

GABRIELA FRANÇA OLIVEIRA

**REGRESSÃO QUANTÍLICA APLICADA À SELEÇÃO  
GENÔMICA PARA CARACTERÍSTICAS OLIGOGÊNICAS EM  
MELHORAMENTO DE PLANTAS AUTÓGAMAS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2019

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

T

O48r  
2019  
Oliveira, Gabriela França, 1991-  
Regressão quantílica aplicada à seleção genômica para  
características oligogênicas em melhoramento de plantas  
autógamas / Gabriela França Oliveira. – Viçosa, MG, 2019.  
xi, 39 f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.

Orientador: Ana Carolina Campana Nascimento.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 30-33.

1. Análise de regressão. 2. Melhoramento genético.  
3. Métodos de simulação. I. Universidade Federal de Viçosa.  
Departamento de Estatística. Programa de Pós-Graduação em  
Estatística Aplicada e Biometria. II. Título.

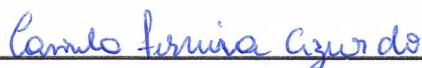
CDD 22. ed. 519.536

GABRIELA FRANÇA OLIVEIRA

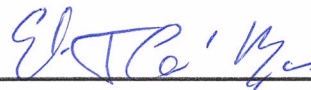
**REGRESSÃO QUANTÍLICA APLICADA À SELEÇÃO GENÔMICA PARA  
CARACTERÍSTICAS OLIGOGÊNICAS EM MELHORAMENTO DE  
PLANTAS AUTÓGAMAS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.


APROVADA: 21 de fevereiro de 2019.



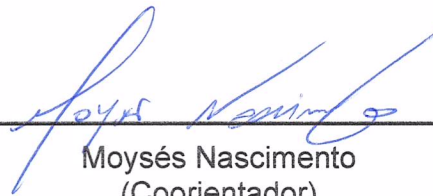
Camila Ferreira Azevedo



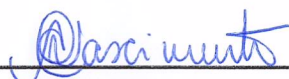
Eveline Teixeira Caixeta



Isabela de Castro Sant'anna  
(Coorientadora)



Moysés Nascimento  
(Coorientador)



Ana Carolina Campana Nascimento  
(Orientadora)

Dedico essa dissertação à minha mãe, mulher guerreira e de fibra que me ensinou a ser forte e batalhar para conquistar todos os meus objetivos. E à minha irmã Fernanda por estar presente em todos os momentos e por ser meu porto seguro.

*“Quem caminha sozinho pode até chegar mais rápido, mas aquele que vai acompanhado, com certeza vai mais longe.” (Clarice Lispector)*

*“Conheça todas as teorias, domine todas as técnicas, mas ao tocar uma alma humana, seja apenas outra alma humana.” (Carl Jung)*

## AGRADECIMENTO

Agradeço a Deus por ter me dado forças para enfrentar todos os obstáculos e por colocar sempre pessoas especiais no meu caminho. Agradeço a Ele por todas as conquistas.

À minha mãe Rochele por ser meu exemplo. Obrigada por estar comigo em todas as minhas escolhas, pelo carinho, pelo amor incondicional, amizade, confiança e por fazer meus sonhos o seu.

À minha irmã Fernanda por estar ao meu lado e ser a melhor companhia em Viçosa. Obrigada pelo incentivo, pelas palavras de carinho, pelos conselhos e por cuidar tão bem de mim. À minha irmã Nágela por ter me dado uns dos melhores presentes da minha vida, o Rafael e o João Henrique. Obrigada por fazerem meus dias mais felizes.

Ao João por ser a melhor pessoa que posso chamar de pai. Obrigada por sempre estar presente em minha vida.

À minha família pela constante torcida, em especial a minha tia Lúcia pela preocupação constante e carinho. Aos meus primos Fernando e a Simone por me acolherem tão bem quando cheguei em Viçosa.

Aos amigos do PPESTBIO, em especial a Carol, Gabriely, Ithalo, Jaquicele, Kaléo, Leísa, Lucas Bianchi, Lucas Silveira, Roberta, Pedro pelos momentos de descontração, companheirismo, pela troca de experiência e incentivo.

Aos meus Amigos da Matemática em especial a Isabella e Mayara, obrigada pelos conselhos, amizade e por se fazerem presente sempre.

A minha orientadora Ana Carolina pela confiança em mim depositada. Obrigada pelos conselhos, incentivo, paciência, preocupação, dedicação e pelos conhecimentos adquiridos. Você foi muito mais que uma orientadora, você foi uma amiga. A você minha eterna gratidão.

Ao professor e coorientador Moysés Nascimento pelos ensinamentos, conselhos, amizade, paciência, disponibilidade e por ser um exemplo de profissional.

A minha coorientadora Isabela pela paciência, amizade, por me acalmar e estar sempre disposta a encontrar soluções.

Ao professor e coorientador Leonardo Bhering e o Juan Romero pela disponibilidade.

Agradeço aos professores do Programa de Pós Graduação em Estatística Aplicada e Biometria, pela minha formação acadêmica, em especial, a professora Camila pela disponibilidade, dedicação, pelos conselhos e por ser um exemplo de profissional.

Aos funcionários do Departamento de Estatística, em especial a Anita e o Junior, pela amizade, paciência e incentivo.

Aos membros da banca examinadora Camila Azevedo Ferreira, Eveline Teixeira Caixeta, Isabela de Castro Sant'Anna e Moysés Nascimento pela disponibilidade e pelas sugestões que tanto contribuíram para o presente trabalho.

Ao LICAE e Bioinformática que tornaram o ambiente de estudos e trabalho mais prazeroso, produtivo e divertido nesses anos.

À CAPES, pelo suporte financeiro para o desenvolvimento deste trabalho.

Enfim, a todos que colaboraram de alguma forma, alguns longe, mas que conseguiram me confortar com um abraço ou com uma palavra de incentivo, para o sucesso deste trabalho.

## **BIOGRAFIA**

GABRIELA FRANÇA OLIVEIRA, filha de Rochele Eulália de Souza e Lázaro Antônio de Oliveira, nasceu em Patos de Minas, Minas Gerais, em 29 de setembro de 1991.

Em março de 2010, ingressou no curso de Licenciatura em Matemática na Universidade Federal de Viçosa, Viçosa – MG, graduando-se em janeiro de 2016.

Em março de 2017, iniciou o curso de Mestrado no Programa de Pós-Graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, submetendo-se à defesa da dissertação em 21 de fevereiro de 2019.

## SUMÁRIO

RESUMO .....	viii
ABSTRACT .....	x
1. INTRODUÇÃO.....	1
2. REFERENCIAL TEÓRICO.....	3
2.1 Melhoramento Genético de Plantas Autógamas.....	3
2.2 Seleção Genômica Ampla.....	4
2.2.1 Regressão Quantílica.....	6
2.2.1.1 Estimação dos Parâmetros.....	7
2.2.1.2 Regressão Quantílica Regularizada (RQR).....	8
2.2.2 Método RR-BLUP.....	9
2.2.3 LASSO Bayesiano (BLASSO).....	10
2.3 Simulação computacional como ferramenta auxiliar em programas de melhoramento.....	12
2.4 Capacidade preditiva e ganho de seleção.....	13
3. MATERIAL E MÉTODOS.....	13
3.1 Simulação das populações.....	13
3.2 Avanço das gerações.....	14
3.3 Predição Genômica.....	15
3.3.1 RR-BLUP.....	15
3.3.2 Lasso Bayesiano (BLASSO).....	16
3.3.3 Regressão Quantílica Regularizada.....	18
3.4 Comparações das metodologias.....	19
3.5 Cenários avaliados.....	20
4. RESULTADOS E DISCUSSÃO.....	20
5. CONCLUSÃO.....	29
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	30
Apêndice: Rotinas computacionais utilizadas para a análise.....	34

## RESUMO

OLIVEIRA, Gabriela França, M.Sc., Universidade Federal de Viçosa, fevereiro de 2019. **Regressão quantílica aplicada à seleção genômica para características oligogênicas em melhoramento de plantas autógamas.** Orientadora: Ana Carolina Campana Nascimento. Coorientadores: Isabela de Castro Sant'Anna, Leonardo Lopes Bhering e Moysés Nascimento.

O constante crescimento populacional e as limitadas áreas agricultáveis requerem o aumento da produtividade das espécies agrônomicas. Nos últimos 50 anos estima-se que 50% do aumento da produtividade mundial dessas espécies foi devido ao melhoramento genético. Para que o melhoramento seja feito de forma eficiente, o conhecimento do sistema reprodutivo das espécies a serem melhoradas é de suma importância para um melhorista, uma vez que tal conhecimento auxilia na adoção de métodos adequados para cada espécie. Em geral, a obtenção de cultivares melhoradas é um processo longo e oneroso. Visando a redução de tempo e custos, além do aumento da acurácia de seleção, a Seleção Genômica Ampla (*Genome Wide Selection* - GWS) que utiliza informações diretas do DNA por meio dos marcadores moleculares para predição do valor genético genômico dos indivíduos, foi proposta. Dentre as diversas metodologias de GWS, recentemente, foi proposto o uso da Regressão Quantílica Regularizada (RQR). A RQR permite ajustar modelos de regressão ao longo de toda distribuição da variável dependente, possibilitando assim uma melhor descrição do fenômeno em estudo, quando comparada a metodologias tradicionais que se baseiam apenas na média condicional. O uso da RQR tem-se mostrado bastante promissor, porém, ainda não foi avaliado em todo o processo de um programa de melhoramento de plantas. Diante do exposto, objetivou-se avaliar o uso da RQR na seleção genômica, considerando dados simulados de plantas autógamas com características oligogênicas. Foi simulada uma população  $F_2$ , com características com duas herdabilidades (0,4 e 0,8) e controladas por quatro genes. Foi realizado o avanço de gerações (até a  $F_6$ ) considerando duas porcentagens de seleção (10% e 20%) e, como critério de seleção, o valor genético genômico obtido por meio da RQR, além da seleção fenotípica e de outros métodos tradicionais de seleção genômica, especificamente RR-BLUP e o BLASSO. Observou-se que o modelo de RQR apresentou, em relação a fixação dos alelos favoráveis, resultados melhores ou iguais aqueles obtidos por todos métodos avaliados. Especificamente, em cenários de herdabilidade 0,4, independente da porcentagem de

seleção, somente a seleção dos indivíduos baseados no modelo de RQR no quantil ( $\tau = 0,5$ ) foi capaz de fixar os alelos favoráveis até a sexta geração. Por outro lado, em cenários de maior herdabilidade (0,8) e com porcentagem de seleção de 10%, a seleção baseada nos métodos RQR ( $\tau = 0,5$ ) e BLASSO permitiram a fixação dos alelos ainda na geração  $F_4$ . Quando a seleção se baseou nos métodos RR-BLUP e seleção fenotípica os alelos favoráveis não foram fixados até a sexta geração em nenhum cenário avaliado. Em relação ao ganho de seleção, a RQR ( $\tau = 0,5$ ) obteve ganhos maiores ou iguais aos métodos tradicionais de seleção genômica em todos os cenários avaliados. Especificamente, os ganhos da RQR ( $\tau = 0,5$ ) foram até 4,5% maiores que aqueles obtidos pelo BLASSO, até 6,1% maiores que os do RR-BLUP e até 4,6% maiores que a seleção fenotípica. Dessa forma verificou-se com aplicação da RQR no melhoramento de plantas considerando populações simuladas de plantas autógamas com características oligogênicas, seria possível uma redução de tempo e consequentemente de custos, devido a diminuição das gerações de autofecundações para a fixação dos alelos favoráveis em todos os cenários avaliados ou a obtenção de genótipos melhorados.

## ABSTRACT

OLIVEIRA, Gabriela França, M.Sc., Universidade Federal de Viçosa, February, 2019. **Quantile regression applied to genomic selection for oligogenic characteristics in autogamous plant breeding.** Advisor: Ana Carolina Campana Nascimento. Co-advisors: Isabela de Castro Sant'Anna, Leonardo Lopes Bhering and Moysés Nascimento.

The constant population growth and the limited agricultural areas require an increase in the productivity of agronomic species. Over the past 50 years it has been estimated that 50% of the world's productivity increase has been due to genetic improvement. For breeding to be done efficiently, the knowledge of the reproductive system of the species to be improved is of paramount importance to a breeder, since this knowledge assists in the adoption of methods suitable for each species. In general, obtaining improved cultivars is a long and costly process. In order to reduce time and costs, in addition to increasing accuracy of selection, Genome Wide Selection (GWS), which uses direct DNA information through molecular markers to predict the genomic genetic value of individuals, was proposed. Among the several methodologies of GWS, recently, the use of Regularized Quantile Regression (RQR) was presented. The RQR allows to adjust regression models throughout the distribution of the dependent variable allowing a better description of the phenomenon under study compared with traditional methodologies based on the conditional average. The use of RQR has shown to be quite promising, but has not yet been evaluated in a complete plant breeding process. In view of the above, the objective was to evaluate the use of RQR in genomic selection, considering simulated data of autogamous plants with oligogenic characteristics. An F<sub>2</sub> population with characteristics determined by four genes considering two heritabilities (0.4 and 0.8) was simulated. Generation (up to F<sub>6</sub>) was carried out considering two percentages of selection (10% and 20%) and, as a selection criterion, the genomic genetic value obtained through the RQR, besides the phenotypic selection and other traditional methods of genomic selection, specifically RR-BLUP and BLASSO. It was observed that the RQR model presented, in relation to the fixation of favorable alleles, better or equal results obtained by all evaluated methods. Specifically, in heritability 0.4 scenarios, regardless of the percentage of selection, only the selection of individuals based on the RQR model ( $\tau = 0.5$ ) was able to fix the favorable alleles up to the sixth generation. On the other hand, in scenarios with higher heritability (0.8) and with selection percentage of 10%, the selection based on the RQR ( $\tau = 0.5$ ) and BLASSO

methods allowed the alleles to be fixed in the F<sub>4</sub> generation. When selection was based on the RR-BLUP and phenotypic selection methods the favorable alleles were not fixed until the sixth generation in any evaluated scenario. In relation to the selection gain, the RQR ( $\tau = 0.5$ ) obtained gains greater than or equal to the traditional methods of genomic selection in all scenarios evaluated. Specifically, the gains from RQR ( $\tau = 0.5$ ) were up to 4.5% higher than those obtained by BLASSO, up to 6.1% higher than RR-BLUP and up to 4.6% higher than phenotypic selection. In this way, it was verified with the application of RQR in the improvement of plants considering simulated populations of autogamous plants with oligogenic characteristics, it would be possible a reduction of time and consequently of costs, due to the decrease of generations of self-fertilization for the establishment of favorable alleles in all assessed scenarios or obtaining improved genotypes.

