

JUAN VICENTE ROMERO

**SELEÇÃO GENÔMICA USANDO GENOTIPAGEM DE BAIXA
SATURAÇÃO NO MELHORAMENTO GENÉTICO DO CAFEIEIRO**

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

R763s
2017

Romero, Juan Vicente, 1979-
Seleção genômica usando genotipagem de baixa saturação
no melhoramento genético do cafeeiro / Juan Vicente Romero. –
Viçosa, MG, 2017.
viii, 94f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Leonardo Lopes Bhering.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Cafeeiro - Melhoramento genético. 2. *Coffea arabica*.
I. Universidade Federal de Viçosa. Departamento de Biologia
Geral. Programa de Pós-graduação em Genética e
Melhoramento. II. Título.

CDD 22 ed. 633.738

JUAN VICENTE ROMERO

**SELEÇÃO GENÔMICA USANDO GENOTIPAGEM DE BAIXA
SATURAÇÃO NO MELHORAMENTO GENÉTICO DO CAFEIEIRO**

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: 21 de fevereiro de 2017.

Antonio Carlos Baião de Oliveira

Moysés Nascimento

Eveline Teixeira Caixeta
Coorientadora

Cosme Damião Cruz
Coorientador

Leonardo Lopes Bhering
Orientador

À minha família: Maria Amelia, Jenny e Alana.

AGRADECIMENTOS

Gostaria de expressar minha grande gratidão às pessoas e instituições que tem contribuído para minha formação durante este ciclo no Brasil:

A Deus, quem faz tudo possível.

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

Ao programa CAPES/PEC-PG, pela concessão da bolsa de estudo.

Ao meu orientador, professor Leonardo Lopes Bhering pelo grande apoio.

Aos meus coorientadores Eveline Teixeira Caixeta e Cosme Damião Cruz, pelas valiosas sugestões, disponibilidade e atenção.

Aos amigos do laboratório de Biometria.

Aos meus ex-chefes, mas sempre amigos Dra. Pilar e Dr. Cortina.

Aos amigos Jorge William e Juan Carlos.

Aos professores e colegas do curso de pós-graduação em genética.

Aos membros da banca, Drs. Moisés Nascimento e Antonio Baião, pela disponibilidade e colaboração.

A todos os que direta ou indiretamente colaboraram para a realização deste trabalho.

À cidade de Viçosa, ao Brasil e seu povo por nos acolher à minha família e a mim.

À Jenny e a Alana, que mesmo diante todas as dificuldades, me acompanharam dando todo o apoio necessário para realização deste sonho.

A todos eles: Muchas gracias!

SUMÁRIO

RESUMO	vii
ABSTRACT	viii
1. INTRODUÇÃO GERAL	1
2. OBJETIVOS	5
2.1. Objetivo Geral	5
2.2. Objetivos específicos.....	5
3. REVISÃO DE LITERATURA.....	6
3.1. Plantas Autógamas	6
3.1.1. Estrutura genética das populações autógamas	6
3.1.2. Melhoramento de autógamas	8
3.2. Características Oligogênicas	9
3.3. Desequilíbrio de Ligação.....	11
3.3.1. Quantificação do DL.....	12
3.3.2. Redução do desequilíbrio de ligação ao longo das gerações	13
3.4. Seleção genômica	14
3.4.1. Acurácia da GS	16
3.4.2. Densidade de marcadores	18
3.4.3. Métodos de estimação dos efeitos dos marcadores e de predição de valores genômicos.....	19
4. REFERÊNCIAS.....	24
CAPÍTULO I.....	34
SIMULAÇÃO E VALIDAÇÃO DE POPULAÇÕES.....	34
1. INTRODUÇÃO	35
2. MATERIAL E MÉTODOS	36
2.1. População Base (F ₂)	36
2.1.1. Simulação da população base	37

2.1.2.	Validação da população base.....	38
2.2.	Gerações avançadas (F ₃ a F ₆)	41
2.2.1.	Simulação de gerações avançadas	41
2.2.2.	Validação de gerações avançadas	42
3.	RESULTADOS E DISCUSSÃO.....	43
3.1.	População base	43
3.1.1.	Validação da simulação da população base F ₂	45
3.1.2.	Validação dos valores genotípicos e fenotípicos da população base ..	47
3.2.	Validação de simulações de populações.....	49
3.2.1.	Validação de autofecundação e seleção fenotípica.....	50
4.	CONCLUSÕES	52
5.	REFERÊNCIAS.....	52
	CAPÍTULO II	57
	ANÁLISES DE CENÁRIOS COM GENOTIPAGEM DE BAIXA SATURAÇÃO PARA A SELEÇÃO GENÔMICA DE CARACTERÍSTICAS OLIGOGÊNICAS EM <i>Coffea arabica</i> L.	57
1.	INTRODUÇÃO	58
2.	MATERIAL E MÉTODOS	59
2.1.	Simulação de populações	59
2.2.	Análise de seleção genômica.....	61
2.3.	Variação da distância de ligação entre loco marcador e loco controlador	63
2.4.	Cenários avaliados.....	65
3.	RESULTADOS.....	66
3.1.	Validação da autofecundação.....	66
3.2.	Avaliação de métodos preditivos, intensidade de seleção e seleção de marcadores.	67
3.3.	Avaliações do tamanho populacional.....	71

3.4.	Avaliação da não associação de marcadores a <i>QTL</i>	73
3.5.	Avaliações do tipo de marcador	74
4.	DISCUSSÃO	76
4.1.	Aplicação em programas de melhoramento de café.....	79
5.	REFERÊNCIAS.....	81
	APÊNDICE A. Validação de rotinas para simulação de genótipos e valores genotípicos e fenotípicos executáveis no programa <i>R</i>	85
	APÊNDICE B. Rotinas executáveis no <i>R</i> para simulação e seleção de genótipos.....	89

RESUMO

ROMERO, Juan Vicente, D.Sc., Universidade Federal de Viçosa, fevereiro de 2017. **Seleção genômica usando genotipagem de baixa saturação no melhoramento genético do cafeeiro.** Orientador: Leonardo Lopes Bhering. Coorientadores: Eveline Teixeira Caixeta e Cosme Damião Cruz.

A eficiência da seleção genômica (*GS*) utilizando relativa baixa cobertura de marcadores moleculares e tamanho populacional reduzido foi avaliada para o melhoramento genético de *Coffea arabica*, considerando características oligogênicas. Foram analisadas populações simuladas, até a sexta geração de seleção e autofecundação, com características envolvendo quatro genes com 40% ou 80% de herdabilidade, em cenários com sete densidades de marcadores e cinco tamanhos populacionais. Os valores genéticos genômicos dos indivíduos foram preditos com os métodos *BLASSO* e *RR-BLUP* e foram analisadas duas intensidades de seleção e o uso de marcadores dominantes versus codominantes. Em cada cenário foram estimados: coeficiente de endogamia, média e variância genotípica, capacidade preditiva e capacidade seletiva. Foram considerados viáveis os cenários em que os alelos favoráveis foram fixados até a sexta geração e foi procurado o uso mínimo de marcadores e menor tamanho populacional. Quatro resultados iniciais aumentaram a eficiência e eficácia seletiva: o uso de marcadores codominantes, o uso do método *RR-BLUP*, o aumento da intensidade de seleção e a seleção dos marcadores que maximizaram a acurácia preditiva. Sob estas condições a seleção foi eficaz, sendo que nas características de alta e media herdabilidade foram fixados alelos na sexta geração, usando densidades de marcadores com distância entre marcas de até 6 cM e populações de pelo menos 200 indivíduos. Com 400 indivíduos foi possível a fixação na geração F_5 , com as mesmas distâncias entre marcadores. Marcadores dominantes e densidades em que regiões codificadoras não sejam marcadas são os principais inconvenientes para a utilização da *GS* nos cenários avaliados.

ABSTRACT

ROMERO, Juan Vicente, D.Sc., Universidade Federal de Viçosa, February, 2017. **Genomic selection with low-density marker in coffee breeding.** Adviser: Leonardo Lopes Bhering. Co-advisers: Eveline Teixeira Caixeta and Cosme Damião Cruz.

The efficiency of genomic selection (*GS*) using relative low density molecular markers and reduced population size was evaluated for breeding of *Coffea arabica*, considering oligogenic traits. Simulated population were analyzed until the sixth generation of selection and selfing, with traits involving four genes with 40% or 80% heritability, in scenarios with seven marker densities and five population sizes. The genomic breeding values of the individuals were predicted using the *RR-BLUP* and *BLASSO* methods. Two selection intensities were analyzed, as well as dominant versus codominant markers. The following variables were estimated in each scenario: inbreeding coefficient, genotypic mean and genotypic variance, predictive capacity, and selective capacity. The scenario was considered viable when it used the minimum number of markers and smaller population size, and when the favorable alleles were fixed until the sixth generation. The strategies that increased efficiency and selective efficacy were: use of codominant markers, use of the *RR-BLUP* method, increase in the intensity of selection, and selection of markers that maximized the predictive accuracy. In intermediate and high heritability traits, the favorable alleles were fixed in F_6 generation, using densities with distances between markers of up to 6 cM, and training populations of at least 200 individuals. Four hundred individuals would allow the fixation in the F_5 generation, with the same distances between markers. Therefore, *GS* methods can be used to select oligogenic traits in the scenarios evaluated; however, dominant markers and densities where coding regions are not marked are the main drawbacks for the use of these methods.

1. INTRODUÇÃO GERAL

O café é uma das bebidas mais apreciadas mundialmente, cujo consumo continua crescendo de forma ativa. É um dos produtos mais negociadas do mundo e as vendas no varejo ultrapassam os US\$ 70 bilhões por ano. Além disso, este produto é socialmente importante, enquanto estima-se que 125 milhões de pessoas dependem do café para a sua subsistência nos países produtores (ICO, 2014).

Coffea arabica L., altamente valorizada pela qualidade da sua bebida, é a espécie mais importante economicamente dentro do seu gênero, responsável por 60% da produção mundial (ICO, 2017). Esta espécie é cultivada em vários países de América latina, além da África e Ásia (Wintgens, 2011), sendo o Brasil o maior produtor mundial com 43,38 milhões de sacas em 2016, produzidas em 1,78 milhões de hectares (CONAB, 2017).

Os cafeicultores, além de estarem expostos a um mercado dinâmico, tanto em exigências quanto em oportunidades, continuam enfrentando os problemas básicos da produção. Neste cenário, o melhoramento genético é determinante na geração de genótipos para suprir a crescente demanda e para dar continuidade à rentabilidade do negócio, contribuindo para o aumento de rendimentos, na melhoria da qualidade da bebida e na resistência a estresses bióticos e abióticos, entre outros elementos da produção.

No entanto, como em outras espécies perenes, o processo de melhoramento em *C. arabica* é demorado, sendo necessários em média 20 anos para o lançamento de uma nova variedade. Portanto, para encurtar os ciclos e custos de melhoramento, a procura de ferramentas que aumentem a eficiência na avaliação e seleção de genótipos tem sido indispensável. Neste contexto, as tecnologias de marcadores moleculares são reconhecidas por melhorar a precisão de seleção, aumentar a intensidade de seleção e possibilitar a prática de seleção precoce, além de reduzir o esforço de testes de campo.

No entanto, o uso de marcadores moleculares em *C. arabica* tem a dificuldade de que esta espécie tem baixa taxa de polimorfismo molecular. Isso é devido à sua origem relativamente recente e pelo método reprodutivo, dado que *C. arabica* é uma espécie autógama, o que contribui para baixos níveis de heterozigosidade (Herrera et al., 2009; Vieira et al., 2010; Geleta et al., 2012). Além disso, por contextos históricos ou seletivos, um pequeno número de plantas é a base de todos os programas de melhoramento de *C. arabica*. No Brasil, por exemplo, a base genética de 110 cultivares

foi definida por treze ancestrais e só sete contribuíram com o 98,8% dos genes (Setotaw et al., 2013), fato que se repete na maioria dos países produtores (Van der Vossen et al., 2015).

Apesar disso, os marcadores moleculares tem sido relevantes para *C. arabica* no estudo da sua origem (Lashermes et al., 1999), no auxílio da identificação da variação entre e dentro de populações (Chaparro et al., 2004; Moncada & McCouch, 2004; Setotaw et al., 2010; Vieira et al., 2010; Missio et al., 2011), no desenvolvimento de mapas de ligação (Pearl et al., 2004; Teixeira-Cabral et al., 2004; Oliveira et al., 2007; Pestana et al., 2015) e na identificação de regiões genômicas associadas às características de interesse, para a seleção assistida de genótipos (Gichuru et al., 2008; Brito et al., 2010; Diola et al., 2011; Pereira et al., 2011; Prakash et al., 2011; Romero et al., 2014; Moncada et al., 2016; Alkimim et al., 2017). Para esses trabalhos têm sido usados tanto marcadores dominantes, quanto codominantes.

Recentemente, as tecnologias de sequenciamento de *DNA* têm permitido a produção de grande número de dados genômicos. Isso viabilizou a análise de genomas completos e mudou a escala das análises. Uma das aplicações dessas informações é a estimação de do valor genético de cada indivíduo de uma população em melhoramento de forma acurada, permitindo fazer seleção eficaz (Grattapaglia & Resende, 2011). Esta abordagem recebe o nome de seleção genômica ampla (*Genome-Wide Selection – GWS*) ou simplesmente seleção genômica (*Genomic Selection – GS*), inicialmente promovida para o melhoramento animal (Meuwissen et al., 2001), mas com aplicação recente nas espécies vegetais, principalmente em culturas de ciclo curto (Zhong et al., 2009; Hayes et al., 2013; Bernardo, 2014; Crossa et al., 2014; Spindel et al., 2015; Ma et al., 2016) e algumas perenes (Resende Jr et al., 2012; Resende et al., 2012b; Kumar et al., 2014; Cros et al., 2015).

A *GS* está baseada no princípio do desequilíbrio de ligação (*DL*) entre marcadores e regiões controladoras da característica em estudo. O genoma é coberto com centenas ou milhares de marcadores, de forma que a maioria dos alelos de interesse sejam associados com um ou mais marcadores, de maneira que capturem a maioria da variação quantitativa. O efeito de cada marcador é estimado com base em dados fenotípicos de uma população de estimação, gerando o modelo de predição dos valores genéticos genômicos (*VGG*). A acurácia do modelo é obtida numa população de validação e,

posteriormente, é aplicado em populações de seleção (Crossa et al., 2011; Kainer et al., 2015).

Como a *GS* inclui simultaneamente todos os efeitos dos marcadores, grande parte da variação da característica pode ser explicada, o que produz uma estimativa acurada do valor genético dos indivíduos. Assim, a expectativa sob a *GS* é grande, especialmente no aumento dos ganhos em seleção por unidade de tempo e custo, principalmente em plantas perenes como o café.

Contudo, para muitas espécies, o ponto crítico da aplicação rotineira da *GS* é o custo da genotipagem por sequenciamento que ainda não é suportado por todos os programas de melhoramento. (Byrne, 2007; Zhang et al., 2011; Wellmann et al., 2013; Kainer et al., 2015; Ma et al., 2016). Além disso, em *C. arabica* é especialmente trabalhoso alcançar alta densidade de marcadores, pela baixa taxa de polimorfismo molecular. Neste sentido, uma redução do número de marcadores utilizados, mantendo uma acurácia adequada, facilitaria a genotipagem e reduziria custos, fazendo aproveitáveis os métodos de *GS* no melhoramento de *C. arabica*.

Um limitante adicional para a *GS* em *C. arabica* é o custo da manutenção de grandes populações no campo por longos períodos. Portanto, as grandes populações necessárias na *GS* para estimação dos efeitos (> 1000 indivíduos), raramente poderiam ser mantidas.

Mas, apesar de essas dificuldades, a utilização da seleção genômica no melhoramento de *C. arabica* foi explorada por Pestana (2013), avaliando a resistência à ferrugem (*Hemileia vastatrix* Berk. et Br.), em uma população F_2 de 245 indivíduos com 191 marcadores. Observou-se nesse trabalho que a capacidade preditiva, usando o método *BLASSO*, foi 34% superior à herdabilidade baseada em valores fenotípicos. Este resultado positivo incentivou o estudo mais detalhado da eficiência do uso das análises de *GS* para a estimação dos *VGG*, em populações pequenas e com genotipagens limitadas, para os padrões da seleção genômica.

Assim, trabalhando nos limites dos requisitos da *GS*, este método poderia ser útil em casos específicos, como na seleção de características oligogênicas. Estas características apresentam herança governada por poucos genes de efeito maior e alta herdabilidade. Estas têm se mostrado de grande importância no melhoramento de *C. arabica*, como na resistência a pragas como o bicho mineiro (*Perileucoptera coffeella*) (Guerreiro Filho et

al., 1999), na resistência a doenças como ferrugem ou a doença do fruto (*Colletotrichum kahawae*), entre outras (Rodrigues Jr et al., 1975; Gichuru et al., 2008; Van der Vossen & Walyaro, 2009; Pestana et al., 2015).

Em vista das vantagens que oferecem os métodos de seleção genômica, mas, sabendo das limitações que esta técnica pode trazer se não são usados tamanhos populacionais ou densidades de marcadores adequados, o presente trabalho buscou quantificar os limites em que a seleção genômica pode ser útil para o melhoramento de *C. arabica*, nas condições de baixa saturação de marcadores e pequenas populações de estimacão, limitando as análises a características oligogênicas. Além disso, foi considerado o uso de marcadores dominantes e codominantes para gerar os modelos preditivos da GS, dado que ambos os tipos de marcadores são frequentemente usados em *C. arabica*.

Para realizar essa quantificação a alternativa considerada foi a análise de conceitos teóricos por meio da simulação computacional. A simulação tem a finalidade de gerar informação suficiente para compreender o funcionamento do sistema, o que permite a previsão do desempenho e a produção de recomendações, antes de passar à experimentação de campo (Dachs, 1988; Ferreira, 2001; Sun *et al.*, 2011). Além disso, a simulação é conveniente por facilitar a comparação das estratégias, demandando menos recurso e tempo. E ainda, os experimentos podem ser replicados em um grande número de vezes, aumentando a precisão das inferências (Gurgel, 2004; Sun *et al.*, 2011; Wang, 2012).

