

GISLÂYNE MAÍRA PEREIRA DE FARIA

**APLICAÇÃO DA SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES
AUTÓGAMAS E ALÓGAMAS**

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
2017

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

T

F224a
2017
Faria, Gislâyne Maira Pereira de, 1984-
Aplicação da seleção genômica ampla em populações
autógamas e alógamas / Gislâyne Maira Pereira de Faria. –
Viçosa, MG, 2017.
ix, 77f. : il. (algumas color.) ; 29 cm.

Orientador: Cosme Damião Cruz.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Biologia molecular. 2. Melhoramento genético.
3. Genômica. I. Universidade Federal de Viçosa. Departamento
de Biologia Geral. Programa de Pós-graduação em Genética e
Melhoramento. II. Título.

CDD 22 ed. 572.8

GISLÂYNE MAÍRA PEREIRA DE FARIA

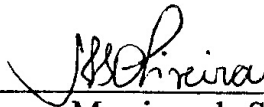
**APLICAÇÃO DA SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES
AUTÓGAMAS E ALÓGAMAS**

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*.

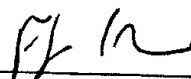
Aprovada: 16 de março de 2017



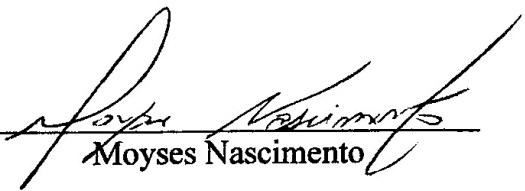
Isis Fernanda de Almeida



Marciane da Silva Oliveira



Felipe Lopes da Silva



Moyes Nascimento



Cosme Damião Cruz
(Orientador)

AGRADECIMENTOS

À Deus, por ter iluminado meu caminho e me abençoado com as graças alcançadas;

Aos meus pais, José e Inês, pelos exemplos de determinação, humildade, carinho, confiança, por quem tenho amor incondicional;

Aos meus irmãos Pábulo, Wallas e Dieime, pelo amor, amizade e incentivo;

A minha querida sobrinha Lorys, por tornar nossas vidas mais alegre;

À Universidade Federal de Viçosa (UFV), pela oportunidade oferecida para a realização deste trabalho;

A CAPES e ao CNPq pelo auxílio financeiro;

Ao Departamento de Genética e Melhoramento, pelo suporte acadêmico;

Ao Professor, Cosme Damião Cruz pela sua orientação, amizade, disponibilidade, dedicação e ensinamento;

Aos Professores, Marcos Deon Vilela de Resende, Moysés Nascimento e Felipe Lopes da Silva pelas suas orientações, no desenvolvimento desta tese.

A Dra. Marciane da Silva Oliveira pela disponibilidade em participar da banca e por todo conhecimento dividido e pela imensa amizade;

A Dra. Ísis Fernanda de Almeida pela disponibilidade em participar da banca e pela imensa amizade;

A todos os professores e colegas do Programa de Pós-Graduação em Genética e Melhoramento, em especial àqueles que contribuíram de alguma forma para minha formação acadêmica e que contribuem para a qualidade do programa;

Aos secretários do Programa de Pós-Graduação em Genética e Melhoramento, Marco Tulio e Odilon, pela amizade, eficiência e disponibilidade em ajudar.

Aos amigos, Aninha, Caio, Luiz Fernando, Bruno Brasileiro, Afonso, Marciane, Camila, Isís, Gustavo e todos do laboratório de bioinformática, pelo carinho e força nessa tão importante etapa da minha vida;

Aos meus amigos que não moram aqui em Viçosa, mas que sempre torceram por mim de longe;

A todos que, direta ou indiretamente, contribuíram para o meu crescimento profissional e para a realização deste trabalho, deixo aqui o meu sincero

MUITO OBRIGADA!!!

BIOGRAFIA

GISLÂYNE MAÍRA PEREIRA DE FARIA, filha de José Pereira dos Santos e Inês Barbosa de Faria Santos, nasceu em Juiz de Fora, Minas Gerais, em 29 de maio de 1984.

No município de Mercês, Minas Gerais, cursou o ensino Primário e Fundamental na Escola Estadual Sena Figueiredo, de 1991 a 1999. De 2000 a 2002, completou o ensino Médio no Colégio Centro Educacional Mercesano.

Em 2005, iniciou o curso superior na Universidade Federal de Viçosa – UFV, em Viçosa, Minas Gerais, colando grau em Julho de 2010 como Engenheira Agrônoma.

Em agosto de 2010, iniciou o curso de Mestrado no Programa de Pós-Graduação em Genética e Melhoramento da UFV, na área de Melhoramento de Plantas e Genética Quantitativa, obtendo o respectivo título em agosto de 2012.

Neste mesmo ano, ingressou no curso de Doutorado no Programa de Pós-Graduação em Genética e Melhoramento da UFV, com ênfase em Biometria, Genética Quantitativa e Seleção Genômica Ampla, obtendo o respectivo título em março de 2017.

SUMÁRIO

RESUMO	VI
ABSTRACT	VIII
INTRODUÇÃO GERAL	10
REFERÊNCIAS BIBLIOGRÁFICAS	13
CAPÍTULO 1	14
CARACTERIZAÇÃO GENÉTICA DAS POPULAÇÕES SIMULADAS DE AUTOFECUNDAÇÃO E ACASALAMENTO AO ACASO	14
INTRODUÇÃO	17
MATERIAL E MÉTODO	20
SIMULAÇÃO DE DADOS	20
SIMULAÇÃO DO GENOMA	20
SIMULAÇÃO DOS GENITORES	20
SIMULAÇÃO DOS INDIVÍDUOS DAS POPULAÇÕES F ₂ E SUAS GERAÇÕES AVANÇADAS POR AUTOFECUNDAÇÃO E ACASALAMENTO AO ACASO	20
SIMULAÇÃO DOS DADOS GENOTÍPICOS E FENOTÍPICOS	21
RESULTADOS E DISCURSSÃO	31
AVALIAÇÃO DAS CARACTERÍSTICAS QUANTITATIVAS	31
TESTE DE NORMALIDADE DAS POPULAÇÕES GERADAS	41
DIVERSIDADE ENTRE E DENTRO DE POPULAÇÕES	43
DESEQUILÍBRIO DE LIGAÇÃO	45
CONCLUSÃO	48
REFERÊNCIAS BIBLIOGRÁFICAS	49
CAPÍTULO 2	53
SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES DERIVADAS DE AUTOFECUNDAÇÃO E DE ACASALAMENTO AO ACASO	53
INTRODUÇÃO	56
MATERIAL E MÉTODOS	58
OBTENÇÃO DAS POPULAÇÕES	58
OBTENÇÃO DOS VALORES GENOTÍPICOS E FENOTÍPICOS	59

METODOLOGIA DE ANÁLISE DA GWS.....	62
AVALIAÇÃO DOS DADOS OBTIDOS NA GWS.....	64
APLICATIVOS COMPUTACIONAIS UTILIZADOS PARA AS ANÁLISES DOS DADOS.....	64
RESULTADOS E DISCUSSÃO.....	65
EFEITO DOS MARCADORES.....	65
ESTIMATIVAS DE ACURÁCIA EM QUE A GENOTIPAGEM OCORREU SOMENTE NA F ₂	67
ESTIMATIVAS DE ACURÁCIA COM O USO DA GENOTIPAGEM DEFASADA NO TEMPO OU CONTEMPORÂNEA.....	72
CONCLUSÃO.....	83
REFERÊNCIAS BIBLIOGRAFICAS.....	84

RESUMO

FARIA, Gislâyne Maira Pereira de, D.Sc., Universidade Federal de Viçosa, março de 2017. **Aplicação da seleção genômica ampla em populações autógamas e alógamas.** Orientador: Cosme Damião Cruz. Coorientadores: Marcos Deon Vilela de Resende e Leonardo Lopes Bhering.

Com o desenvolvimento da biologia molecular, foi possível o desenvolvimento de metodologias que reduziriam o tempo, para obtenção de genótipos superiores. A seleção não se basearia em apenas dados fenotípicos, com o desenvolvimento da genotipagem, seria possível o estudo completo do genoma da espécie em trabalho, desta forma, ter uma maior utilização destas informações genéticas, geradas no melhoramento de plantas. Essa nova metodologia desenvolvida em 2001 foi denominada de Seleção Genômica Ampla (GWS), que foi criada para predizer o fenótipo futuro de uma população baseando em informações de marcadores moleculares, cujos os efeitos genéticos aditivos, que estes explicam, já foram previamente estimados. A seleção genômica ampla tornou-se uma metodologia importante no melhoramento genético de plantas e animais o que pode permitir melhores acurácias e seleção precoce. Com isto, o objetivo deste trabalho em primeiro momento foi realizar um estudo sobre os principais fatores responsáveis por modificarem a estrutura genética de uma população tendo em vista o acasalamento ao acaso e sucessivas gerações de autofecundação e também avaliar a eficácia do processo de simulação em gerar as populações avaliadas. Foram simulados dados de indivíduos e de informações moleculares das populações F₂, derivadas de genitores homocigotos contrastantes. As populações geradas foram de tamanho de 1000 indivíduos e considerou-se seis grupos de ligação com níveis de saturação de 201 marcas moleculares, totalizando 1206 marcadores codominantes. As populações foram submetidas a cinco gerações de autofecundação e acasalamento ao acaso. Para as populações avaliadas, a estrutura genética foi preservada durante o processo de simulação, tornando assim, a simulação um método eficaz. Fatores como o desequilíbrio de ligação, dominância, endogamia, afetaram de forma diferencial, quanto ao tipo de sistema de acasalamento. Na segunda etapa do trabalho foram simuladas populações com estrutura populacional apresentando dados fenotípicos e genotípicos de cada indivíduo dentro da população, imitando alguns dos cenários em que a Seleção Genômica Ampla é aplicada. Com isso, o trabalho teve como objetivo avaliar a acurácia da seleção genômica em diferentes cenários de distribuição de efeitos genéticos e populações de validação. Uma vez geradas as populações na primeira etapa com um número de 1000 indivíduos, para os dados genotípicos considerou-se 1206 locos marcadores, sendo 30 locos que controlava o caráter com herdabilidade de 30 e 60 % e grau médio de dominância de 0.5 e 1.0. Distintos

efeitos de QTL foram simulados, um de acordo com uma distribuição uniforme, outro com distribuição binomial e outro com uma distribuição exponencial com o objetivo de retratar diferentes arquiteturas genéticas. A metodologia RR-BLUP foi utilizada a fim de se computar a acurácia da seleção genômica nos diferentes cenários estabelecidos. A simulação foi eficiente em preservar a estrutura genética das populações, os resultados mostram que o efeito de dominância age como um agente perturbador e que a ação do ambiente também afeta o processo seletivo, para fins de recomendação de uso da GWS.

ABSTRACT

FARIA, Gislâyne Maira Pereira de, D.Sc., Universidade Federal de Viçosa, March, 2017. **Appication of the Genomic Wide Selection in self-fertilization and random mating.** Adviser: Cosme Damião Cruz. Co-advisers: Marcos Deon Vilela de Resende and Leonardo Lopes Bhering

With the development of molecular biology, it was possible to develop methodologies that would reduce the time to obtain superior genotypes. The selection would not be based on phenotypic data only, with the development of genotyping, it would be possible to study the genome of the species in the work, thus, to have a greater use of this genetic information generated in the breeding of plants. This new methodology, developed in 2001, was called Genomic Wide Selection (GWS), which was created to predict the phenotype of a population based on information from molecular markers, whose additive genetic effects, which they explain, have already been previously estimated. Broad genomic selection has become an important methodology in the genetic improvement of plants and animals, which may allow better accuracy and early selection. The main objective of this work was to study the main factors responsible for modifying the genetic structure of a population in order to randomly mating and successive generations of self-fertilization, as well as to evaluate the effectiveness of the simulation process in generating The populations evaluated. Data from individuals and molecular information of the F₂ populations derived from contrasting homozygous parents were simulated. The populations generated were of 1000 individuals and six binding groups were considered with saturation levels of 201 molecular tags, totaling 1206 codominant markers. The populations were submitted to five generations of self-fertilization and random mating. For the populations evaluated, the genetic structure was preserved during the simulation process, thus making simulation an effective method. Factors such as linkage disequilibrium, dominance, inbreeding, affected differentially, as to the type of mating system. In the second stage of the work, populations with population structure were simulated presenting phenotypic and genotypic data of each individual within the population, imitating some of the scenarios in which the broad genomic selection is applied. Thus, the objective of this work was to evaluate the accuracy of genomic selection in different scenarios of distribution of genetic effects and validation populations. Once the populations in the first stage with a number of 1000 individuals were generated, 1206 loco markers were considered for the genotypic data, 30 of which were loco that controlled the heritability of 30 and 60% and a mean degree of dominance of 0.5 and 1.0. Different effects of QTL were simulated, one according to a uniform

distribution, another with binomial distribution and another with an exponential distribution aiming to portray different genetic architectures. The RR-BLUP methodology was used in order to compute the accuracy of the genomic selection in the different established scenarios. The simulation was efficient in preserving the genetic structure of the populations, the results show that the dominance effect acts as a disturbing agent and that the action of the environment also affects the selection process, for purposes of GWS use recommendation.

INTRODUÇÃO GERAL

Para Fehr (1987), o melhoramento de plantas é a arte e a ciência de modificar as plantas a fim de torná-las mais úteis ao homem. Desta forma, o melhoramento de plantas se tornou indispensável por atuar de forma que as plantas se tornem mais produtivas por meio da obtenção de materiais genéticos, cultivares, linhas e híbridos, com maior potencial de produção. O melhoramento genético, seja animal ou vegetal, tem, há vários anos, proporcionado níveis de produtividade cada vez maiores e proporcionado melhoria no desempenho de várias outras características de importância nos diversos campos da agricultura.

O grande problema é que o processo de desenvolvimento de novas cultivares, por meio do melhoramento de plantas, é uma atividade que demanda longo período, esforço humano e recursos físicos e financeiros. Assim, é fundamental que o melhorista preveja quais são as necessidades futuras e desenvolva metodologias para a redução destes longos períodos e destes custos.

Ao longo dos anos, as metodologias utilizadas pelos melhoristas, para o aumento da produtividade e melhoria das outras características de importância econômica, eram baseadas em informações fenotípicas e, em alguns casos, de genealogia, como forma de prever o valor genético do indivíduo. Com o desenvolvimento dos marcadores moleculares e o avanço em técnicas de biologia molecular, criou-se a expectativa de que as informações genotípicas dos marcadores moleculares, uma vez correlacionados com características fenotípicas de interesse, pudessem ser amplamente utilizadas na obtenção e seleção de indivíduos com maior valor genético (Resende Júnior, 2010).

O principal atrativo da genética molecular em benefício do melhoramento genético é a utilização direta das informações de DNA na seleção, de modo a permitir alta eficiência seletiva, rapidez na obtenção de ganhos genéticos com a seleção e baixo custo (Resende et al., 2014). Com a perspectiva de aumento nos ganhos de seleção e redução

nos ciclos de melhoramento via seleção assistida por marcadores, muitas pesquisas foram feitas e QTL (locos controladores de características quantitativas) foram detectados e mapeados nas mais variadas culturas (Frary et al.,2000; Yano et al.,2000; Liu et al.,2002).

Com os avanços de tecnologias de genotipagem em larga escala, tornou-se possível uma cobertura completa do genoma, o que levou os pesquisadores a criar uma nova forma de utilização dessa informação genética gerada (Almeida et al.,, 2016). Meuwissen et al. (2001) propuseram um novo método de seleção denominado de seleção genômica ampla (GWS), a qual pode ser aplicada em todas as famílias, ou estruturas populacionais, em avaliação nos programas de melhoramento genético de espécies autógamas e alógamas.

Com o advento da Seleção genômica ampla, conseguiram abrir espaços onde a seleção assistida por marcadores (SAM) tinha limitações, isto através do uso de modelos de predição que utiliza como base as informações genômicas. Pois a seleção genômica ampla captura todos os genes que afetam um caráter quantitativo sem a necessidade de identificar previamente os marcadores com efeitos significativos e de mapear QTLs, como ocorre com a SAM.

Com a utilização de sistemas de genotipagem em alta densidade, foi predominante para que a seleção genômica fosse viável, pois de acordo com Shengqiang et al., 2009, a genotipagem aumentava a probabilidade de cada um dos *locus* da característica estarem em desequilíbrio de ligação (LD) com um número considerável de *locus* de marcadores em toda a população alvo.

A metodologia de seleção genômica ampla para ser executada primeiramente obtém-se o polimorfismo dos marcadores, e logo em seguida, os seus efeitos são estimados baseados em dados fenotípicos de uma população conhecida como população de estimação ou treinamento. Após ser estimados os efeitos dos marcadores na população de treinamento é então possível prever os valores genotípicos dos indivíduos candidatos

a seleção que não foram fenotipados. Sendo que este segundo grupo de indivíduos, que foram apenas genotipados compõem a população de seleção.

Para a implementação da seleção genômica impõe alguns desafios estatísticos e computacionais, contudo não existe um método estatístico unanime, pois, cada uma destas metodologias pode assumir diferentes pressuposições. Entre elas, o método RR-BLUP (*Ridge Regression Best Linear Unbiased Prediction*) estima simultaneamente os efeitos de todas as marcas, sendo estas consideradas efeitos aleatórios com variância comum.

A metodologia de seleção genômica foi um grande diferencial no melhoramento de genético, pois conseguimos a seleção de genótipos antes mesmo da fenotipagem, o que viabiliza um aumento expressivo em um programa de melhoramento genético, já que em condições normais, é preciso de longos períodos para poder ter a tomada de decisão com base apenas em informações fenotípicas (Resende et al.2012).

O objetivo deste trabalho foi em primeiro momento estudar o comportamento das populações simuladas por acasalamento ao acaso e nas gerações de autofecundação e com base nos valores fenotípicos e genotípicos obtidos, gerar os parâmetros genéticos e obter informações quanto o grau médio de dominância, grau de endogamia, diversidade entre e dentro de populações, conteúdo de informação de polimorfismo e desequilíbrio de ligação. Em uma segunda etapa, avaliar a acurácia da seleção genômica em diferentes cenários de distribuição de efeitos genéticos e populações de validação, sob efeito de dominância. Uma vez que o estudo foi baseado em dados simulados, também se objetivou analisar a eficácia do processo de simulação em reproduzir os princípios genéticos da genética quantitativa e da genética de populações nas populações geradas.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, I. F.; CRUZ, C. D.; RESENDE, M. D. V. Validação e correção de fenótipos na seleção genômica ampla. PESQUISA AGROPECUÁRIA BRASILEIRA (ONLINE), v. 51, p. 1973-1982, 2016.

FEHR, W.R. Principles of cultivar development. New York: Macmillan Publishing Company, 1987. 536p.

FRARY, A.; NESBITT, T.C.; GRANDILLO, S.; KNAAP, E.; CONG, B.; LIU, J.; MELLER, J.; ELBER, R.; ALPERT, K.B.; TANKSLEY, S.D. Fw2.2: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. Science. 289:85-88, 2000.

LIU, J.; VAN ECK, J.; CONG, B.; TANKSLEY, S.D. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc Natl Acad Sci USA. 99: 13302-13306, 2002.

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

RESENDE JÚNIOR, M.F.R. Seleção genômica ampla no melhoramento vegetal. Viçosa: Universidade Federal de Viçosa, 2010. 76 p. (Dissertação - Mestrado em Genética e Melhoramento de Planta).

RESENDE, M.F. R; MUÑOZ, P; ACOSTA, J. J, PETER, G.F, DAVIS J. M, GRATTAPAGLIA, D; RESENDE, M. D. V, KIRST M. . Accelerating the domestication of trees using genomic selection: Accuracy of prediction models across ages and environments. New Phytologist 193: 617–624, 2012.

RESENDE, M.D.V.; SILVA, F. F; AZEVEDO, C. F. Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categorias e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTI-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência. 1ª Ed. Viçosa: UFV. 2014. 881p.

SHENGQIANG Z, DEKKERS, J. C. M, FERNANDO, R. L, JANNINK, J. L. 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. Genetics 182: 355–364.

YANO, M.; KATAYOSE, Y.; ASHIKARIB, M.; YAMANOUCHI, U.; MONNA, L.; FUSE, T.; BABA, T.; YAMAMOTO, K.; UMEHARA, Y.; NAGAMURA, Y.; SASAKIA, T. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the arabidopsis flowering time gene CONSTANS. Plant Cell. 182: 2473-2483, 2000.