## 1. INTRODUÇÃO

Em meados de 2017 a população mundial atingiu 7,6 bilhões de habitantes e a previsão é de que a mesma deva subir para 8,6 bilhões em 2030 (ONU - Organização das Nações Unidas). Desta forma, existe a necessidade de serem produzidos mais alimentos para atender ao constante crescimento dessa população, apesar de serem cada vez mais limitadas as áreas agricultáveis (BORÉM; MIRANDA, 2013). Nesse sentido, a busca pelo aumento da produtividade das espécies agronômicas torna-se relevante e, neste cenário, o melhoramento de plantas é uma estratégia importante, pois, estima-se que 50% do aumento da produtividade mundial das espécies agronômicas dos últimos 50 anos deve-se ao melhoramento (BORÉM; MIRANDA; FRITSCHÉ-NETO, 2017).

Para que o melhoramento seja feito de forma eficiente, o conhecimento do sistema reprodutivo das espécies a serem melhoradas é de suma importância para um melhorista. Tal conhecimento auxilia na adoção de métodos adequados para cada espécie (BESPALHOK; GUERRA; OLIVEIRA, 2007). Dentre os principais sistemas, aqueles baseados em autofecundação, como por exemplo, arroz, aveia, cevada, café, feijão, soja, pêssego, tomate, trigo e acasalamento ao acaso como, batata, cana-de-açúcar, mandioca, maçã, milho. Estes sistemas apresentam destaque visto que são sistemas de cruzamentos de importantes culturas (BORÉM; MIRANDA; FRITSCHÉ-NETO, 2017).

Em geral e independente do sistema reprodutivo da espécie, a obtenção de cultivares melhoradas é um processo longo e oneroso, sendo um dos desafios do melhoramento a busca de procedimentos capazes de fornecer genótipos superiores em um menor tempo e, conseqüentemente, a um menor custo. Por exemplo, em espécies perenes o tempo para obtenção de uma cultivar melhorada pode ser superior a 30 anos enquanto que em espécies anuais, esse processo pode levar em média 12 anos (BESPALHOK; GUERRA; OLIVEIRA, 2007).

Visando a redução de tempo e aumento da acurácia de seleção Meuwissen et al. (2001) propuseram a Seleção Genômica Ampla (*Genome Wide Selection - GWS*). Na GWS são utilizadas informações diretas do DNA por meio dos marcadores moleculares e fenótipos para predição do valor genético genômico dos indivíduos. A principal vantagem em se utilizar a GWS é que ela permite estimar o valor genético genômico dos indivíduos que ainda não tiveram seus fenótipos coletados, gerando assim uma

redução no intervalo de gerações e aumento do ganho genético (RESENDE et al., 2008; CROSSA et al., 2011; CROSSA et al., 2017).

As potencialidades da GWS no melhoramento de plantas foram exploradas na literatura, especificamente, para sistema de autofecundação, Heffner et al. (2011) verificaram que a acurácia de predição da GWS é superior à *Marker Assisted Selection* (MAS) e à Seleção Fenotípica (SF) em culturas de trigo. Pestana (2013) avaliou a eficiência da GWS em prever os valores genômicos dos cafeeiros para resistência à ferrugem, e observou que a GWS conseguiu prever com eficiência esses valores mesmo em populações com tamanho amostral reduzido e baixa densidade de marcadores para os padrões ideais da seleção genômica. Romero (2017), avaliou cenários de diferentes densidades e tamanhos populacionais para o melhoramento de características oligogênicas em *Coffea arabica*, com intuito de estabelecer em qual geração ocorre a fixação dos alelos favoráveis. Como resultado, os autores observaram que em pequenas populações (comumente utilizadas em programas de melhoramento de café) os alelos favoráveis foram fixados na sexta geração ( $F_6$ ), enquanto que em grandes populações essa fixação ocorreu na quinta geração ( $F_5$ ). A GWS também foi aplicada com êxito em outras culturas autógamas, como: arroz (SPINDEL et al., 2015), aveia (ASORO et al., 2011) e cevada (ZHONG et al., 2009).

Uma metodologia alternativa e ainda pouco explorada para estudos de seleção genômica é a Regressão Quantílica (RQ) (KOENKER e BASSET, 1978). Tal metodologia permite ajustar modelos de regressão ao longo de toda a distribuição da variável dependente, ou seja, a metodologia permite analisar a influência de um conjunto de variáveis explicativas em diferentes quantis da variável resposta. Outras características importantes da RQ é não requerer pressuposições sobre a distribuição do erro e o fato de ser robusta a pontos discrepantes. Especificamente, quando aplicada à estudos de GWS, a RQ permite: i) ajustar modelos em diferentes níveis (quantis) ao longo de toda a distribuição dos valores fenotípicos e, ii) a obtenção dos valores genéticos genômicos nos diferentes quantis de interesse (BARROSO et al., 2017; NASCIMENTO et al., 2017; NASCIMENTO et al., 2018). Para solucionar problemas de alta dimensionalidade e uma possível multicolinearidade, a Regressão Quantílica Regularizada (RQR) foi proposta por Li e Zhu (2008). A RQR aplicado a GWS é uma metodologia que combina propriedades dos métodos tradicionais da GWS com as características desejáveis da RQ.

A RQR foi proposta e avaliada para estudos de seleção genômica por Nascimento et al. (2017). Estes autores utilizaram RQR para estimar o valor genético genômico em dados simulados com cenários apresentando diferentes graus de assimetria da distribuição do fenótipo. Os resultados da RQR foram comparados aos advindos do método BLASSO (*Bayesian Least Absolute Shrinkage and Selection Operator*) e os autores observaram que a RQR apresentou um menor erro quadrático médio em relação ao BLASSO, indicando que esta é uma alternativa viável para análises de GWS, inclusive em cenários onde não se observa assimetria na distribuição dos fenótipos. Santos et al. (2018) também utilizaram a RQR e BLASSO para estimar o mérito genético em suínos para características assimétricas relacionadas à carcaça e observaram que a RQR apresentou acurácias iguais ou superiores ao BLASSO para todas as características avaliadas. Neste contexto o uso da RQ tem-se mostrado bastante promissora, porém, ainda não foi avaliada em todo processo de um programa de melhoramento, sendo, atualmente utilizada somente para ranqueamento e cálculo de acurácia dos indivíduos.

Diante do exposto, o presente trabalho tem por objetivo verificar se a utilização da Regressão Quantílica Regularizada na seleção genômica, considerando dados simulados de plantas autógamas com características oligogênicas permite a fixação dos alelos favoráveis em gerações anteriores, acarretando uma redução de custos e tempo para a seleção de indivíduos geneticamente superiores. Os resultados obtidos por meio da RQR serão comparados com aqueles advindos da seleção fenotípica e de outras abordagens de seleção genômica, especificamente RR-BLUP e o BLASSO. Finalmente, serão calculados o ganho de seleção obtido por cada abordagem avaliada.

## **2. REFERENCIAL TEÓRICO**

### **2.1 Melhoramento Genético de Plantas Autógamas**

O melhoramento genético de plantas visa a obtenção de indivíduos geneticamente superiores, e o desenvolvimento de cultivares melhoradas para as necessidades dos agricultores e consumidores. Existem alguns objetivos em comum nos programas de melhoramento, independente da cultura que está sendo trabalhada, como: aumento de produtividade, qualidade, adaptação em novas áreas e obtenção de

cultivares resistentes a pragas e doenças (BESPALHOK; GUERRA; OLIVEIRA, 2007).

O melhoramento genético via seleção é uma das principais ferramentas de um melhorista, independentemente do método de melhoramento utilizado. A seleção busca aumentar a frequência de genes associados às características de interesse em cada geração. O processo de seleção pode ser utilizado tanto para espécies autógamas como espécies alógamas. Plantas autógamas são plantas que se reproduzem por autofecundação, como por exemplo: trigo, arroz, amendoim, feijão, alface, berinjela, café, soja e pêssego (VALLEJO et al., 2010; BORÉM e MIRANDA, 2013). Enquanto as alógamas são plantas que se reproduzem por fecundação cruzada, como por exemplo: alfafa, cana-de-açúcar, eucalipto, maçã, mandioca, milho e uva (BORÉM e MIRANDA, 2013).

O melhoramento genético é um processo bastante longo e oneroso. Por exemplo, em espécies perenes esse processo pode ser superior a 30 anos, já em espécies anuais esse processo dura em média 12 anos (BESPALHOK; GUERRA; OLIVEIRA, 2007).

A introdução de alelos de interesse nas linhagens superiores em programas de melhoramento de plantas autógamas pode ser realizada por meio do processo de hibridação. Para o cruzamento, são utilizados indivíduos parentais geneticamente distintos e complementares para as características desejáveis. Os híbridos resultantes deste cruzamento podem ser conduzidos por sucessivas autofecundações e seleções de genótipos superiores até atingirem a homozigotidade (BORÉM e MIRANDA, 2013). A homozigotidade é atingida quando ocorre a fixação dos alelos favoráveis, ou seja, ocorre a fixação das características desejáveis, pois após as sucessivas gerações de autofecundações ocorre um aumento do número de indivíduos homozigotos e a diminuição dos indivíduos heterozigotos.

## **2.2 Seleção Genômica Ampla**

O avanço da genética molecular trouxe vários benefícios para o melhoramento vegetal e animal. Esse avanço possibilitou a utilização de informações diretas do DNA para a seleção de indivíduos geneticamente superiores. Neste contexto, os marcadores moleculares são ferramentas úteis para identificar variações no genoma, permitindo assim a distinção de indivíduos geneticamente diferentes e também a seleção de

genótipos de indivíduos sem a necessidade da ocorrência da expressão fenotípica (RESENDE et al., 2008; CROSSA et al., 2011).

Atualmente, apesar de existir uma diversidade de marcadores moleculares, os que mais se destacam são os SNPs (*Single Nucleotide Polymorphisms*), por serem amplamente distribuídos no genoma e ser codominantes, além da facilidade de automação. Diante da abundância destes marcadores, Meuwissen et al. (2001) idealizaram a Seleção Genômica Ampla (*Genome Wide Selection – GWS*), cujo modelo geral é definido por:

$$y_i = \mu + \sum_{j=1}^p x_{ij}g_j + e_i, \quad (1)$$

em que:

$y_i$  é o valor do fenótipo para o indivíduo  $i$  ( $i = 1, 2, \dots, n$ );

$\mu$  é a média da característica de interesse;

$g_j$  é o efeito do marcador SNP ( $j = 1, 2, \dots, p$ );

$x_{ij}$  são os elementos da matriz de incidência do marcador  $j$  no indivíduo  $i$ ;

$e_i$  é a  $i$ -ésima observação do erro aleatório.

O principal propósito da GWS é a análise de uma grande quantidade de marcadores moleculares distribuídos no genoma, afim de estimar os seus efeitos sobre as características fenotípicas, predizer o valor genético genômico dos indivíduos para posterior seleção e capturar os genes que afetam o caráter quantitativo. Devido a abundância destes marcadores, supõe-se que eles ou parte deles estejam em desequilíbrio de ligação com locos de características quantitativas (*Quantitative Trait Loci – QTL*) (RESENDE et al., 2014).

Para que a seleção genômica ocorra de maneira eficaz, deve-se atentar para a escolha adequada do método estatístico a ser utilizado e esta depende da quantidade de marcadores utilizados, do tipo de marcador, da intensidade do desequilíbrio de ligação, do nível de relação entre a população de treinamento e a população a ser melhorada, entre outros (NAKAYA e ISOBE, 2012). Desta forma, a escolha de metodologias adequadas e a comparação das mesmas é imprescindível para a eficácia da seleção genômica nos programas de melhoramento.

Nas próximas seções serão apresentadas algumas metodologias que são utilizadas em estudos em seleção genômica e que serão empregadas neste trabalho.

### 2.2.1 Regressão Quantílica

Os modelos tradicionais de regressão, são abordagens bastante utilizadas, por serem de fácil implementação e interpretação. Na aplicação destes modelos alguns pressupostos devem ser atendidos para que a estimação seja confiável, mas nem sempre eles são válidos para dados reais. Por exemplo, se o pressuposto de normalidade dos erros for violado, a média condicional pode ser uma medida de tendência central inadequada, pois ela pode estar sendo influenciada por valores discrepantes. Essa violação acarreta em testes de hipóteses inválidos devido ao viés nos valores  $p$  (HAO e NAIMAN, 2007). Outro pressuposto que nem sempre é atendido é o de homocedasticidade dos erros. Neste caso, por mais que os estimadores de mínimos quadrados obtidos sejam não viesados os mesmos não possuirão variância mínima (MONTGOMERY et al., 2012). Portanto, os estimadores de mínimos quadrados obtidos não serão BLUE (*Best Linear Unbiased Estimator*).

Visando contornar essas limitações Koenker e Bassett (1978) propuseram um método mais geral, a Regressão Quantílica (RQ). Esse método é baseado no método dos erros absolutos ponderados que, diferentemente dos modelos tradicionais de Regressão, que estimam somente um modelo para todo conjunto de dados, a RQ permite ajustar modelos de regressão ao longo de toda a distribuição da variável dependente, fornecendo assim uma análise estatística mais completa da relação das variáveis estudadas.

Geralmente, os quantis são definidos de acordo com finalidade do estudo. Por exemplo, se o interesse está em examinar uma localização na cauda inferior pode-se utilizar o quantil 0,1 ou quantil 0,9 se o interesse recai sobre a calda superior (HAO e NAIMAN, 2007).

Além disso, de acordo com Koenker (2005), os modelos de RQ são robustos a pontos discrepantes além de serem capazes de lidar com uma possível heterocedasticidade, que pode ser detectada a partir da variação das estimativas dos coeficientes dos parâmetros para os diversos quantis.

O modelo de regressão quantílica proposto por Koenker e Bassett (1978) que descreve a relação funcional entre a variável dependente e as variáveis independentes, pode ser descrito conforme a equação:

$$y_i = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \beta_2(\tau)x_{i2} + \dots + \beta_p(\tau)x_{ip} + e_i(\tau),$$

em que  $\beta_0(\tau)$  é a constante de regressão no quantil  $\tau$ ,  $\beta(\tau)$  são os coeficientes da regressão,  $e_i(\tau)$  são os erros aleatórios independentes e identicamente distribuídos com o quantil de ordem  $\tau$  igual a zero e  $\tau$  refere-se ao quantil assumido ( $\tau \in (0,1)$ ).