Deste modo, o estudo foi realizado com a análise de populações simuladas em vários cenários, monitorando os processos durante várias gerações de seleção e autofecundação, procurando a homozigose de alelos favoráveis (fixação).

O documento expõe o trabalho em dois capítulos, o primeiro explica os procedimentos de simulação de genomas, permutações e autofecundações, e a sua respectiva validação. No segundo capítulo são aplicadas as simulações em diferentes cenários, para responder os vários questionamentos, que em conjunto resolvem o objetivo do trabalho.

2. OBJETIVOS

2.1. Objetivo Geral

Determinar a viabilidade dos métodos de seleção genômica, com baixa densidade de marcadores moleculares e limitado tamanho populacional, para o melhoramento de características oligogênicas em *Coffea arabica*.

2.2. Objetivos específicos

- Estabelecer a densidade de marcadores e tamanho populacional mínimo para conduzir a homozigose os alelos favoráveis em uma população.
- Analisar o resultado da falta de ligação entre marcadores moleculares e o total das regiões codificadoras, em condições de baixa densidade de marcadores.
- Comparar o efeito do tipo de marcador usado para a seleção genômica de características oligogênicas.

3. REVISÃO DE LITERATURA

3.1. Plantas Autógamas

As plantas autógamas são aquelas que se reproduzem sexuadamente por autofecundação. Pode-se dar pelo mecanismo da cleistogamia, no qual a polinização do estigma ocorre antes da abertura do botão floral, evitando a entrada de pólen externo. Também pode ser favorecida pela distribuição dos estigmas e estames na flor, de tal forma que a autopolinização é quase garantida (Borém & Miranda, 2013).

No entanto, a autogamia absoluta não é comum, ocasionalmente pode ocorrer fecundação cruzada, pela ação de insetos polinizadores, vento ou pela proximidade entre plantas. Dessa forma, Allard (1999) considerou autógamas aquelas plantas com menos de 5% de alogamia, porém, do ponto de vista prático, o melhoramento das culturas são dirigidas segundo o predomínio do sistema reprodutivo.

Entre as culturas com predomínio de autogamia podem-se citar: cereais, como o trigo, cevada, aveia e arroz; leguminosas, como a soja, ervilha, amendoim, feijoeiro comum e feijão caupi; olerícolas, como o tomate, quiabo, alface e beringela e perenes, como o café arábica, pêsego e nectarina (Vallejo *et al.*, 2010; Borém & Miranda, 2013).

3.1.1. Estrutura genética das populações autógamas

W. L. Johannsen (1903) forneceu as bases fundamentais para a seleção em espécies autógamas. Em seus experimentos com feijão, utilizou um lote de sementes de diferentes tamanhos no qual investigou o efeito da seleção sobre o peso médio das sementes das progênes. Após a análise de seus resultados definiu as linhas puras como toda descendência por autofecundação de um indivíduo homozigoto e estabeleceu três princípios: (i) existem variações herdáveis e variação causada pelo ambiente, (ii) a seleção é eficiente quando é feita sobre as variações herdáveis e (iii) a seleção não gera variabilidade (Borém & Miranda, 2013).

As populações de espécies autógamas estão frequentemente constituídas por mistura de genótipos homozigotos, não necessariamente puros. Mas, de acordo com Johannsen, após vários ciclos de autofecundação é possível isolar diversas linhas puras. Pode-se selecionar entre linhas, mas não tem sentido selecionar entre indivíduos da mesma linha porque todas as plantas terão o mesmo genótipo, já que as variações dentro de cada linha serão devidas a efeitos ambientais (Vallejo & Estrada, 2002).

As linhas puras permanecerão homozigotas para todos os locos indefinidamente, desde que sejam mantidas por autofecundação. Isto porque, com as sucessivas gerações de autofecundação, há aumento no número de homozigotos, frente a uma diminuição no número de heterozigotos (50% a cada geração). Com as sucessivas gerações de autofecundação, os indivíduos homozigotos produzirão indivíduos homozigotos, diferente dos heterozigotos que produzirão tanto homozigotos quanto heterozigotos. Depois de várias gerações, a proporção de homozigotos tende a ser 100% em contraposição dos heterozigotos (Tabela 1). Este fenômeno é conhecido como efeito da endogamia.

Tabela 1. Proporção de indivíduos homozigotos e heterozigotos em uma população submetida a sucessivas gerações de autofecundação.

Geração	V ₁ (AA) x V ₂ (aa)		
	AA	Aa	aa
F1	0	1	0
F2	(1/4)	(1/2)	(1/4)
F3	(3/8)	(1/4)	(3/8)
F4	(7/16)	(1/8)	(7/16)
:	:	:	:
F _n	$[1-(1/2)^n]/2$	$(1/2)^n$	$[1-(1/2)^n]/2$
:	:	:	:
F _∞	(1/2)	0	(1/2)

A endogamia é definida como o efeito do cruzamento natural ou artificial entre indivíduos aparentados. O sistema de acasalamento que mais rapidamente conduz à endogamia é a autofecundação (Ramalho *et al.*, 2012). Seus efeitos em uma espécie autógena são:

- Aumento progressivo da homozigose, o que conduz à fixação de alelos e, portanto, dos padrões fenotípicos.
- Alteração da frequência genotípica, enquanto a frequência alélica permanece inalterada.
- Formação de 2^n linhas uniformes, sendo n o número de genes segregantes na população.
- Ampliação e redistribuição da variância genética. Aumento entre linhas e redução dentro delas.

A endogamia é medida por meio do coeficiente de endogamia, geralmente simbolizado pela letra F , mas não deve ser confundido com o símbolo F da geração filial. Em um indivíduo diploide, este é definido como a probabilidade de que seus alelos sejam idênticos por ascendência, ou seja, que tenham a mesma origem ancestral.

Em uma população submetida a sucessivas gerações de autofecundação, e usando a segunda geração como referência, o coeficiente de endogamia esperado é calculado com a expressão:

$$F = 1 - (1/2)^{t-2}; \text{ sendo } t \text{ a geração analisada (Ramalho } et al., 2012).$$

Por outro lado, conhecendo as frequências genotípicas de um loco em uma geração particular, o coeficiente de endogamia realizado é estimado a partir da expressão:

$$F = (H_e - H_o) / H_e$$

Sendo: $H_e = 1 - \sum p_i^2$; $p_i = (n_{ii} + (n_{ij}/2)) / N$; $H_o = n_{ij} / N$

em que H_e é a frequência de heterozigotos esperados e H_o dos observados, p_i a frequência estimada do alelo i , n_{ii} é o número de homozigotos do alelo i , n_{ij} o número de heterozigotos, N o número total de indivíduos.

3.1.2. Melhoramento de autógamias

A hibridação artificial de espécies autógamias tem como objetivo reunir, em uma nova linhagem pura, alelos favoráveis presentes em duas ou mais cultivares comerciais, em linhagens elites de programas de melhoramento, em plantas introduzidas ou em espécies relacionadas. Conforme mencionado anteriormente, em uma população autógama, as frequências alélicas se mantêm constantes ainda quando as frequências dos genótipos mudem, a menos que seja feita seleção contra um determinado genótipo. Portanto, os híbridos resultantes podem ser conduzidos por autofecundações e em seguida realizar a seleção de linhagens superiores, até atingirem a homozigosidade (Vallejo & Estrada, 2002; Ramalho *et al.*, 2012; Borém & Miranda, 2013)

Em termos gerais, os passos necessários para o desenvolvimento de linhagens puras por meio da hibridação são os seguintes:

1. Seleção de parentais
2. Hibridação e geração da população F_1
3. Condução de populações segregantes

4. Seleção de plantas individuais
5. Avaliação de linhagens puras em gerações avançadas
6. Produção comercial de sementes do novo cultivar

Os métodos de melhoramento por hibridação diferem entre si, principalmente pela metodologia de avanço de gerações para fixação de genótipos homozigotos para seleção das linhas puras superiores.

3.2. Características Oligogênicas

As características genéticas a serem melhoradas em uma espécie agrícola, podem ser qualitativas ou quantitativas. Uma grande parte das características agrônômicas que precisam ser melhoradas apresenta herança quantitativa, tais como produção, arquitetura, ciclo vegetativo, entre outras. Caracteres quantitativos são controlados por muitos genes, influenciados pelo ambiente de forma pronunciada, havendo uma variação contínua do fenótipo.

Não menos importantes, os caracteres qualitativos são aqueles governados por um (característica monogênica) ou poucos genes (oligogênica). Apresentam variação discreta e são menos influenciados pelo ambiente, e por isso que têm alta herdabilidade. As características oligogênicas apresentam herança governada por genes de efeito maior. Geralmente, os estudos da herança e da variação baseiam-se na análise de gerações, separando os indivíduos em classes e avaliando suas proporções (Vallejo & Estrada, 2002; Vallejo *et al.*, 2010). O conhecimento da genética desse caráter é imprescindível na escolha de estratégias mais adequadas de seleção.

Com o propósito de melhoramento, na literatura têm sido analisadas diversas características com herança oligogênica, algumas relacionadas com a composição ou a arquitetura da planta e outras com relação à resposta a fatores ambientais ou biológicos.

Blair *et al.* (2009; 2012) estudaram em grãos de feijão comum a herança de concentração de fitatos, compostos que fornecem fosfatos durante a germinação, mas não são digeridos pelos humanos, retendo micronutrientes tais como ferro e zinco. Acharam três *QTL* (Quantitative Trait Locus) como base da herança oligogênica para a concentração de fitato. Esse resultado permitirá o melhoramento para a biodisponibilidade de micronutrientes em feijão, pela redução das quantidades de fitato.

Por outro lado, componentes da arquitetura da planta têm sido analisados individualmente e identificada sua herança oligogênica. Por exemplo, no feijão-caupi o comprimento do pedúnculo, componente importante para o melhoramento da colheita mecânica ou para facilitar a colheita manual, tem sido identificado que é controlado por cinco genes, que tem efeito gênico aditivo (Rocha *et al.*, 2009). Outras características têm sido analisadas no feijão-caupi, como o tamanho da semente, controlada por efeitos gênicos aditivos e envolvendo cinco genes (Lopes *et al.*, 2003); a textura do grão, controlada por dois genes complementares e a pigmentação roxa em partes vegetativas da planta condicionada por dois genes (Singh & Ishiyaku, 2000).

No entanto, a herança oligogênica é mais comumente associada a características de resistência ou tolerância. São encontrados estudos em que a tolerância a fatores abióticos é determinada por poucos genes, segundo análises de gerações ou dialélicas. Um exemplo é a tolerância á toxicidade por alumínio em milho, em que os resultados evidenciam herança oligogênica (Almeida, 2007), ou em arroz a tolerância ao efeito de baixas temperaturas, que revelaram herança por um ou dois alelos dominantes e dois genes complementares com alelos recessivos (Cruz *et al.*, 2010).

Contudo, os estudos de herança oligogênica relacionados com fatores bióticos são mais frequentes, sendo principalmente associados à resistência a doenças. Como a resistência ao *plum pox* vírus em damascos (Salava *et al.*, 2005), em que a segregação em razão 1:7 (resistente: suscetível), indicou que a resistência foi controlada por três genes complementares dominantes independentes, sendo necessários os três genes dominantes para a expressão da resistência. Outro caso é a resistência a fusarium em trigo (Liu *et al.*, 2005), em que gerações F₈ e F₉ de *RILs* foram usadas para determinar o número mínimo de genes, encontrando que quatro genes condicionam a resistência á doença. Já o trabalho de Vida *et al.* (2009), é um exemplo da incorporação de seis genes de resistência a ferrugem em cultivares de trigo, usando seleção assistida por marcadores.

Em *C. arabica*, estudos de herança oligogênica têm sido feitos principalmente na resistência a doenças. No caso da resistência do cafeeiro à ferrugem (*Hemileia vastatrix*) têm demonstrado que, segundo a raça do patógeno, ela pode ser governada por um a quatro genes, com formas alélicas dominantes e/ou recessivas (Pereira, 1995; Oliveira *et al.*, 2004; Capucho *et al.*, 2009; Pestana, 2010; Pestana *et al.*, 2015). Também tem sido confirmada a herança oligogênica da resistência a *Colletotrichum*

kahawae, uma doença altamente limitante da produção, que ainda é restrita ao continente Africano (Van der Vossen & Walyaro, 2009).

3.3. Desequilíbrio de Ligação

O desequilíbrio de ligação, em relação a um par de locos, é qualquer desvio das frequências alélicas em relação às frequências esperadas, sob a hipótese de independência, indicando a existência de associação não-aleatória, ou preferencial, entre alelos de diferentes locos em uma população. O desequilíbrio de ligação (*DL*) entre marcadores moleculares e regiões controladoras das características é essencial para a seleção genômica ampla, como será explicado posteriormente. Nessa abordagem pode ser usado o *DL* devido à distância física entre marcador e *QTL* e o *DL* entre locos não ligados fisicamente (Resende *et al.*, 2012a).

De forma geral, qualquer fator que altere as frequências alélicas pode interferir na dinâmica do *DL*. Assim, fatores como mutação, migração, seleção e tamanho efetivo populacional reduzido (deriva genética devida à amostragem) podem ocasionar a associação não-aleatória dos alelos nos indivíduos (Lewontin, 1988; Resende Jr *et al.*, 2013).

Alguns fatores favorecem o *DL* de uma população, como o tamanho populacional pequeno, o isolamento genético entre linhagens ou a subdivisão populacional, em que pode ter perda de alelos, que conduzem a associações preferenciais. Além disso, pode ocorrer deriva genética, que promove a flutuação de alelos ao longo das gerações, fazendo que locos se associem de maneira não aleatória. A seleção artificial e natural de indivíduos leva à perda de alelos para locos específicos e, portanto, poucas combinações alélicas são passadas às seguintes gerações, podendo gerar maior *DL*. O sistema de acasalamento também determina os níveis de *DL*, por exemplo, em espécies de reprodução autógama o *DL* tende a ser maior do que na espécie de polinização cruzada (Flint-Garcia *et al.*, 2003; Gupta *et al.*, 2005).

Por outro lado, a redução do *DL* ocorre em caso de mistura populacional, resultando na alteração das frequências alélicas, com isso, o *DL* cai rapidamente quando há acasalamentos ao acaso. No entanto, a recombinação é o principal fator que provoca a diminuição do *DL* intra-cromossômico, e a segregação independente que diminui *DL* inter-cromossômico. A mutação exerce papel duplo, sendo importante tanto na criação de *DL* quanto na redução ao longo das gerações (Flint-Garcia *et al.*, 2003).

3.3.1. Quantificação do DL

As duas estatísticas mais usadas para quantificar o DL são o coeficiente padronizado de desequilíbrio (D') (Lewontin, 1964) e o quadrado do coeficiente de correlação entre dois locos (r^2) (Hill & Robertson, 1968). Essas estatísticas medem o grau de associação entre locos de acordo com a variação dos seus alelos.

Considerando dois locos, um com alelos “A” e “a” e o outro com alelos “B” e “b”, e frequências p_A , p_a , p_B , p_b , respectivamente. Quando não ocorre recombinação, a frequência gamética do haplótipo $AB = \pi_{AB} = p_A p_B$, ou seja, é igual ao produto das frequências individuais. Logo, a frequência gamética entre cromossomos que sofrem recombinação seria $AB = p_A p_B + D$, sendo D o desvio da frequência esperada (coeficiente de desequilíbrio de ligação). Então, o componente básico para o cálculo de desequilíbrio é a diferença entre a frequência esperada (não recombinantes) e a observada (recombinantes) dos gametas, dada por: $D = (\pi_{AB} - p_A p_B)$.

Também é possível expressar o D como a diferença entre o produto de frequências dos haplótipos em acoplamento (AB e ab) e o produto das frequências dos haplótipos em repulsão (Ab e aB): $D = \pi_{AB} \pi_{ab} - \pi_{Ab} \pi_{aB}$.

D sofre influência de mudanças nas frequências dos alelos e, portanto, faz necessário recorrer a medidas relativas, para possibilitar a comparação do DL de diferentes pares de locos. Assim, uma medida de desequilíbrio é D' , que reescala o D entre os valores de 0 e 1. É definido como:

$$D' = \frac{D}{D_{max}}$$

Quando D for positivo, $D_{max} =$ maior valor de $p_A p_b$ e $p_a p_B$, e quando D for negativo, $D_{max} =$ maior valor de $-p_A p_B$ e $-p_a p_b$ (Hartl & Clark, 1997).

Outra medida de desequilíbrio de ligação é fornecida por:

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

Onde r^2 também pode ser considerando como o quadrado do coeficiente de correlação entre dois locos, codificados como 0 e 1 (Hartl & Clark, 1997).

Segundo Resende *et al.* (2012a), o r^2 tem três interpretações: (i) o desvio da frequência observada de haplótipos em relação à esperada segundo segregação independente; (ii) o

quadrado da correlação (r) entre alelos; (iii) a proporção da variação no *QTL* explicada pelo marcador.

Ambas estatísticas, D' e r^2 pressupõem que os polimorfismos são bialélicos, no entanto, representam diferentes aspectos das associações gaméticas. D' capitaliza eventos de recombinação, enquanto r^2 capitaliza eventos de mutação e recombinação ocorridos historicamente. O estatístico r^2 é mais adequado para o mapeamento fino, porque está mais relacionado com a distância. Além disso, D' é fortemente afetado por amostras pequenas, de modo que a comparação entre locos contendo baixas frequências alélicas torna-se inadequada (Flint-Garcia *et al.*, 2003; Resende Jr *et al.*, 2013).