CAPÍTULO 1

CARACTERIZAÇÃO GENÉTICA DAS POPULAÇÕES SIMULADAS DE AUTOFECUNDAÇÃO E ACASALAMENTO AO ACASO

Resumo - O melhoramento de plantas tem como uma das principais funções tornar as plantas mais produtivas, isto por meio da obtenção de cultivares com maior potencial de produção. Entretanto, é uma atividade que demanda um longo tempo. Uma alternativa, consiste na utilização dos recursos computacionais como aliados na obtenção de bancos de dados confiáveis, tanto para estudos ligados ao melhoramento genético e testar modelos, métodos e estratégias mais eficazes em condições práticas. O sucesso em um programa de melhoramento depende do pleno conhecimento do comportamento das populações as quais esteja trabalhando. Diante do exposto, neste trabalho foram utilizados conceitos de genética que estão relacionados com a constituição genética dos indivíduos isoladamente ou dentro das populações, para assim discorrer sobre os principais fatores que modificam a estrutura genética de uma população tendo em vista os acasalamentos ao acaso e sucessivas gerações de autofecundação. Também teve como objetivo, avaliar a eficácia do processo de simulação em gerar populações e gerações em que o princípio genético seja preservado. As populações geradas foram de tamanho de 1000 indivíduos e considerou-se seis grupos de ligação com níveis de saturação de 201 marcas moleculares, totalizando 1206 marcadores codominantes analisados. A partir da população F₂ foram submetidas cinco gerações de autofecundação e de acasalamento ao acaso. Após obtenção das 11 populações foram feitas várias análises para entender o comportamento destas populações. Conclui-se que o processo de simulação biológico foi eficaz em preservar a estrutura genética de populações e descrever as propriedades dos principais parâmetros genéticos dentro da dinâmica estabelecida por sucessivas gerações de acasalamento ao acaso e de autofecundação e que a base teórica de genética de populações e quantitativa, em modelos de um único gene, foi devidamente retratada em dados gerados *in silico*, envolvendo mais de mil locos gênicos, e diferentes blocos de ligação.

Abstract - The improvement of plants has as one of the main functions to make the plants more productive, this through the obtaining of cultivars with greater potential of production. However, it is an activity that requires a long time. One alternative is to use computational resources as allies in obtaining reliable databases for both studies related to genetic improvement and to test models, methods and strategies that are more effective on practical conditions. Success in a breeding program depends on full knowledge of the behavior of the populations you are working with. Considering the above, in this work genetic concepts were used that are related to the genetic constitution of the individuals either alone or within the populations, so to discuss the main factors that modify the genetic structure of a population for random and successive mating Generations of self-fertilization. It also aimed to evaluate the effectiveness of the simulation process in obtaining populations and generations in which the genetic principle is preserved. The populations generated were of 1000 individuals and six binding groups were considered with saturation levels of 201 molecular markers, totaling 1206 codominant markers analyzed. From the F₂ population, five generations of self-fertilization and random mating were submitted. After obtaining the 11 populations, several analyzes were made to understand the behavior of these populations. It is concluded that the biological simulation process was effective in preserving the genetic structure of populations and describing the properties of the main genetic parameters within the dynamics established by successive generations of mating and self-fertilization and that the theoretical basis of population genetics and quantitative analysis, in single gene models, was duly portrayed in data generated *in silico*, involving more than a thousand gene loci, and different binding blocks.

INTRODUÇÃO

Para o desenvolvimento de novas espécies agrícolas melhoradas, é necessário que o melhorista realize um estudo das características genéticas da população de interesse, sendo estas características quantitativas ou qualitativas. Como citado por Falconer e Mackay (1996), os caracteres qualitativos são de herança monogênica, ou seja, condicionados por um ou poucos genes e tem pouca ou nenhuma influência do ambiente e os caracteres quantitativos são de herança poligênica e são governados por muitos genes, sendo que cada gene apresenta segregação conforme as leis Mendelianas. Quando se analisa uma população segregante, observa-se que os caracteres quantitativos apresentam distribuição contínua de fenótipos. O grande desafio dos pesquisadores é que a maioria das características de interesse agrônomo apresenta herança quantitativa, e sendo os caracteres quantitativos muito influenciados pela variação do ambiente, há dificuldade na identificação dos genótipos com base apenas no fenótipo observado.

Para que o melhoramento genético seja eficiente, é fundamental a realização de um estudo amplo do comportamento das populações e, desta forma, o melhorista consiga gerar informações fidedignas e ser eficaz ao realizar as mudanças em magnitude e sentido desejado.

O conhecimento da estrutura da população inclui a obtenção de informações da frequência dos alelos que compõem os diferentes genótipos das famílias e das frequências genotípicas resultantes do inter cruzamento considerando diferentes sistemas de acasalamento. Portanto, o melhoramento de plantas se baseia na manipulação de caracteres quantitativos por meio da manipulação de populações considerando os processos de acasalamento ao acaso, cruzamentos endogâmicos e hibridações. Sobre estas populações é praticada a seleção que constitui a estratégia essencial para qualquer programa de melhoramento genético que, basicamente, tenha o mesmo objetivo: identificar, acumular e perpetuar genes favoráveis (Cruz e Carneiro, 2003). O

conhecimento em genética quantitativa permite o entendimento sobre as mudanças da população, em especial no seu potencial e em sua variabilidade, que a torna tão relevante para o melhoramento de plantas.

As estratégias de melhoramento genético aplicada a características quantitativas dependem fundamentalmente dos sistemas de acasalamento. Neste sentido, o estudo de caracteres de herança quantitativa não se baseia somente no indivíduo, mas na população como um todo, o que obriga o pesquisador a estender seus estudos a grupos maiores de indivíduos. Alguns fatores devem ser considerados neste processo seletivo, destacando o efeito do ambiente, a dominância e a epistasia. Desta forma, estes fatores podem tanto aumentar quanto diminuir a manifestação fenotípica de um caráter, a média de um conjunto de indivíduos será uma medida mais confiável, pois alguns efeitos, como o caso do efeito ambiental, tendem a se cancelar. Outra medida usada para definir uma população, além da média, é a variância (Cruz, 2012).

Para descrever sobre os principais fatores que modificam a estrutura genética de uma população tendo em vista os acasalamentos ao acaso e as sucessivas gerações de autofecundação são utilizados conceitos de genética que estão relacionados com a constituição genética dos indivíduos isoladamente ou dentro das populações.

Os estudos da estrutura das populações, mediante as informações dos parâmetros genéticos, podem proporcionar importantes esclarecimentos ao desenvolvimento de um programa de melhoramento. Atualmente, têm sido desenvolvidas pesquisas com modernos métodos de avaliação genética e de estimação de parâmetros genéticos, para diferentes critérios de seleção; outro enfoque aborda somente as estruturas das populações (Malhado et al. 2008). Assim, a disponibilidade de dados, reais e simulados, passa a ser uma necessidade constante.

Desta maneira, a utilização da simulação computacional tem sido de grande utilidade em estudos genéticos sob vários contextos, como o de populações, do indivíduo

ou do próprio genoma. Segundo Cruz (2001), a simulação computacional demanda dos geneticistas o desenvolvimento de modelos biológicos que retratem, da melhor maneira possível, os fenômenos de interesse, e dos programadores as rotinas para o processamento adequado, apesar de impor restrições, para que a influência de certos fatores possa ser avaliada.

O emprego de simulação computacional, além de demandar menor recurso e tempo, pode ser generalizado com mais facilidade. De acordo com Cruz (2006), uma das grandes contribuições da informática é possibilitar que o estudo de determinadas situações seja realizado por meio da simulação de uma situação complexa, em que são estabelecidos parâmetros e restrições, de forma que os efeitos de certos fatores sejam convenientemente estudados.

É prudente, em certos estudos da área de melhoramento genético, que cada característica simulada tenha a propriedade de descrever uma variável com média, herdabilidade e precisão experimental estabelecidas pelo pesquisador (Cruz, 2006), pois tais informações são rotineiramente trabalhadas pelo pesquisador e conseguem captar o potencial genético do material estudado, sua variabilidade e a influência.

Pelo exposto, realizou-se este trabalho com o objetivo de avaliar o comportamento das populações simuladas por acasalamento ao acaso e nas gerações de autofecundação e com base nos valores fenotípicos e genotípicos obtidos, gerar os parâmetros genéticos e obter informações quanto o grau médio de dominância, grau de endogamia, diversidade entre e dentro de populações, conteúdo de informação de polimorfismo e desequilíbrio de ligação, isto, utilizando diferentes cenários de distribuição de efeitos genéticos e também avaliar a eficácia do processo de simulação em gerar as populações e gerações cujos princípios genéticos sejam preservados conforme princípios fundamentados da genética e de populações.

MATERIAL E MÉTODO

Simulação de dados

Para a geração dos dados foi utilizado o módulo de simulação do aplicativo computacional GENES (Cruz, 2016), o qual é amplamente utilizado em análises de modelos aplicados ao melhoramento de plantas e animais.

Desta maneira, foram simulados genomas parentais e amostras de populações F_2 , que foram utilizadas como base para obtenção de gerações avançadas por acasalamento ao acaso e autofecundação, e genotipadas em relação a marcadores codominantes.

Simulação do genoma

Para este estudo foi tomada como referência uma espécie diploide fictícia com $2n=2x= 12$ cromossomos, cujo comprimento total do genoma foi estipulado em 600 cM. Foram gerados genomas com níveis de saturação de 201 marcas moleculares espaçadas a 0,5cM por grupo de ligação. Cada genoma foi composto por 6 grupos de ligação, 100 cM em cada grupo, perfazendo total de 1206 marcadores analisados.

Simulação dos genitores

Após obtenção do genoma, este foi utilizado para obter os genitores. Foi simulado a situação em que os pais são homozigotos, sendo um dominante e outro recessivo. Também foi concebido que todos os pares de locos estivessem em aproximação na população F_1 heterozigota.

Simulação dos indivíduos das populações F_2 e suas gerações avançadas por autofecundação, acasalamento ao acaso

Os procedimentos para a simulação das populações F_2 e as gerações avançadas de autofecundação e acasalamento ao acaso seguiram os seguintes passos:

1- A partir do genoma simulado, foram construídos os genótipos parentais homocigotos e contrastantes para os marcadores, de forma que a geração F₁ estivesse em aproximação para todos os marcadores.

2- A partir do genótipo da geração F₁, foram gerados gametas para a formação dos indivíduos das populações F₂. Posteriormente a população F₂ foi submetida a cinco ciclos de acasalamento ao acaso e cinco ciclos de autofecundações para obter as gerações avançadas. Para fins práticos e didáticos as populações analisadas foram denominadas de A_j (j= 1,2...5) para as gerações derivadas por acasalamento ao acaso, S_j (j= 1,2...5) para as gerações derivadas por autofecundação.

Simulação dos dados genotípicos e fenotípicos

Para processar as análises foram simulados dados genotípicos e fenotípicos considerando dominância parcial (gmd=d/a=0.5) e dominância completa (gmd=d/a=1.0). Foram obtidos 12 características quantitativas com herdabilidades iguais a 0.30 e 0.60 controlados cada um por 30 locos, considerando dois alelos por loco, com distribuição da importância de cada loco dado por pesos estabelecidos a partir de distribuições uniforme, binomial e exponencial.

Os fenótipos dos indivíduos (i) serão gerados segundo o modelo F_i= G_i + M_i, em que G_i é o efeito genético dado pelo somatório dos efeitos genéticos em cada loco e M_i o efeito ambiental, gerado segundo uma distribuição normal com média e variância compatível com a herdabilidade do caráter simulado.

Em que a herdabilidade é dada por:

$$h^2 = \frac{\sigma_g^2}{\sqrt{\sigma_g^2 + \sigma_e^2}}$$

Sendo h^2 a herdabilidade, σ_g^2 é a variância genética e σ_e^2 é a variância ambiental entre os indivíduos.

O Valor genético total expresso por um determinado indivíduo pertencente à população A_j e S_j foi estimado a partir da expressão:

$$G_i = \mu + a_i + d_i$$

Em que:

$$a_i = \sum_{j=i}^{30} p_j \alpha_j$$

$$d_i / a_i = 0,5$$

$$d_i / a_i = 1,0$$

Sendo α_j é o efeito do alelo favorável no loco j , considerando igual a 1, 0 ou -1 para as classes genóticas AA, Aa e aa, respectivamente, e p_j é a contribuição do loco j para a manifestação da característica considerada, no trabalho, como tendo distribuição uniforme, binomial e exponencial. O valor de d_i foi estabelecido a partir do grau médio da dominância manifestado em cada característica.

A partir das 11 populações foram feitas várias análises para entender o comportamento das populações. É importante o conhecimento dos parâmetros genéticos e estatísticos das consequências do uso de métodos de melhoramento genético.

Desta forma, com base nos valores genotípicos e fenotípicos em toda as gerações estudadas, foram estimados os parâmetros ou geradas informações sobre:

a- Heterozigose observada e esperada sob equilíbrio de Hardy-Weinberg

É através da frequência do heterozigoto que conseguimos definir o coeficiente de endogamia. É esperado que em uma população de autofecundação, a cada geração a frequência observada dos heterozigotos será reduzida à metade. Em quanto uma geração de acasalamento ao acaso se a população está em Equilíbrio de Hardy-Weinberg, não teremos demonstração de endogamia nas gerações avaliadas.

b- Grau de endogamia

A endogamia é definida como todo sistema de acasalamento que promove o aumento de homozigose nas descendências, como por exemplo, o cruzamento entre parentes.

O Coeficiente de endogamia (F) é a probabilidade de dois alelos, em um único loco, tomados ao acaso de uma população, serem idênticos por ascendência. Segundo Falconer (1989), o coeficiente de endogamia refere-se à probabilidade de que os alelos de um loco de um indivíduo sejam idênticos por ascendência e, esses alelos são idênticos quando derivam ou são cópias de um alelo comum encontrado nos ancestrais daquele indivíduo. Para Wrigth, endogamia ocorre em consequência da identidade dos gametas que se unem, sendo expressa pela correlação entre os valores dos gametas que formam a progênie derivada de uma população. Cruz (2012) descreve que, em uma população panmítica, onde ocorre o acasalamento ao acaso, essa correlação é nula, porém, em uma população endogâmica, a correlação não é nula e é proporcional ao coeficiente de endogamia

O coeficiente de endogamia depende do tamanho da população e, quanto menor for o tamanho da população, em gerações anteriores, maior será o número de ancestrais comuns e maior será o coeficiente de endogamia. O coeficiente de endogamia varia de acordo com o sistema de acasalamento:

Autofecundação (S) – as autofecundações sucessivas são o método mais rápido para atingir endogamia máxima. Em programas de híbridos de linhagens, a obtenção de linhagens puras é por meio de fecundações sucessivas.

Acasalamento ao acaso (A) – ressaltamos que após uma geração de acasalamento ao acaso, a endogamia desaparece. Essa afirmativa é válida para uma população suficientemente grande, de forma que a probabilidade do encontro gamético de alelos idênticos por ascendência é nula. Para uma população finita (N), em que os indivíduos

são monoicos e ocorre o acasalamento ao acaso, surge, a cada geração, um efeito da endogamia.

Quanto à estimação da endogamia, pode ser obtida, pelas frequências de heterozigoto. Os valores endogamia serão obtidos por:

$$F = 1 - \frac{H_o}{H_e}$$

Em que,

$$H_o = 1 - \sum P_{ii}$$

$$H_e = 1 - \sum P_i^2$$

Onde, P_{ii} é a frequência dos genótipos homozigotos e P_i^2 é a frequência alélica estimada do i -ésimo alelo.

Como consequência da endogamia, temos o aumento da homozigose nas descendências oriundas de cruzamentos endogâmicos. Em algumas espécies, principalmente em animais e plantas alógamas, a endogamia pode levar à depressão endogâmica. A depressão endogâmica é a perda de vigor na descendência ocasionado pelo aparecimento de genes deletérios e letais em condição homozigota nas descendências (caráter com algum nível de dominância).

Um bom exemplo do efeito da depressão endogâmica é o milho. Plantas de milho obtidas por autofecundação apresentam desempenho muito inferior a plantas não endogâmicas. Elas apresentam menor altura, menor produtividade e menor fertilidade. Para o melhoramento de plantas, a principal importância da endogamia é a fixação genética de um genótipo. Em outras palavras, a descendência será idêntica à planta mãe. Isto acontece quando é feita a autofecundação de uma planta homozigota.

Nas espécies alógamas, a tolerância à endogamia varia de espécie para espécie. Em alfafa, uma espécie autotetraplóide, e cenoura os efeitos são mais prejudiciais, seguindo-se de milho, diplóide. Por outro lado, a cebola, o girassol, as curcubitáceas, e outras espécies, são menos sensíveis.

c- Porcentagem de locos em Equilíbrio de Hardy-Weinberg

A lei de Hardy-Weinberg nós possibilita o entendimento do que acontece às frequências gênicas e genotípicas das populações estudadas. Se as frequências gênicas e genotípicas permanecem constantes de uma geração para a seguinte, quando isso acontece é dito que está em EHW. Isso é observado em uma população de cruzamento ao acaso, em que não há seleção, mutação, migração ou deriva genética (Nicholas, 2011).

De acordo com Mariz (2010), a seleção, migração, mutação e deriva genética são os quatro fatores que podem interferir no EHW de uma população, interferindo desta forma, na variabilidade genética população, em que a seleção atua de maneira negativa na variabilidade genética, pois promove a fixação de genes na mesma, com a utilização de poucos indivíduos para o acasalamento, enquanto a migração está relacionada com a entrada de novos genes em uma determinada população, já a mutação é uma variação brusca e hereditária, que pode ocorrer em algum gene, por fim a deriva genética é uma alteração natural na frequência de alelos que ocorre em populações pela perpetuação de genes por meio da reprodução, podendo ocasionar rápida redução da variabilidade genética e aumento de homozigosidade.

Segundo Menezes (2005; 2006), qualquer desvio significativo no EHW indicará que a população está subdividida, que existe uma endogamia ou fluxo de genes de outra população.

Uma população sob cruzamentos aleatórios encontram-se no estado de “Equilíbrio perfeito” e tendo como suas frequências $[(p + q)^2 = p^2 + 2pq + q^2 = 1]$, para cada loco gênico.

d- *Média e variância considerando valores genotípicos iguais a 1,0 e -1 para AA, Aa e aa, respectivamente, para cada loco;*

A base teórica de genética quantitativa relata que a média genotípica, em uma população em equilíbrio, é dada por $M_o = \mu + a(p-q) + 2pqd$. Quando existe endogamia em uma população, esta altera a média da população com as sucessivas gerações. Desta forma, a média da população endogâmica pode ser obtida por $M_F = a(p-q) + 2pqd - 2pqdF$, quando ao considerar o efeito combinado de todos os locos que influenciam o caráter e que os valores genotípicos dos locos combinam-se aditivamente, a média da população será obtida pela soma da contribuição dos locos separados como descrito a seguir:

$$M_F = M_o - 2F \sum_i d_i p_i q_i$$

Em que,

M_o = média da população original

Através destas equações matemáticas podemos observar em quais circunstância um caráter métrico mostrara mudança no valor médio. A mudança na média, em que é devida a endogamia, é causada por $-2Fpqd$.

e- *Diversidade entre e dentro de populações*

O sucesso de um programa de melhoramento reside na existência de variabilidade na população de trabalho. Melhoristas têm recomendado, para a formação de uma população-base (ou população de melhoramento), o intercruzamento entre genitores contrastantes.

Portanto, conhecer a variabilidade genética entre e dentro de populações é um precioso instrumento que poderá ajudar na adoção de práticas mais eficientes de conservação e na identificação de características de interesse para o melhoramento genético das espécies (Quintal, 2009).

As medidas de dissimilaridade são de grande importância em estudos de diversidade genética em que se procura identificar genitores a serem utilizados em programas de hibridação. Em outras, situações, estudos sobre a diversidade genética têm sido realizados com o intuito de identificar grupos de populações com maior similaridade (Cruz, 2012). Diversos métodos para estimar a variação genética de estrutura populacional já foram desenvolvidos, com aplicações variadas em níveis individuais, intrapopulacional e interpopulacional (Cruz et. al., 2011).

A caracterização genética com o uso de marcadores moleculares tem demonstrado ser um instrumento muito eficaz para quantificação da diversidade genética entre os indivíduos. Os marcadores genéticos são loci que apresentam características detectáveis que diferenciam os indivíduos de determinada população, demonstrando variações entre os indivíduos.

Para a realização da diversidade entre e dentro das populações estudadas, fizemos com base na distância euclidiana e distância de Hedrick. De acordo com Cruz (2008), a distância Euclidiana tem como finalidade quantificar a diversidade genética entre duas populações a partir de suas frequências alélicas, já a distância de Hedrick se baseia em quantificar a dissimilaridade entre as populações por meio das frequências genotípicas.

f- Conteúdo de informação de polimorfismo (PIC)

Botstein et al. (1980) descreve o conteúdo de Informação Polimórfica (PIC), como um indicador da qualidade do marcador em estudos genéticos (segregação, identificação de populações e controle de paternidade). De acordo com Cruz (2006), o PIC é utilizado para quantificar o polimorfismo de uma população, sendo que o valor de PIC fornece uma estimativa do poder discriminatório do marcador por considerar não somente o número de alelos por loco, mas também a frequência relativa desses alelos.