Os parâmetros da RQ são estimados por meio de algoritmos de programação linear (SILVA e PORTO, 2006; MISSIO et al., 2009).

As potencialidades da Regressão Quantílica já foram exploradas em algumas áreas do conhecimento, como por exemplo, na economia (NASCIMENTO et al., 2012; NASCIMENTO et al., 2017), medicina humana (BEYERLEIN et al., 2011) e também no melhoramento genético (BARROSO et al., 2017; NASCIMENTO et al., 2017; NASCIMENTO et al., 2018; SANTOS et al., 2018).

### 2.2.1.1 Estimação dos Parâmetros

Nos modelos de RQ, a estimação dos parâmetros se dá a partir do método dos erros absolutos ponderados, neste caso, a distância dos pontos observados à reta ajustada é medida minimizando a soma da média ponderada das distâncias verticais, onde se atribui um peso  $1 - \tau$  para os pontos abaixo da reta e para os pontos acima da reta o peso atribuído é  $\tau$  (HAO e NAIMAN,2007). A escolha do valor de  $\tau$  resultará em uma função ajustada do quantil condicional. Então, para os diferentes  $\tau$ 's utilizados, é necessário encontrar estimadores com as propriedades desejáveis para cada um deles.

O objetivo é encontrar os valores de  $\beta_0(\tau), \beta_1(\tau), \dots, \beta_p(\tau)$  que minimizem a equação:

$$\sum_{i=1}^n d_{\tau}(y_i, \hat{y}_i) = \sum_{y_i \geq \hat{y}_i} \tau |y_i - [\hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)x_{i1} + \hat{\beta}_2(\tau)x_{i2} + \dots + \hat{\beta}_p(\tau)x_{ip}]| + \sum_{y_i < \hat{y}_i} (\tau - 1) | -[\hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)x_{i1} + \hat{\beta}_2(\tau)x_{i2} + \dots + \hat{\beta}_p(\tau)x_{ip}]|,$$

em que  $d_{\tau}$  é a distância entre  $y_i$  e  $\hat{y}_i$ ;  $\hat{y}_i = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)x_{i1} + \hat{\beta}_2(\tau)x_{i2} + \dots + \hat{\beta}_p(\tau)x_{ip}$ . Ao minimizar esta equação tem-se como resultados as estimativas dos coeficientes da regressão do quantil de interesse. A minimização desta função exige a utilização de algoritmos de programação linear como o *Método Simplex* que é apresentado em detalhes por Koenker (2005).

### 2.2.1.2 Regressão Quantílica Regularizada (RQR)

A Regressão Quantílica Regularizada (RQR) foi proposta por Li e Zhu (2008) e consiste em um método capaz de solucionar problemas de alta dimensionalidade e uma possível multicolinearidade, decorrentes da quantidade superior de parâmetros a serem estimados em relação a quantidade de observações. Além disso, é também indicado quando a variável de interesse apresenta distribuição assimétrica. A RQR combina propriedades dos métodos tradicionais da GWS (por exemplo, penalização via encurtamento dos coeficientes de regressão- *shrinkage*) com as propriedades desejáveis da RQ (NASCIMENTO et al., 2017).

Considerando o modelo geral de GWS (1), a obtenção dos estimadores dos coeficientes de regressão se dá a partir da solução do seguinte problema de otimização:

$$\hat{g} = \underset{g}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \rho_{\tau} \left( y_i - \mu - \sum_{j=1}^p x_{ij} g_j \right) + \lambda \sum_{j=1}^p |g_j| \right\}$$

em que  $\sum_{j=1}^p |g_j|$  é a soma dos valores absolutos dos coeficientes de regressão,  $\lambda$  é o parâmetro de penalização e  $\rho_{\tau}(\cdot)$ , denotada função *check* por Koenker e Bassett (1978), é definida por:

$$\rho_{\tau} \left( y_i - \mu - \sum_{j=1}^p x_{ij} g_j \right) = \begin{cases} \tau \left( y_i - \mu - \sum_{j=1}^p x_{ij} g_j \right), & \text{se } y_i - \mu - \sum_{j=1}^p x_{ij} g_j \geq 0 \\ (\tau - 1) \left( y_i - \mu - \sum_{j=1}^p x_{ij} g_j \right), & \text{caso contrário} \end{cases}$$

em que  $\tau \in (0,1)$  indica o quantil de interesse.

O valor genético genômico estimado (GEBV), via RQR, é obtido pela expressão:

$$\hat{y}(\tau) = \sum_j x_{ij} \hat{g}_j(\tau)$$

em que  $x_{ij}$  são os elementos da matriz de incidência do marcador  $j$  no indivíduo  $i$  e  $\hat{g}_j(\tau)$  é o efeito do  $j$ -ésimo marcador SNP considerando o quantil  $\tau$ .

### 2.2.2 Método RR-BLUP

O método RR-BLUP (*Random Regression/ Ridge Regression Best Linear Unbiased Predictor*), consiste em um método de regressão aleatória, proposto por Meuwissen et al. (2001). Essa abordagem é utilizada para modelar os efeitos dos marcadores, contornando os problemas de dimensionalidade e multicolinearidade. O RR-BLUP utiliza preditores do tipo BLUP (*Best Linear Unbiased Prediction*), estima simultaneamente o efeito de todos os marcadores e assume que os efeitos dos marcadores são covariáveis de efeitos aleatórios (RESENDE et al., 2014).

Para contornar o problema de dimensionalidade, o RR-BLUP utiliza em seu procedimento um método de penalização ou *shrinkage*. Os estimadores associados ao modelo produzem um parâmetro de penalização  $\lambda$ . Quando  $\lambda$  é desconhecido, tem-se a regressão *Ridge Regression*. Em contrapartida, se esse parâmetro estiver relacionado à  $\lambda = \frac{\sigma_e^2}{\left(\frac{\sigma_g^2}{n_Q}\right)}$  tem-se a regressão aleatória BLUP. Em que  $\sigma_e^2$  é a variância residual;  $\sigma_g^2$  é a variância genética aditiva da característica;  $n_Q$  é o número de possíveis QTLs, uma quantidade desconhecida *a priori*, mas que pode ser inferida por meio da seguinte expressão:  $n_Q = 2 \sum_i^n p_i(1 - p_i)$ , onde  $p_i$  é a frequência alélica de A do marcador  $i$ .

A estimação dos efeitos genéticos dos marcadores via RR-BLUP é baseada no seguinte modelo:

$$y = Wb + Xg + e,$$

em que:

$y$  é o vetor de observações fenotípicas ( $I \times 1$ , sendo  $I$  o número de indivíduos genotipados e fenotipados);

$b$  é o vetor de efeitos fixos (média geral), com matriz de incidência  $W$  ( $I \times p$ );

$g$  é o vetor de efeitos aleatórios dos marcadores ( $J \times 1$ , sendo  $J$  o número de marcadores moleculares) associado a matriz de incidência  $X$ ;

$e$  é o vetor de erros aleatórios.

Matricialmente, as equações de estimação dos efeitos genéticos dos marcadores via RR-BLUP podem ser escritas conforme a seguir:

$$\begin{bmatrix} W'W & W'X \\ X'W & X'X + I \frac{\sigma_e^2}{\sigma_g^2/n_Q} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ W'y \end{bmatrix},$$

em que se pressupõe *a priori* que todos os marcadores explicam quantidades iguais da variação genética aditiva ( $\sigma_g^2$ ). Para estimar os componentes de variância ( $\sigma_e^2$ ) e ( $\sigma_g^2$ ), utiliza-se o método *Restricted Maximum Likelihood* (REML).

O valor genético genômico estimado (GEBV) dos indivíduos, via RR-BLUP, é obtido pela expressão:

$$\hat{y}_i = \sum_j x_{ij} \hat{g}_j,$$

em que  $x_{ij}$  são os elementos da matriz de incidência do marcador  $j$  no indivíduo  $i$  e  $\hat{g}_j$  é o efeito do  $j$ -ésimo marcador SNP.

### 2.2.3 LASSO Bayesiano (BLASSO)

A versão bayesiana da regressão LASSO (*Bayesian Least Absolute Shrinkage and Selection Operator*- BLASSO) foi proposta por Park e Casella (2008) e adaptado para seleção genômica por de Los Campos et al. (2009). A estimação dos efeitos de marcadores via BLASSO é baseado no seguinte modelo:

$$y = Wb + Xg + e,$$

em que:

$y$  é o vetor de observações fenotípicas ( $I \times 1$ , sendo  $I$  o número de indivíduos genotipados e fenotipados);

$b$  é o vetor de efeitos fixos (média geral), com matriz de incidência  $W$  ( $I \times p$ );

$g$  é o vetor de efeitos aleatórios dos marcadores ( $J \times 1$ , sendo  $J$  o número de marcadores moleculares) associado a matriz de incidência  $X$ ;

$e$  é o vetor de erros aleatórios.

Na inferência bayesiana, os parâmetros são tratados como quantidades aleatórias e toda a informação *a priori* pode ser representada por modelos probabilístico. Desta forma, é assumido distribuições de probabilidade para os dados e para todas as quantidades desconhecidas (parâmetros), respectivamente, dadas por:

$$y|g_1, g_2, \dots, g_j, \sim \text{Normal}(Wb + Xg, I\sigma^2);$$

$$e|\sigma^2 \sim \text{MVN}(0, I\sigma^2);$$

$$\sigma^2 \sim \chi^{-2}(v_e S_e^2);$$

em que *MVN* refere-se à distribuição normal multivariada,  $\sigma^2$  tem como distribuição a

*priori* uma qui-quadrado invertida escalonada com hiperparâmetros  $v_e$  e  $S_e^2$ .

Empregando uma formulação em termos de um modelo hierárquico aumentado, tem-se:

$$g_j | \tau \sim N(0, D\sigma^2)$$

$$p(\tau^2 | \lambda^2) = \prod_i \left( \frac{\tau^2}{2} \right) e^{\left[ \frac{-\tau^2 \lambda^2}{2} \right]}$$

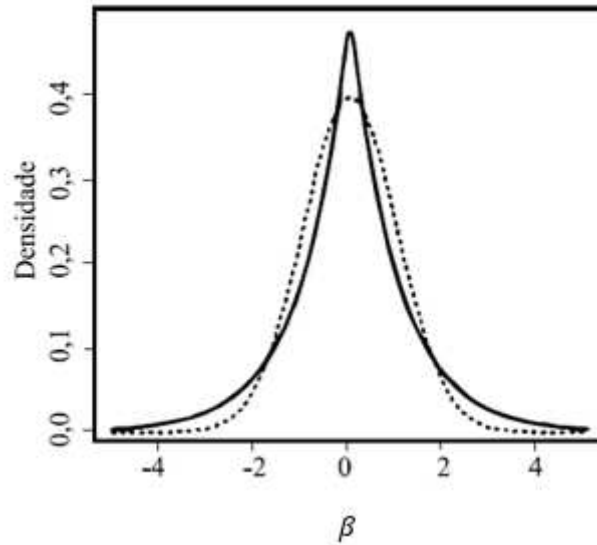
em que  $D = \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_j^2)$ ,  $\lambda$  é o parâmetro de penalização que pode ser estimado por meio do método MCMC (*Markov Chain Monte Carlo*). Segundo Park e Casella (2008), essa formulação conduz a uma distribuição exponencial dupla para os efeitos dos marcadores, ou seja:

$$p(g_j | \lambda^2) = \int_{\mathbb{R}} N(0, \sigma^2 \tau^2) \text{Exp}\left(\frac{\lambda^2}{2}\right) d\tau^2 \propto \frac{1}{2\left(\frac{\sigma}{\lambda^2}\right)} e^{\left(-\frac{g_j}{\sigma/\lambda^2}\right)},$$

$$g_j | \lambda^2 \sim \text{ExpDupla}\left(0, \frac{\sigma}{\lambda^2}\right).$$

Portanto, supõe-se *a priori* que os efeitos dos marcadores seguem uma distribuição exponencial dupla, com variâncias heterogêneas.

Em comparação à distribuição normal, a distribuição exponencial dupla possui maior massa de densidade no zero e possui caldas mais robustas, conforme o observado na Figura 1. De acordo com de Los Campos et al. (2009), o método BLASSO produz uma penalização ou regularização mais forte para os efeitos dos marcadores que são próximos a zero e menor naqueles distantes de zero, sendo essa penalização específica para cada marcador.



**Figura 1.** Densidades das distribuições normal (curva pontilhada) e exponencial dupla (curva cheia), ambas com médias iguais a zero e variâncias iguais a um.

Fonte: Tibshirani (1996).

O valor genético genômico estimado (GEBV) do indivíduo, via BLASSO, é obtido pela expressão:

$$\hat{y}_i = \sum_j x_{ij} \hat{g}_j$$

em que  $x_{ij}$  é o elemento da matriz de incidência relativo ao marcador  $j$  e ao indivíduo  $i$  e  $\hat{g}_j$  é o efeito do  $j$ -ésimo marcador SNP.

### 2.3 Simulação computacional como ferramenta auxiliar em programas de melhoramento

O grande desafio dos programas de melhoramento genético, independente da cultura utilizada, é a busca contínua de cultivares melhorados. Existem diversos métodos de seleção, como por exemplo: seleção fenotípica, seleção via progênie, seleção genômica, entre outros. Apesar das diversas metodologias de seleção disponíveis, a maioria delas demanda a realização de experimentos para obtenção do banco de dados, e essa estratégia exige tempo e recursos (SILVA, 2014).

Com o avanço computacional, um recurso bastante utilizado pelos melhoristas com o intuito de reduzir custos e tempo, e ainda pela facilidade em generalização, é a

simulação computacional. Em linhas gerais, a simulação consiste em reproduzir um comportamento real por meio de modelos que podem ser testados em inúmeros cenários, com o objetivo de compreender o funcionamento do mesmo. A simulação tem o propósito de gerar informações eficientes para a compreensão do funcionamento de um modelo real, antes de aplicá-lo no campo (FERREIRA, 2001; SUN et al., 2011). Uma outra vantagem de utilizar dessa técnica é que os experimentos podem ser replicados diversas vezes, gerando assim inferências mais precisas (GURGEL, 2004; SUN et al., 2011; WANG, 2012).