3.3.2. Redução do desequilíbrio de ligação ao longo das gerações

Devido aos eventos de recombinação a cada geração, uma proporção de haplótipos em acoplamento (AB/ab) se transforma em haplótipos em repulsão (Ab/aB) e vice-versa, reduzindo assim o *DL* a cada geração. O *DL* em uma geração qualquer pode ser calculado com a expressão:

$$DL = (1 - c)^t DLo$$

em que *DLo* é o *DL* inicial, t é o número de gerações de cruzamento ao acaso e c é a fração de recombinação entre os locos *A* e *B* (Hamilton, 2009; Hedric'k, 2010). As frequências alélicas permanecem constantes ao longo das gerações. Pela expressão anterior é possível notar que o *DL* tende a reduzir ao longo das gerações, tendendo a zero, se nenhum outro processo, além da recombinação, estiver agindo na população. Essa redução é função do tempo (t) e da fração de recombinação (c) entre os locos. Assim, quanto menor c , menor é a probabilidade de um evento de recombinação ocorrer, e, portanto, maior é a associação preferencial entre eles devido à ligação física, fazendo com que a diminuição do *DL* ocorra mais lentamente (Hamilton, 2009; Hedric'k, 2010).

A diminuição do *DL* é maior nos cruzamentos de espécies alógamas do que em espécies autóginas, isto é devido à recombinação menos eficaz em espécies de autopolinização, onde os indivíduos são mais propensos a ser homocigotos (Flint-Garcia *et al.*, 2003; Gupta *et al.*, 2005). Isto explica os resultados de Barrera (2013), quem fazendo análises de seleção genômica, por meio de simulação, encontrou que este método pode ser vantajoso nas espécies autóginas, já que os valores de confiabilidade permanecem acima da herdabilidade do caráter após 10 gerações de autofecundação, comparada com espécies com polinização cruzada.

3.4. Seleção genômica

Se um grande número de marcadores espalhados por todo o genoma de um indivíduo são analisados, é provável que alguns destes marcadores estarão em *DL* com os *QTL* de uma característica de interesse. Dessa forma, pode-se trabalhar com todo o conjunto de genes (*QTL*) simultaneamente. Logo, no estudo de uma população, os efeitos destes marcadores podem ser estimados com base em dados fenotípicos, gerando modelos de predição dos valores genéticos genômicos (*VGG*) dos indivíduos (também chamados valor de melhoramento estimado genômicamente - *GEBV* Genomic Estimated Breeding Value). Posteriormente, esses modelos poderiam ser aplicados em populações de seleção sem a necessidade de fenotipagem (Crossa *et al.*, 2011).

Essa é a ideia básica da seleção genômica que foi proposta por Meuwissen *et al.* (2001), e que nos últimos anos tem-se desenvolvido pela possibilidade de saturação dos genomas, sobretudo graças às tecnologias de sequenciamento de *DNA* de nova geração, que permitem a identificação mais eficiente de polimorfismos de nucleotídeo único (Single Nucleotide Polymorphism, *SNP*). Além disso, estes mesmos autores forneceram as primeiras propostas para gerar as estimativas dos efeitos dos marcadores, contornando os problemas de ter uma relação desequilibrada do número informações genômicas e fenotípicas, e problemas por colinearidade derivadas da associação entre marcadores. Esses problemas são mais bem explicados no item 3.4.3.

O método envolve três etapas: (i) Estabelecimento do modelo de predição e estimação dos efeitos dos marcadores, (ii) Validação do modelo, e (iii) Predição dos valores genéticos e seleção dos genótipos.

O sistema mais utilizado para o estabelecimento e validação do modelo de predição usa uma população de referência, que em forma prática será dividida em duas subpopulações: a primeira que servirá para a estimação e a segunda na validação, que geralmente é de menor tamanho. A população de referência é fenotipada em condições normais de cultivo, permitindo uma adequada avaliação fenotípica das características de interesse do melhoramento. Ao mesmo tempo esta mesma população é genotipada com alguns milhares de marcadores de *DNA* (*SNP*, *DArT* ou *SSR*), em todo o genoma.

Na primeira etapa, a partir dos conjuntos de dados genotípicos e fenotípicos, os efeitos de todos os marcadores são estimados e o modelo de predição é gerado, em cada uma das características avaliadas.

A validação cruzada do modelo se faz utilizando a informação da segunda população, que não participou do processo de estimação dos efeitos dos marcadores, e que, portanto, é independente da população de estimação. Isso permite que os erros dos *VGG* preditos e dos valores fenotípicos sejam independentes, e toda correlação entre esses valores (tecnicamente chamada de capacidade preditiva), seja de natureza genética e equivalente à própria acurácia (Legarra *et al.*, 2008; Resende *et al.*, 2008).

O *VGG* de cada indivíduo é o valor predito em um modelo de regressão, onde as variáveis explicativas são o número de um dado alelo em cada marcador. Com isso os coeficientes de regressão são necessariamente o efeito de substituição alélica dos respectivos marcadores, conforme teoricamente apresentado em Falconer e Mackay (1996). De forma simplificada, o *VGG* de cada indivíduo é calculado multiplicando o número de um dos alelos observados em cada um dos marcadores pelo efeito estimado do marcador.

A prática efetiva da seleção genômica é realizada na população de seleção, composta por plantas jovens, geradas, por exemplo, a partir do cruzamento dos mesmos genitores elite que deram origem ao teste de progênie utilizado como população de treinamento (Grattapaglia, 2014). Nesta geração não é necessária a fenotipagem e só será realizada a genotipagem. Aqui as equações de predição, derivadas na população de estimação, são usadas na predição dos *VGG* ou fenótipos futuros dos candidatos à seleção (Resende *et al.*, 2008).

Estudos de simulação sobre *GS* para vários caracteres (Bernardo & Yu, 2007; Resende *et al.*, 2008; Wong & Bernardo, 2008; Heffner *et al.*, 2009; Zhong *et al.*, 2009; Grattapaglia & Resende, 2011; Hayes *et al.*, 2013) demonstraram que a *GS* produz ganhos por unidade de tempo maiores que a seleção fenotípica e a seleção assistida por marcadores (*MAS*). Esses ganhos são devidos à redução do intervalo entre ciclos de melhoramento e à maior acurácia de seleção. Informações baseadas em dados reais têm ratificado estas respostas, tanto em plantas anuais (Lorenzana & Bernardo, 2009; Albrecht *et al.*, 2011; Crossa *et al.*, 2011; Fritsche-Neto, 2011; Heslot *et al.*, 2012) quanto em perenes (Resende Jr, 2010; Cavalcanti *et al.*, 2012; Cros *et al.*, 2015; Muranty *et al.*, 2015).

A diminuição do intervalo de tempo necessário para completar cada ciclo seletivo é possível em virtude da seleção precoce. Em espécies perenes, este fator possibilita

maiores ganhos com uso de *GS* em relação à seleção fenotípica, mesmo com acurácias semelhantes para os dois tipos abordagem (Resende, 2008; Resende *et al.*, 2011).

3.4.1. Acurácia da *GS*

Na *GS* o aumento da acurácia de predição é devido à maior detecção da variância genética entre genótipos, em função da genotipagem com alta densidade de marcadores. A acurácia ($r_{\hat{g}g}$) é definida como a correlação entre valores genômicos verdadeiros (g) e valores genômicos estimados (\hat{g}), e seu quadrado ($r_{\hat{g}g}^2$) é capacidade preditiva da *GS* em estimar os genótipos, que informa o quanto o valor estimado é próximo do valor real (Resende *et al.*, 2012a).

Na prática, a acurácia da seleção genômica é avaliada pela abordagem da validação cruzada, por meio dos procedimentos explicados no item 3.4, onde são usadas duas populações: uma população de estimação e outra de validação. Como a sub-população de validação não foi envolvida na predição dos efeitos dos marcadores, os erros dos *VGG* preditos e dos valores fenotípicos são independentes e, portanto, toda correlação entre estes valores é de natureza genética e equivale à capacidade preditiva da *GS* em estimar os fenótipos ($r_{\hat{g}y}$). Logo, a acurácia preditiva do genótipo está relacionada com a capacidade preditiva do fenótipo pela expressão: $r_{\hat{g}g} = r_{\hat{g}y}/h$, quando são usados valores fenotípicos brutos para o cálculo da correlação, onde h é a raiz quadrada da herdabilidade individual. Para Resende (2008) e Legarra *et al.* (2008) $r_{\hat{g}y} = r_{\hat{y}y}$, onde \hat{y} é o fenótipo predito.

Teoricamente, a acurácia da *GS* ($r_{\hat{g}g}$) depende da proporção da variação genética explicada pelos marcadores (r_{mq}^2) e da confiabilidade da estimativa dos efeitos dos marcadores (r_{mm}^2), segundo a expressão de Resende (2008): $r_{\hat{g}g} = (r_{mq}^2 \cdot r_{mm}^2)^{1/2}$.

Resende *et al.* (2012a) apresentaram a seguinte equação, derivada da expressão de Goddard *et al.* (2011), para o cômputo da acurácia esperada da *GS*:

$$r_{\hat{g}g} = \sqrt{\frac{r_{mq}^2}{1 + \frac{1}{N r_{mq}^2 h^2 / n_Q}}}$$

Em que r_{mq}^2 é a proporção da variação genética explicada pelos marcadores (magnitude do desequilíbrio de ligação), N é o número de indivíduos genotipados e fenotipados, h^2

é herdabilidade no sentido restrito do caráter, e n_Q é o número de locos que controlam o caráter.

A partir da equação anterior foi construída a Figura 1, em que os valores de acurácia preditiva variam em relação ao tamanho populacional de estimação, para uma característica oligogênica (baixo número de genes controladores e media a alta herdabilidade), em quatro cenários com dois fatores: a herdabilidade e a variação genética explicada pelos marcadores. Nos quatro cenários, o tamanho populacional ótimo é rapidamente alcançado, sendo inferior a 500 indivíduos, independentemente do valor de r_{mq}^2 , onde um número maior não vai contribuir significativamente para o aumento da $r_{\hat{g}g}$. Além disso, fica evidente que o principal elemento determinante da acurácia seletiva é a proporção da variação genética explicada pelos marcadores (r_{mq}^2).

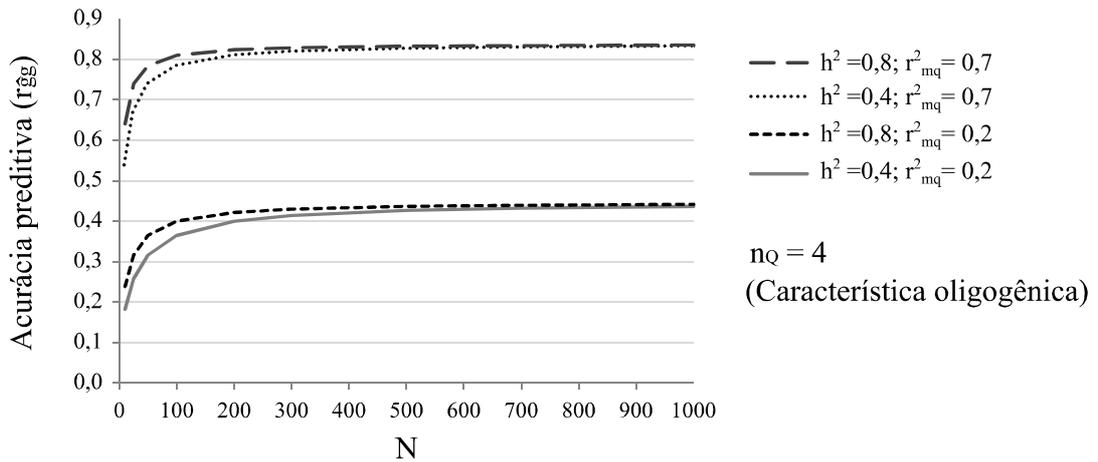


Figura 1. Estimação da acurácia de uma característica oligogênica em quatro cenários de avaliação. h^2 : herdabilidade; n_Q : número de locos que controlam o caráter; r_{mq}^2 : proporção da variação genética explicada pelos marcadores e N : indivíduos da população de estimação.

Em síntese, a acurácia da GS ($r_{\hat{g}g}$) é inversamente proporcional ao número de locos controlando o caráter e diretamente proporcional à herdabilidade do caráter, ao número de indivíduos na população e ao número de marcadores. Mas, especificamente, na avaliação de características oligogênicas, o principal parâmetro a ser considerado é a r_{mq}^2 , mediante o manejo da densidade de marcadores e da observação do padrão de desequilíbrio de ligação que existe na população.

3.4.2. Densidade de marcadores

O número de marcadores para alcançar a ótima cobertura do genoma vai depender da taxa de *DL*. Assim, uma população de melhoramento com tamanho efetivo grande e taxa de recombinação alta vai precisar de uma densidade de marcadores maior (Hill & Robertson, 1968).

Para cobrir efetivamente um genoma com marcadores, o ideal é que cada *QTL* esteja em desequilíbrio de ligação com um marcador. Dessa forma, serão capturados os efeitos de todos os locos para explicar a maioria ou toda a variação genética (Resende *et al.*, 2012a). Além disso, no contexto da *GS* é necessário que o *DL* entre marcadores e *QTL* da população de estimação permaneça nas populações de seleção, de forma duradoura em várias gerações, mantendo as acurácias adequadas (Goddard, 2009). Assim, ter alta densidade de marcadores reduz a perda de acurácia em função da divergência genética ou múltiplas gerações de seleção (Muir, 2007; Zhong *et al.*, 2009).

Por outro lado, também é ideal, do ponto de vista prático, que um número reduzido de marcadores explique a maioria da variação genética. Neste sentido, Habier *et al.* (2009), mediante simulação, analisaram a perda de acurácia dos *VGG* usando duas estratégias para diminuir a densidade de marcadores na *GS*. A primeira foi o uso de marcadores uniformemente espaçados, cujo resultado foi favorável, ainda com uma perda moderada de acurácia. Com esta mesma estratégia um resultado similar foi obtido por Moser *et al.* (2010) em gado leiteiro. Eles com 1000 *SNPs* igualmente espaçados obtiveram o 85% da acurácia gerada com 42500 *SNPs* aleatórios.

A segunda estratégia de Habier *et al.* (2009) foi a seleção de marcadores informativos, em que a perda de acurácia também foi moderada. Isso coincide com os resultados de Resende Jr *et al.* (2012), que avaliaram a acurácia da *GS* em *Pinus*, trabalhando com características relacionadas à resistência à ferrugem. Usando seleção de marcadores com maiores efeitos absolutos, encontraram acurácia preditiva de até 61% maior do que usando o total de marcadores, passando de 4800 a menos de 600. Recentemente, em soja, Ma *et al.* (2016) também avaliaram o efeito de marcadores pré-selecionados na acurácia de predição. Concluíram que a seleção de marcadores por bloco de haplótipos foi adequada para a predição de valores genômicos do rendimento de grãos, melhorando a predição em 4% em comparação com a amostragem de marcador aleatória ou equidistante.

3.4.3. Métodos de estimação dos efeitos dos marcadores e de predição de valores genômicos

Dos métodos de estimação da seleção genômica ressalta-se a estimação simultânea dos efeitos genéticos dos marcadores, sem uso de testes de significância para marcadores individuais. Esses métodos permitem tanto a captura de efeitos genéticos grandes quanto pequenos e, deste modo, capturando quase a totalidade da variância genética aditiva (Heffner *et al.*, 2009; Jannink *et al.*, 2010).

A seleção genômica, para a predição do desempenho de genótipos, deve superar dois problemas: a colinearidade, proveniente do desequilíbrio de ligação entre os marcadores e, sobretudo, a dimensionalidade, já que o número de parâmetros que precisam ser estimados é muito maior que o número de observações fenotípicas disponíveis. Neste sentido, várias abordagens estatísticas de predição de VGG têm sido propostas.

Os principais métodos podem ser divididos em regressão com redução dimensional, regressão implícita e regressão explícita (Resende *et al.*, 2012a). Por sua vez, os métodos de regressão com redução dimensional incluem os de componentes independentes (ICR), quadrados mínimos parciais (PLSR), e de componentes principais (PCR) (Solberg *et al.*, 2009). Na classe de regressão implícita, estão as redes neurais artificiais (RBFNN - radial basis function neural networks) (González-Camacho *et al.*, 2012) e a regressão Kernel não-paramétrica, via modelos aditivos generalizados (Gianola *et al.*, 2006).

Os métodos da classe de regressão explícita foram os primeiros a serem explorados e também tem sido os mais difundidos. Eles podem ser divididos em dois grupos:

(i) métodos de estimação penalizada, como *RR-BLUP* (Random “Ridge” Regression-Best Linear Unbiased Prediction) (Whittaker *et al.*, 2000; Meuwissen *et al.*, 2001), *LASSO* (Least Absolute Shrinkage and Selection Operator) (Tibshirani, 1996), Rede Elástica (Elastic Net) (Zou & Hastie, 2005) e *RR-BLUP-Het* (*RR-BLUP* com heterogeneidade de variância genética) (Resende *et al.*, 2011). Os métodos de estimação penalizada diferem de acordo com as funções de penalização usadas, as quais produzem diferentes graus de redução dimensional. Essa redução previne a super-parametrização e pode conduzir a redução do erro quadrático médio de estimação.

(ii) métodos de estimação bayesiana, tais como *BayesA*, *BayesB* (Meuwissen *et al.*, 2001), *Fast BayesB* (Meuwissen, 2009), *Bayes C*, *BayesC π* , *BayesD*, *BayesD π* (Habier *et al.*, 2011), *LASSO* Bayesiano (Park & Casella, 2008; Campos *et al.*, 2009), *IBLASSO*

(*LASSO* Bayesiano melhorado) (Legarra *et al.*, 2011) e outros. Na estimação bayesiana, a redução das estimativas dos efeitos do modelo é controlada pela distribuição a priori assumida para esses efeitos. Diferentes prioris induzem a diferentes métodos.

De forma geral, os métodos estatísticos proporcionam diferentes suposições sobre o número e distribuição dos efeitos dos *QTL*, apresentando diferentes abordagens sobre o modelo genético associado ao caráter estudado, afetando as respectivas acurácia do *VGG*. A Tabela 2, modificada de Kumar *et al.* (2014), apresenta um resumo das presunções e características de alguns métodos.

Na estimação penalizada, o método *RR-BLUP* tem sido um dos primeiros propostos e uma das abordagens mais usadas. Isso por ser um dos métodos mais simples para modelar o efeito dos marcadores, contornando os problemas de dimensionalidade e multicolinearidade.