Segundo Cruz (2006), quando o $PIC \geq 0.1$ o loco será polimórfico, sendo que, neste caso, o alelo mais frequente tem frequência menor que 0.95. O PIC pode ser obtido através da equação a baixo.

$$PIC = 1 - \sum_{i=1}^{\alpha} p_i^2 - \sum_{i,j=1}^{\alpha} \sum_{(i \neq j)} p_i^2 p_j^2$$

O PCI surgiu para quantificar o valor da informação do polimorfismo de um locus marcador, sendo considerado um parâmetro que apresenta uma dependência do número de alelos e de suas frequências. As classificações dos marcadores com valores de PIC, são considerados muito informativo, mediamente informativo e pouco informativo, com valores superiores a 0.5, entre 0.25 e 0.50 e, inferiores a 0.25 respectivamente (Bolstein et al., 1980).

g- Desequilíbrio de ligação

O desequilíbrio de ligação (LD) ou desequilíbrio de fase gamética é associação não ao acaso de alelos em diferentes locos. Em um conjunto de indivíduos, se dois alelos são encontrados juntos com frequência maior do que aquela esperada com base no produto de suas frequências infere-se que tais alelos estão em desequilíbrio de ligação (Resende, 2008). Inicialmente, o desequilíbrio de ligação está presente na população em uma taxa determinada pela distância genética entre os dois locos e o número de gerações desde que ele surgiu. Existem diferentes metodologias para estimar os níveis de LD, que são amplamente descritas em plantas (Flint-Garcia et al., 2003; Gaut & Long, 2003; Gupta et al., 2005; Simko et al., 2004). As duas estatísticas mais usadas para mensurar o desequilíbrio de ligação são r^2 e D' (Jorde, 2000; Liang et al., 2001; Gorelick & Laubichler, 2004).

Alguns fatores influenciam o LD, como autofecundação, pequenos tamanhos de populações, isolamento genético entre linhas, subdivisão populacional, baixa taxa de recombinação, mistura populacional, seleção artificial e natural, estes são alguns fatores que são responsáveis pelo aumento no LD. Outros fatores que causar a queda ou quebra do LD são: acasalamento ao acaso, elevadas taxas de recombinação, elevadas taxas de mutações, dentro outros. Existem fatores que podem aumentar ou quebrar o LD, ou podem aumentar o LD entre determinado par de alelos e diminuir-lo entre os pares de alelos. Por exemplo, a mutação pode romper o LD entre pares de alelos envolvendo alelos selvagens, e promover LD entre os pares de alelos dos mutantes envolvidos. Outros fatores que afetam o LD incluindo estrutura populacional e epistasia não tem recebido atenção desejada nas revisões realizadas (Resende, 2008; Goddard, 2009; Grattapaglia & Resende, 2011; Abdurakhmonov & Abdurakarimov, 2008).

O desequilíbrio de liquidação é estimado a partir de coeficientes padronizados de desequilíbrio D' (Hedrick, 1987), e o quadrado do coeficiente de correlação entre dois locos r^2 (Weir & Hill, 1986) para pares de locos. Uma medida do desequilíbrio, r^2 , é dada por:

$$r^2 = \frac{(D)^2}{p_A p_a p_B p_b}$$

Entretanto, a menos que os dois locos tenham frequências alélicas idênticas, o valor da correlação igual a 1 não é possível de ser obtida.

Outra medida de desequilíbrio é a estatística D' é calculada conforme descrito a baixo:

$$= |D'| \frac{D^2}{\min(p_A p_b p_a p_B)} \quad D < 0$$

$$= |D'| \frac{D^2}{\min(p_A p_b p_a p_B)} \quad D > 0$$

A estatística D' é baseada nas frequências observadas, e irá variar entre 0 e 1 se as frequências alélicas diferirem entre os locos. D' pode ser menor do que 1, somente se as todos os quatro possíveis gametas forem observados, conseqüentemente assumindo que eventos de recombinação ocorreram entre os locos.

A análise de desequilíbrio de ligação tem sido utilizada para medir a proximidade entre um loco e outro, para mapear características qualitativas e mais recentemente para mapear QTL (Flint-Garcia et al., 2003).

A seleção assistida por marcadores (MAS) e o mapeamento de QTL proposta por Lande e Thompson (1990) e a seleção genômica ampla (GWS) proposta por Weuwissen et al. (2001), são fundamentadas na ocorrência de desequilíbrio de ligação na população estudada. Sendo de grande importância é a extensão desse desequilíbrio de ligação no cromossomo em uma população de seleção que se um marcador e um QTL estão em equilíbrio na população, o conhecimento do genótipo do marcador em um indivíduo não apresenta qualquer valor para a seleção. Os alelos dos marcadores informam sobre a presença e efeitos dos locos que governam os caracteres quantitativos, fornecendo meios para estimação dos efeitos dos locos dos QTLs e para seu eficiente uso na seleção genética (Almeida et al., 2016).

São apresentadas formas de visualização da extensão do desequilíbrio de ligação entre pares de locos. Os gráficos de declínio do desequilíbrio de ligação são usados para visualizar a razão na qual o desequilíbrio de ligação diminui em função da distância genética ou física, nas sucessivas gerações de acasalamento ou autofecundação. Segundo Resende (2008), o decréscimo do desequilíbrio de ligação com o aumento da distância indica quantos marcadores e fenótipos são necessários para a acurada predição no contexto da seleção genômica ampla e da detecção de QTL usando LD em nível populacional. Desta forma, o estudo do desequilíbrio de ligação das populações é uma

ferramenta de grande importância para discorrer sobre o quanto os tipos de acasalamento alteram o desequilíbrio ao longo das gerações.

Além da estimação de parâmetros, as populações foram caracterizadas quanto à distribuição de seus dados sendo feito testes de normalidade e obtenção de coeficientes de simetria (s) e curtose (k) de todas as gerações.

RESULTADOS E DISCUSSÃO

Avaliação das características quantitativas

Os sistemas reprodutivos abordados neste trabalho, autofecundação e acasalamento ao acaso, resultam em dinâmicas alélicas e genotípicas particulares, amplamente relatado na literatura em exemplo de segregação de um modelo simples de único gene com dois alelos, e que pode ser visto nos resultados apresentados a seguir considerando vários locos distribuídos em seis grupos de ligação. O entendimento destas dinâmicas permite aos melhoristas desenvolver metodologias que fundamentam os diferentes métodos de melhoramento vegetal, empregados atualmente, nas diferentes culturas de importância econômica.

Sabe-se que a ocorrência de cruzamentos e de autofecundações, levam a estruturas populacionais bem distintas em gerações avançadas, e podemos comprovar este fenômeno por meio da média, da variância, da distribuição dos valores e da endogamia. Na Tabela 1 e Figura 1, estão apresentados os coeficientes de endogamia (F), heterozigose, e PIC, sendo estes, alguns parâmetros de importância no estudo das populações. Quanto à estimação da endogamia, destaca-se o fato que ela pode ser obtida pelas frequências de heterozigoto. Neste trabalho realizamos os estudos da endogamia através da frequência de heterozigoto, onde avaliamos cinco gerações avançadas por autofecundação (S_j , sendo $j= 1, \dots, 5$) e acasalamento ao acaso (A_j sendo $j= 1, \dots, 5$).

Podemos observar (Tabela 1) nas populações, ou gerações estudadas que no processo de acasalamento e por autofecundação constata-se um acréscimo no grau de endogamia, sendo observado o valor igual a 0.015183, que traduz um ligeiro déficit de homozigotos, na geração F₂, chegando até valor de 0.969728 na geração S₅. Desta forma, constatamos que o processo adotado para estimação do coeficiente de endogamia, utilizando a heterozigosidade, foi eficiente, já que este processo foi feito para a avaliação de apenas um loco, conseguimos comprovar a existência do aumento da endogamia com o passar das gerações de autofecundação utilizando 1206 locos gênicos. Consequentemente podemos destacar a importância da simulação em estudos genéticos, sobre vários contextos, incluindo estudos de populações, do indivíduo ou do próprio genoma.

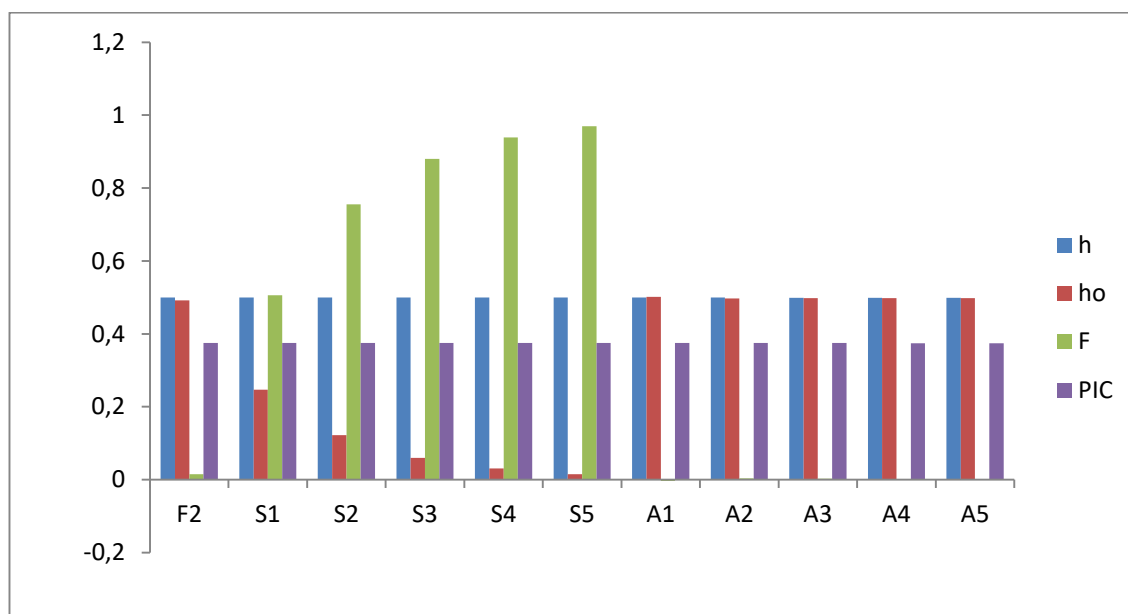
Tabela 1: Heterozigose esperada sob a hipótese de Equilíbrio de Hardy –Weinberg (He); Heterozigose observada (Ho), coeficiente de endogamia (F) e PIC para gerações avançadas por autofecundação (S_j) e acasalamento ao acaso (A_j).

Geração	He	Ho	F	PIC
F ₂	0.499722	0.492135	0.015183	0.374861
S ₁	0.499593	0.246873	0.505852	0.374797
S ₂	0.499528	0.122196	0.755376	0.374764
S ₃	0.499501	0.060074	0.879731	0.374750
S ₄	0.499491	0.030455	0.939028	0.374745
S ₅	0.499479	0.015122	0.969728	0.374739
A ₁	0.499511	0.501131	-0.003243	0.374755
A ₂	0.499426	0.497464	0.003929	0.374713
A ₃	0.499255	0.498162	0.00219	0.374627
A ₄	0.499002	0.498122	0.001765	0.374500
A ₅	0.498754	0.498256	0.001000	0.374375

Estudos realizados por Oliveira et al. (2012) avaliaram a associação entre o nível de endogamia e o potencial produtivo e atributos relacionados à qualidade da raiz de mandioca (*Manihot esculenta Crantz*). Os autores avaliaram 357 acessos de mandioca para seis características de importância econômica para a cultura da mandioca como

produtividade de raízes frescas e de amido, teor de amido, amilose e compostos cianogênicos nas raízes e peso da parte aérea. Concluíram que a maioria dos genótipos avaliados de mandioca (77%) possuía alto nível de endogamia (acima de 0,50) e que através da análise de regressão, constataram que não havia associação entre o alto nível de endogamia com baixo potencial produtivo dos acessos de mandioca.

Figura 1: Heterozigose esperada sob a hipótese de Equilíbrio de Hardy –Weinberg (He); Heterozigose observada (Ho), coeficiente de endogamia (F) e PIC para gerações avançadas por autofecundação (S_j) e acasalamento ao acaso (A_j).



Queiroz et al. (2000) conduziram estudos para avaliar o efeito da endogamia e suas possíveis influências sobre o crescimento de bovinos de corte. Utilizaram 27.406 observações de peso à desmama, padronizado aos 205 dias (PD), e 14.317 informações de peso ao ano, padronizado aos 365 dias (PA), de animais da raça Gir, no período de 1970 a 1995. Concluíram que, apenas 11,65% das vacas, 19,84% dos touros e 9,23% dos bezerros da raça Gir estudados apresentaram algum grau de endogamia, mostrando assim, uma depressão endogâmica para ambas as características (PD e PA). Desta forma, é preciso evitar os acasalamentos, intencionalmente, endogâmicos e monitorar os

acasalamentos com base no parentesco entre os indivíduos, visando manter baixas as taxas de endogamia.

Entretanto, a autofecundação aumenta a homozigose média das plantas e pode acarretar um efeito conhecido como “depressão endogâmica” em espécies alógamas. De acordo com Miranda (2001), a depressão é uma diminuição na expressão de caracteres quantitativos, em decorrência do aumento da homozigose causada pela endogamia. A causa da depressão endogâmica foi associada com a presença de alelos deletérios e letais em genótipos homozigóticos.

Quanto as gerações de acasalamento ao acaso, praticamente não foi detectado endogamia nas populações, indo de valores negativos (-0.003243) a valores positivos (0.00100). Os valores obtidos retratam o que seria esperado em uma população de acasalamento ao acaso. Segundo Cruz (2006), valores negativos de endogamia podem ser considerados valores nulos, ou representativos de excesso de heterozigotos na população desta forma, observa-se que não houve endogamia na geração F₂ e suas derivadas por acasalamento ao acaso, os valores muito baixos de endogamia indicam que o tamanho populacional empregado neste trabalho não foi o suficiente para gerar endogamia significativa. Enfim, não houve endogamia na população nem fixação de formas genotípicas homozigotas e nem alteração nos valores genotípicos conforme consta na Tabela 4 e 5.

Nas populações de autofecundação é esperado, ao longo das gerações, que seu coeficiente de endogamia tenha aumento e, como consequência, sua heterozigosidade seja reduzida a cada geração (Tabela 1). Esse comportamento também foi observado durante as 5 gerações de autofecundação, em que os valores de heterozigosidade observada transitaram de 0.246873 para 0.015122. De maneira geral, nota-se que a simulação foi fiel ao retratar o que seria esperado na estrutura dessas cinco populações ao longo dos cruzamentos.

Almeida et al., (2016), ao avaliar o grau de endogamia entre os progenitores utilizados para obtenção da sua geração de irmãos completos, constatou que não existia endogamia na geração um, pois não havia relação de parentesco entre os progenitores. Por outro lado, em consequência do acasalamento entre os indivíduos aparentados da geração um (irmãos completos), verificaram um aumento no grau de endogamia com o decorrer dos ciclos.

Porto et. al., (2008), ao avaliar o teor de óleo em genótipos autofecundados, determinaram que, em média, é menor do que para materiais de polinização livre. Estes autores destacaram que as reduções do teor de óleo podiam ser explicados pelos efeitos da endogamia nos genótipos autofecundados, sendo semelhante ao que foi verificado por Krieger et al. (2006). Em populações de polinização livre, verificou redução da produtividade em 10% com uma geração de autofecundação.

De acordo com observado na Tabela 1, os marcadores utilizados foram polimórficos, em que apresentaram valores de PIC (conteúdo de informação de polimorfismo) de 0.37, ou seja, superior ao valor de $PIC \geq 0.1$. O valor de PIC se manteve constante nas gerações de autofecundação ou de acasalamento ao acaso, pois tais sistemas de cruzamento não são capazes de alterar a frequência gênica da população, visto que as populações estavam em equilíbrio de Hardy-Weinberg.

Menezes et al., (2006) avaliaram a utilização de 27 microssatélites para caracterização genética das raças caprinas nativas do Brasil. Os autores constataram que todos os microssatélites analisados se mostraram polimórficos para o estudo de caracterização genética das raças nativas brasileiras. A heterozigidade média esperada foi de 0.672 e a observada, de 0.582. Todos os microssatélites analisados foram informativos, à exceção do marcador MAF209 (PIC: 0.042). Os resultados obtidos indicaram a capacidade desses marcadores em identificar a variabilidade genética em raças caprinas nativas brasileiras.

Na Tabela 2 são apresentados os resultados relativos ao equilíbrio da população estudada. As consequências dos sistemas de acasalamento (autofecundação, hibridação e acasalamento ao acaso) já são relatadas na literatura em modelos envolvendo informações de um loco gênico. A caracterização da existência, ou não, de equilíbrio é fundamental para a postulação de hipóteses genéticas e dados, mesmos que simulados, que contemplam estas informações são de grande importância.

De acordo com a Tabela 2, foi preservado no processo de simulação e constatado para um conjunto relativamente grande de locos gênicos (1206) o Equilíbrio de Hardy-Weinberg.

Tabela 2: Número de locos com hipótese de equilíbrio rejeitadas e suas respectivas porcentagens de locos fora do Equilíbrio de Hardy-Weinberg, para cada geração de autofecundação e acasalamento ao acaso.

Gerações	Número de locos rejeitados	Porcentagem
F ₂	130	10.77
S ₁	1206	100
S ₂	1206	100
S ₃	1206	100
S ₄	1206	100
S ₅	1206	100
A ₁	67	5.55
A ₂	70	5.81
A ₃	57	4.72
A ₄	50	4.15
A ₅	58	4.81

Verificam-se, nas Tabela 3 e 4, as médias com base nos valores fenotípicos e genotípicos nas gerações avançadas de uma população F₂ por autofecundação e acasalamento ao acaso utilizando um grau médio de dominância de 0.5 e 1.0 em diferentes cenários. Para este estudo, foram atribuídos valores genotípicos arbitrários para que se pudesse gerar um valor hipotético de média e de variância genotípica, sem significado prático per si, mas de grande importância para acompanhar a dinâmica nas sucessivas

gerações de acasalamento acaso ou autofecundação. Tais informações, geralmente de natureza fenotípica, tem sido objeto de inúmeros estudos na área de melhoramento em que o objetivo comum é a mudança no potencial médio da população sob um montante de variabilidade que permita sempre obter novos ganhos por seleção.

Desta forma, podemos observar com base tanto nos valores fenótipos quanto nos valores genotípicos que os processos de simulação foram capazes de retratar o comportamento genético esperado para a média populacional.

No nosso estudo acompanhamos as mudanças na média e na variância reproduzida pela ação dos 1206 locos cujos arranjos genotípicos foram estabelecidos por processos de simulação. As médias, com o passar dos ciclos de autofecundação, apresentaram decréscimo de suas magnitudes (Tabela 3 e 4), em consequência do estabelecimento de valor de grau médio da dominância parcial utilizada para obtenção dos dados. Quanto aos ciclos de acasalamento ao acaso, a média nas cinco gerações permaneceu constante, como demonstrado.

Uma vez que estabelecemos dominância parcial e dominância completa no controle do caráter, o comportamento da média para a geração de autofecundação está de acordo com o esperado, independentemente do tipo de herdabilidade e distribuição de efeitos de QTL. Podemos observar para a característica com herdabilidade de 0.30 após as sucessivas gerações em S_5 (102.09) é inferior à média da geração F_2 (126.20), isto para o cenário com distribuição uniforme, mas esta redução da média perante aos outros cenários, também ocorreu. Quanto às gerações de acasalamento ao acaso, os valores das médias não houve diferença significativa com relação à geração F_2 . Verifica-se que com herdabilidade de 0.60, a geração A_5 (151.83) não houve diferença significativa para a geração F_2 (150.90), para dominância completa com distribuição binomial. Neste estudo, mostramos que independentemente da distribuição dos efeitos de QTL e herdabilidade utilizado na simulação, quando o valor do heterozigoto diferir do valor médio do

homozigoto, os locos em heterozigosidade irão influenciar nas médias das gerações em autofecundação e nas médias das gerações em acasalamento ao acaso não terá grande influência por estar em Equilíbrio de Hardy-Weinberg (Tabela 3 e 4).

Sabemos que existe variabilidade entre os indivíduos, em todas as amostras ou populações e que a endogamia fixa formas homozigotas e, conseqüentemente, à ampliação da variabilidade genética total na população como pode ser comprovado nos valores obtidos e descritos nas Tabela 3 e 4. Através dos estudos biométricos, comprova-se que a variância genética total em uma população derivada de n gerações de autofecundação é expressa por meio de:

$$\sigma_{Gn}^2 = (1 + F_n)\sigma_A^2 + (1 - F_n^2)\sigma_D^2$$

Desta forma, na geração em que o coeficiente de endogamia é igual a 0, a variância aditiva total é $\sigma_A^2 + \sigma_D^2$, porém como foi considerado no processo de simulação a variância de dominância (σ_D^2) diferente de zero, com isto, a variância genética total vai depender do coeficiente de endogamia (F_n).

Reis et. al. (2009), ao avaliar quais seriam as causas na redução da σ_A^2 em população com acasalamento aleatório, contataram que estava relacionada principalmente com o aumento em F e em uma população sob seleção, esse declínio era em consequência do aumento da endogamia e das mudanças direcionais nas frequências gênicas, o que refletia no aumento da média genética dessa população.

De acordo as Tabelas 3 e 4, com o aumento do coeficiente de endogamia com o passar das gerações de autofecundação, houve aumento no valor da variância genética, sendo o oposto para as gerações de acasalamento ao acaso, por consequência de não ter endogamia.