## **2.4 Capacidade preditiva e ganho de seleção**

Visando avaliar e comparar as diferentes abordagens de GWS no processo de seleção, nos estudos geralmente calculam-se a capacidade preditiva e o ganho de seleção obtidos em cada geração.

A capacidade preditiva corresponde ao coeficiente de correlação entre valores fenotípicos observados ( $y$ ) e os valores genômicos estimados ( $\hat{y}$ ), ou seja,  $r_{y,\hat{y}} = cor(y, \hat{y})$  (RESENDE et al., 2012).

O ganho de seleção (GS) é obtido em função do diferencial de seleção (DS) e a herdabilidade de acordo com a seguinte expressão:

$$GS = DS * h^2$$

em que  $DS$  corresponde a diferença entre a média da população selecionada em relação a média da população original, onde, foi realizado o processo de seleção e  $h^2$  corresponde a herdabilidade.

## **3. MATERIAL E MÉTODOS**

### **3.1 Simulação das populações**

Para o presente estudo foi simulado no programa GENES (CRUZ, 2013) uma população base com um genoma de 1040 centimorgans (cM) de extensão. O genoma simulado contém oito grupos de ligação com marcadores espaçados a cada 1 cM. Foram definidos quatro locos controladores da característica, localizados em quatro diferentes grupos de ligações, com efeito uniforme e ausência de dominância e epistasia.

A população  $F_1$  utilizada foi obtida pelo produto de parentais contrastantes, gerando assim gametas para a formação da população  $F_2$ , composta de 625 indivíduos. Uma vez gerado o genoma base, foram simulados os valores genotípicos e dois conjuntos de valores fenotípicos, com herdabilidades ( $h^2$ ) de 40% e 80%. Para determinar os valores genotípicos ( $vg_i$ ), foi utilizada a seguinte equação:

$$vg_i = \mu + a_i + d_i, \quad (2)$$

em que:

$vg_i$  é o valor genotípico do indivíduo  $i$ ,  $\mu$  é a média genotípica populacional ( $\mu = 1,0$ );  $a_i$  é o efeito aditivo, onde  $a_i = \sum_{k=1}^4 \rho_k \alpha_k$  sendo  $\rho_k = 2,5$  o efeito do alelo favorável com a mesma contribuição para todo loco  $k$  e  $\alpha_k$  é a contribuição do loco  $k$  (1, 0 ou -1 para classes genotípicas AA, Aa e aa, respectivamente);

$d_i$  é o efeito devido a dominância, considerado nulo neste estudo ( $d_i = 0$ ).

E para determinação dos valores fenotípicos ( $vf_i$ ), foi utilizada a seguinte equação:

$$vf_i = vg_i + \varepsilon_i,$$

em que:

$vf_i$  é o valor fenotípico do indivíduo  $i$ ;

$vg_i$  é o valor genotípico do indivíduo  $i$ ;

$\varepsilon_i$  é o efeito ambiental gerado segundo uma distribuição normal com média e variância compatível com a herdabilidade da característica especificada ( $\varepsilon_i \sim N(0, \sigma_e^2)$ ), sendo

$\sigma_e^2 = \frac{\sigma_g^2(1-h^2)}{h^2}$ , onde  $\sigma_g^2$  é a variância genética.

### 3.2 Avanço das gerações

As gerações  $F_3, F_4, F_5$  e  $F_6$  foram obtidas por autofecundação considerando a geração ( $F_2$ ) como base. Para geração  $F_2$  o modelo de GWS foi ajustado por diferentes metodologias e obtidos os valores genéticos genômicos (GEBV) dos indivíduos. A partir dos GEBVs, foram selecionados indivíduos que possuem maiores valores genéticos genômicos baseados no modelo ajustado na  $F_2$ . A quantidade de indivíduos depende da porcentagem de seleção, neste trabalho utilizou-se porcentagens de seleção de 10% e 20% e, a partir dos genótipos dos indivíduos selecionados, as gerações subsequentes ( $F_3$  até a  $F_6$ ), de 200 indivíduos cada, foram simuladas no programa R (R

Development Core Team, 2018), conforme descrito no Apêndice B e adaptado de ROMERO (2017).

### 3.3 Predição Genômica

O modelo geral de seleção genômica a ser ajustado pode ser descrito como (MEUWISSEN et al., 2001):

$$y_i = \sum_{j=1}^{1040} x_{ij}g_j + e_i \quad (2)$$

em que  $y_i$  é a  $i$ -ésima observação do fenótipo  $y$  com dimensão  $I \times 1$ , sendo  $I = 625$ ,  $g_j$  é o efeito do  $j$ -ésimo marcador ( $j = 1, 2, \dots, 1040$ ),  $x_{ij}$  são os elementos da matriz de incidência do marcador  $j$  no indivíduo  $i$ , com parametrização 1, 0 e  $-1$  e  $e_i$  é a  $i$ -ésima observação do erro aleatório,  $e_i \sim N(0, \sigma_e^2)$ .

Os parâmetros do modelo (2) foram estimados usando três diferentes metodologias: LASSO Bayesiano (BLASSO) (DE LOS CAMPOS et al., 2009), RR-BLUP (MEUWISSEN et al., 2001) e Regressão Quantílica Regularizada (RQR) (LI e ZHU, 2008).

#### 3.3.1 RR-BLUP

O método RR-BLUP (*Random Regression/ Ridge Regression Best Linear Unbiased Predictor*), consiste em um método de regressão aleatória, proposto por Meuwissen et al. (2001). O RR-BLUP utiliza preditores do tipo BLUP (*Best Linear Unbiased Prediction*) para estimar simultaneamente o efeito de todos marcadores, assumindo que os efeitos dos marcadores são covariáveis de efeitos aleatórios (RESENDE et al., 2014).

A estimação dos efeitos genéticos dos marcadores via RR-BLUP é baseado no seguinte modelo:

$$y = 1\mu + Xg_{rr} + e,$$

em que:

$y$  é o vetor de observações fenotípicas ( $I \times 1$ , sendo  $I$  o número de indivíduos genotipados e fenotipados);

$\mu$  é a média geral e  $1$  é o vetor com dimensão  $(I \times 1)$  cujos seus elementos são iguais a 1);

$g$  é o vetor de efeitos aleatórios dos marcadores  $(J \times 1)$ , sendo  $J$  o número de marcadores moleculares) associado a matriz de incidência  $X$  com parametrização -1, 0 e 1;

$e$  é o vetor de erros aleatórios.

Matricialmente, as equações de estimação dos efeitos genéticos dos marcadores via RR-BLUP podem ser escritas conforme a seguir:

$$\begin{bmatrix} 1'1 & 1'X \\ X'1 & X'X + I\lambda_{RR} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{g}_{rr} \end{bmatrix} = \begin{bmatrix} 1'y \\ X'y \end{bmatrix},$$

em que  $\lambda_{RR} = \frac{\sigma_e^2}{\sigma_g^2} = \frac{\sigma_e^2}{\frac{\sigma_g^2}{n_Q}}$ , é chamado de parâmetro de penalização sendo  $\sigma_e^2$  a variância

residual,  $\sigma_g^2$  a variância genética aditiva do caráter e  $n_Q$  o número de QTL. A estimação dos coeficientes da regressão foi obtida através da função *mixed.solve* do pacote *rrBLUP* (ENDELMAN, 2011). Após a estimação dos coeficientes da regressão (efeitos dos marcadores), os valores genéticos genômicos no modelo RR-BLUP foram obtidos por meio da seguinte equação:

$$\hat{y}_i = \sum_j x_{ij} \hat{g}_{rrj},$$

em que  $x_{ij}$  são os elementos da matriz de incidência do marcador  $j$  no indivíduo  $i$  e  $\hat{g}_j$  é o efeito do  $j$ -ésimo marcador SNP.

### 3.3.2 Lasso Bayesiano (BLASSO)

A versão bayesiana da regressão LASSO (*Bayesian Least Absolute Shrinkage and Selection Operator*- BLASSO), proposta por Park e Casella (2008) e adaptada para seleção genômica por de Los Campos et al. (2009), inclui um termo de variância comum para os efeitos genéticos de marcadores e outro de efeitos residuais. A estimação dos efeitos de marcadores via BLASSO é baseado no seguinte modelo:

$$y = Wb + Xg + e,$$

em que:

$y$  é o vetor de observações fenotípicas  $(I \times 1)$ , sendo  $I$  o número de indivíduos

genotipados e fenotipados);

$b$  é o vetor de efeitos fixos (média geral), com matriz de incidência  $W$  ( $I \times p$ );

$g$  é o vetor de efeitos aleatórios dos marcadores ( $J \times 1$ , sendo  $J$  o número de marcadores moleculares) associado a matriz de incidência  $X$ ;

$e$  é o vetor de erros aleatórios.

Para implementação do método, supõe-se as seguintes distribuições *a priori* para os parâmetros do modelo:

$$e|\sigma^2 \sim MVN(0, I\sigma^2);$$

$$g_j|\lambda, \sigma^2 \sim \prod_{j=1}^p \frac{\lambda}{2\sigma} \exp\left(\frac{-\lambda|g_j|}{\sigma}\right)$$

em que *MVN* refere-se à distribuição normal multivariada,  $\sigma^2$  tem como distribuição *a priori* uma qui-quadrado invertida escalada  $\lambda$  é o parâmetro de regularização que é estimado por meio conjunto de dados via método MCMC usando uma *priori* não informativa.

Empregando-se uma formulação em termos de um modelo hierárquico aumentado, tem-se:

$$g_j|\tau \sim N(0, D\sigma^2)$$

$$p(\tau^2|\lambda^2) = \prod_i \left(\frac{\tau^2}{2}\right) e^{\left[\frac{-\tau^2\lambda^2}{2}\right]}$$

em que  $D = \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_j^2)$ . Segundo Park e Casella (2008), essa formulação conduz a uma distribuição exponencial dupla para os efeitos dos marcadores, ou seja:

$$g_j|\lambda^2 \sim \text{ExpDupla}\left(0, \frac{\sigma}{\lambda^2}\right).$$

O valor genético genômico estimado (GEBV) do indivíduo, via BLASSO, é obtido pela expressão:

$$\hat{y}_i = \sum_j x_{ij} \hat{g}_j$$

em que  $x_{ij}$  é o elemento da matriz de incidência relativo ao marcador  $j$  e ao indivíduo  $i$  e  $\hat{g}_j$  é o efeito do  $j$ -ésimo marcador SNP.

Neste trabalho os modelos bayesianos foram ajustados utilizando a função *bglr* implementada no pacote *BGLR* (CAMPOS e RODRIGUEZ, 2015) com 100.000 iterações para os algoritmos MCMC (*Markov chain Monte Carlo*), das quais 10.000

foram descartadas (*burn-in*) para garantir o aquecimento da cadeia e um *thin* de 5. A análise de convergência foi realizada baseando-se nos critérios propostos por Raftery e Lewis (1992) e Heidenberg e Welch (1983).

### 3.3.3 Regressão Quantílica Regularizada

A Regressão Quantílica Regularizada (RQR) foi proposta por Li e Zhu (2008) e, ao contrário dos métodos tradicionais, que se baseiam em médias condicionais, baseia-se em quantis condicionais. Então, é possível ajustar diferentes modelos de acordo com o interesse do pesquisador. Neste trabalho foram ajustados modelos considerando os quantis  $\tau = 0,1; 0,5$  e  $0,9$ . A obtenção dos estimadores dos coeficientes da regressão nos diferentes quantis de interesse, por meio da RQR, se deu a partir da solução do seguinte problema de otimização:

$$\hat{g} = \underset{g}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \rho_{\tau} \left( y_i - \mu - \sum_{j=1}^p x_{ij} g_j \right) + \lambda \sum_{j=1}^p |g_j| \right\}$$

em que  $\sum_{j=1}^p |g_j|$  é a soma dos valores absolutos dos coeficientes de regressão,  $\lambda$  é o parâmetro de penalização,  $n = 625$ ,  $p = 1040$  e  $\rho_{\tau}(\cdot)$  denotada função *check* por Koenker e Bassett (1978), que é definida por:

$$\rho_{\tau} \left( y_i - \mu - \sum_{j=1}^p x_{ij} g_j \right) = \begin{cases} \tau \left( y_i - \mu - \sum_{j=1}^p x_{ij} g_j \right), & \text{se } y_i - \mu - \sum_{j=1}^p x_{ij} g_j \geq 0 \\ (\tau - 1) \left( y_i - \mu - \sum_{j=1}^p x_{ij} g_j \right), & \text{caso contrário} \end{cases}$$

em que ( $\tau = 0,1; \tau = 0,5$  e  $\tau = 0,9$ ).

A função utilizada para estimação dos parâmetros das regressões nos três quantis foi *rq* do pacote *quantreg* (KOENKER, 2015).

Para comparação de metodologias, o parâmetro de penalização  $\lambda$  da RQR foi definido como sendo a metade do parâmetro de penalização obtido pelo método BLASSO. De acordo com Santos et al. (2018), a utilização do mesmo  $\lambda$  para ambas metodologias não é recomendado, uma vez que a força de penalização do método BLASSO é menor que da RQR, por isso a utilização de um  $\lambda$  menor para RQR.

Após a estimação dos coeficientes da regressão (efeitos dos marcadores), os valores genéticos genômicos foram obtidos por meio da seguinte equação:

$$\hat{y}_i(\tau) = \sum_j x_{ij} \hat{g}_j(\tau),$$

em que  $x_{ij}$  são os elementos da matriz de incidência do marcador  $j$  no indivíduo  $i$  e  $\hat{g}_j(\tau)$  é o efeito do  $j$ -ésimo marcador SNP considerando o quantil  $\tau$ .

### 3.4 Comparações das metodologias

Baseando-se na população  $F_2$  simulada, estipulou-se que 80% dos indivíduos pertenceriam à população de estimação e 20% à população de validação. Para obter as estimativas dos valores genéticos genômicos dos indivíduos (GEBVs), adotou-se a RQR considerando diferentes quantis (0,1; 0,5 e 0,9), RR-BLUP e o BLASSO.

Para cada método, o modelo preditivo foi validado com informações dos indivíduos pertencentes a população de validação. Foi estimada a capacidade preditiva dos métodos ( $r_{y,\hat{y}}$ ) que é a medida coeficiente de correlação entre valores fenotípicos observados ( $y$ ) e os valores genômicos estimados ( $\hat{y}$ ).