O método usa a regressão “*Ridge*”, com a qual se pretende equilibrar a bondade de ajuste do modelo com a complexidade do mesmo. A medida de complexidade é definida pela soma de quadrados dos efeitos dos marcadores (β) que são regularizados através de um parâmetro de penalização (λ). O estimador penalizado evita sobreajuste do modelo e estabiliza as estimações em comparação com a regressão por mínimos quadrados, devido a um fator de encolhimento ou “*shrinkage*” (Piepho, 2009). Neste caso o parâmetro de penalização equivale ao cociente entre as variâncias do resíduo e dos efeitos: $\sigma_e^2/\sigma_\beta^2$ (Ruppert *et al.*, 2003).

Se o parâmetro de regressão for associado a $\lambda = \sigma_e^2/\sigma_g^2/n$ tem-se a regressão aleatória (*Random Regretion*) BLUP, em que σ_e^2 e σ_g^2 são a variância residual e variância genética do caráter, respectivamente, e n é o número de locos controlando o caráter, o qual é desconhecido *a priori*. Esse valor pode ser tomado como, $n = [2 \sum p_i (1-p_i)]$, em que p_i é a frequência de um dos alelos do loco i . Alternativamente, o parâmetro de penalização λ pode ser expresso como $(1-h^2)/(h^2/n)$, sendo h^2 a herdabilidade da característica, ou ser determinado por via iterativa, escolhendo-se aquele que maximiza a correlação entre valor fenotípico e valor genético predito na validação cruzada (Resende *et al.*, 2010).

Tabela 2. Suposições de alguns métodos para obter a equação de predição para o cálculo de valores genéticos genômicos na seleção genômica (modificada de Kumar *et al.*, 2014).

Método	Referência	Suposições-chave e características
<i>RR-BLUP</i>	Meuwissen <i>et al.</i> (2001)	Distribuição normal dos efeitos dos <i>SNPs</i> com uma variância comum para o efeito do marcador; um número muito grande de <i>QTLs</i> com pequeno efeito.
<i>Bayes A</i>	Meuwissen <i>et al.</i> (2001)	Distribuição-t dos efeitos dos <i>SNP</i> (<i>B_j</i>); a maior parte dos efeitos dos <i>QTL</i> está perto do zero, mas alguns têm o efeito muito grande. Cada efeito apresenta uma distribuição normal com variância específica.
<i>Bayes B</i>	Meuwissen <i>et al.</i> (2001)	Distribuição de mistura de efeitos nulos, com probabilidade $1 - \pi$, e distribuição-t de efeitos, com probabilidade π ; <i>SNP</i> com efeitos diferentes de zero escolhem-se como em <i>Bayes A</i> .
<i>Bayes C Pi</i>	Habier <i>et al.</i> (2011)	Mistura de distribuição de efeitos nulos e distribuição normal de efeitos; assume uma variação de efeito de marcador comum de todos os marcadores com efeitos diferentes de zero, mas em vez de usar π fixo, estima π .
<i>LASSO</i> Bayesiano	De los Campos <i>et al.</i> (2009)	Distribuição exponencial dupla para os efeitos de marcadores, com variâncias heterogêneas. Tem uma proporção muito grande de <i>SNPs</i> com efeitos perto de zero e pequena proporção com moderado a grande efeito.
Bayes Empírico	Xu (2007)	Semelhante a <i>Bayes A</i> ; em vez de usar <i>MCMC</i> para estimar os parâmetros de variação, um algoritmo de maximização usa-se para reduzir o tempo de cômputo.
<i>RKHS</i>	Gianola <i>et al.</i> (2006)	Usa uma função central para converter a matriz de marcadores no grupo de distâncias entre pares de observações. Num cenário de dimensão finita, isso equivale a modelar o vetor dos valores genéticos como normal multivariada.
Aprendizagem da máquina	Moser <i>et al.</i> (2009)	Estes são métodos essencialmente não paramétricos.
<i>PLSR</i>	Moser <i>et al.</i> (2009)	Não necessita que uma distribuição prévia de efeitos se especifique; computacionalmente menos exigente já que usa dimensões reduzidas de dados.

Esse método utiliza preditores *BLUP*, mas os algoritmos computacionais são diferentes em relação ao *BLUP* tradicional, visto que os efeitos de marcadores são ajustados como covariáveis e não como variáveis classificatórias (Resende *et al.*, 2010). Sua formulação através de um modelo misto permite estimar todos os efeitos alélicos simultaneamente, e tem como vantagem poder estimar diretamente tanto os componentes de variância, quanto o parâmetro de penalização usando REML (*Restricted Maximum Likelihood*) (Paccapelo, 2015). Outra vantagem é que não há necessidade de uso da matriz de parentesco, porque essa matriz baseada em pedigree, usada no *BLUP* tradicional, é substituída pela matriz de parentesco estimada pelos próprios marcadores (Resende *et al.*, 2012a).

Este enfoque assume que os efeitos dos marcadores são aleatórios, com distribuição normal, média zero (0) e variância comum σ_p^2 . Essas pressuposições são equivalentes ao modelo infinitesimal proposto por Fisher (Resende *et al.*, 2010). Contudo, na realidade é provável que as marcas individuais não possuam a mesma variância, com isso, o *RR-BLUP* trata incorretamente todos os efeitos igualmente (Xu, 2003). Por essa razão vários autores têm proposto métodos que formulam um fator de *shrinkage* específico para cada marca, como Bayes A e Bayes B de Meuwissen *et al.* (2001) ou o Bayesian LASSO de Park e Casella (2008).

Então, de acordo com a teoria, o *RR-BLUP* torna-se boa alternativa quando muitos *QTL* controlam a característica de interesse e nenhum deles é de grande efeito. Entretanto, os métodos bayesianos assumem a existência de poucos genes de grandes efeitos e muitos genes com pequenos efeitos. Logo, *a priori* muitos dos efeitos de marcadores são assumidos como zero (*Bayes B*) ou quase zero (*Bayes A*), o que reduz o tamanho do genoma, permitindo o enfoque em regiões do genoma onde realmente existem *QTL*.

Por outro lado, visando a análise de características controladas por poucos genes, os métodos que assumem a mesma variância para todas as marcas podem não ser apropriados. Isto porque algumas das marcas apresentariam grandes efeitos e muitas delas estarão em regiões que não influenciam o valor mensurado da característica.

O método *LASSO* e especificamente o método *BLASSO* (Bayesian *LASSO*), proposto por Park e Casella (2008) e adaptado para seleção genômica por De Los Campos *et al.* (2009), é adequado quando é considerado que alguns genes de efeito maior dominam a característica e que a maioria dos marcadores tenham efeitos pertos de zero. Este método pressupõe que o efeito dos marcadores segue a distribuição de Laplace

(exponencial dupla), com variâncias heterogêneas, ou seja, considera que existam algumas marcas com maior efeito associado à característica e faz seleção direta de covariáveis, uma vez que a predição do efeito da maioria dos marcadores é próxima de zero (Resende *et al.*, 2012a).

As distribuições exponencial e normal podem ser comparadas na Figura 2, na qual se observa que a densidade de Laplace coloca mais massa no zero e tem caudas mais robustas do que a distribuição normal. Conforme a interpretação De Los Campos *et al.* (2009), o método *BLASSO* produz uma penalização mais forte para os efeitos dos marcadores próximos a zero e menor naqueles com maiores valores absolutos.

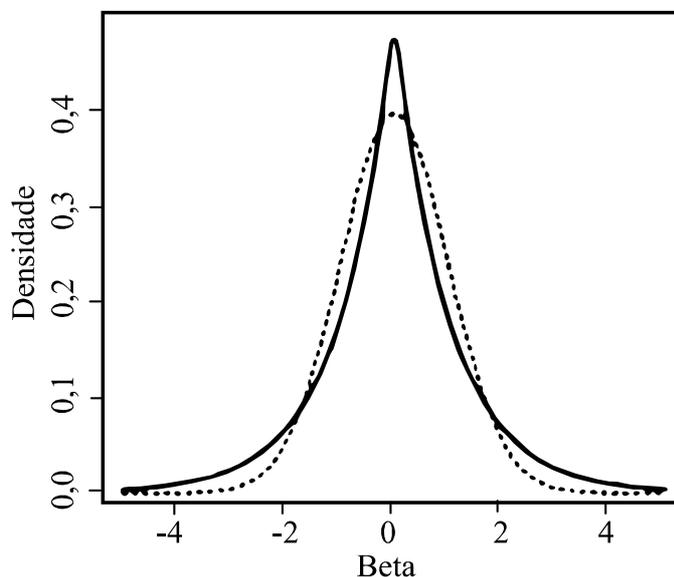


Figura 2. Densidades das distribuições normal (curva pontilhada) e exponencial dupla (curva cheia), ambas com médias iguais a zero e variâncias iguais a um.

Uma qualidade do *BLASSO* é que aplica um fator de penalização específico para cada marca. A debilidade da abordagem, segundo Legarra *et al.* (2011), é que ele inclui um termo de variância comum para modelar ambos os termos, os resíduos e os efeitos genéticos dos marcadores, não sendo necessariamente iguais, segundo a decomposição de variâncias propostas na genética quantitativa.

4. REFERÊNCIAS

- ALBRECHT, T.; WIMMER, V.; AUINGER, H.; ERBE, M.; KNAAK, C.; OUZUNOVA, M.; SIMIANER, H.; SCHÖN, C. C. Genome-based prediction of testcross values in maize. **Theoretical and Applied Genetics**, v. 123, n. 2, p. 339-350. 2011.
- ALKIMIM, E. R.; CAIXETA, E. T.; SOUSA, T. V.; PEREIRA, A. A.; DE OLIVEIRA, A. C. B.; ZAMBOLIM, L.; SAKIYAMA, N. S. Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease. **Molecular Breeding**, v. 37, n. 1, p. 6. 2017.
- ALMEIDA, R. V. **Parâmetros genéticos e alterações nas frequências alélicas em três ciclos de seleção divergente para tolerância ao alumínio em milho**. 2007. 51 f. Tese (MSc Genética e melhoramento)- Universidade Federal de Viçosa, Viçosa-MG, 2007.
- BARRERA, C. F. **Seleção genômica ampla em gerações avançadas de cruzamentos ao acaso ou autofecundação**. 2013. 76 f. Tese (Doutorado em Genética e Melhoramento)- Universidade Federal de Viçosa, Viçosa MG, 2013.
- BERNARDO, R. Genomewide selection when major genes are known. **Crop Science**, v. 54, n. 1, p. 68-75. 2014.
- BERNARDO, R.; YU, J. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v. 47, n. 3, p. 1082-1090. 2007.
- BLAIR, M.; ASTUDILLO, C.; GRUSAK, M. A.; GRAHAM, R.; S.E., B. Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). **Journal of Molecular Breeding**, v. 23, n. 2, p. 197-207. 2009.
- BLAIR, M.; HERRERA, A.; SANDOVAL, T. A.; CALDAS, G. V.; FILLEPPI, M.; SPARVOLI, F. Inheritance of seed phytate and phosphorus levels in common bean (*Phaseolus vulgaris* L.) and association with newly-mapped candidate genes. **Journal of Molecular Breeding**, v. 30, n. 3, p. 1265-1277. 2012.
- BORÉM, A.; MIRANDA, G. **Melhoramento de Plantas**. Viçosa-MG: Editora UFV, 2013.
- BRITO, G. G.; CAIXETA, E. T.; GALLINA, A. P.; ZAMBOLIM, E. M.; ZAMBOLIM, L.; DIOLA, V.; LOUREIRO, M. E. Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. **Euphytica**, v. 173, n. 1, p. 255-264. 2010.
- BYRNE, D. H. Molecular marker use in perennial plant breeding. **Acta Horticulturae**, v. 751, p. 163-167. 2007.
- CAMPOS, G. D. L.; NAYA, H.; GIANOLA, D.; CROSSA, J.; LEGARRA, A.; MANFREDI, E.; WEIGEL, K.; COTES, J. M. Predicting quantitative traits with regression models for dense molecular markers. **Genetics**, v. 182, n. 1, p. 375-385. 2009.

CAPUCHO, A. S.; CAIXETA, E. T.; ZAMBOLIM, E. M.; ZAMBOLIM, L. Herança da resistência do Híbrido de Timor UFV 443-03 à ferrugem-do-cafeeiro. **Pesquisa Agropecuária Brasileira**, v. 44, n. 3, p. 276-282. 2009.

CAVALCANTI, J. J. V.; RESENDE, M. D. V.; SANTOS, F. H. C.; PINHEIRO, C. R. Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. **Revista Brasileira de Fruticultura**, v. 34, n. 3, p. 840-846. 2012.

CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira: V. 4 Café. Safra 2017, primeiro levantamento**. Brasília. 2017. p. 98.

CROS, D.; DENIS, M.; SÁNCHEZ, L.; COCHARD, B.; FLORI, A.; DURAND-GASSELIN, T.; NOUY, B.; OMORÉ, A.; POMIÈS, V.; RIOU, V.; SURYANA, E.; BOUVET, J.-M. Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). **Theoretical and Applied Genetics**, v. 128, n. 3, p. 397-410. 2015.

CROSSA, J.; PÉREZ, P.; DE LOS CAMPOS, G.; MAHUKU, G.; DREISIGACKER, S.; MAGOROKOSHO, C. Genomic Selection and Prediction in Plant Breeding. **Journal of Crop Improvement**, v. 25, n. 3, p. 239-261. 2011.

CROSSA, J.; PEREZ, P.; HICKEY, J.; BURGUENO, J.; ORNELLA, L.; CERON-ROJAS, J.; ZHANG, X.; DREISIGACKER, S.; BABU, R.; LI, Y.; BONNETT, D.; MATHEWS, K. Genomic prediction in CIMMYT maize and wheat breeding programs. **Heredity**, v. 112, n. 1, p. 48-60. 2014.

CRUZ, R. P.; DUARTE, I. T. L.; CABREIRA, C. Inheritance of rice cold tolerance at the seedling stage. **Scientia Agricola**, v. 67, n. 6, p. 669-674. 2010.

CHAPARRO, A. P.; CRISTANCHO, M. A.; CORTINA, H. A.; GAITAN, A. L. Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. **Genetic Resources and Crop Evolution** v. 51, n. 1, p. 291-297. 2004.

DACHS, J. N. W. **Estatística computacional: uma introdução ao Turbo Pasca**. Rio de Janeiro: LTC, 1988. 236 p.

DIOLA, V.; DE BRITO, G. G.; CAIXETA, E. T.; MACIEL-ZAMBOLIM, E.; SAKIYAMA, N. S.; LOUREIRO, M. E. High-density genetic mapping for coffee leaf rust resistance. **Tree Genetics & Genomes**, v. 7, n. 6, p. 1199-1208. 2011.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. Edinburgh: Longman Group Limited, 1996. 464 p.

FERREIRA, D. F. Uso de simulação no melhoramento. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. D.; VALADARES-INGLIS, M. C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis -Brasil: Fundação-MT, 2001. p. 1119-1141.

FLINT-GARCIA, S.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of linkage disequilibrium in plants. **Annual Review of Plant Biology**, v. 54, n. 1, p. 357-374. 2003.

FRITSCHÉ-NETO, R. **Seleção genômica ampla e novos métodos de melhoramento do milho**. 2011. 39 f. Tese (Doutorado em genética e melhoramento de plantas)- Universidade Federal de Viçosa, Viçosa, 2011.

GELETA, M.; HERRERA, I.; MONZÓN, A.; BRYNGELSSON, T. Genetic diversity of arabica Coffee (*Coffea arabica* L.) in Nicaragua as estimated by simple sequence repeat markers. **The Scientific World Journal**, v. 2012, p. 11. 2012.

GIANOLA, D.; FERNANDO, R. L.; STELLA, A. Genomic-assisted prediction of genetic value with semiparametric procedures. **Genetics**, v. 173, n. 1, p. 1761-1776. 2006.

GICHURU, E. K.; AGWANDA, C. O.; COMBES, M. C.; MUTITU, E. W.; NGUGI, E. C. K.; BERTRAND, B.; LASHERMES, P. Identification of molecular markers linked to a gene conferring resistance to coffee berry disease (*Colletotrichum kahawae*) in *Coffea arabica* L. **Plant Pathology**, v. 57, n. 6, p. 1117-1124. 2008.

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

GODDARD, M. E.; HAYES, B. J.; MEUWISSEN, T. H. E. Using the genomic relationship matrix to predict the accuracy of genomic selection. **Journal of Animal Breeding and Genetics**, v. 128, n. 6, p. 409-421. 2011.

GONZÁLEZ-CAMACHO, J. M.; CAMPOS, G. D. L.; PEREZ, P.; GIANOLA, D.; J., C.; MAHUKU, G.; BABU, R.; CROSSA, J. Genome-enabled prediction of genetic values using radial basis function. **Theoretical and Applied Genetics**, v. 125, n. 1, p. 759-771. 2012.

GRATTAPAGLIA, D. **Status atual da genômica aplicada ao melhoramento florestal e o novo paradigma da seleção genômica ampla**. In: 3º Encontro Brasileiro de Silvicultura, Campinas, SP, Brasil. 2014. p. 99-126.

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

GUERREIRO FILHO, O.; SILVAROLLA, M. B.; ESKES, A. B. Expression and mode of inheritance of resistance in coffee to leaf miner *Perileucoptera coffeella*. **Euphytica**, v. 105, n. 1, p. 7-15. 1999.

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.

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

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. Genomic selection using low-density marker panels. **Genetics**, v. 182, n. 1, p. 343-353. 2009.

HABIER, D.; FERNANDO, R. L.; KIZILKAYA, K.; GARRICK, D. J. Extension of the bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, n. 186, 12 p. 2011. Disponível em: <http://www.biomedcentral.com/1471-2105/12/186>. Acesso em: 1 Abr 2015.

HAMILTON, M. B. **Population genetics**. Oxford: Wiley-Blackwell, 2009. 407 p.

HARTL, D. L.; CLARK, A. G. **Principles of population genetics**. Sunderland, Massachusetts: Sinauer Associates Inc, 1997. 542 p.

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.

HEDRICK, P. W. **Genetics of populations**. Sudbury, MA: Jones and Bartlett Publishers, 2010. 675 p.