Tabela 3: Média e Variância dos valores genotípicos e fenotípicos nas gerações avançadas de uma população F₂ por autofecundação e acasalamento ao acaso, para dois caracteres de herdabilidade 30 e 60% com distribuição dos efeitos de QTL uniforme, binomial e exponencial, com grau médio de dominância 0.5.

h ²	Distribuição dos efeitos de QTL	Valores Fenotípicos e Genotípicos											
		F ₂	S ₁	S ₂	S ₃	S ₄	S ₅	A ₁	A ₂	A ₃	A ₄	A ₅	
30	Uniforme	Média	126.20	113.85	107.39	104.29	102.92	102.09	126.41	125.92	125.99	126.62	126.95
		σ_g^2	607.58	669.98	706.91	723.35	729.73	735.18	279.14	209.65	200.01	204.40	204.30
		σ_f^2	2086.50	2165.48	2254.16	2542.49	2383.82	2476.50	932.23	691.55	708.78	680.33	687.50
	Binomial	Média	126.56	113.86	107.70	104.57	103.21	102.30	127.79	126.55	126.87	127.28	127.19
		σ_g^2	1620.72	1814.42	1941.42	1971.79	2023.11	2032.57	774.63	657.06	657.06	623.57	604.62
		σ_f^2	5365.41	5901.43	6395.18	6764.59	6848.11	6573.47	2560.12	2139.81	2139.81	2034.03	1986.80
	Exponencial	Média	124.45	111.50	104.09	101.03	99.66	98.72	124.18	124.80	124.44	124.88	126.25
		σ_g^2	3275.25	3651.65	3893.01	3990.31	4080.93	4105.83	1855.02	1477.64	1359.05	1393.15	1421.20
		σ_f^2	10813.19	12585.14	13422.77	12950.95	13125.72	13794.51	6431.74	4901.43	4564.56	4587.17	4588.55
60	Uniforme	Média	126.20	113.86	107.39	104.29	102.92	102.09	126.41	125.92	125.99	126.62	126.95
		σ_g^2	607.58	669.98	706.91	723.35	729.73	735.18	279.14	209.65	200.01	204.40	204.30
		σ_f^2	1028.16	1123.01	1158.94	1164.77	1226.96	1213.72	449.93	341.63	324.74	324.10	347.24
	Binomial	Média	126.56	113.86	107.70	104.57	103.21	102.30	127.79	126.55	126.87	127.28	127.19
		σ_g^2	1620.72	1814.42	1941.42	1971.79	2023.11	2032.57	774.63	657.06	657.06	623.57	604.62
		σ_f^2	2796.23	2870.22	3281.28	3439.94	3430.29	3508.01	1294.96	1084.78	1111.17	986.11	1009.82
	Exponencial	Média	124.45	111.50	104.09	101.03	99.66	98.72	124.18	124.80	124.44	124.88	126.25
		σ_g^2	3275.25	3651.65	3893.01	3990.31	4080.93	4105.83	1855.02	1477.64	1359.05	1393.15	1421.20
		σ_f^2	5367.24	6150.23	6727.13	6830.68	6541.42	6630.26	3221.67	2360.29	2278.07	2276.16	2470.85

Tabela 4: Média e Variância dos valores genotípicos e fenotípicos nas gerações avançadas de uma população F₂ por autofecundação e acasalamento ao acaso, para dois caracteres de herdabilidade 30 e 60% com distribuição dos efeitos de QTL uniforme, binomial e exponencial, com grau médio de dominância 1.0.

h ²	Distribuição dos efeitos de QTL	Valores Fenotípicos e Genotípicos											
		F ₂	S ₁	S ₂	S ₃	S ₄	S ₅	A ₁	A ₂	A ₃	A ₄	A ₅	
30	Uniforme	Média	150.83	126.15	113.45	107.27	104.40	102.85	151.63	150.83	150.85	151.38	151.38
		σ_g^2	749.76	737.59	740.28	739.67	737.64	740.15	343.90	266.34	260.34	266.27	266.27
		σ_f^2	2456.04	2441.18	2449.95	2390.39	2420.60	2377.08	1161.20	832.96	829.79	861.55	861.55
	Binomial	Média	150.90	126.12	113.62	107.36	104.61	102.99	153.06	151.33	151.71	152.05	151.83
		σ_g^2	2033.49	2010.79	113.62	2012.63	2052.49	2050.46	936.18	841.34	835.54	813.28	793.87
		σ_f^2	7138.12	6690.95	2038.78	6733.17	6729.99	6528.01	3198.60	2766.51	2705.89	2662.49	2586.88
	Exponencial	Média	148.92	123.56	109.97	104.14	101.07	99.48	149.55	150.47	149.08	149.29	151.05
		σ_g^2	4176.99	4114.06	4139.76	4141.78	4155.59	4155.20	2390.44	1929.25	805.50	1866.38	1876.81
		σ_f^2	14094.03	13644.21	13880.12	13816.71	14129.10	14325.17	8116.45	6603.62	6059.37	6013.41	6448.46
60	Uniforme	Média	150.83	126.15	113.45	107.27	104.40	102.85	151.63	150.83	150.85	151.38	151.38
		σ_g^2	749.76	737.59	740.28	739.67	737.64	740.15	343.90	266.34	260.34	266.27	261.27
		σ_f^2	1244.99	1243.04	1232.52	1232.52	1217.33	1258.17	536.61	447.79	452.45	468.37	427.10
	Binomial	Média	150.90	126.12	113.62	107.36	104.61	102.99	153.06	151.33	151.71	152.05	151.83
		σ_g^2	2033.49	2010.79	113.62	2012.63	2052.49	2050.46	936.18	841.34	835.54	813.28	793.87
		σ_f^2	3357.86	3276.13	3587.38	3314.38	3431.66	3493.11	1548.97	1385.22	1394.96	1351.41	1312.67
	Exponencial	Média	148.92	123.56	109.97	104.14	101.07	99.48	149.55	150.47	149.08	149.29	151.05
		σ_g^2	4176.99	4114.06	4139.76	4141.78	4155.59	4155.20	2390.44	1929.25	805.50	1866.38	1876.81
		σ_f^2	7156.14	6598.52	6974.24	6900.04	7379.30	6813.07	3814.07	3280.65	2800.30	3088.56	3065.89

Teste de normalidade das populações geradas

Um padrão que deve ser observado, nas gerações avançadas de autofecundação e acasalamento ao acaso, é a distribuição dos dados de características simuladas considerando ação poligênica e distribuição normal. Quando avaliamos a normalidade dos fenótipos (Tabela 5) através do teste de Lilliefors, observa-se que estes possuem uma aproximação razoável da distribuição normal com curtose e simetria, em todas as gerações de acasalamento ao acaso e autofecundação. Os valores da simetria foram estatisticamente iguais a zero, consistente com a simetria da distribuição normal. Valores abaixo de zero indicam assimetria da curva. Os valores de curtose variam entre 3.317 a 2.400. Independentemente da herdabilidade a curvatura e simetria foram conservadas e próximas da distribuição normal, estes resultados foram obtidos tanto para as gerações de autofecundação, quanto para as gerações de acasalamento ao acaso.

Sanchez (2013) realizou uma verificação de normalidade utilizando o teste de Lilliefors em suas populações, as quais eram compostas de dez gerações de acasalamento ao acaso e dez gerações de autofecundação sob ausência de dominância. O autor observou um padrão de distribuição consistente normal, com curtose (2.6218 ± 0.6) e simetria (-0.1223 ± 0.1) em todas as gerações de autofecundação e acasalamento ao acaso, utilizando diferentes características de herdabilidade.

Almeida et al., (2016) trabalhou com duas gerações, sendo uma constituída por irmãos completos e a outra geração foi confeccionada através do acasalamento ao acaso dos irmãos completos dentro de cada família. Obteve valores de curtose entre 2.87 e 2.97. E sua simetria foi estatisticamente igual à zero, consistente com a distribuição normal.

Tabela 5: Valores de Curtose e simetria populacional das gerações avançadas de uma população F_1 por acasalamento ao acaso e autofecundação, para dois caracteres de herdabilidade 30 e 60% com distribuição dos efeitos de QTL uniforme, binomial e exponencial.

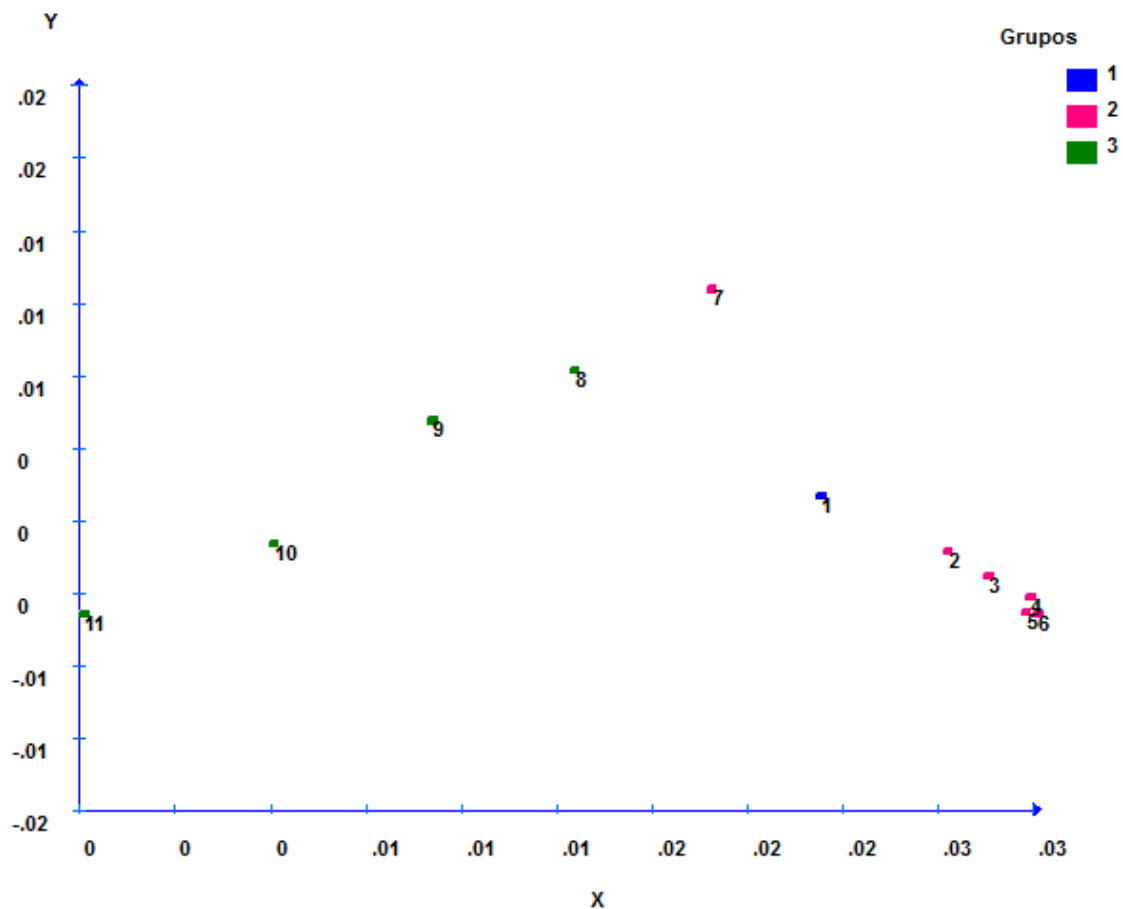
h^2	Distribuição		Valores Fenotípicos										
	dos efeitos de QTL		F2	S ₁	S ₂	S ₃	S ₄	S ₅	A ₁	A ₂	A ₃	A ₄	A ₅
30	Uniforme	curtose	2.995	2.812	3.152	2.695	2.691	2.973	2.995	2.843	2.926	3.122	2.835
		simetria	-0.041	0.094	-0.055	-0.067	-0.048	0.007	-0.021	-0.189	-0.128	-0.112	0.033
	Binomial	curtose	3.195	2.870	2.911	2.731	2.880	2.720	2.917	3.086	2.918	2.941	3.034
		simetria	-0.173	-0.106	0.019	-0.026	-0.031	-0.039	-0.124	-0.108	0.101	0.065	-0.007
	Exponencial	curtose	2.953	3.011	2.699	2.903	2.896	2.886	2.883	3.220	2.873	2.888	3.112
		simetria	0.001	-0.149	0.076	0.030	0.033	-0.017	-0.071	0.082	-0.157	-0.098	-0.154
60	Uniforme	curtose	3.183	3.141	2.910	2.908	3.151	3.050	2.850	2.946	3.073	3.380	2.857
		simetria	-0.070	-0.043	-0.031	-0.115	-0.011	-0.182	-0.015	-0.025	-0.037	0.025	-0.175
	Binomial	curtose	2.956	3.164	2.913	2.908	2.750	2.928	2.926	3.047	2.921	2.976	3.161
		simetria	-0.274	-0.085	-0.107	-0.035	-0.016	0.036	-0.102	-0.282	-0.283	-0.123	-0.244
	Exponencial	curtose	2.732	2.517	2.461	2.504	2.401	2.459	3.055	3.125	2.997	2.646	2.905
		simetria	-0.359	-0.059	-0.067	-0.060	-0.096	0.049	-0.309	-0.293	-0.256	-0.214	-0.267

Diversidade entre e dentro de populações

Na Figura 2 temos a Projeção da dissimilaridade entre as 11 populações, expressa pela distância de Hedrick, no espaço bidimensional, com isso podemos verificar que este procedimento foi eficaz na distinção das populações de autofecundação, através da dispersão das populações em relação aos eixos mas, como esperado, não há diferença entre as populações derivadas por acasalamento ao acaso quando se considera as diferenças nas frequências genotípicas. Resultado diferente foi observado na Figura 3, onde a dissimilaridade entre as populações foi expressa pela distância euclidiana. A distância variou de 0 a 0.03 indicando que, por esta medida, não há possibilidade de distinção das populações fato também esperado tendo em vista que os sistemas de cruzamentos avaliados não são capazes de mudar a frequência alélica das populações. Verifica-se que há diferentes formas de quantificar e realçar as diferenças genéticas entre populações e, portanto, o pesquisador deve estar atento às possíveis causas da diferenciação e adotar o procedimento biométrico apropriado que seja capaz de detectar estas diferenças.

Loiola et. al., (2012), ao avaliar a divergência genética entre os indivíduos da população de coqueiro-gigante-do-Brasil-da-Praia-do-Forte, determinaram uma baixa diversidade genética dentro da população alógama. Foram analisados 150 indivíduos, em que a divergência genética foi calculada através da distância Euclidiana média. Gonçalves et al., (2014), ao avaliar a diversidade genética de acessos tradicionais de feijoeiros via características das sementes de 65 acessos do Banco Ativo de Germoplasma de *Phaseolus* da Unemat, realizaram as avaliações baseado na distância Euclidiana. Em que concluíram que, os acessos avaliados apresentam ampla variabilidade genética e que alguns são promissores a serem utilizados em futuros trabalhos de melhoramento. E os métodos de agrupamento (UPGMA e Tocher) foram parcialmente concordantes, onde os acessos que apresentaram similaridade foram ordenados no mesmo grupo.

Figura 3: Projeção da dissimilaridade entre as 11 populações, expressa pela distância euclidiana, no espaço bidimensional. Grupos: 1 (população F₂); 2 (populações de autofecundação codificadas de 2 a 6); 3 (populações de acasalamento ao acaso codificadas de 7 a 12).

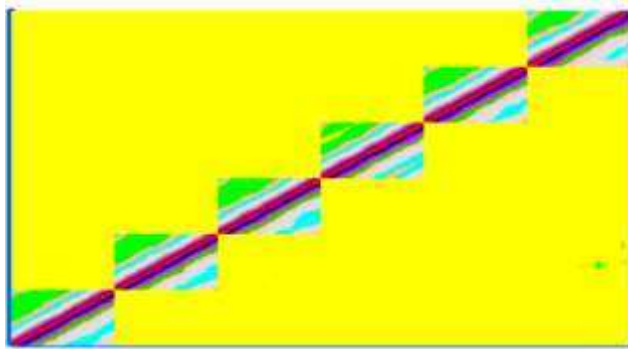


Desequilíbrio de ligação

No presente trabalho podemos visualizar, através das Figuras 4, 5 e 6, o LD nas gerações F₂, e nos conjuntos de populações de autofecundação e de acasalamento ao acaso, respectivamente.

Na Figura 4 temos a geração F₂, e observamos que o desequilíbrio de ligação que ocorre, que por sua vez é proporcionado pela ligação fatorial com a formação de seis blocos de desequilíbrio que corresponde a seis grupos de ligação simulados para o genoma. A intensidade das cores nos diferentes grupos de ligação, variando de mais clara a mais escura, demonstra também a intensidade do desequilíbrio de ligação. Para todo o desequilíbrio de ligação atribuído a essa população está relacionado a ligação gênica como pode ser observado há desequilíbrio para todos os seis grupos de ligação.

Figura 4: Desequilíbrio de ligação da geração F₂.



Quanto ao desequilíbrio de ligação nas gerações de acasalamento ao acaso, verificamos que a quebra do desequilíbrio de ligação já ocorre na primeira geração. Podemos comprovar através da análise da Figura 5, que o processo de simulação foi eficiente, pois a geração F₂ foi derivada de uma geração F₁ de pais que eram contrastantes, o que evidencia que todo o desequilíbrio está relacionado com a ligação gênica. A intensidade das cores nos diferentes grupos de ligação, variando de mais clara a mais escura, demonstra também a intensidade do desequilíbrio de ligação, sendo a mais clara a parte em que tem pouco LD, já a parte mais escura é onde tem maior concentração de LD. Desta forma, podemos visualizar a quebra do desequilíbrio de ligação ao longo das gerações e os seis grupos de ligação não podemos visualizá-los ao fim da A₅ (figura 5). Entretanto, o desequilíbrio de ligação é mantido durante as sucessivas gerações de autofecundação, o qual pode ser observado na Figura 6, em que os seis grupos de ligação permaneceram durante as cinco gerações de autofecundação. Quanto à intensidade das cores, houve uma predominância das cores mais claras, indicando desta forma, a queda no desequilíbrio de ligação. Os resultados demonstrados neste trabalho têm grande relevância para os estudos posteriores na Seleção genômica ampla (GWS), pois esta metodologia depende da existência de desequilíbrio de ligação entre os marcadores e os QTLs. De acordo com Daetwyler et al (2007) e Dekkers (2004), quanto maior forem as

estimativas do desequilíbrio de ligação maior será a quantidade de diversidade genética mantida ao longo do tempo e maiores serão os ganhos a longo prazo.

Figura 5: Desequilíbrio de ligação nas gerações de acasalamento ao acaso ($A_1 \dots A_5$).