Além desta, foi calculado o ganho de seleção que é dado em função do diferencial de seleção ( $DS$ ) e a herdabilidade ( $GS = DS * h^2$ ). O  $DS$  corresponde a diferença entre a média da população selecionada em relação a média da população original, onde foi realizado o processo de seleção e  $h^2$  corresponde a herdabilidade. Além disso, foram determinadas as médias e as variâncias genotípicas.

Pode-se dizer que houve a fixação dos alelos favoráveis em determinada geração quando a população atinge a média genotípica 11 (equação 2) e variância igual zero. Dessa forma, o processo de seleção é considerado mais eficiente na medida em que forem necessárias menos gerações para fixação dos alelos favoráveis, ou seja, para conduzir os QTL's a homozigose. Outra maneira de avaliar a fixação dos alelos favoráveis é por meio do ganho de seleção máximo, ou seja, se o método atingiu o valor máximo para o ganho de 3,958 para herdabilidade 0,4 e 7,917 para herdabilidade 0,8 indica que ocorreu a fixação dos alelos favoráveis na geração correspondente.

### 3.5 Cenários avaliados

Na análise foram utilizadas populações com 625 indivíduos e cada cenário foi repetido 10 vezes, com arranjos aleatórios nas populações de estimação e validação. Os cenários avaliados consideraram duas porcentagens de seleção (10% e 20%) e duas herdabilidades ( $h^2 = 0,4$  e  $0,8$ ) conforme Tabela 1, buscando-se avaliar em qual geração ocorre a fixação dos alelos favoráveis em cada caso.

**Tabela 1.** Métodos e condições experimentais de cada cenário avaliado

Cenário	$h^2$	P.I.S. <sup>a</sup>	Métodos
1	0,4	10%	RQR ( $\tau = 0,1; \tau = 0,5$ e $\tau = 0,9$ ), BLASSO, RR-BLUP e S.F.
2	0,4	20%	RQR ( $\tau = 0,1; \tau = 0,5$ e $\tau = 0,9$ ), BLASSO, RR-BLUP e S.F.
3	0,8	10%	RQR ( $\tau = 0,1; \tau = 0,5$ e $\tau = 0,9$ ), BLASSO, RR-BLUP e S.F.
4	0,8	20%	RQR ( $\tau = 0,1; \tau = 0,5$ e $\tau = 0,9$ ), BLASSO, RR-BLUP e S.F.

<sup>a</sup> Porcentagem de indivíduos selecionados adotado a partir da  $F_3$ . Na  $F_2$  foram selecionados 50 indivíduos. Abreviações:  $h^2$ , herdabilidade.

## 4. RESULTADOS E DISCUSSÃO

Neste trabalho avaliou-se o uso de seleção genômica na seleção de indivíduos visando o avanço de gerações em espécies autógamas, considerando características oligogênicas. Foram utilizadas, além das abordagens tradicionais, BLASSO e RR-BLUP, a Regressão Quantílica Regularizada. Além disso, os resultados obtidos foram comparados com aqueles provenientes de um processo de seleção fenotípica. Os quatro cenários estudados (considerando duas diferentes herdabilidades e duas porcentagens de seleção) permitiram avaliar, ao longo de cinco gerações, diferentes abordagens estatísticas de seleção genômica, com o intuito de estabelecer qual ou quais dessas abordagens permitem a fixação dos alelos favoráveis de forma mais rápida. Nos casos avaliados, a fixação dos alelos favoráveis ocorre quando a média e variância genotípica forem iguais a 11 e a 0 respectivamente (Tabela 2 e Figuras 2 e 3).

De modo geral, observou-se o aumento das médias e a diminuição das variâncias genotípicas ao longo das gerações para todos os métodos avaliados, exceto quando o processo é baseado em seleção fenotípica (Figuras 2 e 3 Tabela 2). O método de

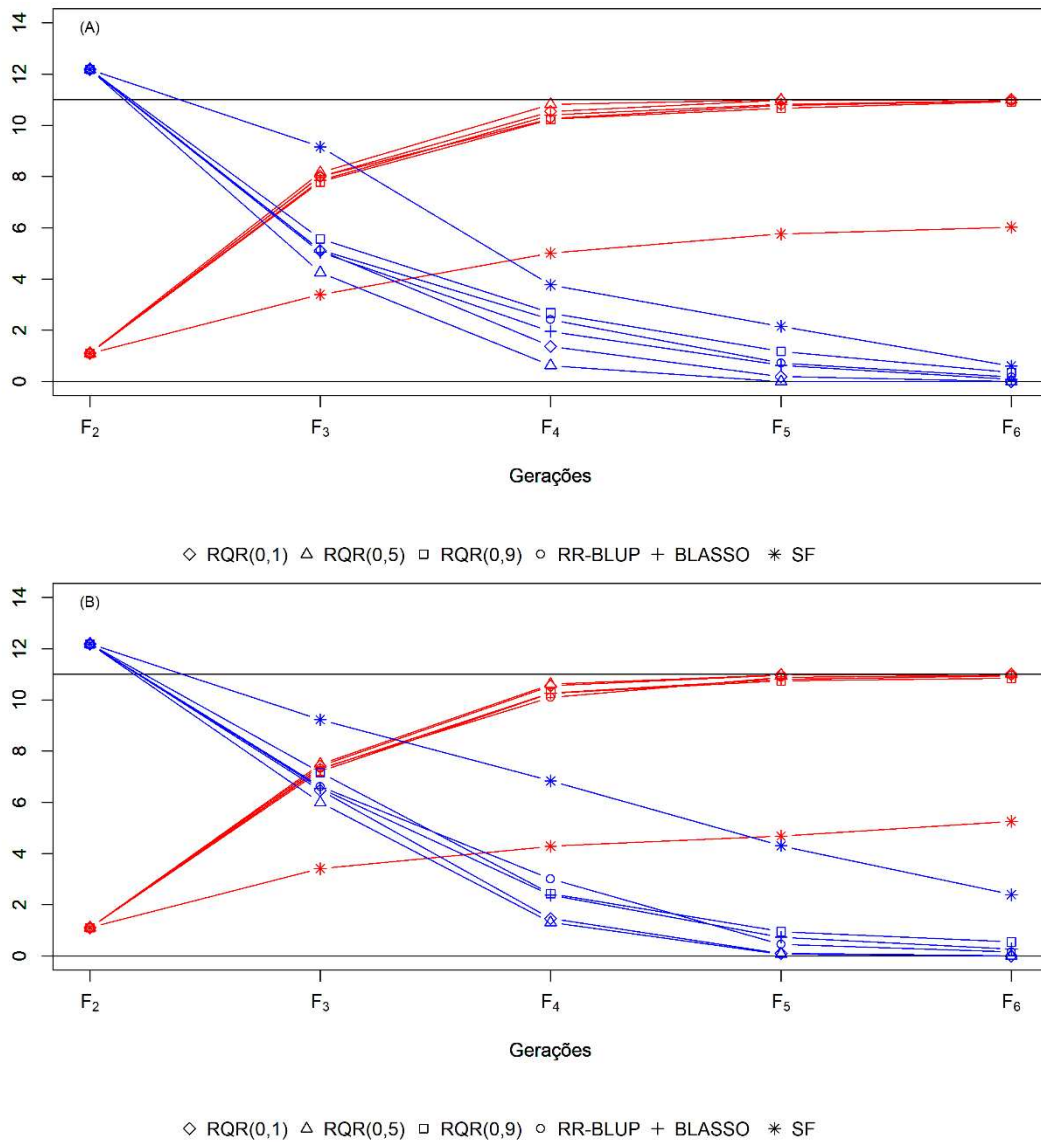
seleção fenotípica foi aquela que obteve as menores médias e maiores variâncias (Figuras 2 e 3 Tabela 2), indicando a necessidade de uma quantidade maior de gerações para uma possível fixação dos alelos favoráveis.

Com relação aos resultados de fixação ou não dos alelos favoráveis, estes foram diferentes nos cenários avaliados. Quando se considerou característica com herdabilidade 0,4, independente da porcentagem de seleção, somente a RQR nos quantis ( $\tau = 0,1$ ) e ( $\tau = 0,5$ ) foi capaz de fixar os alelos favoráveis até a sexta geração (Tabela 2 e Figuras 2 e 3). Especificamente, quando se utilizou a RQR ( $\tau = 0,1$ ), os alelos favoráveis foram fixados na sexta geração para as duas porcentagens de seleção, enquanto a RQR( $\tau = 0,5$ ), com a diminuição da porcentagem dos indivíduos selecionados de 20% para 10%, passou a fixar os alelos ainda na quinta geração (Figuras 2 e 3 e Tabela 2). Por outro lado, os métodos tradicionais de seleção genômica (BLASSO e RR-BLUP) não fixaram os alelos até a sexta geração (Figuras 2 e 3 e Tabela 2). Estes resultados são similares aqueles observados por Romero (2017), em que o autor comparou dois métodos tradicionais de seleção genômica - RR-BLUP e BLASSO – utilizando uma população simulada de *Coffea arabica* para característica oligogênica, com diferentes herdabilidades e duas porcentagens de seleção, também com o objetivo de verificar a fixação dos alelos favoráveis. O autor verificou que nos cenários com herdabilidade 0,4 e diferentes porcentagens de seleção os métodos tradicionais não conseguiram fixar os alelos favoráveis até a sexta geração.

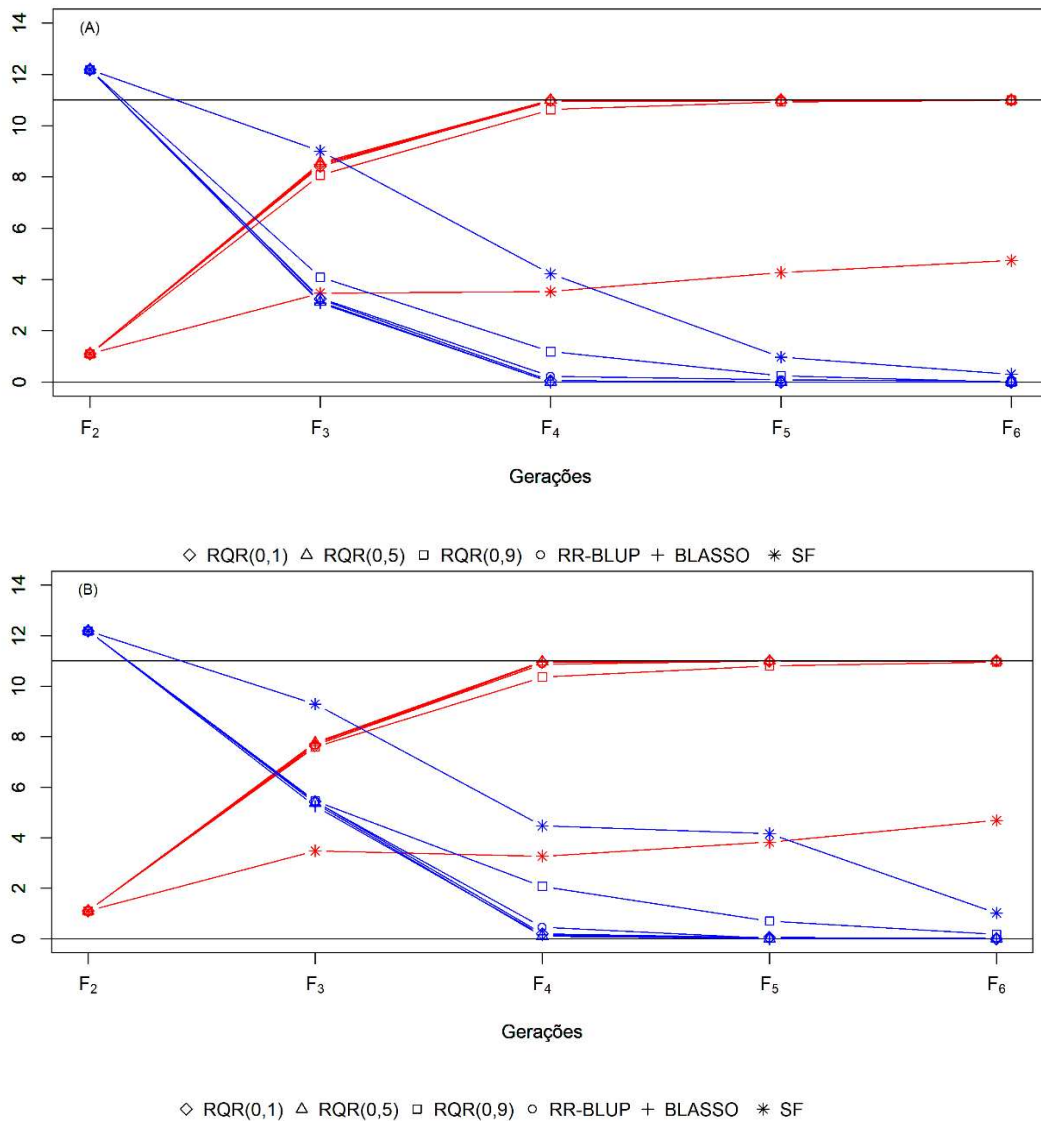
Nos cenários em que se considerou a característica com herdabilidade 0,8, os alelos favoráveis foram fixados pelos métodos BLASSO e RQR ( $\tau = 0,1$  e  $\tau = 0,5$ ) com as duas porcentagens de seleção. Com a diminuição da porcentagem de seleção de 20% para 10% os métodos BLASSO, RQR ( $\tau = 0,1$  e  $\tau = 0,5$ ) diminuíram uma geração na fixação dos alelos. Sendo que os métodos BLASSO, RQR ( $\tau = 0,5$ ) passaram a fixar os alelos na quarta geração e a RQR ( $\tau = 0,1$ ) passou a fixar na quinta geração. Ademais, a RQR ( $\tau = 0,9$ ) só fixou os alelos na sexta geração com a porcentagem de seleção de 10%, e a seleção fenotípica não fixou os alelos em nenhum cenário avaliado (Figuras 2 e 3 e Tabela 2). Assim, a diminuição da porcentagem de seleção teve efeito tanto na eficácia, ou seja, a fixação dos alelos favoráveis até a sexta geração, quanto na eficiência, que está relacionada com o uso de menos gerações para fixação dos alelos.

Dentre os métodos tradicionais de seleção genômica, somente o método RR-BLUP não fixou os alelos, em nenhum cenário avaliado, até a sexta geração (Tabela 2 e

3). Este resultado pode ser explicado pelo fato de que o RR-BLUP é um método mais apropriado para características poligênicas (RESENDE et al., 2014) e, neste estudo, foram simuladas características oligogênicas. Oligogênes são melhores explicados por meio de metodologias que consideram a existência de um número menor de genes controladores da mesma.