HEFFNER, E. L.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection for crop improvement. **Crop Science**, v. 49, n. 1, 12 p. 2009. Disponível em: <http://dx.doi.org/10.2135/cropsci2008.08.0512>. Acesso em: 12 May 2015.

HERRERA, J. C.; ALVARADO A, G.; CORTINA G, H. A.; COMBES, M.-C.; ROMERO G, G.; LASHERMES, P. Genetic analysis of partial resistance to coffee leaf rust (*Hemileia vastatrix* Berk & Br.) introgressed into the cultivated *Coffea arabica* L. from the diploid *C. canephora* species. **Euphytica**, v. 167, n. 1, p. 57-67. 2009.

HESLOT, N.; YANG, H. P.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection in plant breeding: a comparison of models. **Crop Science**, v. 52, n. 1, p. 146-160. 2012.

HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, v. 38, n. 1, p. 226-231. 1968.

ICO Organização Internacional do Café. Disponível em: <http://www.ico.org/documents/50%20anniversary/background-info-wcc-salvador-p.pdf>. Acesso em: 14 nov 2014.

ICO Organização Internacional do Café. Disponível em: <http://www.ico.org/prices/production.pdf>. Acesso em: 19 de Janeiro 2017.

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

KAINER, D.; LANFEAR, R.; FOLEY, W. J.; KÜLHEIM, C. Genomic approaches to selection in outcrossing perennials: focus on essential oil crops. **Theoretical and Applied Genetics**, v. 128, n. 12, p. 2351-2365. 2015.

KUMAR, S.; VOLZ, R. K.; CHAGNÉ, D.; GARDINER, S. Breeding for Apple (*Malus × domestica* Borkh.) Fruit Quality Traits in the Genomics Era. In: TUBEROSA, R.; GRANER, A.; FRISON, E. **Genomics of Plant Genetic Resources**. Netherlands: Springer, 2014. Cap. 16, p. 387-416.

- LASHERMES, P.; COMBES, M.; ROBERT, J.; TROUSLOT, P.; D'HONT, A.; ANTHONY, F.; CHARRIER, A. Molecular characterization and origin of the *Coffea arabica* L. genome. **Molecular Genetics and Genomics**, v. 261, n. 1, p. 259-266. 1999.
- LEGARRA, A.; ROBERT-GRANIÉ, C.; MANFREDI, E.; ELSEEN, J.-M. Performance of Genomic Selection in Mice. **Genetics**, v. 180, n. 1, p. 611-618. 2008.
- LEGARRA, A.; ROBERT-GRANIÉ, C.; CROISEAU, P.; GUILLAUME, F.; FRITZ, S. Improved Lasso for genomic selection. **Genetics Research**, v. 93, n. 1, p. 77-87. 2011.
- LEWONTIN, R. C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. **Genetics**, v. 49, n. 1, p. 49-67. 1964.
- LEWONTIN, R. C. On measures of gametic disequilibrium. **Genetics**, v. 120, n. 3, p. 849-852. 1988.
- LIU, S.; ABATE, Z. A.; L., M. A. Inheritance of Fusarium head blight resistance in the soft red winter wheat Ernie. **Theoretical and Applied Genetics**, v. 110, n. 3, p. 454-461. 2005.
- LOPES, F. C. D. C.; GOMES, R. L. F.; FREIRE FILHO, F. R. Genetic control of cowpea seed sizes. **Scientia Agricola**, v. 60, n. 2, p. 315-318. 2003.
- LORENZANA, R. E.; BERNARDO, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and Applied Genetics**, v. 120, n. 1, p. 151-161. 2009.
- MA, Y.; REIF, J. C.; JIANG, Y.; WEN, Z.; WANG, D.; LIU, Z.; GUO, Y.; WEI, S.; WANG, S.; YANG, C.; WANG, H.; YANG, C.; LU, W.; XU, R.; ZHOU, R.; WANG, R.; SUN, Z.; CHEN, H.; ZHANG, W.; WU, J.; HU, G.; LIU, C.; LUAN, X.; FU, Y.; GUO, T.; HAN, T.; ZHANG, M.; SUN, B.; ZHANG, L.; CHEN, W.; WU, C.; SUN, S.; YUAN, B.; ZHOU, X.; HAN, D.; YAN, H.; LI, W.; QIU, L. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.). **Molecular Breeding**, v. 36, n. 8, p. 113. 2016.
- MEUWISSEN, T. H. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. **Genetics Selection Evolution**, v. 41, n. 1, 9 p. 2009. Disponível em: <http://www.gsejournal.org/content/41/1/35>. Acesso em: 12 Feb 2015.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 1, p. 1819-1829. 2001.
- MISSIO, R. F.; CAIXETA, E. T.; ZAMBOLIM, E. M.; PENA, G. F.; ZAMBOLIM, L.; DIAS, L. A. S.; SAKIYAMA, N. S. Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. **Genetics and Molecular Research**, v. 10, n. 4, p. 2366-2381 2011.
- MONCADA, M.; TOVAR, E.; MONTROYA, J. C.; GONZÁLEZ, A.; SPINDEL, J.; MCCOUCH, S. A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. **Tree Genetics & Genomes**, v. 12, n. 5. 2016.

MONCADA, P.; MCCOUCH, S. Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome**, v. 47, n. 3, p. 501-509. 2004.

MOSER, G.; KHATKAR, M. S.; HAYES, B. J.; RAADSMA, H. W. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. **Genetics Selection Evolution**, v. 42, n. 37, 15 p. 2010. Disponível em: <http://www.gsejournal.org/content/42/1/37>. Acesso em: 12 fev 2015.

MOSER, G.; TIER, B.; CRUMP, R. E.; M.S., K.; RAADSMA, H. W. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. **Genetics Selection Evolution**, v. 41, n. 56, 16 p. 2009. Disponível em: <http://www.gsejournal.org/content/41/1/56>. Acesso em: 14 Feb 2015.

MUIR, W. M. Comparison of genomic and traditional BLUP- estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal Animal and Beeding Genetics**, v. 124, n. 1, p. 342-355. 2007.

MURANTY, H.; TROGGIO, M.; SADOK, I. B.; RIFAI, M. A.; AUWERKERKEN, A.; BANCHI, E.; VELASCO, R.; STEVANATO, P.; VAN DE WEG, W. E.; DI GUARDO, M.; KUMAR, S.; LAURENS, F.; BINK, M. C. A. M. Accuracy and responses of genomic selection on key traits in apple breeding. **Horticulture Research**, v. 2, p. 15060. 2015.

OLIVEIRA, A. C. B.; SAKIYAMA, N. S.; CAIXETA, E. T.; CAPUCHO, A. S.; RUFINO, R. **Análise genética da resistência de cafeeiros do Híbrido de Timor à raça II de *Hemileia vastatrix*** In: Congresso brasileiro de pesquisas cafeeiras, São Lourenço, Brasil. 2004.

OLIVEIRA, C. B.; SAKIYAMA, N. S.; CAIXETA, E. T.; ZAMBOLIM, E. M.; RUFINO, R. N.; ZAMBOLIM, L. Mapa parcial de ligação gênica de *Coffea arabica* L. e recuperação do genitor recorrente em progênies de retrocruzamentos. **Crop Breeding and Applied Biotechnology**, v. 7, n. 1, p. 196-203. 2007.

PACCAPELO, M. V. **Modelos de selección genómica para caracteres cuantitativos basados en marcadores moleculares aplicados al mejoramiento de maíz**. 2015. 87 f. Tese (Magíster en Estadística Aplicada)- Universidad Nacional de Córdoba, Argentina, 2015.

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

PEARL, H. M.; NAGAI, C.; MOORE, P. H.; STEIGER, D. L.; OSGOOD, R. V.; MING, R. Construction of a genetic map for arabica coffee. **Theoretical and Applied Genetics**, v. 108, n. 5, p. 829-835. 2004.

PEREIRA, A. A. **Herança da resistência a *Hemileia vastatrix* Berk. et Br. Em cafeeiros derivados do Híbrido de Timor**. . 1995. 66 f. Tese (Doctor Scientiae)- Universidade Federal de Viçosa, Viçosa-MG, 1995.

PEREIRA, G.; PADILHA, L.; PINHO, E.; TEIXEIRA, R.; CARVALHO, C.; MALUF, M.; CARVALHO, B. Microsatellite markers in analysis of resistance to coffee leaf

miner in Arabica coffee. **Pesquisa Agropecuária Brasileira**, v. 46 n. 1, p. 1650-1656. 2011.

PESTANA, K. **Caracterização fenotípica e molecular da resistência do cafeeiro híbrido de timor A *Hemileia vastatrix***. 2010. 60 f. Tese (MSc. Fitotecnia)- Universidade Federal de Viçosa, Viçosa, 2010.

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

PESTANA, N. K.; CAPUCHO, A. S.; CAIXETA, E. T.; ZAMBOLIM, E. M.; ALMEIDA, D.; CRUZ, C. D.; ZAMBOLIM, L.; PEREIRA, A. A.; OLIVEIRA, A. C. B.; SALIYAMA, N. S. Inheritance study and linkage mapping of resistance loci to *Hemileia vastatrix* in Híbrido de Timor UFV 443-03. **Tree Genetics & Genomes**, v. 11. 2015.

PIEPHO, H. P. Ridge regression and extensions for genomewide selection in maize. **Crop Science**, v. 49, n. 4, p. 1165-1176. 2009.

PRAKASH, N. S.; MUNISWAMY, B.; HANUMANTHA, B. T.; SREENATH, H. L.; SUNDARESHA, K. D.; SURESH, N.; SANTHOSH, P.; SOUMYA, P.; ASHA, B. M.; BHAT, S. S.; JAYARAMA. Marker assisted selection and breeding for leaf rust resistance in coffee (*Coffea arabica* L.). **Indian Journal of Genetics & Plant Breeding** v. 71, n. 2, p. 185-189. 2011.

RAMALHO, M. A.; ABREU, A. D. F.; SANTOS, J.; NUNES, J. A. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. Lavras - MG - Brasil: UFLA, 2012.

RESENDE JR, M. F. R. **Seleção genômica ampla no melhoramento vegetal**. 2010. 78 f. Tese (Mestrado em Genética e Melhoramento de Plantas)- Universidade Federal de Viçosa, Viçosa MG, 2010.

RESENDE JR, M. F. R.; A., A. A.; BARRERA, C. F.; RESENDE, M. D. V.; CRUZ, C. D. Seleção genômica ampla. In: CRUZ, C. D.; SALGADO, C. C.; BHERING, L. L. **Genômica aplicada**. Viçosa-MG: UFV, 2013. Cap. 10, p. 375-424.

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

RESENDE, M. D. V. **Genômica quantitativa e seleção no melhoramento de plantas perenes e animais**. Colombo, PR: Embrapa Florestas, 2008. 324 p.

RESENDE, M. D. V.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal**, v. 56, n. 1, p. 63-77. 2008.

RESENDE, M. D. V.; SILVA, F. F.; LOPES, P. S.; AZEVEDO, C. F. **Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana**

(MCMC), **regressão aleatória multivariada e estatística espacial**. Viçosa, MG: Editora UFV, 2012a. 291 p.

RESENDE, M. D. V.; SILVA, F. F.; VIANA, J. M. S.; PETERNELLI, L. A.; RESENDE JÚNIOR, M. F. R.; VALLE, P. M. **Métodos estatísticos na seleção genômica ampla**. Colombo, PR: Embrapa Florestas, 2011.

RESENDE, M. D. V.; RESENDE JÚNIOR, M. F. R.; AGUIAR, A. M.; ABAD, J. I. M.; MISSIAGGIA, A. A.; SANSALONI, C.; PETROLI, C.; GRATTAPAGLIA, D. **Computação da Seleção Genômica Ampla (GWS)**. Colombo, PR: Embrapa Florestas, 2010. 79 p.

RESENDE, M. D. V.; RESENDE, M. F. R.; SANSALONI, C. P.; PETROLI, C. D.; MISSIAGGIA, A. A.; AGUIAR, A. M.; ABAD, J. M.; TAKAHASHI, E. K.; ROSADO, A. M.; FARIA, D. A.; PAPPAS, G. J.; KILIAN, A.; GRATTAPAGLIA, D. Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, v. 194, n. 1, p. 116-128. 2012b.

ROCHA, M. M.; CARVALHO, K. J. M.; FREIRE FILHO, F. R.; LOPES, Â. C. D. A.; GOMES, R. L. F.; SOUSA, I. S. Controle genético do comprimento do pedúnculo em feijão-caupi. **Pesquisa Agropecuária Brasileira**, v. 44, n. 3, p. 270-275. 2009.

RODRIGUES JR, C. J.; BETTENCOURT, A. J.; RIJO, L. Races of the pathogen and resistance to coffee rust. **Annual Review of Phytopathology**, v. 13, p. 49-70. 1975.

ROMERO, G.; VÁSQUEZ, L. M.; LASHERMES, P.; HERRERA, J. C. Identification of a major QTL for adult plant resistance to coffee leaf rust (*Hemileia vastatrix*) in the natural Timor hybrid (*Coffea arabica* x *C. canephora*). **Plant Breeding**, v. 133, n. 1, p. 121-129. 2014.

RUPPERT, D.; WAND, M. P.; CARROLL, R. J. **Semiparametric Regression**. Cambridge University Press, 2003. 416 p.

SALAVA, J.; POLÁK, J.; KRŠKA, B. Oligogenic inheritance of resistance to *plum pox* virus in Apricots. **Czech J. Genet. Plant Breed.**, v. 41, n. 4, p. 167-170. 2005.

SETOTAW, T. A.; CAIXETA, E. T.; PENA, G. F.; ZAMBOLIM, E. M.; PEREIRA, A. A.; SAKIYAMA, N. S. Breeding potential and genetic diversity of "Híbrido do Timor" coffee evaluated by molecular markers. **Crop Breeding and Applied Biotechnology**, v. 10, p. 298-304. 2010.

SETOTAW, T. A.; CAIXETA, E. T.; PEREIRA, A. A.; BAIÃO DE OLIVEIRA, A. C.; CRUZ, C. D.; ZAMBOLIM, E. M.; ZAMBOLIM, L.; SAKIYAMA, N. S. Coefficient of Parentage in *Coffea arabica* L. Cultivars Grown in Brazil. **Crop Science**, v. 53, n. 4, p. 1237-1247. 2013.

SINGH, B. B.; ISHIYAKU, M. F. Genetics of rough seed coat texture in cowpea. **Journal of Heredity**, v. 91, n. 2, p. 170-174. 2000.

SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. E. Reducing dimensionality for prediction of genome-wide breeding value. **Genetics**

- Selection Evolution**, v. 41, n. 29, 8 p. 2009. Disponível em: <http://www.gsejournal.org/content/41/1/29>. Acesso em: 12 May 2015.
- SPINDEL, J.; BEGUM, H.; AKDEMIR, D.; VIRK, P.; COLLARD, B.; REDOÑA, E.; ATLIN, G. N.; JANNINK, J.-L.; MCCOUCH, S. R.; MAURICIO, R. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. **PLoS Genetics**, v. 11(2):e1004982–e1004982, 2015. Disponível em: <http://dx.doi.org/10.1371/journal.pgen.1004982>. Acesso em: 12 May 2016.
- SUN, X.; PENG, T.; MUMM, R. H. The role and basics of computer simulation in support of critical decisions in plant breeding. **Molecular Breeding**, v. 28, n. 4, p. 421-436. 2011.
- TEIXEIRA-CABRAL, T. A.; SAKIYAMA, N. S.; ZAMBOLIM, L.; PEREIRA, A. A.; SCHUSTER, I. Single-locus inheritance and partial linkage map of *Coffea arabica* L. **Crop Breeding and Applied Biotechnology**, v. 4, n. 1, p. 416-421. 2004.
- TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. **Journal of the Royal Statistics Society Series B**, v. 58, n. 1, p. 267-288. 1996.
- VALLEJO, F. A.; ESTRADA, E. I. **Mejoramiento de plantas**. Palmira. Col.: UNal, 2002. 402 p.
- VALLEJO, F. A.; ESPITIA, M.; ESTRADA, E. I.; RAMÍREZ, H. **Genética Vegetal**. Cali: UNAL, 2010.
- VAN DER VOSSEN, H.; BERTRAND, B.; CHARRIER, A. Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. **Euphytica**, v. 204, n. 2, p. 243-256. 2015.
- VAN DER VOSSEN, H. A. M.; WALYARO, D. J. Additional evidence for oligogenic inheritance of durable host resistance to coffee berry disease (*Colletotrichum kahawae*) in arabica coffee (*Coffea arabica* L.) **Euphytica**, v. 165, n. 1, p. 105-111. 2009.
- VIDA, G.; GÁL, M.; UHRIN, A.; VEISZ, O.; SYED, N. H.; FLAVELL, A. J.; WANG, Z.; BEDŐ, Z. Molecular markers for the identification of resistance genes and marker-assisted selection in breeding wheat for leaf rust resistance. **Euphytica**, v. 170, n. 1, p. 67-76. 2009.
- VIEIRA, E. S. N.; VON PINHO, É. V. R.; CARVALHO, M. G. G.; ESSELINK, D. G.; VOSMAN, B. Development of microsatellite markers for identifying Brazilian *Coffea arabica* varieties. **Genetics and Molecular Biology**, v. 33, n. 3, p. 507-514. 2010.
- WANG, J. Modelling and simulation of plant breeding strategies. In: ABDURAKHMONOV, I. Y. **Plant Breeding**. InTech, 2012. Cap. 2, p. 19-40.
- WELLMANN, R.; PREUß, S.; THOLEN, E.; HEINKEL, J.; WIMMERS, K.; BENNEWITZ, J. Genomic selection using low density marker panels with application to a sire line in pigs. **Genetics Selection Evolution**, v. 45, n. 1, 11 p. 2013. Disponível em: <http://www.gsejournal.org/content/45/1/28>. Acesso em: 14 feb 2015.

WHITTAKER, J. C.; THOMPSON, R.; DENHAM, M. C. Marker assisted selection using ridge regression. **Genetical Research**, v. 75, n. 1, p. 249-252. 2000.