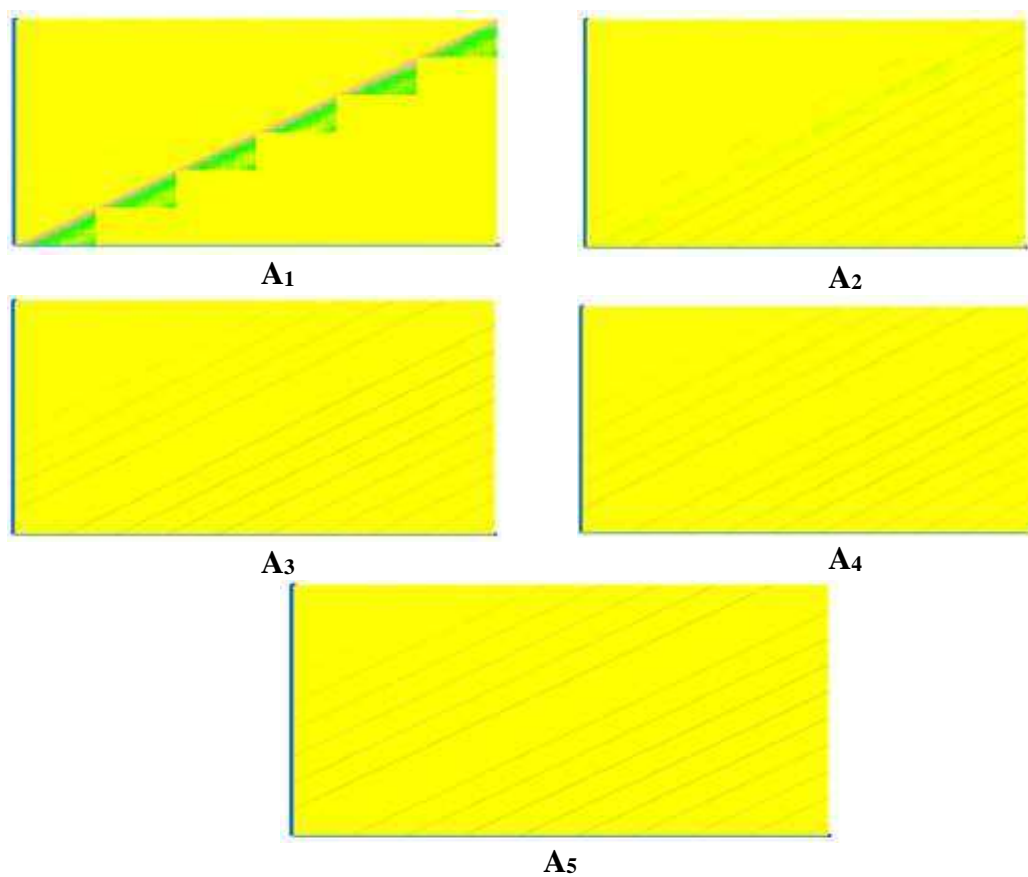
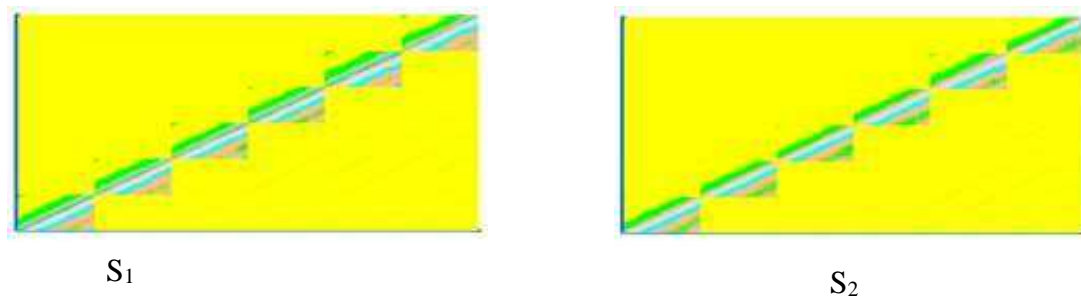
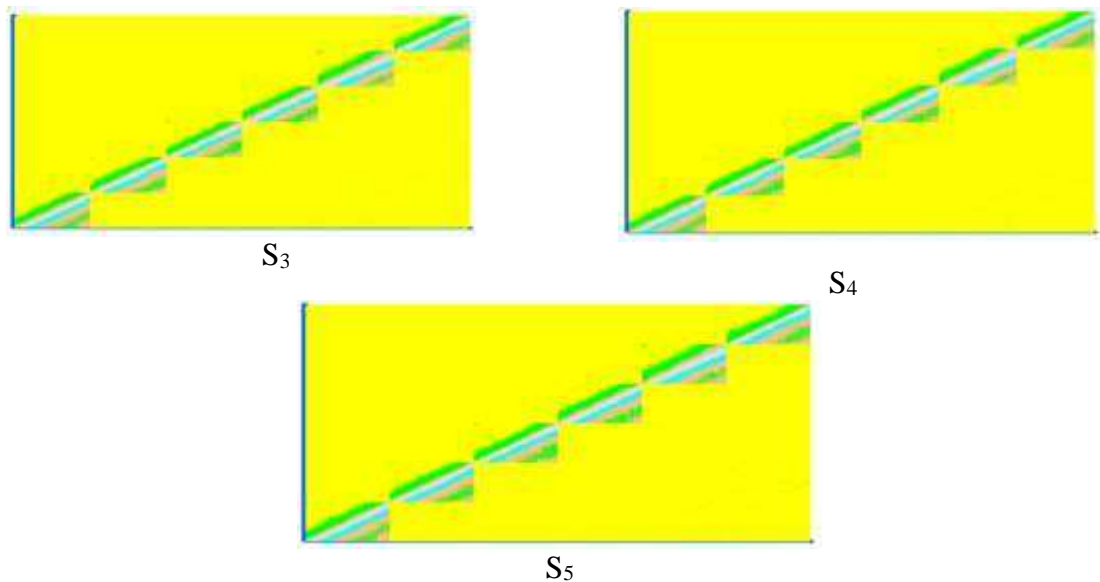


Figura 6: Desequilíbrio de ligação nas gerações de autofecundação ($S_1 \dots S_5$).





CONCLUSÃO

- O processo de simulação biológico foi eficiente em preservar a estrutura genética de populações e descrever as propriedades dos principais parâmetros genéticos dentro da dinâmica estabelecida por sucessivas gerações de acasalamento ao acaso e de autofecundação.
- As populações geradas neste estudo se prestam para estudos genéticos futuros sempre fiéis representante de populações geradas por sistemas de acasalamento ao acaso ou autofecundação.
- A base teórica de genética de populações e quantitativa, em modelos de um único gene, foi devidamente retratada em dados gerados *in silico*, envolvendo mais de mil locos gênicos, e diferentes blocos de ligação.
- O genoma da espécie, o tipo de população e os sistemas de acasalamento afetam de forma diferencial as taxas de desequilíbrio entre pares de locos. Onde destacamos que o LD é mais lenta em autógamias de que em alógamas.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABURAKHMONOV, I. Y; ABDUKARIMOV, A. Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International journal of plant genomics*, 2008.
- ALMEIDA, I. F.; CRUZ, C. D.; RESENDE, M. D. V. Validação e correção de fenótipos na seleção genômica ampla. *PESQUISA AGROPECUÁRIA BRASILEIRA (ONLINE)*, v. 51, p. 1973-1982, 2016.
- BOTSTEIN, D.; WHITE, R.L.; SKOLMICK, H. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics*, v.32, p.314-331, 1980.
- CRUZ, C.D. A informática no melhoramento genético. In: Nass I.I. et al. (Eds.) *Recursos Genéticos e Melhoramento de Plantas*. Rondonópolis: Fundação-MT, 2001. p. 1086-1118.
- CRUZ, C.D. Genes Software – extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum*. v.38, n.4, p.547-552, 2016
- CRUZ, C.D. *Princípios de Genética Quantitativa*. Editora UFV. Viçosa (MG). 2º reimpressão. 394p. 2012.
- CRUZ, C.D. *Programa Genes: Análise multivariada e simulação*. UFV. Viçosa- MG, 2006.
- CRUZ, C.D. *Programa Genes: Diversidade Genética*. UFV. Viçosa- MG, 2008.
- CRUZ, C.D.; CARNEIRO, P. C. S. *Modelos biométricos aplicados ao melhoramento genético (volume 2)*. Viçosa: UFV. 585 p., 2003.
- CRUZ, C.D.; FERREIRA, F.M.; PESSONI, L.A. *Biometria aplicada ao estudo da diversidade genética*. Visconde do Rio Branco: Suprema, 2011. 620p.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. *Modelos biométricos aplicados ao melhoramento genético (volume 1)*. 4ª Ed. Viçosa: UFV. 2012. p. 514.
- DAETWYLER, H.D.; VILLANUEVA, B.; BIJMA, P.; WOOLLIAMS, J.A. Inbreeding in genome-wide selection. *J. Anim Breed Genet*, v.124, p.369-376, 2007.
- DEKKERS, J.C.M. Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science*, v.82, n. 13, p.313-328, 2004.
- FALCONER, D.S. *Introduction to quantitative genetics*. 3rd ed. New York: Longman Scientific and Technical, 1989. 438 p.
- FALCONER, D.S.; MACKAY, T.F.C. *Introduction to quantitative genetics*. 4.ed. Edinburgh : Longman Group Limited, 464p., 1996.
- FLINT-GARCIA, S.A.; THORNSBERRY, J.M.; BUCKLER, E.S. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol*, v.289, n.54, p. 357-374, 2003.

GAUT, B.S.; LONG, A.D. The lowdown on linkage disequilibrium. *The Plant Cell* online, v.15, n.7, p.1502-6, 2003.

GONÇALVES, D. L.; AMBROZIO, V. C.; BARELLI, M. A. A.; NEVES, L. G.; SOBRINHO, S. P.; LUZ, P. B.; SILVA, C. R. Divergência genética de acessos tradicionais de feijoeiros através de características da semente. *Biosci. J.*, Uberlândia, v. 30, n. 6, 2014.

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

GORELICK, R.; LAUBICHLER, M.D. Decomposing multilocus linkage disequilibrium. *Genetics*, V.166, n.3, p. 1581–1583, 2004.

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

GUPTA, P.K.; RUSTGI, S.; KULWAL, P.L. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant molecular biology*, v.57, n.4, p.461-485, 2005.

HEDRICK, P.W. Gametic disequilibrium measures: proceed with caution. *Genetics*, V.117, n. 2, p. 331-341, 1987.

JORDE, L.B. Linkage disequilibrium and the search for complex disease genes. *Genome research*, V.10, n. 10, p. 1435-1444, 2000.

KRIEGER, M.; ZANOTTO, M.D.; MYCZKOWSKI, M.L.; VENCOSKY, R. Efeito da endogamia em progênies de mamona (*Ricinus communis*) In: congresso brasileiro de mamona. Campina Grande: Embrapa Algodão; Embrapa Tabuleiros Costeiros; SAGRI, 2006.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, V.124, n. 3, p.743-756, 1990.

LIANG, K.-Y., HSU, F.-C., BEATY, T.H., BARNES, K.C. Multipoint linkage-disequilibrium-mapping approach based on the case-parent trio design. *Am. J. Hum. Genet.* 68: 937–950, 2001.

LOIOLA, C.M.L.; RAMOS, S.R.R.; ARAGÃO, W.M.; NUNNES, A.O.; NUNES, G. H.S. Divergência genética entre indivíduos da população de coqueiro-gigante-do-Brasil-da-Praia-do-Forte do Banco Internacional de Coco para a América Latina e Caribe. 7º Congresso Brasileiro de Melhoramento de Plantas. Uberlândia- MG.

MALHADO, C.H.M.; CARNEIRO, P.L.S.; PEREIRA, D.G.; MARTINS FILHO, R. Progresso genético e estrutura populacional do rebanho Nelore no Estado da Bahia. *Pesq. agropec. bras.* vol.43 no.9 Brasília Sept. 2008.

MARIZ, T. M. A. Caracterização zoométrica, estrutura populacional e índices reprodutivos da raça Sindi no Brasil. 2010. 121f. Tese (Doutorado em Zootecnia) – Universidade Federal da Paraíba, Areia.

Menezes, M. P. C.; Martinez, A. M.; Ribeiro, M. N.; Filho, E. C. P.; Bermejo, J. V. D. Caracterização genética de raças caprinas nativas brasileiras utilizando-se 27 marcadores microssatélites. R. Bras. Zootec., v.35, n.4, p.1336-1341, 2006.

MENEZES, M. P. C. Variabilidade e relações genéticas entre raças caprinas nativas brasileiras, ibéricas e canárias. 2005. 110f. Tese (Doutorado em Produção Animal) – Universidade Federal da Paraíba, Universidade Federal Rural de Pernambuco, Universidade Federal do Ceará, Aréia.

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

MIRANDA, J.B.F. Endogamia e consanguinidade. In: NASS, L.L. Recursos genéticos e melhoramento: plantas. Rondonópolis: Fundação MT, 2001.p.629-647.

NICHOLAS, F. W. Introdução à genética veterinária. 3. ed. Porto Alegre: Artmed, 2011

OLIVEIRA, E.J., SANTOS, V.S., OLIVEIRA, L.A. Endogamia e características produtivas em germoplasma de mandioca. Congresso Brasileiro de Recursos genéticos. Sociedade Brasileira de Recursos Genéticos, Belém, PA, 2012.

PORTO, M.S., MACEDO, F.C.O., DANTAS, F.V., MILANI, M. Teor de óleo em genótipos de mamona autofecundados. Congresso brasileiro de mamona: energia e ricinoquímica, Salvador, 2008.

QUEIROZ, S.A.; ALBUQUERQUE, L.G.; LANZONI, N.A. Efeito da endogamia sobre características de crescimento de bovinos da raça Gir no Brasil. Revista Brasileira de Zootecnia, v.29, n.4, p.1014-1019, 2000.

QUINTAL, S.S.R. Caracterização e avaliação de um Banco de Germoplasma de mamoeiro para estudos dos parâmetros genéticos e diversidade genética. Campo dos Goytacazes, 2009.168p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, UENF.

REIS, R.L.; MUNIZ, J.A.; SILVA, F.F.; AQUINO, L.H. Estimativas de variância genética aditiva em populações selecionadas e não-selecionadas via simulação Monte Carlo utilizando o software R . Ciênc. agrotec. vol.33 no.1 Lavras, 2009.

RESENDE, M.D.V. Genômica quantitativa e seleção no melhoramento de plantas perenes e animais. Colombo: Embrapa Florestas, p. 330, 2008.

SANCHEZ, C.F.B. Seleção genômica ampla em populações derivadas de acasalamento ao acaso ou de autofecundação. Viçosa: Universidade Federal de Viçosa, 2013. 74 p. (Tese - Doutorado em Genética e Melhoramento de Planta).

I. SIMKO, I; HAYNES, K. G; EWING, E. E.; COSTANZO, S.; CHRIST, B. J; JONES, R. W. Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. Molecular Genetics and Genomics. Vol. 271, n .5, 522–531p.

WEIR, B S; HILL, W G. Nonuniform recombination within the human beta-globin gene cluster. Am J Hum Genet. May; 38(5): 776–781p. 1986

WRIGHT, S. The genetical structure of populations. *Annual Eugen.*, v. 15, p. 323-354, 1951.

CAPÍTULO 2

SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES DERIVADAS DE AUTOFECUNDAÇÃO E DE ACASALAMENTO AO ACASO

Resumo: A seleção genômica ampla (GWS) é preconizada como metodologia que possibilita acelerar e aumentar a eficiência no melhoramento genético, em que a predição genética é um dos grandes desafios em razão das principais características de importância econômica ser de natureza complexa, controlados por um número muito grande de genes de efeito pequeno. Neste contexto, o objetivo deste trabalho foi avaliar a eficiência da GWS na predição do valor genético em gerações avançadas por acasalamento ao acaso e autofecundação com genotipagem utilizando a própria geração e também em gerações imediatamente anteriores. Utilizou-se diferentes cenários relativos a distribuição da importância do efeito de QTL, o efeito da dominância e a influência do ambiente na predição de valores genéticos e estimação de efeitos de marcadores. Para as análises foram utilizados dados simulados, partindo de uma população F_2 e suas gerações avançadas por autofecundação e acasalamento ao acaso. Os resultados demonstraram que o sistema de reprodução afeta a eficiência do uso das estimativas dos efeitos dos marcadores em gerações avançadas. As populações representativas das espécies alógamas, tiveram baixa eficiência no processo de predição, diferente para as espécies autógamas que foi muito eficiente. O efeito de dominância age como um agente perturbador e se tornou mais pronunciado à medida que a herdabilidade da característica foi aumentando. A metodologia RR-BLUP foi eficiente em estimar efeitos e prever valores genéticos independente da existência de genes de pequenos ou grandes efeitos.

Abstract: The broad Genomic Wide Selection (GWS), is advocated as a methodology that allows to accelerate and increase efficiency in genetic improvement, where genetic prediction is one of the great challenges because the main features of economic importance are complex in nature, controlled by a very large number of small-effect genes. In this context, the objective of this work was to evaluate the efficiency of GWS in the prediction of genetic value in advanced generations by random mating and self-fertilization with genotyping using the same generation and also in previous generations. Different scenarios were used for the distribution of the importance of the QTL effect, the effect of dominance and the influence of the environment on the prediction of genetic values and estimation of markers effects. For the analyzes, simulated data were used, starting from an F₂ population and its advanced generations by self-fertilization and random mating. The results demonstrated that the reproduction system affects the efficiency of the use of the estimations of the effects of the markers in advanced generations. The representative populations of the alogamous species had low efficiency in the prediction process, different for the autogamous species, which was very efficient. The dominance effect acts as a disturbing agent and has become more pronounced as the heritability of the trait has increased. The RR-Blup methodology was efficient in estimating effects and predicting genetic values independent of the existence of genes of small or large effects.

INTRODUÇÃO

Ao longo dos anos, as metodologias utilizadas pelos melhoristas para o aumento da produtividade e melhoria das outras características de importância econômica, eram baseadas em informações fenotípicas e, em alguns casos, de genealogia, como forma de prever o valor genético do indivíduo. Com o desenvolvimento dos marcadores moleculares e o avanço em técnicas de biologia molecular, criou-se a expectativa de que as informações genotípicas dos marcadores moleculares, uma vez correlacionados com características fenotípicas de interesse, pudessem ser amplamente utilizadas na obtenção e seleção de indivíduos com maior valor genético (Resende Júnior, 2010).

Um dos principais atrativos da genética molecular em benefício do melhoramento genético é a utilização direta das informações de DNA na seleção, de modo a permitir alta eficiência seletiva, rapidez na obtenção de ganhos genéticos com a seleção e baixo custo (Resende et al., 2008). Com a perspectiva de um aumento nos ganhos de seleção e redução nos ciclos de melhoramento via seleção assistida por marcadores (SAM), muitas pesquisas foram feitas e QTL (locos controladores de características quantitativas) foram detectados e mapeados nas mais variadas culturas (Frery et al., 2000; Yano et al., 2000; Liu et al., 2002).

Entretanto, a metodologia SAM possui limitações que inviabiliza sua utilização, tal como a necessidade de estabelecer uma associação entre os marcadores e os QTL para cada família avaliada e o fato de serem feitas apenas a detecção de um pequeno número de QTL de grande efeito, os quais, devido à natureza poligênica e a alta influência ambiental dos caracteres quantitativos, não explicam suficientemente toda a variação genética, já que esta técnica não é sensível para a detecção de genes de menor efeito (Dekkers, 2004; Desta e Ortiz, 2014).

Com os avanços de tecnologias de genotipagem em larga escala tornou-se possível cobertura completa do genoma, o que levou ao desenvolvimento da seleção genômica

(GS) ou seleção genômica ampla (GWS) proposta por Meuwissen et al. (2001), com uma forma de aumentar a eficiência e acelerar o melhoramento genético, a qual pode ser aplicada em todas as famílias em avaliação nos programas de melhoramento genético de espécies autógamas e alógamas.

A GWS enfatiza a predição simultânea (sem o uso de testes de significância para marcas individuais) dos efeitos genéticos de milhares de marcadores dispersos em todo o genoma de um organismo, de forma a capturar os efeitos de todos os locos de caráter quantitativo (tanto de pequenos quanto de grandes efeitos) e explicar toda a variação genética do caráter (Meuwissen et al., 2001). Desde que haja desequilíbrio de ligação entre os alelos dos marcadores e alelos dos genes que controlam o caráter pode-se prever que haverá ganhos por meio da GWS.

A predição dos efeitos genéticos é realizada com base em dados fenotípicos e genotípicos de indivíduos pertencentes a uma amostra da população de seleção. Esses efeitos genéticos dos marcadores sobre fenótipos de caracteres quantitativos são somados e usados na predição de valores genéticos de indivíduos apenas genotipados, candidatos à seleção em programa de melhoramento genético (Resende et al., 2014)

A GWS permite identificar precocemente os genótipos superiores, sem utilizar métodos muito trabalhosos e que apresentam baixos ganhos com a seleção por unidade de tempo, este método proporciona taxas de ganhos genéticos estimados maiores e mais precisas, isto através da redução do intervalo nas gerações e de uma melhor utilização dos recursos genéticos (Sonesson et al., 2010; Schierenbeck et al., 2011; Pryce et al., 2012). Esses ganhos têm justificado a utilização da GWS em programas de melhoramento animal (Schaeffer, 2006; Goddard & Hayes 2007; Meuwissen et al., 2013) e vegetal (Resende Júnior, 2010; Crossa et al., 2013; Fritsche Neto, 2011; Hayes et al., 2013). No caso do melhoramento vegetal, temos como exemplo os resultados obtidos por Cavalcanti et al. (2012), em cajueiro, que obtiveram 86% de acurácia na seleção sobre o caráter peso de

amêndoas. Heslot et al. (2012), em soja e trigo, também obtiveram elevadas acurácias (83% e 99%, respectivamente).

O presente trabalho teve como objetivo avaliar a eficiência da seleção genômica ampla (GWS) na predição do valor genético em gerações avançadas por autofecundação e acasalamento ao acaso, considerando as particularidades de cada sistema de acasalamento, e verificar os efeitos de dominância e herdabilidade na predição de valores. Como todo o estudo foi fundamentado em dados simulados, constitui-se também como objetivo, constatar a própria eficácia do processo de simulação em gerar populações e gerações cujos os princípios genéticos sejam preservados, conforme a genética quantitativa e de populações.

Material e Métodos

Obtenção das populações

Para a geração dos dados foi utilizado o módulo de simulação do aplicativo computacional GENES (Cruz, 2016). Para realizar o estudo da seleção genômica ampla (GWS) em gerações avançadas de acasalamento ao acaso e autofecundação, realizou-se a simulação de cinco gerações de uma estrutura populacional. Onde, cada estrutura populacional apresenta-se dois conjuntos de informações: genotípica e fenotípica de cada indivíduo dentro da população. Em cada sequência de dados duplicados foram considerados 1206 locos e um tamanho efetivo (N_e) das populações de 1000 indivíduos para todas as populações originadas. Foram simulados genomas parentais e de uma população base (F_2), que foi utilizada como referência para obtenção de cinco gerações avançadas por autofecundação e acasalamento ao acaso. Desta forma, para fins práticos e didáticos as populações analisadas foram denominadas de A_j ($j= 1,2, 3, 4, 5$) para as gerações derivadas por acasalamento ao acaso e representativas das populações alógamas

e S_j ($j= 1, 2, 3, 4, 5$) para a 5 gerações derivadas por autofecundação e representativas de populações autógamas.

