 (0,438:0,548)	0,489 (0,427:0,560)	0,544 (0,483:0,609)	0,513 (0,467:0,580)
5	0,1	0,483 (0,424:0,526)	0,460 (0,381:0,512)	0,529 (0,455:0,583)	0,530 (0,468:0,580)
	1	0,436 (0,373:0,503)	0,428 (0,373:0,467)	0,473 (0,414:0,528)	0,465 (0,406:0,506)
	10	0,310 (0,236:0,407)	0,334 (0,292:0,405)	0,383 (0,305:0,463)	0,341 (0,277:0,426)

**Tabela 6.** Acurácia de predição dos valores genéticos em diferentes estruturas populacionais e tamanhos efetivos ( $N_e$ ).

População	$N_e$	Acurácia por meio do valor fenotípico ( $\sqrt{h^2}$ )			
		0,548 ( $h^2=0,3$ )		0,837 ( $h^2=0,7$ )	
		Capacidade de expansão	Produção de grãos	Capacidade de expansão	Produção de grãos
FS1	10	0,820	0,765	0,883	0,840
	20	0,778	0,757	0,874	0,857
FIC	20	0,695	0,659	0,789	0,760
	38	0,664	0,708	0,781	0,791
FMI	38	0,570	0,596	0,696	0,703
	72	0,559	0,528	0,682	0,672
OP	1000	0,405	0,472	0,556	0,667

FS1 - Família S1; FIC - Família de irmãos completos; FMI - Família de meios irmãos; OP - População de polinização aberta.