**Figura 2.** Médias (linhas vermelhas) e variâncias genóticas médias (linhas azuis) dos modelos avaliados ao longo de cinco gerações considerando herdabilidade 0,4 e duas porcentagens de seleção (PS). (A) PS= 10%; (B) PS= 20%.



**Figura 3.** Médias (linhas vermelhas) e variâncias genóticas médias (linhas azuis) dos modelos avaliados ao longo de cinco gerações considerando herdabilidade 0,8 e duas porcentagens de seleção (PS). (A) PS= 10%; (B) PS= 20%.

**Tabela 2.** Médias e variâncias genotípicas médias dos modelos avaliados ao longo de cinco gerações considerando duas herdabilidade (0,4 e 0,8) e duas porcentagens de seleção (10 e 20%)

Modelos Preditivos	$h^2$	OS	$F_2$		$F_3$		$F_4$		$F_5$		$F_6$	
			MG	VG	MG	VG	MG	VG	MG	VG	MG	VG
BLASSO	0,4	10%	1,104	12,189	7,859	5,043	10,405	1,955	10,818	0,634	10,975	0,085
	0,4	20%	1,104	12,189	7,302	6,557	10,267	2,372	10,813	0,731	10,931	0,271
	0,8	10%	1,104	12,189	8,490	3,087	11,000	0,000	11,000	0,000	11,000	0,000
	0,8	20%	1,104	12,189	7,723	5,248	10,959	0,151	11,000	0,000	11,000	0,000
RR-BLUP	0,4	10%	1,104	12,189	7,995	5,130	10,264	2,417	10,778	0,723	10,949	0,176
	0,4	20%	1,104	12,189	7,321	6,614	10,098	3,009	10,889	0,445	10,959	0,157
	0,8	10%	1,104	12,189	8,481	3,273	10,943	0,220	10,974	0,084	10,988	0,036
	0,8	20%	1,104	12,189	7,659	5,427	10,868	0,458	10,995	0,019	10,994	0,023
RQR (0,1)	0,4	10%	1,104	12,189	8,008	5,112	10,540	1,369	10,978	0,191	11,000	0,000
	0,4	20%	1,104	12,189	7,411	6,472	10,549	1,459	10,974	0,098	11,000	0,000
	0,8	10%	1,104	12,189	8,419	3,239	10,988	0,049	11,000	0,000	11,000	0,000
	0,8	20%	1,104	12,189	7,655	5,428	10,944	0,201	10,986	0,053	11,000	0,000
RQR (0,5)	0,4	10%	1,104	12,189	8,163	4,252	10,821	0,623	11,000	0,000	11,000	0,000
	0,4	20%	1,104	12,189	7,498	5,993	10,611	1,309	10,981	0,072	11,000	0,000
	0,8	10%	1,104	12,189	8,560	3,120	11,000	0,000	11,000	0,000	11,000	0,000
	0,8	20%	1,104	12,189	7,763	5,366	10,973	0,100	11,000	0,000	11,000	0,000
RQR (0,9)	0,4	10%	1,104	12,189	7,794	5,569	10,240	2,674	10,674	1,175	10,925	0,345
	0,4	20%	1,104	12,189	7,214	7,165	10,261	2,434	10,748	0,950	10,854	0,550
	0,8	10%	1,104	12,189	8,070	4,087	10,630	1,187	10,930	0,239	11,000	0,000
	0,8	20%	1,104	12,189	7,599	5,460	10,359	2,077	10,809	0,702	10,958	0,172
SELEÇÃO FENOTÍPICA	0,4	10%	1,104	12,189	3,396	9,160	5,011	3,771	5,759	2,155	6,030	0,605
	0,4	20%	1,104	12,189	3,413	9,230	4,286	6,835	4,679	4,305	5,259	2,394
	0,8	10%	1,104	12,189	3,465	9,011	3,530	4,220	4,265	0,971	4,744	0,304
	0,8	20%	1,104	12,189	3,478	9,291	3,269	4,475	3,828	4,157	4,686	1,021

Abreviações:  $h^2$ , herdabilidade, MG, Média genotípica;  $h^2$ , PS, porcentagem de seleção. VG, Variância genotípicas.

Outra forma de avaliar a fixação dos alelos favoráveis é através do ganho de seleção máximo. Para os cenários simulados, quando, em uma determinada geração, o método avaliado atingir o valor de 3,958 em cenários de herdabilidade 0,4, o método conseguiu fixar os alelos favoráveis na geração correspondente. Em cenários de herdabilidade 0,8 a fixação ocorre quando o valor máximo é 7,917.

Na Tabela 3 são apresentados os ganhos médios de seleção por geração em relação à geração  $F_2$ . Observa-se que os ganhos de seleção dos métodos avaliados aumentou em todas as gerações observadas (Tabela 3). O método que apresentou os melhores ganhos na terceira geração foi a RQR ( $\tau = 0,5$ ) e nas demais gerações os ganhos foram superiores ou iguais aos demais métodos (Tabela 3). Especificamente, nos cenários com característica de herdabilidade 0,4 e porcentagem de seleção de 10%, na geração  $F_4$  o ganho de seleção da RQR ( $\tau = 0,5$ ) foi 4,5% maior comparado ao BLASSO e 6,1% maior em comparação ao RR-BLUP (Tabela 3). Com o aumento das porcentagens dos indivíduos selecionados de 10% para 20% o ganho de seleção diminuiu na geração  $F_4$ , mas ainda assim a RQR ( $\tau = 0,5$ ) foi superior aos demais métodos, sendo 3,8% maior em comparação ao BLASSO e 5,7% em comparação ao RR-BLUP e 3,2% em comparação à seleção fenotípica (Tabela 3). Nos cenários de herdabilidade 0,8 e porcentagem de seleção de 10% o ganho de seleção da RQR ( $\tau = 0,5$ ) e do BLASSO foram iguais após a geração  $F_3$ , e com porcentagem de 20% o ganho de seleção foi igual para os dois métodos depois da geração  $F_4$ . Já os métodos RR-BLUP e seleção fenotípica tiveram um ganho de seleção inferior à RQR ( $\tau = 0,5$ ) em relação às duas porcentagens de seleção e o método de seleção fenotípica foi o método que obteve os menores ganhos de seleção (Tabela 3).

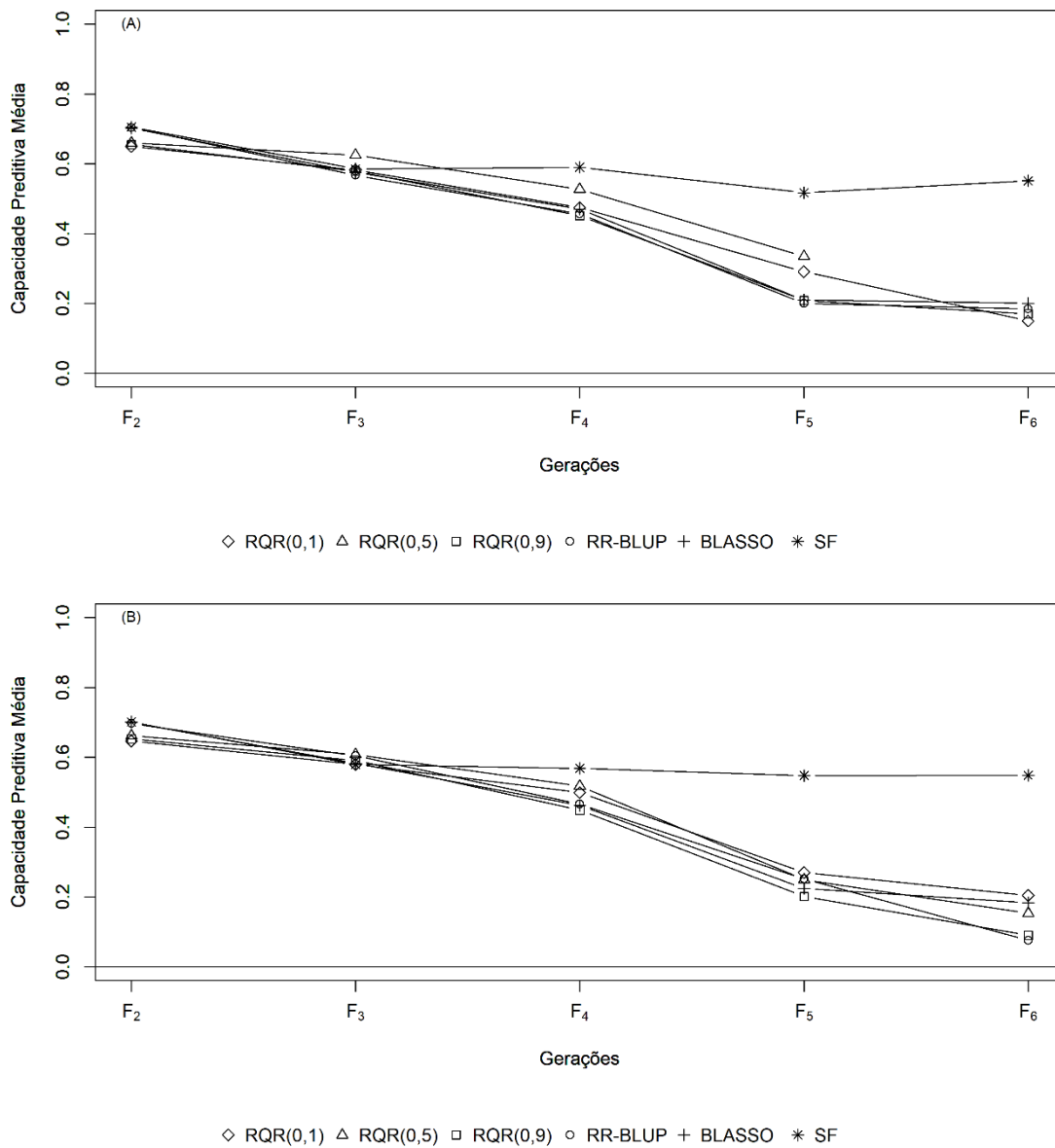
**Tabela 3.** Ganho médio de seleção por geração em relação à geração  $F_2$ 

Modelos Preditivos	$h^2$	PS	Ganho de Seleção			
			$F_3$	$F_4$	$F_5$	$F_6$
BLASSO	0,4	10%	2,702	3,720	3,885	3,948
	0,4	20%	2,479	3,665	3,884	3,931
	0,8	10%	5,909	<b>7,917</b>	7,917	7,917
	0,8	20%	5,295	7,884	<b>7,917</b>	7,917
RRBLUP	0,4	10%	2,756	3,664	3,869	3,938
	0,4	20%	2,487	3,597	3,914	3,942
	0,8	10%	5,902	7,871	7,896	7,907
	0,8	20%	5,244	7,811	7,913	7,912
RQR (0,1)	0,4	10%	2,761	3,774	3,949	<b>3,958</b>
	0,4	20%	2,523	3,778	3,948	<b>3,958</b>
	0,8	10%	5,852	7,907	<b>7,917</b>	7,917
	0,8	20%	5,241	7,872	7,905	<b>7,917</b>
RQR (0,5)	0,4	10%	2,823	3,887	<b>3,958</b>	3,958
	0,4	20%	2,557	3,803	3,951	<b>3,958</b>
	0,8	10%	5,965	<b>7,917</b>	7,917	7,917
	0,8	20%	5,327	7,895	<b>7,917</b>	7,917
RQR (0,9)	0,4	10%	2,676	3,654	3,828	3,928
	0,4	20%	2,444	3,663	3,857	3,900
	0,8	10%	5,573	7,621	7,861	<b>7,917</b>
	0,8	20%	5,196	7,404	7,764	7,883
SELEÇÃO FENOTÍPICA	0,4	10%	0,917	1,563	1,862	1,970
	0,4	20%	0,923	1,273	1,430	1,662
	0,8	10%	1,889	1,941	2,529	2,912
	0,8	20%	1,899	1,732	2,179	2,866

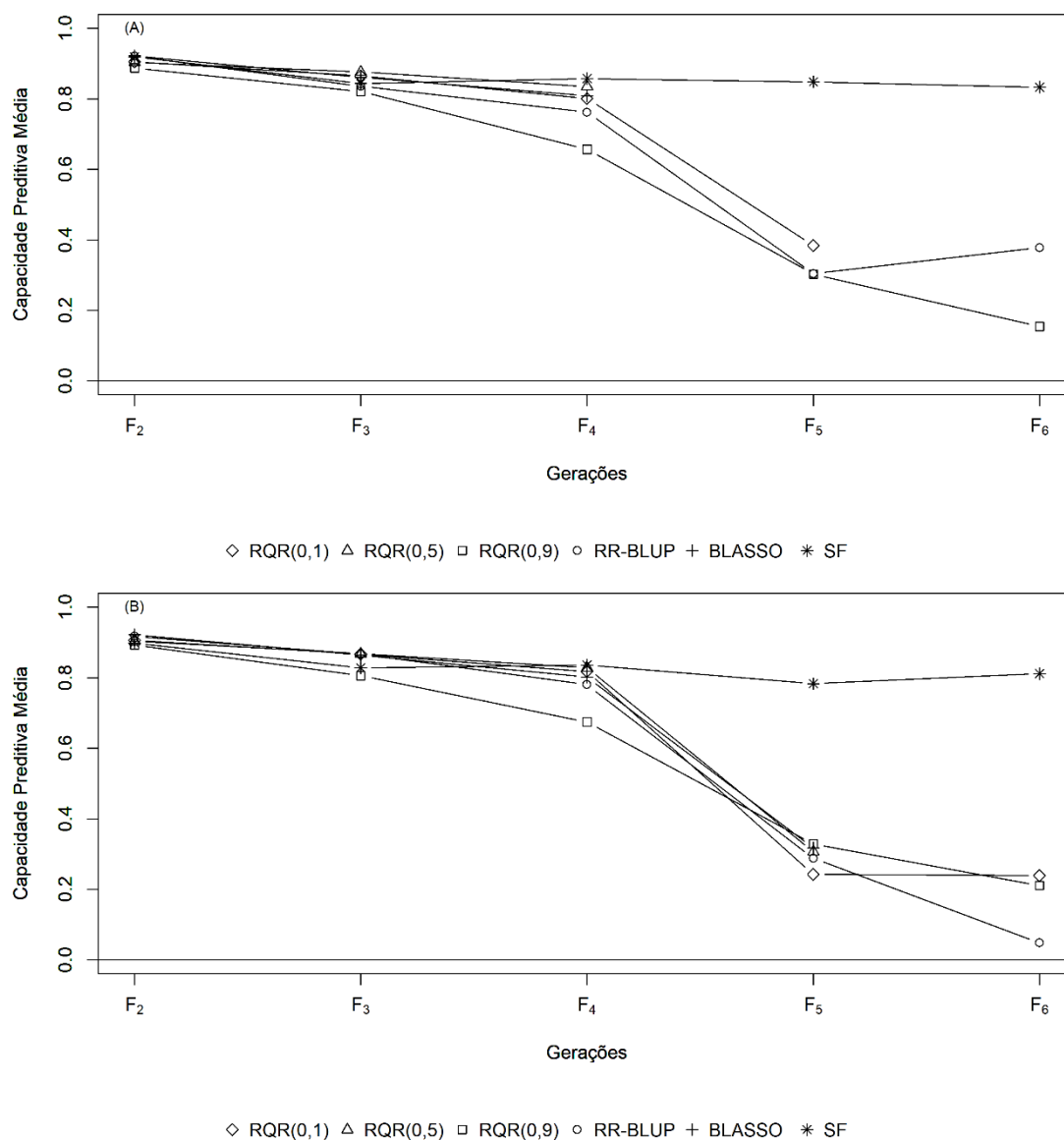
Valores em negrito referem-se ao valor máximo do ganho de seleção considerando duas herdabilidades (0,4 e 0,8). Abreviações:  $h^2$ , herdabilidade, PS, porcentagem de seleção.