Obtenção dos valores genotípicos e fenotípicos

Para processar as análises de GWS foram simulados dados genotípicos e fenotípicos considerando doze caracteres representativos de cenários contemplando três efeitos: (1) relação de dominância: foram estabelecidos valores de dominância parcial ($gmd=d/a=0.5$) e dominância completa ($gmd=d/a=1.0$) sobre o controle da característica. (2) herdabilidade: foram simulados caracteres quantitativos com herdabilidades iguais a 0.30 e 0.60 controlados cada uma por 30 locos, considerando dois alelos por loco e (3) ação gênica: foram simuladas ações diferenciadas do loco associando uma distribuição representativa da importância de cada loco dado por pesos estabelecidos a partir de distribuições uniforme, binomial e exponencial.

Os fenótipos dos indivíduos (i) serão gerados segundo o modelo $F_i = G_i + M_i$, em que G_i é o efeito genético dado pelo somatório dos efeitos genéticos em cada loco e M_i o efeito ambiental, gerado segundo uma distribuição normal com média e variância compatível com a herdabilidade do caráter simulado, dada por:

$$h^2 = \frac{\sigma_g^2}{\sqrt{\sigma_g^2 + \sigma_e^2}}$$

Sendo h^2 a herdabilidade, σ_g^2 é a variância genética e σ_e^2 é a variância ambiental entre os indivíduos.

O Valor genético total expresso por um determinado indivíduo pertencente à população A_j e S_j foi estimado a partir da expressão:

$$G_i = \mu + a_i + d_i$$

Em que:

$$a_i = \sum_{j=1}^{30} p_j \alpha_j$$

$$d_i/a_i = 0,5 \text{ ou } d_i/a_i = 1,0$$

Sendo α_j é o efeito do alelo favorável no loco j , considerando igual a 1, 0 ou -1 para as classes genóticas AA, Aa e aa, respectivamente, e p_j é a contribuição do loco j para a manifestação da característica considerada, no trabalho, como tendo distribuição uniforme, binomial e exponencial (Figura 1). O valor de d_i foi estabelecido a partir do grau médio da dominância manifestado em cada característica.

Sendo que o efeito de dominância ocorre quando os efeitos dos alelos de um determinado loco não são somente aditivos, mais interagem entre si de modo que o valor do genótipo heterozigoto desvia-se a média dos valores dos genótipos homozigotos. Neste caso a e $-a$ são os valores genotípicos dos genótipos homozigotos A_1A_1 e A_2A_2 , assim d é o valor genotípico do genótipo heterozigoto A_1A_2 . Se $d=0$, não há ação dominante no *locos* são puramente decorrentes de efeitos aditivos.

Quanto aos diferentes tipos de distribuição, definimos distribuição uniforme aquela em que é composta por um conjunto de genes, todos eles atuam com a mesma importância, quanto na distribuição binomial existe um conjunto de genes com maior importância do que outros e isto é mais enfatizado na distribuição exponencial.

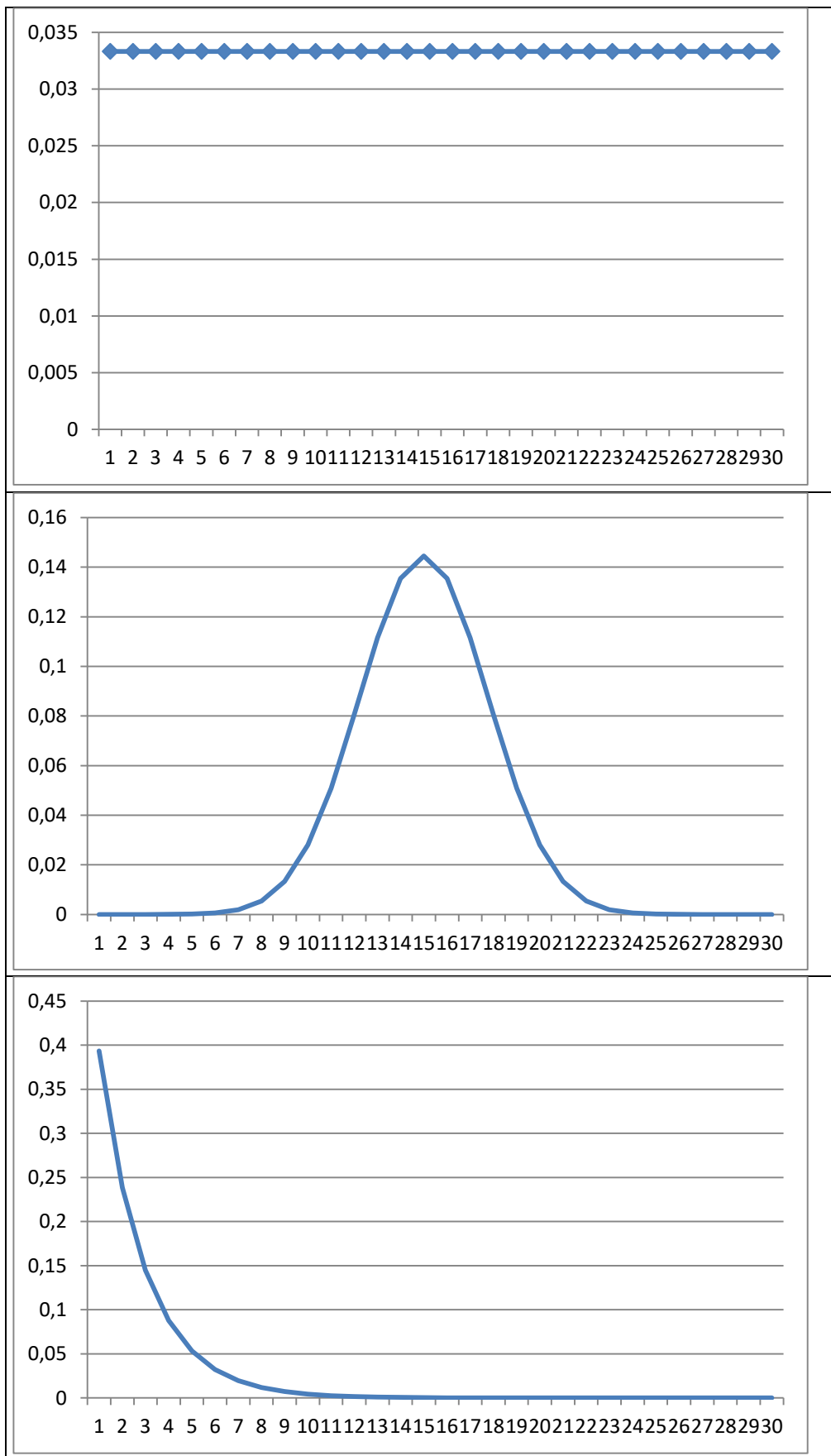


Figura 1: Distribuição uniforme, binomial e exponencial da importância dos efeitos de 30 locos sobre uma característica quantitativa.

Metodologia de análise da GWS

Existem diversos métodos para determinarmos os valores genéticos genômicos (VGG), em geral, produzem resultados semelhantes (Zhang et al., 2011). O melhor método é aquele que consegue refletir de maneira mais acurada a natureza biológica do caráter quantitativo e sua natureza poligênica (Resende et al. 2010).

Quanto aos métodos de aplicação da GWS podem ser divididas entre paramétricas e não paramétricas. Em que os métodos paramétricos está a classe de regressão explícita, como o BLUP (*Best Linear Unbiased Prediction* - Regressão Aleatória) e suas derivações, o LASSO (*Least Absolute Shrinkage and Selection Operator*) e os métodos Bayesianos, como o Bayes A e Bayes B, entre outros. Quanto a classe de regressão implícita está o método RKHS (*Reproducing Kernel Hilbert Space*) e Redes Neurais, já entre os métodos que utilizam redução e dimensionalidade estão as metodologias de componentes principais, componentes independentes e quadrados mínimos parciais (Desta e Ortiz, 2014).

Apesar de inúmeros distinções entres estes métodos mencionados anteriormente, e que são utilizados na GWS, todos enfrentam de maneiras positiva, as principais dificuldades da GWS, que é estimar um grande número de efeitos com base em um número limitado de observações (Wimmer et al. 2013). Mais atualmente, estas dificuldades que limitava a seleção genômica, foram contornadas com a ajuda de técnicas computacionais e o aprimoramento destas metodologias (Fernando et al.,2014).

Desta forma, utilizou-se neste estudo a metodologia RR-BLUP. Esta metodologia é denominada regressão aleatória ou regressão de cumeeira (*Ridge Regression- Best Linear Unbiased Prediction- RR-BLUP*). Os coeficientes de regressão *ridge* são definidos como aqueles que minimizam a soma de quadrados. Esse método foi inicialmente proposto para uso na seleção baseada em marcadores por Whittaker et al. (2000) e foi um dos primeiros propostos para a seleção genômica (Whittaker et al., 2000; Endelman et al.,

2011). Este método, utiliza preditores do tipo BLUP, mas os efeitos dos marcadores são definidos como variáveis explicativas em vez de variáveis classificatórias. Assim, são variáveis regressivas ajustadas como covariáveis de efeitos aleatórios, isto é, os fenótipos são regredidos com base nestas covariáveis. (Meuwissen et al., 2001, Resende et al., 2010; Jannink et al., 2010). Portanto, são preditos os fenótipos a partir da regressão simultânea de grande quantidade de marcas espalhadas por todo o genoma (De Los Campos et al., 2013).

Os efeitos dos marcadores são estimados pelo modelo linear misto geral, conforme Resende et al., (2012).

$$y = Xb + Zh + e,$$

Em que:

- y é o vetor de observações fenotípicas;
- b é o vetor de efeitos fixos (média geral);
- h é o vetor dos efeitos aleatórios dos marcadores;
- e refere-se ao vetor de resíduos aleatórios;
- X e Z são as matrizes de incidência para b e h .

A estrutura de médias e variâncias no modelo em questão é definida como:

$$h \sim N(0, G)$$

$$E(y) = Xb$$

$$e \sim N(0, R = I\sigma_e^2)$$

$$Var(y) = V = ZGZ' + R$$

$$G = \sum_i^N I\sigma_{gi}^2 = I\sigma_g^2/n$$

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I \frac{\sigma_e^2}{(\sigma_g^2/n)} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

De posse dos efeitos de marcadores foram estimados os efeitos dos indivíduos por meio do seguinte estimador:

$$VGG = \hat{y} = \sum_i^n Z_i \hat{h}_i$$

As equações de predição apresentadas acima foram modeladas com σ_g^2 comum. Desta forma, a variância genética explicada por cada loco é dada por σ_g^2/n , em que σ_g^2 é a variância genética total e n é o número de marcadores utilizados em cada um dos locos testados e σ_e^2 é a variância residual. Essa estratégia foi implementada por Meuwissen et al., (2001), Bernardo et al. (2007) e Zhang et al. (2011).

Avaliação dos dados obtidos na GWS

Na GWS é necessário o uso de uma população de estimação para obter os efeitos dos marcadores e uma população de validação, para quantificar a eficiência da estimação destes efeitos na recuperação do valor genômico em uma população independente. Adotou-se a estratégia de considerar uma população com 1000 indivíduos de cada geração a utilizar 80% dos dados para fins de treinamento (ou estimação) e o restante como dados de indivíduos de população de validação. Assim, a fenotipagem e genotipagem dos indivíduos ocorreram na mesma geração. Outra estratégia foi a de usar dados da fenotipagem e genotipagem dos indivíduos da geração anterior e validação em gerações avançadas por autofecundação ou acasalamento ao acaso. Assim, na fase de treinamento houve a predição dos valores genômicos e a estimação dos efeitos dos marcadores que foram utilizados em modelos aplicados na população de validação para prever o valor genômico.

Para determinarmos a eficiência da GWS numa dada população, foi quantificada a acurácia, ou acurácia de predição, obtida pelo quadrado da correlação do valor genético predito com o valor genotípico real (ou simulado) dos indivíduos.

Aplicativos computacionais utilizados para as análises dos dados

Para todas as simulações e análises estatísticas foram implementadas através dos recursos computacionais do programa GENES (Cruz, 2016).

RESULTADOS E DISCUSSÃO

Efeito dos Marcadores

A seleção genômica (GS) foi recentemente proposta como nova estratégia de seleção para comparar com a MAS com base em alguns QTL identificados. É um paradigma inovador no melhoramento de planta, tendo-se adotado extensamente no melhoramento animal (Sun et al., 2014; Coninck et al., 2014).

Nos programas de melhoramento genético, a parte principal para obter ganhos com a seleção é a predição dos valores genéticos. Desta forma, para deduzir o genótipo do QTL, basta tratar as marcas como se fossem o QTL e, assim, estimar os efeitos dos alelos marcadores ou genótipos. Segundo Goodard e Hayes (2007), o parâmetro chave é a proporção da variância do QTL explicada pelas marcas. Conforme já tinha sido estabelecido pela simulação apenas os 30 locos marcadores, tomados aleatoriamente no genoma, estavam associados aos QTLs.

Através das Figuras 2, 3 e 4, foi retratada a eficiência da metodologia de análise, RR-BLUP, em reconhecer na distribuição dos efeitos dos 1206 marcadores estimados, o padrão da distribuição dos 30 locos de marcadores selecionados como associados à característica quantitativa, conforme simulação. Verifica-se que os diferentes cenários de distribuição, onde foi possível demonstrar as diferentes arquiteturas genômica dos marcadores, em que ilustramos os grandes (na distribuição exponencial), médios (na distribuição binomial), pequenos (na distribuição uniforme) e nenhum efeito destes 1206 marcadores. As características quantitativas com distribuição uniforme apresentaram comportamento em que se teve menor variação dos efeitos dos marcadores estimados (-0.1 a 0.3), ou seja, apresentaram efeitos reduzidos de QTL dentro dos 30 locos simulados, por outro lado, a binomial (variação de -0.4 a 0.6) e exponencial (variação de -0.5 a 1.5) apresentaram efeitos variados dentro de cada loco, com alguns QTLs de maior efeito. As estimativas dos efeitos de marcadores foram semelhantes ao tipo de distribuição dos efeitos genéticos (uniforme, binomial e exponencial).

Almeida et al., (2016) demonstrou em suas análises que ao estimar os efeitos dos marcadores, estes foram semelhantes independente do tipo de distribuição dos efeitos genéticos (uniforme ou exponencial), isto para as duas metodologias de análises, RR-Blup e Blasso. No presente trabalho verificou-se que a técnica de estimação é capaz de captar as particularidades dos efeitos dos marcadores estudados proporcionando modelo

que, certamente, será adequado para fins de predição em populações de treinamento ou validação.

Característica de $h^2= 0.30$

Característica de $h^2= 0.60$

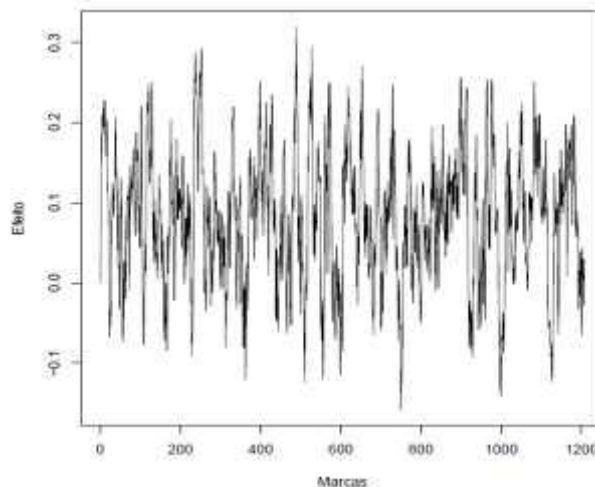
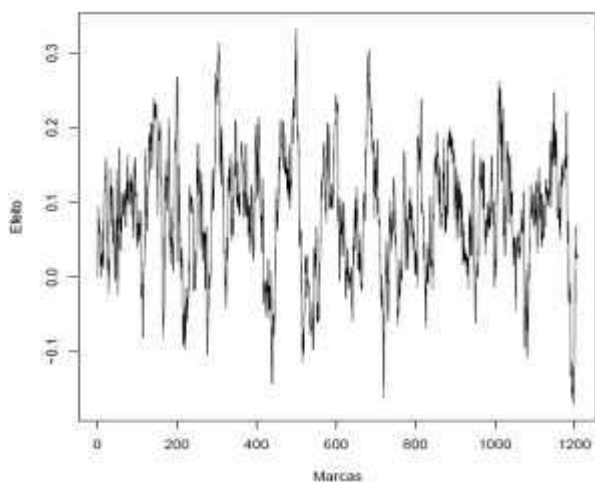


Figura 2: Efeito dos 1206 marcadores, estimado por modelos de RR-BLUP, em característica quantitativa controlada por 30 locos, dispostos aleatoriamente no genoma, cuja importância foi estabelecida com base na distribuição uniforme.

Característica de $h^2= 0.30$

Característica de $h^2= 0.60$

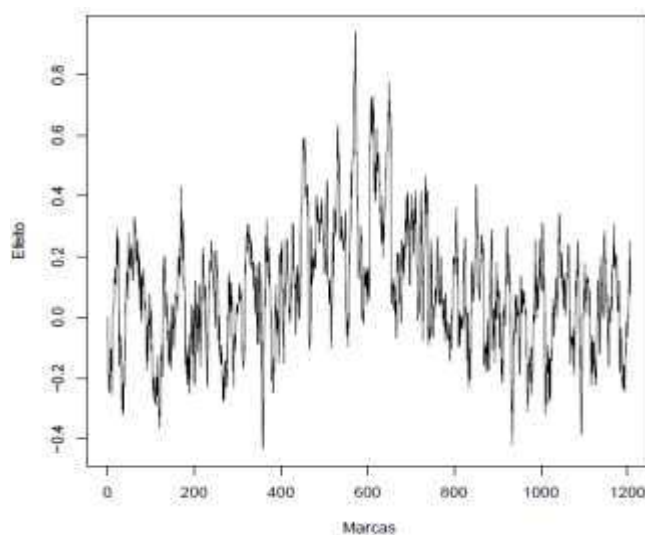
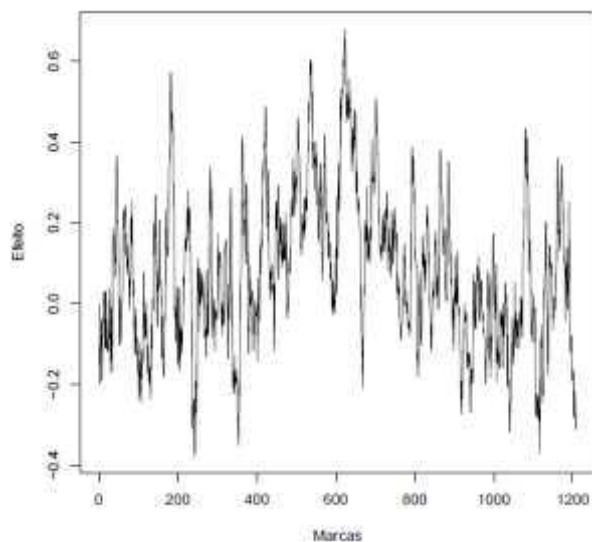


Figura 3: Efeito dos 1206 marcadores, estimado por modelos de RR-BLUP, em característica quantitativa controlada por 30 locos, dispostos aleatoriamente no genoma, cuja importância foi estabelecida com base na distribuição binomial.

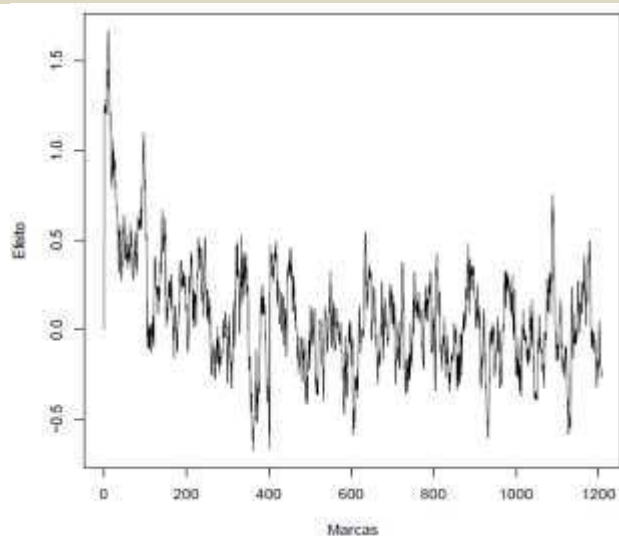
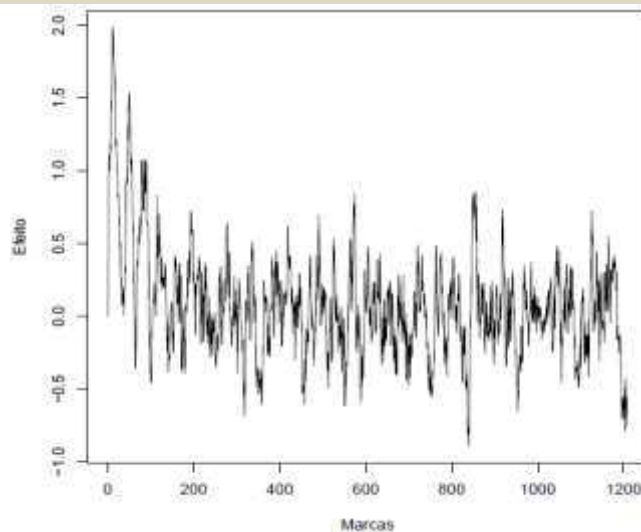
Característica de $h^2= 0.30$ Característica de $h^2= 0.60$ 

Figura 4: Efeito dos 1206 marcadores, estimado por modelos de RR-BLUP, em característica quantitativa controlada por 30 locos, dispostos aleatoriamente no genoma, cuja importância foi estabelecida com base na distribuição exponencial.