WINTGENS, J. N. **Coffee: Growing, processing, sustainable production: A guidebook for growers, processors, traders, and researchers**. Weimheim: Wiley - VCH, 2011. 983 p.

WONG, C. K.; BERNARDO, R. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical and Applied Genetics**, v. 116, n. 6, p. 815-824. 2008.

XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, v. 163, n. 2, p. 789-801. 2003.

XU, S. An empirical Bayes method for estimating epistatic effects of quantitative trait loci. **Biometrics**, v. 63, n. 1, p. 513-521. 2007.

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. 1, p. 3642-3650. 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**, v. 182, n. 1, p. 355-364. 2009.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society B**, v. 67, n. 1, p. 301-320. 2005.

CAPÍTULO I

SIMULAÇÃO E VALIDAÇÃO DE POPULAÇÕES

1. INTRODUÇÃO

Para verificar a viabilidade do melhoramento de características oligogênicas em *Coffea arabica*, com a aplicação de métodos de seleção genômica (ou *Genomic Selection – GS*), usando marcação de baixa cobertura, foram propostas análises de populações simuladas, variando tamanhos populacionais, número de marcadores e valores de herdabilidade.

O uso de análises por simulação é conveniente por facilitar a comparação das estratégias, demandando menor recurso e tempo e pode ser generalizada para diferentes culturas. Ainda, os experimentos podem ser replicados em grande número de vezes, aumentando a precisão das inferências (Gurgel, 2004; Sun *et al.*, 2011; Wang, 2012).

A simulação consiste em imitar o comportamento de um sistema real por meio de modelos, que logo podem ser testados em vários cenários, com a finalidade de gerar informação suficiente para compreender o funcionamento do sistema. Isso permite a previsão do desempenho e a produção de recomendações (Dachs, 1988; Ferreira, 2001; Sun *et al.*, 2011).

Com base nos resultados das experiências de simulação, os melhoristas podem aperfeiçoar seus métodos e aumentar a eficiência do programa. A simulação tem servido para a comparação de estratégias, para assim decidir ou não a sua adoção (Casali & Tigchelaar, 1975; Muehlbauer *et al.*, 1981; Van Oeveren & Stam, 1992; Wang *et al.*, 2003). Além disso, tem sido utilizada para facilitar a seleção de parentais ou prever o desempenho de linhas (Ferreira *et al.*, 2000; Veiga *et al.*, 2000; Gurgel, 2004; Liu *et al.*, 2004; Sun *et al.*, 2011). Na era genômica, as simulações têm sido amplamente usadas na resolução de questionamentos nos estudos de mapeamento com marcadores moleculares (Xu, 2003; Yu *et al.*, 2005; Silva *et al.*, 2007; Bhering & Cruz, 2008; Yu *et al.*, 2008; Rocha *et al.*, 2010; Brito *et al.*, 2014; Peixoto *et al.*, 2014) e recentemente na projeção e delineamento da seleção genômica (Meuwissen *et al.*, 2001; Bernardo & Yu, 2007; Resende *et al.*, 2008; Habier *et al.*, 2009; Zhong *et al.*, 2009; Resende Jr, 2010; Hickey & Gorjanc, 2012; Lorenz, 2013; Rutkoski *et al.*, 2013; Estaghvirou *et al.*, 2014; Ceron-Rojas *et al.*, 2015; Montesinos-López *et al.*, 2016).

O uso da simulação no melhoramento genético deve ser baseada em modelos biológicos, que representem satisfatoriamente os fenômenos de interesse, com estrutura simples e desempenho comparável ao modelo real, para ser manipulada e interpretada

adequadamente (Cruz, 2001). Neste sentido, a teoria da genética quantitativa fornece grande parte da estrutura para a definição dos modelos e a análise da informação, mas a abordagem pode ser diferente dependendo dos objetivos.

Como qualquer outra técnica, na simulação existem erros potenciais associados. Portanto, para o adequado processamento das informações que a simulação produz, previamente é necessário certificar que os valores gerados apresentem coerência com o sistema analisado (Ferreira, 2001). Esta validação é obtida através da comparação de resultados com uma referência apropriada. Para este trabalho, essas referências foram os princípios genéticos que caracterizam os processos de recombinação e segregação e o desenvolvimento de populações endogâmicas.

Diante do exposto, este capítulo tem como objetivo a descrição dos processos de simulação na geração de populações, e a sua respectiva validação, que posteriormente foram usados nas análises dos cenários desenvolvidos no capítulo II.

2. MATERIAL E MÉTODOS

A estratégia geral de análises do trabalho consistiu na comparação de cenários, em que diferentes versões de uma única população F_2 foram usadas para determinar os modelos preditivos da *GS*, e com eles a estimação dos valores genéticos genômicos, para selecionar os indivíduos com os maiores valores em cada cenário. A seleção foi praticada até a F_6 e cada geração de seleção foi simulada a partir do genótipo dos indivíduos selecionados, avançando por meio de autofecundação (Figura 1).

Para simular as populações foram usados dois procedimentos: o primeiro para gerar uma população base, da qual foram adaptadas as diferentes versões, segundo densidade de marcadores e/ou tamanho populacional, usadas nos cenários discutidos no capítulo II. O segundo procedimento foi usado para simular as gerações descendentes, produto das seleções feitas nos diferentes cenários.

2.1. População Base (F_2)

A população base correspondeu a uma geração F_2 com características apropriadas para *GS*, em relação à densidade de marcadores e número de indivíduos. Esta população foi simulada com os módulos respectivos do programa *GENES* (Cruz, 2013).

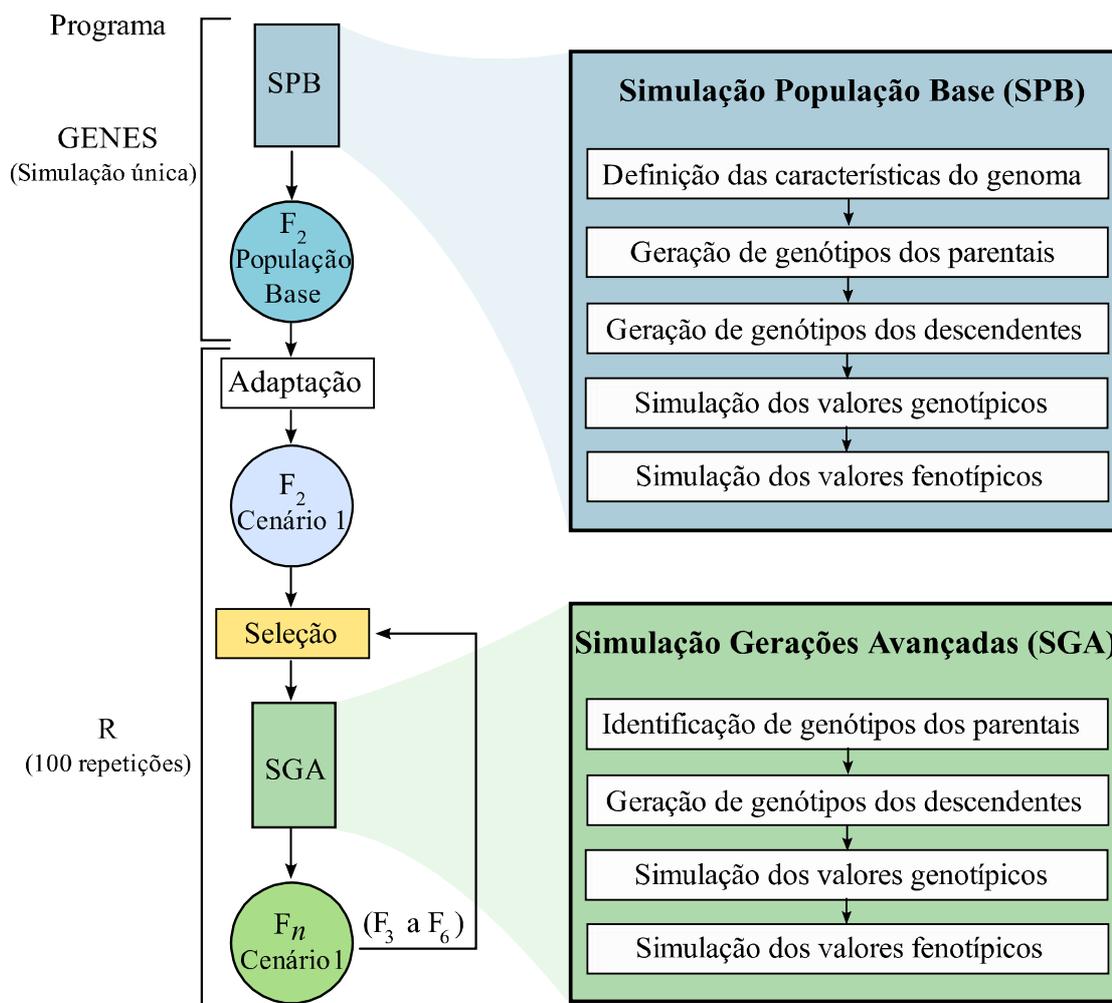


Figura 1. Esquema geral da estratégia de simulação. A simulação da população base (SPB) foi feita uma única vez, adaptando-se para cada cenário segundo tamanho da população e número de marcadores. A simulação de gerações avançadas (SGA) foi feita com o genótipo dos indivíduos selecionados, a cada geração.

2.1.1. Simulação da população base

A população base foi simulada com um genoma de 1040 cM de extensão, com oito grupos de ligação de 130 cM, com saturação de um marcador por centimorgan. Os 1040 locos representaram alelos codominantes, sendo definidos quatro deles como controladores da característica, localizados em quatro grupos de ligações diferentes. A simulação do genoma foi feita com o módulo de *Simulação de Genoma Complexo* do programa *GENES*.

Na estratégia de simulação, as populações foram o produto da hibridação de pais contrastantes, que, por consequência, produziram uma geração F_1 com todos os locos em fase de acoplamento. A partir do genótipo da geração F_1 , foram gerados os gametas para a formação dos indivíduos da população F_2 , composta de 625 indivíduos.

A produção de gametas foi feita simulando-se o pareamento dos cromossomos homólogos e realizando-se permutas ao longo deles, desconsiderando interferência nas regiões delimitadas por dois marcadores adjacentes. A probabilidade de ocorrência de recombinação numa região entre marcadores adjacentes foi dada de acordo com a distância destes marcadores no genoma simulado. Finalmente, para a simulação do genótipo de cada indivíduo o *GENES* considera o encontro aleatório de gametas mediante sorteio (Silva, 2005).

Os valores genotípicos e fenotípicos da população base F_2 foram simulados no módulo de *Simulação Fenotípica* do programa *GENES*. Para a simulação dos valores genotípicos (VG_i) foi definido o controle da característica por quatro locos com igual contribuição (efeito uniforme), e ausência de dominância, usando a expressão:

$$VG_i = \mu + a_i + d_i$$

Sendo:

μ = Constante ou média genotípica populacional ($\mu = 1,0$).

a_i = Efeito aditivo, onde $a_i = \sum_{j=1}^4 p_j \alpha_j$, sendo $p_j = 2,5$ a contribuição de todo loco j para a manifestação da característica e α_j o efeito do alelo favorável no loco j , considerando igual a 1, 0 ou -1 para as classes genotípicas *AA*, *Aa* e *aa*, respectivamente.

d_i = Efeito devido à dominância, considerado nulo neste trabalho ($d_i = 0$).

Os valores fenotípicos (VF_i) foram obtidos segundo o modelo:

$$VF_i = VG_i + \varepsilon_i$$

Em que ε_i é o efeito ambiental gerado segundo uma distribuição normal com média e variância compatível com a herdabilidade (h^2) do caráter, que corresponde ao quadrado da correlação dos valores fenotípicos e genotípicos.

Foram simulados dois conjuntos de dados fenotípicos, com herdabilidade de 40% e 80%, respectivamente. Adicionalmente, foi analisada a informação genotípica simulada sem influência ambiental (herdabilidade 100%), como informação de referência.

2.1.2. Validação da população base

A validação da população base consistiu em verificar a concordância dos parâmetros estabelecidos para a simulação e os resultados do programa. Para isso foi caracterizada a

população simulada, pelo teste de segregação dos marcadores, mapeamento genético e análises das regiões controladoras de características oligogênicas.

A razão de segregação de cada marcador foi verificada com o teste de Qui-quadrado a 5% de probabilidade, com a hipótese de segregação específica de 1:2:1 ($AA:Aa:aa$) para cada loco, conforme segue:

$$\chi^2 = \sum [(O_i - E_i)^2 / E_i]$$

Sendo O_i e E_i os valores observados e esperados, respectivamente, para a i -ésima classe fenotípica ($i= 1,2$ e 3).

No mapeamento genético, foi verificada a porcentagem de recombinação entre pares de marcadores pelo método da máxima verossimilhança. Os grupos de ligações foram obtidos usando como critérios de agrupamento a frequência máxima de recombinação (r_{max}) de 30% e o valor mínimo de LOD score de 3,0. O LOD score (*logarithm of the likelihood odds ratio*) é indicativo da probabilidade de ligação entre dois locos, calculado com a expressão:

$$LOD = \text{Log}_{10}(L(\hat{r}; n_{ij})/L(r = 0,5; n_{ij}))$$

A qual contrasta duas estimativas de verossimilhança (L) em função da frequência de recombinação (r) e as ocorrências genotípicas da progênie (n_{ij}). No numerador r toma o valor da frequência de recombinação estimada ($r = \hat{r}$), que é contrastada com uma $r = 0,5$, no denominador, que expressa ausência de ligação.

Com essas análises foi possível comprovar o número de grupos de ligação, o número de marcadores por grupo de ligação, o tamanho dos grupos, a distância média entre marcadores e a sua variância.

Também foi analisado o grau de distorção da distância entre par de marcas adjacentes obtidas no mapeamento em relação à distância esperada, por meio do coeficiente de estresse (S), de forma similar ao usado em estudos de divergência genética (Cruz *et al.*, 2014), dado em porcentagem por:

$$S = \sqrt{\frac{\sum_{k=1}^{m-1} (d_k - d_k')^2}{\sum_{k=1}^{m-1} d_k^2}}$$

Sendo d_k a distância entre os k marcadores adjacentes no genoma simulado; d_k' a distância entre os respectivos marcadores adjacentes no genoma analisado e m o total de marcadores.

Após estimar as frequências de recombinação entre marcadores adjacentes, a ordem mais provável dos marcadores foi determinada usando o critério SARF (*Sum of Adjacent Recombinant Fractions*), que se baseia na escolha da melhor ordem considerando a menor soma de recombinações adjacentes (Falk, 1989). A correlação de Spearman (*rank*) foi usada para quantificar a alteração na ordem das marcas, expressando o grau de concordância na ordem do genoma analisado com respeito ao simulado. Um valor de um (1) indica a exata concordância na ordem, enquanto um valor de menos um (-1) caracteriza a completa inversão da ordem.

A correlação de Spearman (r_s) foi estimada por grupo de ligação com a expressão:

$$r_s = 1 - \frac{6 \sum_i^m d_i^2}{(m^3 - m)}$$

Em que d é a diferença entre a posição do marcador i no grupo de ligação simulado e a posição do mesmo marcador no grupo de ligação analisado e m é o número de marcadores por grupo de ligação.

Na análise dos *QTL* (*Quantitative trait loci*) foi verificada a associação entre marcadores e *QTL* e a posição relativa dos quatro *QTL*, mediante teste de razão de verossimilhança calculando o *LOD Score* e *LOD máximo*, no método de intervalo simples. Também foi estimado o efeito de cada *QTL* pelo cálculo do quadrado do coeficiente de determinação (r^2).

O modelo de regressão usando no método de intervalo simples é:

$$Y_j = \mu + a x_j^* + d z_j^* + \varepsilon_j$$

Sendo:

Y_j = o valor da característica Y no indivíduo j ;

μ = a média geral da característica da população;

a = o efeito aditivo do loco que está sendo estudado sobre a característica;

d = o efeito de dominância do loco que está sendo estudado sobre a característica;

x_j^* e z_j^* = as variáveis condicionadoras e dependentes dos genótipos dos marcadores que flanqueiam o *QTL*, no indivíduo j ; e

ε_j = o erro aleatório.

Além da estimação de parâmetros genéticos, foram caracterizados os resultados fenotípicos quanto á média (μ), variância (σ^2), distribuição dos dados mediante testes de normalidade de Lilliefors e obtenção dos coeficientes de simetria (s) e curtose (k).

Finalmente foi estimado o quadrado da correlação entre o valor fenotípico e genotípico, para verificar a herdabilidade das características simuladas.

2.2. Gerações avançadas (F_3 a F_6)

Em cada cenário das populações analisadas, os indivíduos melhor valorados foram selecionados e utilizados como parentais da seguinte geração. O processo seleção/simulação de descendentes foi repetido até a geração F_6 . Essas etapas foram integradas como rotinas no programa *R* (R Development Core Team, 2016) (Apêndice B), permitindo a continuidade nos ciclos de seleção, simulação e análises nas gerações F_3 a F_6 . Essa estratégia permitiu melhorar a eficiência e a eficácia da simulação ao eliminar a migração de informação entre programas, evitando possíveis erros e ganhando tempo. Para alcançar esta integração foram desenvolvidas rotinas de simulação de genótipos, valores genotípicos e valores fenotípicos análogos às utilizadas no programa *GENES*.

2.2.1. Simulação de gerações avançadas

A seleção de parentais foi realizada com base no resultado do respectivo modelo preditivo de cada cenário avaliado, procedimentos que são explicados no capítulo seguinte. De cada população analisada foram selecionados os indivíduos com maior valor genético genômico (*VGG*), e a partir do genótipo destes indivíduos a seguinte geração foi simulada por autofecundação.

A autofecundação foi simulada gerando gametas recombinantes de cada indivíduo selecionado, cuja informação foi combinada para gerar o genótipo dos novos indivíduos. Especificamente no genoma de cada progenitor, e para cada grupo de ligação, foram sorteados os pontos nos quais se produziram as permutas entre homólogos. Um dos homólogos recombinado foi selecionado em um segundo sorteio para ser incluído no gameta. Este processo se repetiu independentemente, para gerar o segundo gameta e ao parear estas informações se estabeleceu o genoma do novo indivíduo. Cada nova geração foi composta por 200 indivíduos e cada progenitor selecionado aportou o número de descendentes proporcional para alcançar esse número.