Ainda visando avaliar e comparar as diferentes abordagens de GWS no processo de seleção, foram calculadas as capacidades preditivas para cada método, ao longo das gerações. Como resultados, tem-se que as estimativas da capacidade preditiva (CP) variaram entre 0,64 e 0,92 na geração  $F_2$  (Figuras 4 e 5). Nas gerações subsequentes, as estimativas de CP foram decaindo até valores próximos a zero na geração  $F_6$ , exceto nos cenários onde os alelos favoráveis foram fixados e no método de seleção fenotípica (Figuras 4 e 5). Essa redução nas estimativas de CP em cada geração de autofecundação está de acordo com o observado por Sant'Anna et al. (2019) que avaliaram o efeito do desequilíbrio de ligação e a capacidade preditiva em multigerções e diferentes tipos de acasalamentos. Segundo os autores, a redução na capacidade preditiva ao longo das gerações de autofecundação pode estar relacionada a diminuição do desequilíbrio de ligação em cada geração. A seleção fenotípica foi o

método que obteve as maiores estimativas da CP (Figuras 4 e 5). Entretanto, quando observamos os valores de média e variância genóticas (Tabela 2) observa-se que não houve a fixação dos alelos favoráveis até a sexta geração por este método. Desta forma, apenas a medida da CP não é suficiente para concluir sobre qual o melhor método estatístico para o processo de seleção que busca a fixação de alelos favoráveis.



**Figura 4.** Capacidade preditiva média dos modelos avaliados ao longo de cinco gerações considerando herdabilidade 0,4 e duas porcentagens de seleção. (A) PS=10%; (B) PS= 20%.



**Figura 5.** Capacidade preditiva média dos modelos avaliados ao longo de cinco gerações considerando herdabilidade 0,8 e duas porcentagens de seleção. (A) PS= 10%; (B) PS= 20%.

De maneira geral, a regressão quantílica regularizada, apesar de ser um método ainda pouco explorado no melhoramento genético, tem-se mostrado bastante promissor em estudos de associação e seleção genômica tanto no melhoramento animal como no de plantas (BARROSO et al., 2017; NASCIMENTO et al., 2017; NASCIMENTO et al., 2018; SANTOS, et al 2018). No presente estudo, a RQR também apresentou resultados melhores ou iguais àqueles tradicionais de seleção genômica nos cenários avaliados. Este foi o único método que permitiu alcançar a fixação dos alelos favoráveis até sexta geração em todos os cenários avaliados e juntamente com o método BLASSO, a RQR ( $\tau = 0,5$ ) fixou os alelos favoráveis na quarta geração ( $F_4$ ) no cenário com

herdabilidade 0,8 e proporção de seleção 10%. Especificamente, a RQR ( $\tau = 0,5$ ) foi o modelo que obteve os melhores resultados, sendo o único método que permitiu fixar os alelos favoráveis em todos os cenários avaliados e também obteve o ganho de seleção máximo em todos os cenários. Esse resultado indica que a RQR é uma alternativa viável e eficaz para análise, e que o seu uso possibilitaria uma redução de tempo, conseqüentemente custos além do aumento no ganho de seleção. A eficiência observada com o uso da RQR pode ser explicada, porque, ao contrário dos métodos tradicionais, que se baseiam em médias condicionais, a RQR permite ajustar modelos em diferentes níveis (quantis) da distribuição do fenótipo permitindo assim um estudo mais completo do fenômeno de interesse (KOENKER e BASSETT, 1978).

## 5. CONCLUSÃO

A utilização de modelos de Regressão Quantílica Regularizada em estudos de seleção genômica mostrou-se eficaz, uma vez que em todos cenários analisados a mesma fixou os alelos em gerações anteriores ou em gerações iguais aos métodos tradicionais, até a sexta geração de autofecundação. A RQR ( $\tau = 0,5$ ) foi o único modelo que conseguiu fixar os alelos em todos os cenários avaliados. Especificamente, nos cenários onde se considerou herdabilidade 0,4, somente a seleção baseada na RQR fixou os alelos favoráveis até a sexta geração. Quando se considerou herdabilidade 0,8 os métodos BLASSO e RQR ( $\tau = 0,5$ ) obtiveram os mesmos resultados, com porcentagem de seleção de 10% a seleção baseada nos dois métodos conseguiram fixar os alelos na quarta geração ( $F_4$ ). A seleção realizada com base nos métodos RR-BLUP e de seleção fenotípica não levou à fixação dos alelos favoráveis até a sexta geração em nenhum cenário avaliado. Em relação ao ganho de seleção, a RQR ( $\tau = 0,5$ ) obteve ganhos maiores ou iguais aos métodos tradicionais de seleção genômica em todas as gerações e em todos os cenários avaliados. Estes ganhos foram até 6,1% maiores do que aqueles obtidos pelas demais metodologias avaliadas. Dessa forma, verificou-se que com aplicação da RQR no melhoramento de plantas considerando populações simuladas, seria possível uma redução de tempo e conseqüentemente de custos, devido a diminuição das gerações de autofecundações, para a fixação dos alelos favoráveis em todos os cenários avaliados ou para a obtenção de genótipos melhorados

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- ASORO, F. G. et al. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. **The Plant Genome**, Madison, v. 4, n. 2, p. 132-144, 2011.
- BESPALHOK F., J.C.; GUERRA, E.P.; OLIVEIRA, R. Introdução ao Melhoramento de Plantas. In: BESPALHOK F., J.C.; GUERRA, E.P.; OLIVEIRA, R. **Melhoramento de Plantas**. Disponível em: <<http://www.bespa.agrarias.ufpr.br/paginas/livro/introducao.pdf>>. Acessado em: 7 mai. 2018.
- BEYERLEIN, A. et al. Genetic Markers of Obesity Risk: Stronger Associations with Body Composition in Overweight Compared to Normal Weight Children. **PLoS one**, Bethesda, v. 6, n. 4, 2011. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019057>>. Acessado em: 20 set. 2018.
- BORÉM, A.; MIRANDA, G. **Melhoramento de Plantas**. Viçosa: Editora UFV, 2013.
- BORÉM, A.; MIRANDA, G. V.; FRITSCHÉ-NETO, R. **Melhoramento de plantas**. 7. ed. rev. e ampl. Viçosa: Editora UFV, 2017. 523 p.
- CAMPOS e RODRIGUEZ, **Bayesian Generalized Linear Regression**. URL: <http://cran.r-project.org/web/packages/BGLR/index.html>, 2015.
- CROSSA, J. et al. Genomic Selection and Prediction in Plant Breeding. **Journal of Crop Improvement**, London, v. 25, n. 3, p. 239-261. 2011.
- CROSSA, J. et al. Genomic selection in plant breeding: methods, models, and perspectives. **Trends in plant science**, Cambridge, v. 22, n. 11, p. 961-975, 2017.
- DE CASTRO SANT'ANNA, I. et al. Multigenerational prediction of genetic values using genome-enabled prediction. **PloS one**, Bethesda, v. 14, n. 1, 2019. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0210531>>. Acessado em: 22 jan. 2019.
- DE LOS CAMPOS, G. et al. Predicting quantitative traits with regression models for dense molecular makers. **Genetics**, Austin, v. 182, p. 375-385, 2009.
- ENDELMAN, J. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. **The Plant Genome**, Madison, v. 4, n. 3, p. 250-255, 2011.
- FERREIRA, D. F. Uso de simulação no melhoramento. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. D.; VALADARES-INGLIS, M. C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação-MT, 2001. p. 1119-1141.
- GUJARATI, D. N. **Econometria básica**. 5. ed. São Paulo: Pearson Makron Books, 2011. 920 p.

GURGEL, F. de L. **Simulação computacional no melhoramento genético de plantas**. 2004. 174 f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras. 2004.

HAO, L.; NAIMAN, D. Q. **Quantile regression**. Sage publications, 2007. 126p.

HEFFNER, ELLIOT L.; JANNINK, JEAN-LUC; SORRELLS, MARK E. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. **The Plant Genome**, Madison, v. 4, n. 1, p. 65-75, 2011.

HEIDELBERGER, P.; WELCH, P.D. Simulation run length control in the presence of an initial transient. **Operations Research Landing**, Catonsville, v.31, n. 6, p.1109-1144, 1983.

KOENKER, R.; BASSET, G. Regression Quantiles. **Econometrica**, v.46, p. 33–50, 1978.

KOENKER, R. **Quantile Regression**. Cambridge: Cambridge University Press, 2005.

KOENKER R. quantreg: Quantile regression. R package version 4.91. <http://CRAN.Rproject.org/package=quantreg>, 2015.

LI, Y.; ZHU, J. L1-Norm Quantile Regression. **Journal of Computational and Graphical Statistics**, London, v. 17, n. 1, p. 163–185, 2008.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, v. 157, n. 1, p. 1819–1829, 2001.

MISSIO, F.; J.; JAYME JR, F. G.; OLIVEIRA, A. M. H. C. **Desenvolvimento financeiro e crescimento econômico: Teoria e evidência empírica para os estados brasileiros (1995-2004)**. Belo Horizonte: UFMG/ Cedeplar, 2009. 33p.

MONTGOMERY, D.C.; PECK, E.A.; VINING, G.G. **Introduction to linear regression analysis**. 5. Ed. New York: John Wiley & Sons, 2012. 645p.

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

NASCIMENTO, A. C. C. et al. Eficiência técnica da atividade leiteira em Minas Gerais: Uma aplicação de regressão quantílica. **Revista Brasileira de Zootecnia**, Viçosa, v. 41, n. 3, p. 783-789, 2012.

NASCIMENTO, A. C. C. et al. Identificando os determinantes da eficiência técnica na produção de café de montanha em Minas Gerais. **Revista Brasileira de Biometria**, Lavras, v. 35, n. 3, p. 461-473, 2017.

NASCIMENTO, M. et al. Quantile regression for genome-wide association study of flowering time-related traits in common bean. **PloS one**, Bethesda, v. 13, n. 1, 2018. Disponível em:

<<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0190303>>. Acessado em: 18 nov. 2018.

NASCIMENTO, M. et al. Regularized quantile regression applied to genome-enabled prediction of quantitative traits. **Genetics and Molecular Research**, Ribeirão Preto, v. 16, n. 1, p. 1-12, 2017.

PARK, T.; CASELLA, G. The Bayesian Lasso. **Journal of the American Statistical Association**, v. 103, n. 482, p. 681-686, 2008.

PESTANA, K.N. **Mapa genético integrado e seleção genômica ampla visando resistência de Coffea arabica L. à ferrugem**. 2013. 109 f. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa. 2013.

RAFTERY, A.L.; LEWIS, S.M. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. **Statistical Science**, Bethesda, v. 7, n. 4, p. 493-497, nov. 1992.

RESENDE, M.D.V. et al. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, Colombo, v. 56, n. 63, p. jan./jun. 2008.

RESENDE, M.D.V.; SILVA, F.F.; AZEVEDO, C.F. **Estatística matemática, biométrica e computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição Sobrevivência**. Viçosa: Suprema, 2014. 881p.

ROMERO, J.V. **Seleção genômica usando genotipagem de baixa saturação no melhoramento genético do cafeeiro**. 2017. 104p. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa. 2017.

SANTOS, P. M. D. et al. Use of regularized quantile regression to predict the genetic merit of pigs for asymmetric carcass traits. **Pesquisa Agropecuária Brasileira**, Brasília, v. 53, n. 9, p. 1011-1017, Sept. 2018.

SILVA, E. N.; PORTO JÚNIOR, S. S. Sistema financeiro e crescimento econômico: Uma aplicação de regressão quantílica. **Economia Aplicada**, Ribeirão Preto, v. 10, n. 3, p. 425-442, jul./set. 2006.

SILVA, G.N. **Redes neurais artificiais: novo paradigma para a predição de valores genéticos**. 2014. 105p. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, Viçosa. 2014.

SPINDEL, J.; BEGUM, H. et al. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. **PLoS genetics**, San Francisco, v. 11, n. 2, 2015. Disponível em: <<https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004982>>. Acessado em: 21 out. 2018.

SUN, X.; PENG, T.; MUMM, R. H. The role and basics of computer simulation in support of critical decisions in plant breeding. **Molecular Breeding**, v. 28, n. 4, p. 421-436, 2011.

TIBSHIRANI, R. Regression shrinkage and selection via the LASSO, J. **R. Stat. Soc. B**, London, v. 58, n. 1, p. 267-288, 1996.

VALLEJO, F.A. et al. **Genética Vegetal**. Cali: UNAL, 2010.

WANG, J. Modelling and simulation of plant breeding strategies. In: ABDURAKHMONOV, I. Y. **Plant Breeding**. London: InTech, 2012. p. 19-40.

ZHONG, S. et al. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, n. 1, 2009. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19299342>>. Acessado em: 18 out. 2018.