Estimativas de acurácia em que a genotipagem ocorreu somente na F_2

Verifica-se por meio da Tabela 1 o quanto a agregação da informação genotípica contribui para o aumento da eficiência da seleção. Considerando inicialmente apenas a geração contemporânea, ou seja, treinamento e validação na própria F_2 , constata-se que os valores obtidos de acurácia sempre superaram a herdabilidade da característica evidenciando os benefícios à prática de seleção, mesmo diante a presença de certos efeitos perturbadores da seleção. Constata-se que características com herdabilidade de 30% possibilitam acurácia de maior que 85% em níveis menores de dominância ($d/a = 0.5$) quando o critério de seleção é o valor genômico estimados a partir de informações resultantes da genotipagem da população.

Conforme Resende et al., 2012, a escolha da metodologia apropriada, viabiliza a seleção com informações de marcadores moleculares antes mesmo da fenotipagem, acelerando o programa de melhoramento, já que em condições normais, o período para tomada de decisão é bem mais longo, isto por ter apenas informações fenotípicas.

Deve-se ter em mente que na agricultura moderna novos desafios têm surgido e os programas de melhoramento genético têm buscado, ao longo dos anos, estratégias eficientes para seleção de genótipos que apresentem desempenho superior. Quando se pratica a seleção com base nos valores fenotípicos dos indivíduos apenas, conforme o melhoramento clássico faz, para distinguir a superioridade de um indivíduo em relação aos outros, tem-se alguns fatores perturbadores tanto genéticos, como a dominância, quanto ambiental. O fator ambiente reduz a associação entre fenótipo e o genótipo tornando a seleção menos efetiva e a magnitude de seu efeito é, geralmente, quantificado pela herdabilidade. Neste trabalho, considerou-se dois níveis de herdabilidade (30 e 60%) e constatou-se que a acurácia seletiva foi ampliada na medida em que se considerou a informação molecular na predição do valor genômico. Desta forma a GWS contribui de forma a melhorar os critérios de seleção acrescentando, aos valores fenotípicos, informações moleculares, vinda de um processo de genotipagem.

Tabela 1: Valores da acurácia da seleção das gerações avançadas por autofecundação (S_1 a S_5) e acasalamento ao acaso (A_1 a A_5) obtidas a partir da fenotipagem e genotipagem da população F_2 para as doze características com herdabilidade iguais a 0.30 e 0.6 repetidos no cenário de grau médio de dominância (gmd) iguais a 0.5 e 1.0, com diferente distribuição dos efeitos de QTL (Uniforme-U, Binomial-B e exponencial-E).

Treinamento	Variável	gmd	F_2	S_1	S_2	S_3	S_4	S_5	A_1	A_2	A_3	A_4	A_5
F_2	U-h= 0.3	0.5	0.87	0.81	0.78	0.76	0.75	0.74	0.57	0.27	0.20	0.22	0.18
F_2	U-h=0.6	0.5	0.90	0.83	0.81	0.79	0.79	0.78	0.58	0.28	0.20	0.22	0.18
F_2	B-h= 0.3	0.5	0.86	0.79	0.75	0.73	0.74	0.72	0.45	0.16	0.11	0.09	0.06
F_2	B-h=0.6	0.5	0.89	0.82	0.78	0.76	0.74	0.74	0.48	0.17	0.13	0.11	0.08
F_2	E-h=0.3	0.5	0.85	0.75	0.71	0.68	0.69	0.66	0.42	0.12	0.07	0.05	0.03
F_2	E-h=0.6	0.5	0.88	0.77	0.73	0.70	0.74	0.68	0.43	0.13	0.05	0.04	0.02
F_2	U-h= 0.3	1	0.70	0.72	0.74	0.74	0.72	0.73	0.47	0.20	0.15	0.15	0.14
F_2	U-h=0.6	1	0.72	0.73	0.74	0.75	0.75	0.74	0.49	0.20	0.14	0.17	0.13
F_2	B-h= 0.3	1	0.67	0.70	0.70	0.69	0.67	0.69	0.34	0.09	0.08	0.07	0.03
F_2	B-h=0.6	1	0.71	0.74	0.73	0.73	0.69	0.73	0.39	0.14	0.10	0.10	0.08
F_2	E-h=0.3	1	0.68	0.66	0.66	0.65	0.65	0.65	0.33	0.11	0.03	0.04	0.03
F_2	E-h=0.6	1	0.72	0.71	0.70	0.69	0.69	0.69	0.35	0.10	0.05	0.05	0.03

De uma forma mais ampla, Bastiaansen et al. (2012) relatam que os principais fatores que são responsáveis por afetarem a acurácia na estimação dos valores genômicos são a herdabilidade da característica, o tamanho da população de referência, o desequilíbrio de ligação entre as marcas e as características quantitativas (QTL) e a distribuição dos efeitos dos QTL. A ação de alguns destes efeitos, tais como herdabilidade e dominância, podem ser apreciados nos resultados apresentados na Tabela 1, em que o comportamento diferenciado das acurácias resultante destes fatores que afetam os seus valores.

Através da Tabela 1, verificou-se que o efeito de dominância tornou-se mais evidente à medida que a herdabilidade da característica aumentava. Observa-se, na F_2 em que a acurácia apresentada foi de 0.85 com herdabilidade de 0.30 com grau de dominância de 0.5, porém com um grau de herdabilidade de 0.60, o valor da acurácia foi de 0.72 com um cenário de dominância igual a 1.

Assim, apesar de não ter sido utilizado neste trabalho, é importante no uso da GWS, o uso de matrizes de incidência apropriadas que também sejam capazes de captar estes efeitos de dominância. Porém, Guimarães (2016), ao realizar estudos de metodologias que incluam não somente os efeitos de dominância como também o efeito da interação dominância por ambiente, concluíram que apesar do ligeiro aumento na predição, sua performance não foi estatisticamente diferente do modelo que incluía apenas os componentes aditivos. Almeida Filho, (2016) a inclusão da informação oriunda do pedigree no modelo genômico, não produziu acurácias mais elevadas quando comparado com os modelos que usaram apenas informações de marcadores.

A Tabela 1 também possibilita concluir sobre a eficácia da predição do valor genético em populações de validação diferente daquela de treinamento em que os efeitos dos marcadores foram estimados. Esta eficácia tem sido relatada em populações relacionadas em que se conserva parte de background genética das populações de

treinamento e de validação. Entretanto, no melhoramento vegetal uma geração descendente pode ser derivada por diferente mecanismo de acasalamento que impactam de diferente intensidade o desequilíbrio de ligação em marcadores e locus QTLs.

De maneira geral é possível observar que o aumento na herdabilidade, causa aumento do valor da acurácia do processo, isto para todos os sistemas de populações estudadas. Obviamente deve-se ter precaução ao interpretar este aumento da acurácia, isto porque, este valor que foi aumentado pela GWS pode não ser compensador tendo em vista a herdabilidade da característica.

Para características com baixa herdabilidade, muitos fenótipos (e genótipos) são necessários para estimar acuradamente os efeitos dos marcadores (Meuwissen et al., 2001; Calus et al., 2008; Goddard, 2009) principalmente quando se objetiva prever gerações avançadas com maior número de ciclos de acasalamento. Segundo Resende et al, (2008) a GWS é excelente para caracteres de baixa herdabilidade, diferente da MAS (seleção assistida por marcadores), que não é útil para características de baixa herdabilidade. Simulações comparando GWS e MAS, em programa de melhoramento de milho (*Zea mays L.*), identificaram maior eficiência de GWS em elevar a resposta à seleção, principalmente para caracteres de baixa herdabilidade (Bernardo & Yu, 2007). O interessante é lembrar que o importante é ter valor de acurácia, por meio de valores genômicos, que sejam superiores a herdabilidade estabelecida com os valores fenotípicos observados.

Na Tabela 1, evidência que agregar informações moleculares não iria conduzir a um aumento significativo da acurácia, isto, tendo em vista a herdabilidade da característica, ao considerar a prática de seleção em sistemas de acasalamento ao acaso. Ao contrário, quando se avalia gerações derivadas do sistema de acasalamento típico das espécies autógamas (S_1 a S_5) observamos resultados que evidenciam uma alta eficiência no processo de predição dos valores genômicos. Constatamos, que para as doze

características avaliadas, para as gerações de autofecundação apresentaram valores acima da herdabilidade previamente estabelecida da característica observada. Assim, tal como esperado tendo em vista a preservação do desequilíbrio gamético, a GWS proporciona resultados mais vantajosos nas espécies autógamas, principalmente devido a redução mais lenta deste valor de desequilíbrio de ligação entre os locos controladores das características. Quando avaliamos as populações de acasalamento ao acaso (A_1 a A_5), observa-se que o processo de predição não foi eficiente, isto para todas as características avaliadas, demonstrando que os valores genéticos genômicos apresentam uma maior relação com os valores fenotípicos, sendo assim, podendo inferir que as informações moleculares podem não ter grandes vantagens considerando apenas este contexto de seleção e ganho em uma única característica.

De acordo com Bernardo e Yu (2007) ao compararem a seleção genômica em diferentes cenários de herdabilidade também verificaram que a acurácia decresceu com menores herdabilidade, assim como Almeida (2013) ao avaliar três diferentes valores de herdabilidade. Desta forma, concluímos que caracteres com baixa herdabilidade são um problema tanto na GWS quanto na seleção fenotípica, pois ambos levam a um decréscimo na acurácia.

Zhong et al. 2009 examinaram os efeitos de diferentes parâmetros na acurácia da seleção genômica e verificaram que dobrar o tamanho da população de estimação aumentaria os valores de acurácia. Estudos com bovinos têm indicado que caracteres com baixa herdabilidade requerem grandes populações de estimação para se atingir altas acurácias (Hayes et al.,2009). Segundo Misztal (2011), em gado leiteiro, com grande número de progênies, é necessário cerca de 2000 genótipos para alcançar um aumento significativo na acurácia. Número maior seria necessário se grupos de progênie são pequenos e herdabilidades são baixas.

Estimativas de acurácia com o uso da genotipagem defasada no tempo ou contemporânea

A predição genômica tornou-se bastante útil no auxílio ao melhoramento vegetal (Riedelsheimer et al., 2013; Technow et al., 2014 e Daetwyler et al., 2015) quanto no melhoramento animal (Sun et al., 2014 e Coninck et al., 2014). Desta forma, inclusão de efeitos não aditivos, é uma alternativa para aumentar a capacidade de predição. De acordo com Da et al., (2014) a estimativa dos efeitos de dominância e seu desempenho frente aos diferentes ambiente de avaliação é uma alternativa promissora, isto porque, a inclusão dessas informações promoveria uma melhoria na capacidade de predição dos modelos selecionados, em razão de determinar com maior segurança a contribuição genética para o fenótipo e minimiza o problema da herdabilidade perdida. Portanto, demonstra-se nas Tabelas (2, 3, 4, 5, 6 e 7), o quanto o efeito de dominância teve influência nos valores de acurácia e ilustra-se o quanto é um agente perturbador, na predição de valores genômicos, o que se tornou mais nítido, à medida que a herdabilidade aumentava, isto em todos os cenários utilizados e nas diferentes espécies avaliadas neste trabalho.

Nos cenários em que a dominância era de 0.5 (Tabela 2, 3 e 4), ao considerarmos as populações representativas das espécies alógamas, os valores de acurácia são sempre superiores a herdabilidade estabelecida com os valores genotípicos analisáveis. Valores estes que variaram de 0.79 a 0.35, em que observamos pequenas variações ao longo das gerações, algumas vezes com acréscimo, em que destacamos que os valores não foram tão afetados pela redução do desequilíbrio de ligação, já que apesar dele ser quebrado após a primeira geração de acasalamento ao acaso, sabemos que ainda existe resquícios de desequilíbrio nas sucessivas gerações. Segundo Sturion e Resende, (2005), quanto maior os valores de acurácia na avaliação de um indivíduo, maior será a confiança na avaliação e no valor genético predito do indivíduo. Assim como esperado, à medida que aumentamos a herdabilidade, os valores de acurácia também foram crescendo,

demonstrando que o processo foi bem-sucedido, isto tanto para o treinamento e validação da população na sua própria geração (diagonal), quanto o treinamento nas gerações anteriores e validação na geração atual (leitura na horizontal). Ressaltamos que os resultados obtidos para as características de baixa herdabilidade ($h^2=0.30$), foram em torno de 65%, isto para todas as populações treinadas e validadas em suas próprias gerações, em quanto que para as características de maior herdabilidade ($h^2=0.60$), os valores obtidos foram em torno de 79%. Portanto, destacamos a eficiência na genotipagem contemporânea (valores na diagonal), a qual proporcionaria ao pesquisador maiores ganhos, ao aplicar recursos para fins de obtenção de informações moleculares dos indivíduos sob seleção. Contudo, ao avaliarmos a predição dos valores genéticos obtidos através do treinamento na F_2 , foram inferiores a predição, realizada pela população em gerações imediatamente anteriores. Tomando como base, a herdabilidade de 0.60 no cenário com distribuição exponencial, na população F_2 teve valores que variaram de (0.43 a 0.02), em quanto que nas gerações imediatamente anteriores variaram de (0.67 a 0.59).

Bernardo e Yu (2007), ilustraram através de simulação o potencial e as perspectivas da utilização da GWS para características quantitativas do milho. Estes autores verificaram ao simular a performance test-cross de duplos haploides, utilizando três ciclos de seleção, isto baseado em informações de marcadores, em que foi usado 20, 40 e 100 QTLs que controlavam as características quantitativas, isto sob diferente herdabilidade, concluíram que a resposta a seleção foi de 18 % a 43% maior via GWS, que via seleção recorrente assistida por marcadores moleculares. Devemos ter em mente que dobrar o valor de uma herdabilidade via técnicas convencionais não é tão fácil, pois devido ao custo elevado que será para ampliar a variabilidade genética, isto através do aumento do número de genótipos para avaliação, redução da influência ambiental que faz através de delineamentos, e por fim, controles de fatores bióticos e abióticos.

Para as espécies autógamas a medida que as gerações avançam, verificamos que os valores de acurácia se apresentaram constates em todos os cenários simulados, sendo observados valores nas gerações S₁ e S₅ (0.81±0.00; 0.79 ±0.02; 0.77±0.04;) para herdabilidade de 0.30, já para herdabilidade de 0.60 (0.87 ±0.01; 0.86±0.00; 0.85±0.03;) Desta forma, destacamos que as populações representativas de autógamas quando treinamos e validamos em suas próprias gerações os valores obtidos de acurácia substancialmente foram alto.

De maneira semelhante ao verificado com as populações representativas de espécies alógamas, ao utilizar informações fenotípicas e genotípicas das gerações anteriores, observamos valores superiores em relação a suas herdabilidade. Entretanto, uma análise mais precisa, indica que a GWS proporciona resultados mais vantajosos em espécies autógamas que as espécies alógamas e um dos motivos é devido ao LD em locos controladores das características quantitativas, pois a sua redução ao longo das gerações é muito mais lenta. Assim, de acordo com Lewontin, (1988), os fatores evolutivos, como o sistema de cruzamento, seleção, mutação, deriva, migração, ou fatores como coancestralidade podem ocasionar a associação não aleatória dos alelos nos indivíduos, ocasionando uma correlação entre eles na formação dos haplótipos, interferindo na dinâmica genética das populações e conseqüentemente na dinâmica equilíbrio-desequilíbrio de ligação. Fatores que podemos demonstrar, através das populações simuladas de acasalamento ao acaso e autofecundação o quanto são distintas quanto aos valores de acurácia predita. Concluímos que as populações de autógamas indicam que a inclusão de informações moleculares associadas as técnicas biométricas propicia grandes informações aos pesquisadores, no processo de seleção (Tabela 2, 3 e 4).

Nos cenários em que a dominância era de 1.0 (Tabela 5, 6 e 7), verificamos o quanto o efeito de dominância agiu como um agente perturbador, pois ao compararmos os valores de acurácia obtidos quando utilizávamos dominância incompleta, os valores foram

inferiores. Observa-se que o efeito de dominância se tornou mais acentuado à medida que a herdabilidade da característica aumentava, isto para todos os cenários de distribuição dos efeitos dos genes. Ilustramos como exemplo mais crítico a acurácia na F_2 , que foi de 0.70 com herdabilidade de 0.30 no cenário com grau médio de dominância de 0.5 e distribuição uniforme, no entanto com dominância completa (1.0) foi de 0.90 com herdabilidade de 0.60. Desta forma, verificamos que é preciso o uso de modelagens que incluem os efeitos de dominância, para fins de utilização no melhoramento pela GWS. De acordo com Ertl et al., 2014, até recentemente eram escassos os estudos sobre os desvios de dominância, isto porque não tinha informações genômicas disponíveis, além da indisponibilidade de grandes conjuntos de dados, que tinham número suficiente de indivíduos com níveis de dominância não nulos, como exemplo família de irmãos completos. Segundo Mistztal et al. (1998) a razão de poucos estudos que visa identificar a real contribuição do efeito de dominância está também geralmente associada ao nível de complexidade computacional usado nestas análises e também a imprecisão nas estimativas dos componentes de variância. Adicionalmente seu efeito e a variância proporcionada pela dominância são negligenciadas por não possuir aplicabilidade para prever a resposta a seleção. Entretanto, Varona et al., (1998) ressaltam que esta variância representa cerca de 10% da variação fenotípica total.

Guimarães (2016), ao incluir efeitos de dominância e seu respectivo efeito de interação com o ambiente utilizando modelos de G-BLUP, verificou que há ligeiro incremento na habilidade de predição. O autor utilizou dados simulados e dados reais, e verificou que os modelos utilizados não apresentaram diferenças acentuadas. Segundo Mistztal et al. (1998) uma explicação para isto, é devido o componente aditivo do modelo usado, não ser capaz de capturar parte do efeito de dominância, sendo assim, modelos com apenas componentes aditivo não seria penalizado pela omissão do efeito de dominância.

Enfim, com a utilização de modelos que incluem apenas efeitos de dominância já seria um grande ganho na predição de valores, e com estes modelos poderíamos utilizar para fornecer informações necessárias na condução de cruzamentos visando, a maximização de ganhos com efeitos de dominância. Verificamos também, o quanto a ação do ambiente, agiu como um agente perturbador, no processo de obtenção dos valores de acurácia, em que os melhores resultados, para fins de recomendação de uso da GWS, está relacionado a situações de baixa herdabilidade e característica, onde o efeito de dominância seja o mínimo possível.

Portanto, tanto as populações representativas das espécies alógamas ou autógamas a inclusão de informações a partir de genotipagem e fenotipagem na própria geração irá propiciar acréscimo no valor da confiabilidade e, conseqüentemente, do ganho a ser obtido pela seleção. Quando usamos a genotipagem nas populações imediatamente anteriores, verificamos que as populações representativas de alógamas, eram inferiores as populações treinadas na população F_2 e ainda eram inferiores as herdabilidade apresentada pela característica usada no treinamento, desta forma, indicam que não teremos ganhos com o processo de genotipagem nestas populações.

Quanto a arquitetura genética, distribuição uniforme, binomial e exponencial, não influenciou de modo significativo na diferenciação dos valores de acurácia, desta forma, os valores para cada distribuição de efeitos genéticos foram semelhantes dentro de cada herdabilidade, a qual realizamos os estudos (Tabela 2, 3, 4, 5, 6 e 7). Sob o aspecto da genética quantitativa, em que o caráter quantitativo é regulado por muitos genes este de efeitos semelhantes e de efeitos pequenos, este conceito clássico de genética, foi demonstrado pela distribuição uniforme, com a inclusão da distribuição binomial e exponencial, podemos ilustrar que existia genes de efeitos maiores, conforme a genética genômica preconiza. Aparentemente sob o aspecto de predição de valores genômicos, não existe diferença nenhuma, tanto para o conceito clássico de genética, quanto para o

genômico, mesmo quando imaginamos muitos genes de efeitos pequenos, ou poucos de efeitos maiores os resultados são semelhantes. Para a herdabilidade de 0.30 a acurácia média para os três cenários foi em torno de 50% e para a herdabilidade de 0.60 foi de 60%, isto para as populações de espécies alógamas onde a predição era em gerações imediatamente anteriores (Tabela 2, 3 e 4). Enquanto nas Tabelas (5, 6, 7) a herdabilidade de 0.30 a acurácia média para os três cenários foi em torno de 70% e para a herdabilidade de 0.60 foi de 80%, isto para as populações de espécies autógamas onde a predição era em gerações imediatamente anteriores. O que leva a observar, que está divergência de comportamento pode estar relacionada com a própria herdabilidade que se aplica em cada cenário, como já relatamos anteriormente, com o aumento da herdabilidade, por consequência os valores de acurácia se eleva.