Uma vez que o genoma da nova população foi gerado, os valores genotípicos (VG_i) e fenotípicos (VF_i) foram simulados, no programa *R*, usando as expressões descritas na simulação da população base.

2.2.2. Validação de gerações avançadas

A validação das populações F_3 a F_6 teve dois enfoques: O primeiro consistiu em verificar, no programa *R*, se as rotinas de simulação de genomas e de valores genotípicos e fenotípicos reproduziram os eventos indicados. O segundo enfoque foi a validação da simulação da autofecundação no transcurso das gerações e da integração dos processos de simulação, seleção e avaliação.

As rotinas de simulação de populações desenvolvidas ou adaptadas ao *R* foram verificadas com a geração e avaliação de uma população F_2 . Essa população não fez parte do núcleo de análises do trabalho, só foi feita com o propósito de garantir a capacidade de simular adequadamente genótipos, valores genotípicos e valores fenotípicos.

A população foi composta de 500 indivíduos, procedente de um cruzamento inicial de progenitores homozigotos contrastantes. O genótipo foi gerado com 520 locos equidistantes a um cM, com quatro grupos de ligação de 130 cM e cada um deles com um *QTL* com efeito uniforme e ausência de dominância. Foram avaliadas duas características: uma correspondente aos valores genotípicos e outra com herdabilidades de 80%. A validação foi realizada seguindo os mesmos procedimentos descritos na validação da população base.

Uma vez verificadas as rotinas de simulação de populações, essas foram usadas para simular um cenário de seleção fenotípica. Esse cenário foi usado para validar o processo de autofecundação até a sexta geração, além de servir como ponto de referência para as análises posteriores.

Foram analisadas 100 populações de 200 indivíduos, amostradas da população base de forma aleatória. De cada população e em cada geração foram selecionados os 40 melhores fenótipos, cujos genótipos foram usados para simular a geração seguinte, até a F_6 , segundo os métodos descritos anteriormente. Foram estimadas a herdabilidade realizada e a média e variância genotípicas.

Foi considerado que populações submetidas a ciclos de autofecundação têm um aumento na frequência de genótipos homozigotos e redução na frequência de genótipos heterozigotos. Conseqüentemente, o monitoramento desta dinâmica em cada geração serviu para verificar o adequado processo de simulação da autofecundação. Foi

calculado o coeficiente de endogamia (F), que relaciona esses fatos a partir da expressão:

$$F = (H_e - H_o) / H_e$$

Sendo:

$$H_e = (\sum(1 - \sum p_i^2)) / M;$$
$$p_i = (n_{ii} + (n_{ij} / 2)) / N; \text{ e}$$
$$H_o = (\sum n_{ij}) / N.M$$

em que n_{ii} é o número de homozigotos do alelo i , n_{ij} o número de heterozigotos, p_i a frequência estimada do alelo i , N o número total de indivíduos e M o número total de marcadores.

Os valores teóricos dependem da diminuição na frequência de heterozigotos (50% a cada geração) em relação à frequência deles em condições de equilíbrio de Hardy-Weinberg. O coeficiente F nestas circunstâncias é previsível a partir da expressão: $F_N = 1 - (1/2)^{N-g}$, sendo g a geração usada como referência, onde $F=0$, e N o número de gerações autofecundadas. Portanto, teoricamente espera-se que o coeficiente de endogamia populacional atinja os seguintes valores, da segunda até a sexta geração: 0,0, 0,5, 0,75, 0,875 e 0,937. As análises foram feitas nos programas *GENES* (Cruz, 2013), *RBIO* (Bhering, 2017) e *R* (R Development Core Team, 2016).

3. RESULTADOS E DISCUSSÃO

3.1. População base

A simulação da população base produziu uma matriz 625 x 1043, correspondente a 625 indivíduos e 1040 marcadores mais três colunas dos valores genotípicos e os valores fenotípicos das duas herdabilidades (40% e 80%). O mapa de ligação resultante da simulação está apresentado na Figura 2.

Coffea arabica é uma espécie tetraploide, composta por dois sub-genomas, mas comporta-se como diploide funcional (Medina, 1950; Mendes, 1950; Lashermes *et al.*, 1999). Não obstante, as avaliações foram feitas com uma versão simplificada do genoma, facilitando o computo das análises sem afetar as respostas, além de permitir que as conclusões possam ser extensíveis a outras espécies. Assim, o genoma simulado correspondeu a uma espécie diploide com oito grupos de ligação, em que quatro dos grupos continham um loco controlador e outros quatro grupos foram incluídos como um sub-genoma sem influencia na característica.

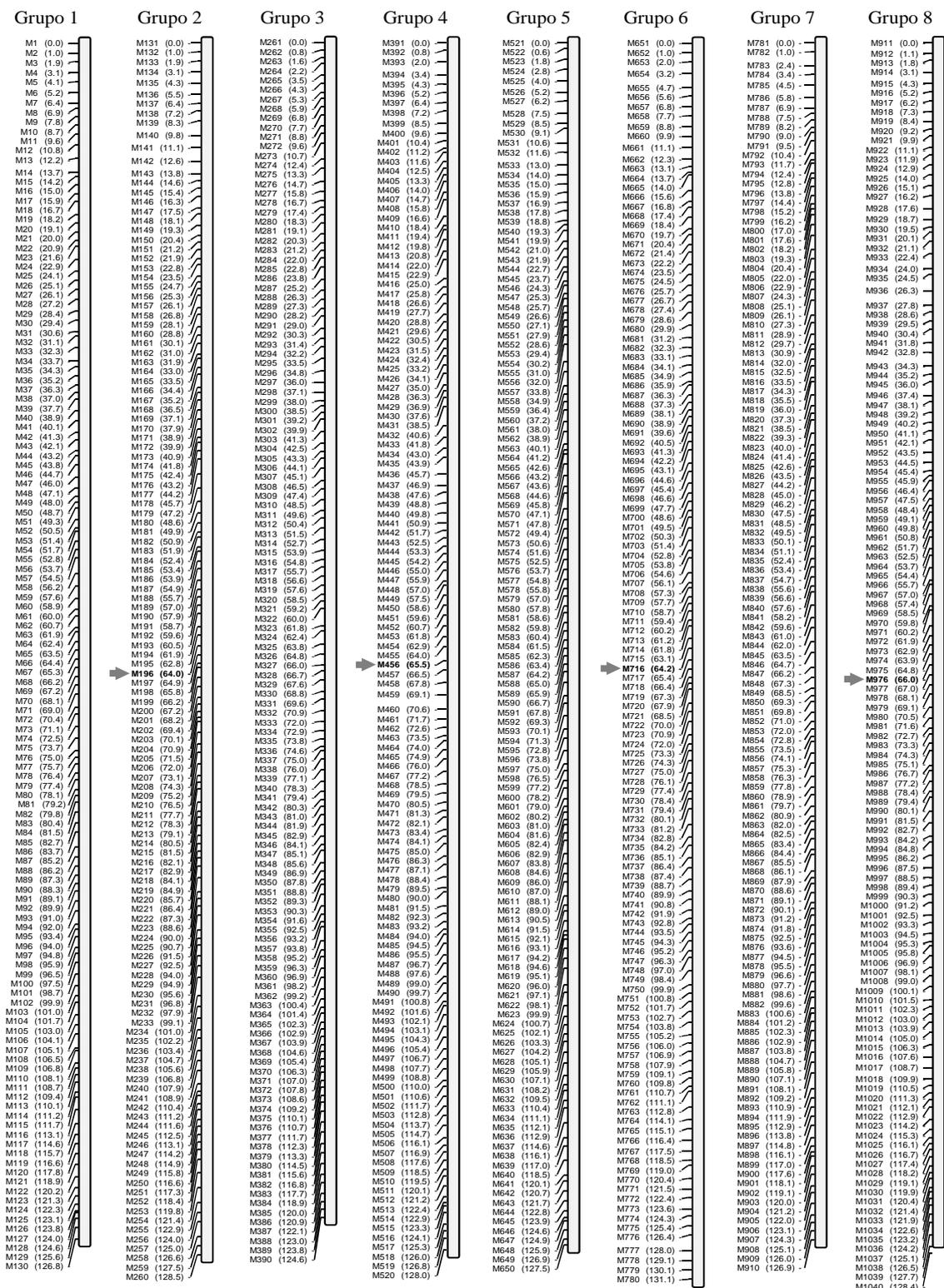


Figura 2. Mapa de ligação do genoma simulado. O nome do marcador está indicado pela letra “M” seguida de um número e entre parênteses a distância genética em cM. As setas apontam os marcadores com ligação completa com os locos controladores da característica.

A população F_2 simulada, além de ser a base para o desenvolvimento de diferentes cenários de análise, foi definida como referência em relação à saturação de marcadores e tamanho populacional. Ainda que as distâncias entre marcadores menores a 0,5 cM são recomendadas para uma ótima predição de VGGs (Resende *et al.*, 2012), a saturação de um centimorgan foi estabelecida como ponto de partida, para a exploração do efeito da baixa densidade de marcadores na acurácia seletiva, já que essa saturação também tem originado adequados valores de acurácia (Heffner *et al.*, 2011).

Por outro lado, de acordo com a equação de Goddard *et al.* (2011), para o cálculo da acurácia esperada da GS (definida na revisão de literatura), o tamanho populacional ótimo para uma característica controlada por quatro locos e herdabilidade de 80% é próxima a 500 indivíduos. Então, a população base F_2 foi composta de 625 indivíduos, com o propósito de usar 500 indivíduos (80%) como população de estimação e 125 indivíduos (20%) como população de validação, na predição dos VGGs.

3.1.1. Validação da simulação da população base F_2

Na validação verificou-se o desempenho da simulação, seguindo os princípios genéticos. Este processo foi feito por meio de testes de hipóteses e comparação de parâmetros. O primeiro princípio verificado foi a segregação de locos individuais pelos testes de Qui-quadrado (χ^2) a 5% de probabilidade, com a hipótese de segregação de 1:2:1. Todos os locos apresentaram a razão esperada, sendo este resultado, o primeiro indício da qualidade da simulação.

O mapeamento genético foi efetuado com todos os marcadores (Figura 2). A recuperação de grupos de ligação foi bem-sucedida, uma vez que o número de grupos obtidos foi de acordo com o esperado (Tabela 1). Igualmente, o número de marcas obtidas em cada grupo de ligação correspondeu ao número especificado.

As estimações da variância entre marcas e de estresse ficaram no intervalo de 0,062 - 0,087 e 0,250 - 0,296, respectivamente. O valor esperado para ambos os parâmetros é de zero, portanto a simetria da distância entre marcas não foi exata, mas trabalhos nos quais foram simulados diferentes níveis de saturação (Cruz, 2006; Silva *et al.*, 2007; Brito *et al.*, 2014) coincidem com esse resultado. Nesses trabalhos os genomas mais saturados apresentaram o maior efeito de estresse. Não obstante, a variação foi considerada aceitável, dado que 90% das distâncias entre marcas adjacentes variaram de 0,6 cM a 1,4 cM e, além disso, a distância média e o tamanho médio dos grupos de ligação foram próximos aos valores definidos (Tabela 1).

Tabela 1. Parâmetros de simulação esperados e observados por grupo de ligação.

Grupo de ligação	Número de Marcas	Tamanho (cM)	Distância média entre marcas (cM)	Variância da distância entre marcas	Estresse (proporção)
1	130	126,8	0,983	0,078	0,289
2	130	128,5	0,996	0,086	0,296
3	130	124,6	0,966	0,062	0,265
4	130	128,0	0,992	0,083	0,293
5	130	127,5	0,988	0,087	0,250
6	130	131,1	1,016	0,070	0,292
7	130	126,9	0,984	0,085	0,280
8	130	128,4	0,995	0,086	0,293
Valor médio observado	130	127,7	0,990	0,079	0,282
Valor paramétrico	130	130,0	1,000	0,000	0,000

Em relação à ordem das marcas no genoma, segundo a correlação de Spearman, não houve alteração na disposição delas nos grupos de ligação, comparada com a ordem do genoma de referência (aquele que segue todos os parâmetros de simulação).

Após o mapa do genoma ter sido estabelecido, a posição de cada *QTL* foi determinada com o teste da razão de verossimilhança, expresso através do *LOD* score. Os marcadores associados corresponderam com aqueles preestabelecidos (*M196 = QTL1*; *M456 = QTL2*; *M716 = QTL3*; *M976 = QTL4*) e se localizaram na região mediana dos respectivos grupos de ligação, segundo foi definido. O pico indica a maior significância estatística e, por conseguinte, a posição do *QTL* (Figura 3). Essas informações em conjunto mostraram a correta detecção dos locos controladores das características consideradas.

Foi definido que os *QTL* teriam um efeito uniforme sobre cada uma das características. Portanto, para verificar se o efeito simulado foi correspondente ao esperado, foi conferida a proporção da variação fenotípica explicada por cada um dos *QTL*, por meio do coeficiente de determinação da regressão. O resultado indica que os efeitos foram simulados corretamente, dado que nas três herdabilidades os valores simulados se aproximaram aos valores esperados (Tabela 2).

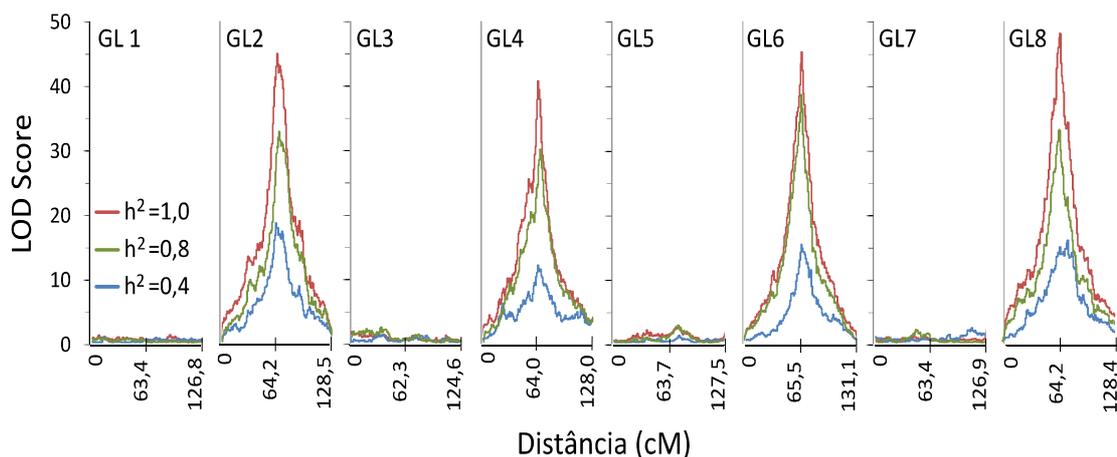


Figura 3 Localização dos *QTL* pelo método de Intervalo Simples para três características definidas por sua herdabilidade. GL: grupos de ligação.

Tabela 2 Comparação dos valores esperados e simulados da variação fenotípica explicada pelos *QTL*.

Marcador	Esperados			Simulados		
	$h^2=1,0$	$h^2=0,8$	$h^2=0,4$	$h^2=1,0$	$h^2=0,8$	$h^2=0,4$
M196 = QTL1	0,25	0,20	0,10	0,247	0,194	0,115
M 456 = QTL2	0,25	0,20	0,10	0,229	0,180	0,081
M 716 = QTL3	0,25	0,20	0,10	0,248	0,220	0,100
M 976 = QTL4	0,25	0,20	0,10	0,259	0,195	0,098
Total	1,00	0,80	0,40	0,983	0,789	0,394

3.1.2. Validação dos valores genotípicos e fenotípicos da população base

O resumo dos parâmetros e testes que certificam a correta simulação dos valores das três características usadas nas avaliações populacionais está apresentado na Tabela 3. Para as três características, o valor obtido na simulação da média fenotípica foi de 1,1, que corresponde ao valor estabelecido (1,0). Da mesma forma, as variâncias, e em consequência as herdabilidades, apresentaram concordância com os níveis prefixados.

Tabela 3. Caracterização fenotípica segundo a herdabilidade da característica simulada.

	$h^2=1,0$	$h^2=0,8$	$h^2=0,4$
Herdabilidade	1,00	0,802	0,401
Média	1,10	1,10	1,10
Variância	12,19	15,42	30,34
Máximo	8,5	12,25	22,18
Mínimo	-9,0	-10,77	-13,43
Curtose	2,608	2,686	3,104
Coef. Simetria	0,032	0,016	0,242
Lilliefors ¹	0,048**	0,016 ^{ns}	0,013 ^{ns}

¹ Mostra-se o valor *D* com teste de hipóteses (1%) de que os dados seguem distribuição normal. (**) rejeita *H*₀; (ns): não rejeita *H*₀.

A característica controle ($h^2=1,0$) apresentou padrão de distribuição discreta, como é evidente na Figura 4. Esse resultado concorda com o comportamento esperado, já que sem efeito ambiental, os valores fenotípicos expressaram os próprios valores genotípicos, os quais tem distribuição binomial. Considerando que a característica é controlada por quatro pares de alelos ($n=4$) e que não tem efeito de dominância, é possível calcular o número de classes fenotípicas (cf) na F_2 pela expressão:

$$cf = 2n + 1 ;$$

Que neste caso resulta em nove classes: indivíduos com expressão de oito, sete, seis, cinco, quatro, três, dois ou um alelo dominante, além dos indivíduos homocigotos recessivos (Tabela 4).