## Apêndice: Rotinas computacionais utilizadas para a análise.

```
setwd("C:\\Dissertação") # Definição da trilha de dados
#
library("rrBLUP")
library("BLR")
library(quantreg)
#
# Leitura do arquivo de marcadores moleculares e fenótipos
# colunas do arquivo: id Vgen Vfen80 Vfen40 M1 M2 M3 M4 M5...M1040
# codificado 1 0 -1
FG=read.table("Dados_F2_base.txt", h=T)
#
caract<-3 # Herdabilidades
genomF2<-data.frame(FG[-1:-(caract+1)])
GL<- 8 # Grupos de ligação
MGL<- ncol(genomF2)/GL # Marcadores por grupo de ligação.
QTL<-c(196, 456, 716, 976) # número do marcador correspondente ao QTL
nis<- 20 # número de indiv a selecionar
n.des<- 10 # descendentes por individuo
d<-read.table("densidade.txt",h=T)
for(R in 1:10) #Repetições
{
for(s in 1:1) #Indica a saturação de marcadores segundo a lista do objeto "d"
{D<-(switch(s,0,1,2,3,4,5,10)) #<----##### Indica a distância QTL-Marcador.
#
if (length(GLF)>0) {MARC<-na.omit(d[,s]); for (j in GLF) {LI<-((MGL*j)-MGL+1);
LS<-((MGL*j); A<-numeric();B<- c(0)
for (i in 1:(nrow(d))) {A[ (MARC[i]<LI || MARC[i]>LS) ] <- c(MARC[i])
if (A!= B[length(B)]) {B<- c(B,A)}}; MARC<- B}
MARC<- MARC[-1]; rm(LI,LS,A,B,i,j)} else {MARC<- na.omit(d[,s])}
GENOTIPO<-as.matrix(genomF2[,MARC])
#
#Divisão de matrizes para Validação
IVal = sample(1:(nrow(GENOTIPO)), (nrow(GENOTIPO)*0.2)) #corresponde ao 20%
do total
GENOTIPOE <- (GENOTIPO[-IVal,]) #Indivíduos Analisados
GENOTIPOV <- (GENOTIPO[IVal,]) #Indivíduos para Validação
#
#Análise por característica
for(h in 1:caract) # Herdabilidades = 3
```

```

{ H<-(switch(h,100,80,40))
G<-2 # Valor para identificar a F2
VGeno<-(FENOTIPOS[,1]); VGeno<- as.data.frame(VGeno) # Acumula os valores
genotípicos das gerações
VFeno<-(FENOTIPOS[,h]); VFeno<- as.data.frame(VFeno) # Acumula os valores
fenotípicos das gerações
Here<- var(FENOTIPOS[,1])/var(FENOTIPOS[,h])
H2<-as.numeric(); H2<-c(H2,Here) # Acumula os valores de herdabilidade
variable<-(FENOTIPOS[-(IVal),h]) # fenótipo da pob de estimação
variableval<-(FENOTIPOS[IVal,h]) # fenótipo da pob de validação

#
# Análise das variáveis

#RR-BLUP
fitRRB=mixed.solve(variable,Z=GENOTIPOE,K=diag(ncol(GENOTIPOE)))

#BLASSO
BL=BGLR(y=variable,ETA=list(list(X=GENOTIPOE,model='BL')), nIter=10000,
burnIn=4000, thin=3)
lambda <- BL$ETA[[1]]$lambda

#Regressão Quantílica 0,1 - Adaptar para as quantílicas 0.5 e 0.9
tryCatch({fitRQ1<-rq(variable ~ -1+ GENOTIPOE, tau=0.1,method='lasso',
lambda=(lambda/2))}, error=function(e){cat("ERROR :",conditionMessage(e), "\n")})

#
# Estimação do valor genético genômico (VGG)

Ef.Marca<- fitRRB$u # Efeitos de Marcadores - RR-BLUP
Ef.Marca<- BL$ETA[[1]]$b # Efeitos de Marcadores – BLASSO
Ef.Marca<- fitRQ1$coefficients # Efeitos de Marcadores - RRQ

VGG<-GENOTIPOE%*%Ef.Marca # Valores Genômicos
#
# Calculo da Capacidade preditiva

c.pred<-cor(VGG, variable)

#
# Validação

VGG_V<- GENOTIPOV%*%Ef.Marca
VGS_V<- FENOTIPOS[IVal, 1] # valor Genômico Simulado de Validação
c.pred_V<- cor(VGG_V,variableval)
A<- (rep("na", (nrow(VGG)-nrow(VGG_V))))); B<-as.numeric(c(VGG_V,A))
VGGs<- cbind(VGG, B) # Acumula os VGG estimados por geração
rm(A,B)
Capacidade.p<- c(c.pred, c.pred_V) # Acumula as capacidades preditivas por geração
#

```

```

# Ranking de indivíduos

#classificação esperada segundo Valores genômicos simulados
cla.sim_<-as.matrix(FENOTIPOS[,1]); rownames(cla.sim_)<- rownames(FENOTIPOS)
cla.sim.s=sort(as.matrix(cla.sim_)[,],decreasing=TRUE)
cla.sim<- as.numeric(names(cla.sim.s)); rm(cla.sim_, cla.sim.s)

#classificação segundo Valores genômicos estimados
VGe= GENOTIPO%*%Ef.Marca + mean(FENOTIPOS[,h])
VGe1=sort((VGe)[,],decreasing=TRUE)
cla<- as.numeric(names(VGe1))
Ranking<- cbind(cla.sim, cla) # Acumula as listas de classificação de ind por valor
genômico

#
# Medias e Variâncias

mediaG<- mean(FENOTIPOS[,1])
variancaG <- var(FENOTIPOS[,1])
mediaF<- mean(FENOTIPOS[,h])
variancaF <- var(FENOTIPOS[,h])

#
#F3 a F6 - SIMULAÇÃO, CLASIFICAÇÃO E SELEÇÃO

GENO=matrix(nrow(GENOTIPOE), 1040) # ind F2 x Marcad #
GENO[genomF2== 0]<- c(2)
GENO[genomF2== 1]<- c(4)
GENO[genomF2==-1]<- c(1)
for (G in 3:6) # Gerações F3 a F6
{

#SELEÇÃO #

ind.sel = cla[1:nis] # Identificação de genótipos de parentais
Geracao<- GENO [ind.sel,]

#
#SIMULAÇÃO GERAÇÕES AVANÇADAS (SGA)

# Geração do genótipo dos descendentes
genoS<-matrix((nis*n.des), ncol(GENO))
MGL<- ncol(GENO)/GL # Marcadores por grupo de lig.
cont<- 0
for (ind in 1:nis)
{
GI<- (Geracao[ind,]) #genoma de um ind.
for (des in 1:n.des)
{res.ind<-NULL

```

```

for (gl in 0:(GL-1))
{
fd<- (MGL*gl)
Fn<- GI[((1+fd):(MGL+fd))]
# Diferencia homologos
cromo1=matrix(); cromos2=matrix()
cromo1[Fn==1]<- c(0); cromos2[Fn==1]<- c(0)
cromo1[Fn==2]<- c(0); cromos2[Fn==2]<- c(1)
cromo1[Fn==3]<- c(1); cromos2[Fn==3]<- c(0)
cromo1[Fn==4]<- c(1); cromos2[Fn==4]<- c(1)
# recombina
s1<-sample(c(1,2),1); s2<-sample(c(16:115), 1)
if (s1==1) {co1<- c(cromo1[1:s2], cromos2[(s2+1): MGL])}
if (s1==2) {co1<- c(cromos2[1:s2], cromo1[(s2+1): MGL])}
if (s1==1 & s2> 100) {co1<- c(cromos2[1:(s2-100)], co1[(s2-99): MGL])}
if (s1==2 & s2> 100) {co1<- c(cromo1[1:(s2-100)], co1[(s2-99): MGL])}
if (s1==1 & s2< 31) {co1<- c(co1[1: (99+s2)], cromo1[(100+s2): MGL])}
if (s1==2 & s2< 31) {co1<- c(co1[1: (99+s2)], cromos2[(100+s2): MGL])}
s1<-sample(c(1,2),1); s2<-sample(c(16:115), 1)
if (s1==1) {co2<- c(cromo1[1:s2], cromos2[(s2+1): MGL])}
if (s1==2) {co2<- c(cromos2[1:s2], cromo1[(s2+1): MGL])}
if (s1==1 & s2> 100) {co2<- c(cromos2[1:(s2-100)], co2[(s2-99): MGL])}
if (s1==2 & s2> 100) {co2<- c(cromo1[1:(s2-100)], co2[(s2-99): MGL])}
if (s1==1 & s2< 31) {co2<- c(co2[1: (99+s2)], cromo1[(100+s2): MGL])}
if (s1==2 & s2< 31) {co2<- c(co2[1: (99+s2)], cromos2[(100+s2): MGL])}
#co<- cbind(co1,co2); co # resultado da recombinação
res<- (c(((3^co1)+co2))) #codifica como 1 2 3 4
res.ind<- c(res.ind,res)
}
cont<- (cont+1)
genoS[cont,]<- res.ind
}
}
genoA=matrix(); # codifica como 1 0 -1 (Codominante)
genoA[genoS==1]<- c(-1)
genoA[genoS==2]<- c(0)
genoA[genoS==3]<- c(0)
genoA[genoS==4]<- c(1)
genoA<-matrix(genoA, (nis*n.des))
if (G==3) {genomF3<-genoA; }# Reúne os genomas simulados
if (G==4) {genomF4<-genoA; }
if (G==5) {genomF5<-genoA; }
if (G==6) {genomF6<-genoA; }

#


---


# Simulação Valores Genotípicos
VG <- cbind(apply(genoA[,QTL], 1, function(x) 1+(sum (x)*2.5)))

#


---


# Simulação Valores Fenotípicos

```

```

if (H== 100) {
if (var(VG[,1])>0) {VF<-VG[,1]; corrf<- (cor(VG[,1], VF)); Here<- (var(VG[,1])/
var(VF))
} else {VF<-VG[,1]; corrf<- "NA"; Here<- "NA"}}
if (H== 80) {
if (var(VG[,1])>0) { sdv<- (sqrt((var(VG[,1])/0.80)-var(VG[,1])))
repeat
{ VF<-(VG[,1] + rnorm(nrow(VG), mean= 0, sd= sdv))
sdv<-sdv+0.001
corrf<- cor(VG[,1], VF)
Here<- (var(VG[,1])/ var(VF))
if (Here > 0.794 & Here < 0.806) break
} #corrf; Here; sdv
} else {VF<-VG[,1]; corrf<- "NA"; Here<- "NA"}}
if (H== 40) {
if (var(VG[,1])>0) { sdv<- (sqrt((var(VG[,1])/0.40)-var(VG[,1])))
repeat
{ VF<-(VG[,1] + rnorm(nrow(VG), mean= 0, sd= sdv))
sdv<-sdv+0.001
corrf<- cor(VG[,1], VF)
Here<- (var(VG[,1])/ var(VF))
if (Here > 0.394 & Here < 0.406) break
}
} else {VF<-VG[,1]; corrf<- "NA"; Here<- "NA"}}
VF<- as.matrix(VF)
A<- (rep("na", (nrow(VGeno)-nrow(VG))))); B<-as.numeric(c(VG,A)); C<-
as.numeric(c(VF,A))
VGeno<- cbind(VGeno,B) # Acumula os valores genotípicos das gerações
VFeno<- cbind(VFeno,C) # Acumula os valores fenotípicos das gerações
rm(A,B,C)
H2<-c(H2,Here) # Acumula os valores de herdabilidade

# _____

# Estimacão VGG
VGG<-(genoA[,MARC])%*%Ef.Marca
#VGGs<- cbind(VGGs, VGG) # Acumula VGG estimados em cada geração
VGGs<- cbind(VGGs, rbind(VGG, matrix(rep(NA,(nrow(VGGs)-nrow(VGG))),1)))

# _____

# Calculo Da Capacidade preditiva
if (var(VF[,1])>0 && var(VGG)>0) {c.pred<-cor(VF[,1], VGG)} else {c.pred<- "NA"}
Capacidade.p<- c(Capacidade.p, c.pred) # Acumula as capacidades preditivas por
geração

# _____

# Medias e Variâncias
mediaG<- c(mediaG, mean(VG[,1]))

```

```

variancaG<- c(variancaG, var(VG[,1]))
mediaF<- c(mediaF, mean(VF[,1]))
variancaF<- c(variancaF, var(VF[,1]))

#


---



# Ranking de indivíduos

#classificação esperada segundo Valores genômicos simulados
cla.sim_<-VG; rownames(cla.sim_)<- (seq(1,nrow(VG),1))
cla.sim.s=sort(as.matrix(cla.sim_)[,],decreasing=TRUE)
cla.sim<- as.numeric(names(cla.sim.s)); rm(cla.sim_, cla.sim.s)

#classificação segundo Valores genômicos estimados
VGe= VGG + mean(FENOTIPOS[,h]) #(beta=média do fenótipo pob estimação(F2))
VGe<- as.matrix(VGe); rownames(VGe)<- seq(1, nrow(VGe),1)
VGe1=sort(as.matrix(VGe)[,],decreasing=TRUE); VGe1<- as.matrix(VGe1)
cla<- as.numeric(rownames(VGe1)); rm(VGe, VGe1)
A<- (rep("na", (nrow(Ranking)-length(cla))))); B<-as.numeric(c(cla.sim,A)); C<-
as.numeric(c(cla,A))
Ranking<- cbind(Ranking,B,C) # Acumula as listas de classificação de ind por valor
genômico
rm(A,B,C)

#


---



#Preparação arquivos para a seguinte geração

GENO<- genoS
}
#


---



# Salvando resultados

colnames(Ranking)<- c("F2s","F2e","F3s","F3e","F4s","F4e","F5s","F5e","F6s","F6e")
colnames(VGeno)<- c("F2","F3","F4","F5","F6");
colnames(VFeno)<c("F2","F3","F4","F5","F6")
colnames(VGGs)<- c("F2","F2V","F3","F4","F5","F6")
resultados<- paste("D",D,"-H",H,"-R",R,".RData", sep = "")
save(genomF2, genomF3, genomF4, genomF5, genomF6, fitRRB,fitRQ1, BL VGeno,
VFeno, mediaG, variancaG, mediaF,
variancaF, Ranking, VGGs, Capacidade.p, He, file = resultados)
}
}

```