Tabela 2: Valores de acurácia da seleção das gerações avançadas por acasalamento a o acaso (A₁ a A₅) e autofecundação (S₁ a S₅) obtidas a partir do fenotipagem e genotipagem das mesmas gerações (contemporânea na diagonal) ou apenas genotipagem e fenotipagem anteriores (fora da diagonal, leitura na horizontal) para as características de herdabilidade igual a 0.30 e 0.60 no cenário de grau médio de dominância igual a 0.5, com efeito de distribuição Uniforme.

Característica de $h^2 = 0.30$

População	População de Treinamento										
	F ₂	S ₁	S ₂	S ₃	S ₄	S ₅	A ₁	A ₂	A ₃	A ₄	A ₅
F ₂	0.87										
S ₁	0.81	0.81									
S ₂	0.78	0.79	0.80								
S ₃	0.76	0.77	0.79	0.81							
S ₄	0.75	0.77	0.78	0.80	0.77						
S ₅	0.74	0.76	0.77	0.79	0.76	0.81					
A ₁	0.57						0.71				
A ₂	0.27						0.49	0.58			
A ₃	0.20						0.44	0.47	0.62		
A ₄	0.22						0.47	0.48	0.60	0.64	
A ₅	0.18						0.49	0.50	0.58	0.61	0.64

Característica de $h^2 = 0.60$

População	População de Treinamento										
	F ₂	S ₁	S ₂	S ₃	S ₄	S ₅	A ₁	A ₂	A ₃	A ₄	A ₅
F ₂	0.9										
S ₁	0.83	0.87									
S ₂	0.81	0.86	0.87								
S ₃	0.79	0.85	0.86	0.88							
S ₄	0.79	0.84	0.86	0.87	0.85						
S ₅	0.78	0.83	0.85	0.87	0.85	0.86					
A ₁	0.58						0.77				
A ₂	0.28						0.60	0.74			
A ₃	0.2						0.58	0.68	0.77		
A ₄	0.22						0.57	0.67	0.74	0.76	
A ₅	0.18						0.58	0.68	0.70	0.70	0.79

Tabela 3: Valores de acurácia da seleção das gerações avançadas por acasalamento a o acaso (A₁ a A₅) e autofecundação (S₁ a S₅) obtidas a partir do fenotipagem e genotipagem das mesmas gerações (contemporânea na diagonal) ou apenas genotipagem e fenotipagem anteriores (fora da diagonal, leitura na horizontal) para as características de herdabilidade igual a 0.30 e 0.60 no cenário de grau médio de dominância igual a 0.5, com efeito de distribuição binomial.

Característica de $h^2 = 0.30$

População	População de Treinamento										
	F ₂	S ₁	S ₂	S ₃	S ₄	S ₅	A ₁	A ₂	A ₃	A ₄	A ₅
F ₂	0.86										
S ₁	0.79	0.79									
S ₂	0.75	0.76	0.77								
S ₃	0.73	0.73	0.74	0.77							
S ₄	0.74	0.72	0.73	0.77	0.76						
S ₅	0.72	0.72	0.73	0.76	0.76	0.76					
A ₁	0.45						0.65				
A ₂	0.16						0.49	0.58			
A ₃	0.11						0.49	0.52	0.62		
A ₄	0.09						0.36	0.53	0.56	0.60	
A ₅	0.06						0.53	0.53	0.55	0.54	0.62

Característica de $h^2 = 0.60$

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.89										
S1	0.82	0.86									
S2	0.78	0.83	0.86								
S3	0.76	0.81	0.85	0.87							
S4	0.74	0.81	0.84	0.86	0.85						
S5	0.74	0.81	0.84	0.85	0.84	0.86					
A1	0.48						0.79				
A2	0.17						0.67	0.78			
A3	0.13						0.64	0.69	0.79		
A4	0.11						0.63	0.73	0.75	0.79	
A5	0.08						0.67	0.71	0.74	0.72	0.78

Tabela 4: Valores de acurácia da seleção das gerações avançadas por acasalamento a o acaso (A₁ a A₅) e autofecundação (S₁ a S₅) obtidas a partir do fenotipagem e genotipagem das mesmas gerações (contemporânea na diagonal) ou apenas genotipagem e fenotipagem anteriores (fora da diagonal, leitura na horizontal) para as características de herdabilidade igual a 0.30 e 0.60 no cenário de grau médio de dominância igual a 0.5, com efeito de distribuição Exponencial.

Característica de $h^2 = 0.30$

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.85										
S1	0.75	0.77									
S2	0.71	0.74	0.75								
S3	0.68	0.72	0.72	0.75							
S4	0.69	0.71	0.71	0.74	0.74						
S5	0.66	0.71	0.70	0.73	0.73	0.73					
A1	0.42						0.65				
A2	0.12						0.41	0.55			
A3	0.07						0.38	0.45	0.57		
A4	0.05						0.49	0.43	0.48	0.55	
A5	0.03						0.35	0.47	0.48	0.50	0.58

Característica de $h^2 = 0.60$

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.88										
S1	0.77	0.85									
S2	0.73	0.81	0.84								
S3	0.7	0.78	0.82	0.84							
S4	0.74	0.78	0.86	0.83	0.83						
S5	0.68	0.77	0.80	0.83	0.83	0.82					
A1	0.43						0.79				
A2	0.13						0.64	0.75			
A3	0.05						0.59	0.62	0.77		
A4	0.04						0.61	0.62	0.67	0.75	
A5	0.02						0.62	0.66	0.67	0.67	0.76

Tabela 5: Valores de acurácia da seleção das gerações avançadas por acasalamento a o acaso (A₁ a A₅) e autofecundação (S₁ a S₅) obtidas a partir do fenotipagem e genotipagem das mesmas gerações (contemporânea na diagonal) ou apenas genotipagem e fenotipagem anteriores (fora da diagonal, leitura na horizontal) para as características de herdabilidade igual a 0.30 e 0.60 no cenário de grau médio de dominância igual a 1.0, com efeito de distribuição Uniforme.

Característica de $h^2 = 0.30$

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.70										
S1	0.72	0.76									
S2	0.74	0.78	0.80								
S3	0.74	0.79	0.81	0.76							
S4	0.72	0.78	0.81	0.76	0.80						
S5	0.73	0.78	0.80	0.76	0.79	0.78					
A1	0.47						0.56				
A2	0.20						0.30	0.43			
A3	0.15						0.27	0.30	0.44		
A4	0.15						0.28	0.30	0.38	0.50	
A5	0.14						0.30	0.31	0.37	0.39	0.43

Característica de $h^2 = 0.60$

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.72										
S1	0.73	0.81									
S2	0.74	0.82	0.85								
S3	0.75	0.82	0.86	0.86							
S4	0.75	0.83	0.85	0.85	0.86						
S5	0.74	0.82	0.84	0.86	0.86	0.85					
A1	0.49						0.67				
A2	0.2						0.42	0.62			
A3	0.14						0.40	0.46	0.64		
A4	0.17						0.44	0.46	0.50	0.65	
A5	0.13						0.40	0.47	0.49	0.53	0.65

Tabela 6: Valores de acurácia da seleção das gerações avançadas por acasalamento a o acaso (A₁ a A₅) e autofecundação (S₁ a S₅) obtidas a partir do fenotipagem e genotipagem das mesmas gerações (contemporânea na diagonal) ou apenas genotipagem e fenotipagem anteriores (fora da diagonal, leitura na horizontal) para as características de herdabilidade igual a 0.30 e 0.60 no cenário de grau médio de dominância igual a 1.0, com efeito de distribuição binomial.

Característica de $h^2 = 0.30$

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.67										
S1	0.70	0.71									
S2	0.70	0.69	0.73								
S3	0.69	0.69	0.73	0.77							
S4	0.67	0.70	0.72	0.77	0.75						
S5	0.69	0.70	0.72	0.77	0.75	0.75					
A1	0.34						0.52				
A2	0.09						0.34	0.47			
A3	0.08						0.28	0.37	0.43		
A4	0.07						0.31	0.36	0.34	0.49	
A5	0.03						0.28	0.36	0.32	0.30	0.48

Característica de $h^2 = 0.60$

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.71										
S1	0.74	0.80									
S2	0.73	0.79	0.83								
S3	0.73	0.79	0.83	0.82							
S4	0.69	0.79	0.83	0.81	0.86						
S5	0.73	0.78	0.82	0.81	0.85	0.84					
A1	0.39						0.67				
A2	0.14						0.43	0.64			
A3	0.1						0.44	0.48	0.62		
A4	0.1						0.43	0.48	0.51	0.61	
A5	0.08						0.43	0.46	0.49	0.47	0.62

Tabela 7: Valores de acurácia da seleção das gerações avançadas por acasalamento a o acaso (A₁ a A₅) e autofecundação (S₁ a S₅) obtidas a partir do fenotipagem e genotipagem das mesmas gerações (contemporânea na diagonal) ou apenas genotipagem e fenotipagem anteriores (fora da diagonal, leitura na horizontal) para as características de herdabilidade igual a 0.30 e 0.60 no cenário de grau médio de dominância igual a 1.0, com efeito de distribuição Exponencial.

Característica de $h^2 = 0.30$

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.68										
S1	0.66	0.71									
S2	0.66	0.70	0.72								
S3	0.65	0.68	0.72	0.72							
S4	0.65	0.69	0.72	0.74	0.74						
S5	0.65	0.68	0.72	0.74	0.73	0.73					
A1	0.33						0.55				
A2	0.11						0.34	0.46			
A3	0.03						0.26	0.29	0.42		
A4	0.04						0.27	0.32	0.32	0.39	
A5	0.03						0.29	0.35	0.34	0.38	0.47

População	População de Treinamento										
	F ₂	S1	S2	S3	S4	S5	A1	A2	A3	A4	A5
F ₂	0.72										
S1	0.71	0.78									
S2	0.7	0.77	0.81								
S3	0.69	0.75	0.79	0.81							
S4	0.69	0.76	0.79	0.81	0.84						
S5	0.69	0.75	0.79	0.81	0.83	0.82					
A1	0.35						0.66				
A2	0.1						0.42	0.63			
A3	0.05						0.35	0.45	0.59		
A4	0.05						0.35	0.45	0.44	0.60	
A5	0.03						0.36	0.47	0.45	0.46	0.60

CONCLUSÃO

- As populações representativas das espécies alógamas, tiveram baixa eficiência no processo de predição, diferente para as espécies autógamias que foi muito eficiente.

- Treinamento e validação para predição de valores genéticos nas populações em suas próprias gerações (genotipagem contemporânea) torna o processo mais eficiente. O que proporcionaria ao melhorista obter ganhos maiores a partir da aplicação de recursos para fins de obtenção de informações moleculares dos indivíduos sob seleção.

- Quanto a arquitetura genética, distribuição uniforme, binomial e exponencial, não influenciou de modo significativo na diferenciação dos valores de acurácia, desta forma, os valores para cada distribuição de efeitos genéticos foram semelhantes dentro de cada herdabilidade.

- Evidenciamos que o efeito de dominância é um agente perturbador e se tornou mais pronunciado à medida que a herdabilidade da característica foi aumentando. E que a ação do ambiente também é um agente perturbador no processo seletivo. Desta forma, a utilização de um modelo que tenha incluso o efeito de dominância, proporciona uma maior habilidade na predição dos valores.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, I. F.; CRUZ, C. D.; RESENDE, M. D. V. Validação e correção de fenótipos na seleção genômica ampla. PESQUISA AGROPECUÁRIA BRASILEIRA (ONLINE), v. 51, p. 1973-1982, 2016.
- ALMEIDA FILHO, J. E.; GUIMARÃES, J.F.R.; SILVA, F.F.; RESENDE, M.D.V., MUNOZ, P.; KIRST, M.; RESENDE, M.F.R. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity*. Jul; 117(1): 33–41, 2016.
- BASTIAANSEN, J.W.M.; COSTER, A.; CALUS, M.P.L; ARENDONK, J.A.M. van; BOVENHUIS, H. Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. *Genetics Selection Evolution*, v.44, p.1-13, 2012.
- BERNARDO, R.; YU, J. Prospects for Genomewide Selection for Quantitative traits in Maize. *Crop Science*, Madison, v. 47, n. 3, p. 1082-1090, May 2007.
- CALUS, M. P. L.; MEUWISSEN, T. H. E.; ROOS, A. P. W.; VEERKAMP, R. F. Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, v. 178, p. 553-561, 2008
- DE CONINCK A, FOSTIER J, MAENHOUT S, DE BAETS B. DAIRRY-BLUP: a high-performance computing approach to genomic prediction. *Genetics*. 197(3):813-22p. 2014.
- CROSSA, J. et al. Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. *G3*, Bethesda, v. 3, n. 11, p. 1903-1926, Nov 2013.
- CRUZ, C.D. Genes Software – extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum*. v.38, n.4, p.547-552, 2016
- DAETWYLER, HAYDEN, M.J.; SPANGENBERG, G. C.; HAYES, B. J. Selection on optimal haploid value increases genetic gain and preserves more genetic diversity relative to genomic selection. *Genetics* 200: 1341–1348.2015.
- DA Y., WANG C., WANG S., HU G. Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. . *PLoS ONE*, 9(1):e87666. 2014.
- DE LOS CAMPOS G, HICKEY JM, PONG-WONG R, DAETWYLER HD, CALUS MPL. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193: 327–345. 2013.
- DEKKERS, J.C.M. Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science*, v.82, n. 13, p.313-328, 2004.
- DESTA ZA, ORTIZ R. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci*. v.19, n.9, 592-601p.2014

ERTL, J, LEGARRA A, VITEZICA ZG, VARONA L, EDEL C, EMMERLING R, GOTZ, K-U. Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. *Genet Sel Evol* 46: 40. 2014.

ENDELMAN J. B. Ridge regression and other kernels for genomic selection with R package RR-BLUP. *Plant genome*. V. 4. 250-255p. 2011.

FERNANDO RL, GARRICK DJ, LEAFLET ASR. Tree different gibbs samplers for bayesb genomic prediction. Iowa State University. 660, 1-2p.2014.

FRARY A, NESBITT TC, GRANDILLO S, KNAAP E, CONG B, LIU J, MELLER J, ELBER R, ALPERT KB, TANKSLEY SD. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*. V. 289, n.5476, 85-8p.2000.

FRITSCH NETO, R. Seleção genômica ampla e novos métodos de melhoramento do milho. 39 f. Tese (Doutorado em Genética e Melhoramento de Plantas) –Universidade Federal de Viçosa, Viçosa, 2011.

GODDARD, M. E. Genomic selection: prediction of accuracy and maximization of long term response. *Genetica*, 136(2):245-257. 2009.

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

GUIMARÃES, J F R EFEITO DA INTERAÇÃO DOMINÂNCIA X AMBIENTE NA HABILIDADE DE PREDIÇÃO GENÔMICA. 30 f. Tese (Doutorado em Genética e Melhoramento de Plantas) –Universidade Federal de Viçosa, Viçosa, 2016.

.HAYES BJ, BOWMAN PJ, CHAMBERLAIN AJ, GODDARD ME. Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92: 433–443. 2009.

HAYES, B. J.; COGAN, N. O. I.; PEMBLETON, L. W.; GODDARD, M. E.; WANG, J.; SPANGENBERG, G. C.; FORSTER, J. W. Prospects for genomic selection in forage plant species. *Plant Breeding*, v. 132, n. 2, p. 133–143, 2013.

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

LEWONTIN, R. C. On measures of gametic disequilibrium. *Genetics, Heidelberg*. v. 120, n. 3, p. 849-852, 1988.

LIU J, VAN ECK J, CONG B, TANKSLEY SD A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc Natl Acad Sci* . v. 99, n.20, 13302-6p. 2002.

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

MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Accelerating improvement of livestock with genomic selection. *Annual Review of Animal Biosciences*, v. 1, n. 1, p.

221–237, 2013.

MISZTAI. Blackwell Verlag GmbH • J. Anim. Breed. Genet. 128 (2011) 245–246. 2011.

MISZTAL I, VARONA L, CULBERTSON, MBERTRAND, J.K., MABRY J. LAWLOR T.J, VAN TASSEL C.P., GENGLER N. Studies on the value of incorporating the effect of dominance in genetic evaluations of dairy cattle, beef cattle and swine. Biotechnol. Agron. Soc. Environ. 2 (4), 227–233p. 1998.

PRYCE J. E., HAYES B. J., GODDARD M. E. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. J. Dairy Sci. 95: 377–388p. 2012.

RESENDE JÚNIOR, M.F.R. Seleção genômica ampla no melhoramento vegetal. 67 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) –Universidade Federal de Viçosa, Viçosa, 2010.

RESENDE, M.D.V. Genômica quantitativa e seleção no melhoramento de plantas perenes e animais. Colombo: Embrapa Florestas, p. 330, 2008.

RESENDE, M.D.V SELEGEN–REML/BLUP: Sistema estatístico e seleção genética computadorizada via modelos lineares mistos. Colombo, Embrapa Florestas. 361p. 2007.

RESENDE, M. D. V. DE, P. S. LOPES, R. L. SILVA, I. E. Pires. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. Pesquisa Florestal Brasileira, 56:63-78. 2008.

RESENDE MDV, MUÑOZ P, RESENDE MDV, GARRICK DJ, FERNANDO RL, DAVIS JM ET AL. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). Genetics 190: 1503–1510. 2012.

RESENDE, R. M. S., M. CASLER, M. D. V. RESENDE. Genomic Selection in Forage Breeding: Accuracy and Methods. Crop Science, 54:143.2014.

RESENDE MDV, RESENDE JR. MFRR, AGUIAR AM, ABAD JIM, MISSIAGGIA AA, SANSALONI C, PETROLI C & GRATTAPAGLIA D. Computação da seleção genômica ampla (GWS). Colombo, Embrapa Florestas. 79p. 2010.

RIEDELSEIMER, C. et al. Genomic predictability of interconnected biparental maize populations. Genetics, 194. 493-503. 2013.

SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. Journal of animal breeding and genetics. v. 123, n. 4, p. 218–223, 2006.

SHIERENBECK S., PIMENTEL E. C. G., TIETZE M., KOERTE J., REENTS R., et al. Controlling inbreeding and maximizing genetic gain using semi-definite programming with pedigree-based and genomic relationships. J. Dairy Sci. 94: 6143–6152. 2011.

SONESSON, A. K., WOOLLIAMS J. A., MEUWISSEN, T. H. E. Maximising genetic gain whilst controlling rates of genomic inbreeding using genomic optimum contribution selection. Proceedings of the 9th World Congress of Genetics Applied to Livestock

Production, edited by German Society for Animal Science. German Society for Animal Science, Leipzig, Germany. pp. 892–895. 2010.

STURION, J.A.; RESENDE, M.D.V. Seleção de progênies de erva-mate (*Ilex paraguarensis* St. Hil.) para produtividade, estabilidade e adaptabilidade temporal de massa foliar. *Boletim de Pesquisa Florestal*, Colombo, n.50, p.37-51, 2005.

SUN C, VANRADEN PM, COLE JB, O'CONNELL JR. Improvement of Prediction Ability for Genomic Selection of Dairy Cattle by Including Dominance Effects. *PloS one* 9(8): e103934

TECHNOW F, SCHRAG TA, SCHIPPRACK W, BAUER E, SIMIANER H, MELCHINGER AE. Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics*. V. 197, n.4, 1343-55p. 2014.

YANO A, M., KATAYOSE A, Y., ASHIKARIB M, YAMANOUCHI U, MONNAC L, FUSED T, BABAC T, YAMAMOTO K, UMEHARA Y, NAGAMURA Y, SASAKI T. Hd1, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene *CONSTANS*. *The Plant Cell* December, vol. 12 no. 12 2473-2483.2000

VARONA L1, MISZTAL I, BERTRAND JK, LAWLOR TJ. Effect of full sibs on additive breeding values under the dominance model for stature in United States Holsteins. *J Dairy Sci*. v.81, n.4, 1126-35p.1998

WHITTAKER, J.C.; THOMPSON, R.; DENHAM, M.C. Marker-assisted selection using ridge regression. *Genetical Research*, v.75, p.249-252, 2000.

WIMMER V, LEHERMEIR C, ALBRECHT T, AUINGER HJ, WANG Y, SCHON CC. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics*. 195(2):573-87p, 2013.

ZHANG, Z; DING, X.; LIU, J; ZHANG, Q; DE KONING, D. J. Accuracy of genomic prediction using low-density marker panels. *Journal of dairy science*. V. 94, n.7. 3642-3650p. 2011.

ZHONG, S; DEKKERS, J.C. M; FERNANDO, R.L.; JANNINK, J.L. Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. *Genetics*. vol. 182 n. 1, 355-364; 2009.