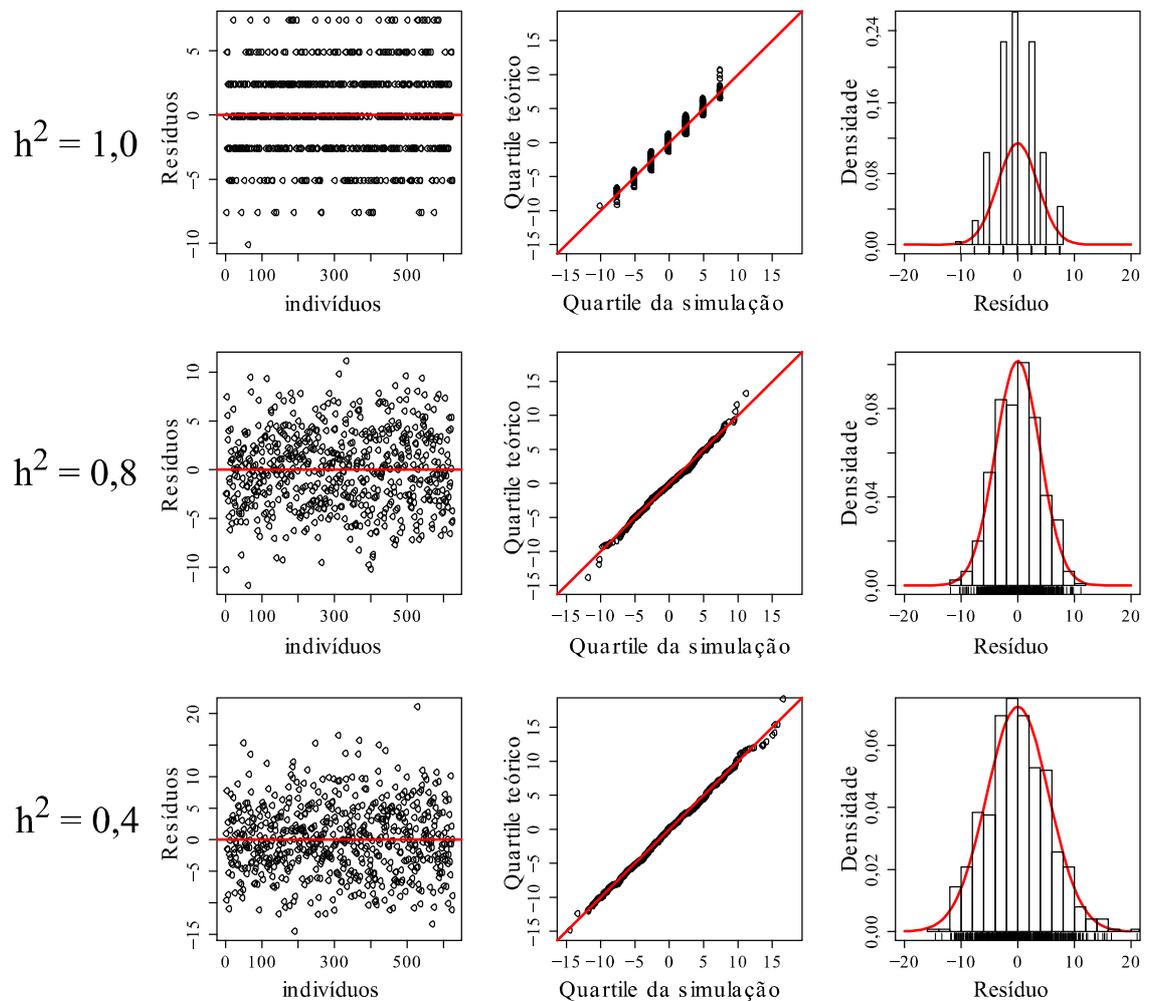


Figura 4. Evidências gráficas da distribuição dos resíduos genotípicos e fenotípicos simulados (Valores dos Resíduos, Normal Q-Q plot e Histogramas).

Tabela 4. Genótipos, número de alelos favoráveis e valor genotípico em cada classe fenotípica para quatro locos com dois alelos.

				AABBccdd					
				AABbCcdd					
				AABbccDd					
			AABbccdd	AAbbCCdd	AABBCcdd				
			AAbbCcdd	AAbbCcDd	AABBccDd				
			AAbbccDd	AAbbccDD	AABbCCdd				
			AaBBccdd	AaBBCcdd	AABbCcDd				
			AaBbCcdd	AaBBccDd	AABbccDD				
			AaBbccDd	AaBBCCdd	AAbbCCDd				
		AAbbccdd	AabbCCdd	AaBbCcDd	AAbbCcDD	AABBCCdd			
		AaBbccdd	AabbCcDd	AaBbccDD	AaBBCCdd	AABBCCdd			
		AabbCcdd	AabbccDD	AabbCCDd	AaBBCcDd	AABBccDD			
		AabbccDd	aaBBCcdd	AabbCcDD	AaBBccDD	AABbCCDd			
		aaBBccdd	aaBbccDd	aaBBCCdd	AaBbCCDd	AABbCcDD			
		aaBbCcdd	aaBbCCdd	aaBBccDd	AaBbCcDD	AAbbCCDD			
Aabbccdd	aaBbccDd	aaBbCcDd	aaBbccDD	aaBbCCDd	AabbCCDD	AaBBCCDd	AABBCCdd		
aaBbccdd	aabbCCdd	aaBbccDD	aaBbCCDd	aaBBCCDd	AaBBCCDd	AaBBccDD	AABBCCdd		
aabbCcdd	aabbCcDd	aabbCCDd	aaBbCcDD	aaBBccDD	AaBbCCDD	AaBbCCDD	AABbCCDD		
aabbccdd	aabbccDd	aabbccDD	aabbCCDD	aaBbCCDD	aaBBCCDD	aaBBCCDD	AaBBCCDD	AABBCCDD	
Número de alelos dominantes									
0	1	2	3	4	5	6	7	8	
Valor genotípico									
-9	-6,5	-4	-1,5	1	3,5	6	8,5	11	

No entanto, no resultado da simulação houve representantes de oito classes, não havendo indivíduos da classe tetra homocigota dominante na geração F_2 , o que é compreensível dado que a frequência dos fenótipos extremos (ffe) é muito baixa:

$$ffe = (1/4)^n = (1/4)^4 \approx 0,004$$

Em condições reais, além dos efeitos de dominância, os efeitos ambientais dificultam a identificação de genótipos superiores. Desconsiderar estes efeitos na característica controle permitiu seu uso como referente teórico para comparar com as características com herdabilidade menor. As características com herdabilidade de 0,8 e 0,4 apresentaram distribuição normal, segundo o teste de Lilliefors, além de valores próximos aos ideais em curtose (3,0) e assimetria (0,0) (Tabela 3). Comparado com o controle, nestas características as classes fenotípicas passaram de uma distribuição discreta a uma contínua, já que o efeito ambiental, em um dado genótipo, pode contribuir para o aumento ou para a diminuição do valor fenotípico. Validação das gerações avançadas de autofecundação ($F_3 - F_6$)

3.2. Validação de simulações de populações

A diferencia da simulação da população base, as gerações avançadas foram simuladas com procedimentos desenvolvidos ou adaptados para rodar diretamente no programa *R*. Estes procedimentos foram validados gerando e avaliando uma população F_2 específica

para esse propósito. Portanto, para não causar possíveis confusões com a população base do trabalho, os resultados são apresentados no Apêndice A.

A validação demonstrou o adequado desempenho dos algoritmos que, seguindo os princípios genéticos, reproduziram satisfatoriamente os eventos biológicos e, portanto, foram considerados confiáveis para sua utilização nas análises do trabalho.

3.2.1. Validação de autofecundação e seleção fenotípica

Na segunda etapa de validação, um cenário de seleção fenotípica foi usado para monitorar o processo de autofecundação por várias gerações e verificar a consistência das simulações.

Na ausência de seleção, mutação e migração, a ocorrência de autofecundações conduz a mudanças nas frequências genotípicas, refletidas no aumento dos valores de endogamia e variância, permanecendo estável o valor da média (Ramalho *et al.*, 2012; Barrera, 2013; Borém & Miranda, 2013). No entanto, se é praticada seleção dirigida, as frequências dos alelos de interesse são alteradas e, em consequência, o valor da média é modificado (aumentando ou diminuindo, segundo a finalidade), além do rápido aumento da endogamia e a diminuição da variância a cada geração, para aqueles alelos.

Portanto, para validar a qualidade da simulação das autofecundações o referente mais apropriado foi a análise das frequências genotípicas dos alelos que não estiveram em seleção, por meio do cálculo do coeficiente de endogamia, visto que o coeficiente desses alelos é previsível, independentemente do processo seletivo.

Os valores observados se aproximaram aos valores esperados nas três características (Tabela 5). Deste modo, se constatou que o incremento no grau de endogamia a cada geração foi reproduzido nas proporções adequadas, demonstrando que os processos de simulação foram capazes de replicar o comportamento genético.

Tabela 5. Coeficiente de Endogamia médio de duas características, estimado com alelos fora de seleção, até a sexta geração de autofecundação.

Geração	Esperado	Observado	
		$h^2=0,8$	$h^2=0,4$
F ₂	0,000	-0,004	-0,004
F ₃	0,500	0,513	0,511
F ₄	0,750	0,768	0,766
F ₅	0,875	0,891	0,887
F ₆	0,937	0,946	0,946

Nesta simulação o objetivo foi fixar na população os alelos favoráveis, que é atingido quando todos os indivíduos se encontram em homozigose para os quatro *QTL*, maximizando o valor genotípico (11,0) e reduzindo a variância genotípica a zero.

Os resultados da seleção fenotípica indicaram progresso nas duas características em direção ao alvo de melhoramento (Tabela 6). Inicialmente as duas características tinham a mesma média (1,1) e variância genotípica (12,2), que corresponderam aos parâmetros da população base. Os valores da média genotípica aumentaram devido à seleção praticada, mas com menores avanços para a característica com $h^2=0,4$, consequência do maior efeito ambiental nessa característica que mascara o valor genotípico, reduzindo a eficiência seletiva.

Tabela 6. Parâmetros populacionais de duas características, simuladas segundo a herdabilidade, até a sexta geração de seleção e autofecundação. Valores médios e desvio padrão entre parênteses, de 100 simulações.

Característica	Gerações	Herdabilidade	Média Genotípica	Variância Genotípica
$h^2=0,4$	F ₂	0,40 (0,04)	1,13 (0,19)	12,19 (0,85)
	F ₃	0,44 (0,06)	4,39 (0,43)	14,23 (1,82)
	F ₄	0,32 (0,07)	7,63 (0,56)	8,69 (1,74)
	F ₅	0,20 (0,06)	9,68 (0,43)	4,48 (1,27)
	F ₆	0,07 (0,05)	10,69 (0,21)	1,26 (0,84)
$h^2=0,8$	F ₂	0,80 (0,02)	1,13 (0,19)	12,19 (0,85)
	F ₃	0,75 (0,06)	5,55 (0,41)	9,55 (0,97)
	F ₄	0,59 (0,07)	9,12 (0,45)	4,53 (0,89)
	F ₅	0,17 (0,09)	10,83 (0,12)	0,61 (0,44)
	F ₆	0,07 (0,06)	10,99 (0,02)	0,02 (0,08)

Nesta simulação, os alelos favoráveis não foram fixados ao longo de seis gerações de seleção fenotípica. A característica com $h^2=0,8$ teve valores próximos, mas o teste de Wilcoxon indicou que eles não foram iguais aos parâmetros de fixação da média ($H_0:\mu=11$, wilcoxon=0, $gl=99$, $p\text{-valor}=0,006$) ou da variância ($H_0:\mu=0$, wilcoxon=55, $gl=99$, $p\text{-valor}=0,006$).

Por outro lado, na característica com $h^2=0,4$ a variância genotípica aumentou na F₃, em relação à geração anterior, aumentando também a herdabilidade (Tabela 6). Isso porque na F₃ foram gerados indivíduos das classes genotípicas inferiores, causado pela baixa eficácia seletiva na F₂, aumentando variância. Com exceção do anterior, os valores de herdabilidade diminuíram a cada geração, até valores próximos zero. Esse foi o reflexo

da seleção, que reduziu as variâncias genóticas e fenotípicas a cada geração, no entanto, a diminuição da variância fenotípica foi menor, devido à variação do efeito ambiental que foi permanente nas gerações.

4. CONCLUSÕES

O genoma da população base gerada ajustou-se aos parâmetros simulados, mostrando correta segregação dos marcadores e adequado ordenamento dos marcadores (mapeamento) e das regiões controladoras das características, além da congruente simulação dos valores genóticos e fenóticos. Os procedimentos de simulação de populações descendentes por autofecundação também produziram os eventos genéticos esperados, ao longo das gerações. Os procedimentos de simulação, seleção e avaliação foram integrados adequadamente. Portanto, os processos de simulação desenvolvidos e as informações geradas são confiáveis para realizar análises e inferências no contexto de características oligogênicas em uma espécie autógama.

Na seleção fenotípica, os alelos controladores das duas características não foram fixados até a sexta geração. Esse resultado é ponto de referência para a interpretação dos resultados do próximo capítulo.

5. REFERÊNCIAS

BARRERA, C. F. **Seleção genômica ampla em gerações avançadas de cruzamentos ao acaso ou autofecundação**. 2013. 76 f. Tese (Doutorado em Genética e Melhoramento)- Universidade Federal de Viçosa, Viçosa MG, 2013.

BERNARDO, R.; YU, J. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, v. 47, n. 3, p. 1082-1090. 2007.

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

BHERING, L. L.; CRUZ, C. D. Tamanho de população ideal para mapeamento genético em famílias de irmãos completos. **Pesquisa Agropecuária Brasileira**, v. 43, n. 3, p. 379-385. 2008.

BORÉM, A.; MIRANDA, G. **Melhoramento de Plantas**. Viçosa-MG: Editora UFV, 2013.

BRITO, S. G.; MELO FILHO, P. A.; SILVA, M. C.; ARCELINO, E. C.; COUTINHO, A. E.; NEDER, D. G. Otimização do mapeamento genético vegetal de populações duplo-haplóide via simulação computacional. **Bioscience Journal**, v. 30, n. 3, p. 311-317. 2014.

CASALI, V. W. D.; TIGCHELAAR, E. C. Computer simulation studies comparing pedigree, bulk, and single seed descent selection in self pollinated populations. **Journal of American Society Horticultural Science**, v. 100, n. 4, p. 364-367. 1975.

CERON-ROJAS, J. J.; CROSSA, J.; ARIEF, V. N.; BASFORD, K.; RUTKOSKI, J.; JARQUÍN, D.; ALVARADO, G.; BEYENE, Y.; SEMAGN, K.; DELACY, I. A. Genomic selection index applied to simulated and real data. **G3: Genes|Genomes|Genetics**, v. 5, n. 10, p. 2155-2164. 2015.

CRUZ, C. D. A informática no melhoramento genético. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. D.; VALADARES-INGLIS, M. C. **Recursos Genéticos e Melhoramento de Plantas**. Rondonópolis: Fundação-MT, 2001. p. 1086-1118.

CRUZ, C. D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v. 35, n. 3, p. 271-276. 2013.

CRUZ, C. D.; CARNEIRO, P. C.; REGAZZI, J. A. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa- Brasil: Editora UFV, 2014. 668 p.

CRUZ, E. M. **Efeito da saturação e do tamanho de populações F2 e de retrocruzamento sobre a acurácia do mapeamento genético**. 2006. 119 f. Tese (Doutorado em Genética e Melhoramento)- Universidade Federal de Viçosa, Viçosa MG-Brasil, 2006.

DACHS, J. N. W. **Estatística computacional: uma introdução ao Turbo Pasca**. Rio de Janeiro: LTC, 1988. 236 p.

ESTAGHVIROU, S. B. O.; OGUTU, J. O.; PIEPHO, H.-P. Influence of outliers on accuracy estimation in genomic prediction in plant breeding. **G3: Genes|Genomes|Genetics**, v. 4, n. 12, p. 2317-2328. 2014.

FALK, C. T. A simple scheme for preliminary ordering of multiple loci: application to 45 CF families. In: ELSTON, R. C.; SPENCE, M. A.; RODGE, S. E.; MCCLUE, J. W. **Multipoint mapping and linkage based upon affected pedigree members**. New York: Liss, 1989. p. 17-22.

FERREIRA, D. F. Uso de simulação no melhoramento. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. D.; VALADARES-INGLIS, M. C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis -Brasil: Fundação-MT, 2001. p. 1119-1141.

FERREIRA, W. D.; RAMALHO, M. A. P.; FERREIRA, D. F.; SOUZA, M. A. Family number in common bean selection. **Genetics and molecular biology**, v. 23, n. 2, p. 403-409. 2000.

GODDARD, M. E.; HAYES, B. J.; MEUWISSEN, T. H. E. Using the genomic relationship matrix to predict the accuracy of genomic selection. **Journal of Animal Breeding and Genetics**, v. 128, n. 6, p. 409-421. 2011.

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

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. Genomic selection using low-density marker panels. **Genetics**, v. 182, n. 1, p. 343-353. 2009.

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

HICKEY, J. M.; GORJANC, G. Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. **G3: Genes|Genomes|Genetics**, v. 2, n. 4, p. 425-427. 2012.

LASHERMES, P.; COMBES, M.; ROBERT, J.; TROUSLOT, P.; D'HONT, A.; ANTHONY, F.; CHARRIER, A. Molecular characterization and origin of the *Coffea arabica* L. genome. **Molecular Genetics and Genomics**, v. 261, n. 1, p. 259-266. 1999.

LIU, P.; ZHU, J.; LU, Y. Marker-assisted selection in segregating generations in self-fertilizing crops. **Theoretical and Applied Genetics**, v. 109, n. 2, p. 370-376. 2004.

LORENZ, A. J. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. **G3: Genes|Genomes|Genetics**, v. 3, n. 3, p. 481-491. 2013.

MEDINA, M. Observacoes citológicas em *Coffea*. XIV Microsporogenese en *Coffea arabica* L. var. Rugosa K. M. C. **Bragantia**, v. 10, n. 2, p. 61-66. 1950.

MENDES, A. Observacoes citológicas em *Coffea*. XV Microsporogenese en *Coffea arabica* L. **Bragantia**, v. 10, n. 3, p. 79-87. 1950.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 1, p. 1819-1829. 2001.

MONTESINOS-LÓPEZ, A.; MONTESINOS-LÓPEZ, O. A.; CROSSA, J.; BURGUEÑO, J.; ESKRIDGE, K. M.; FALCONI-CASTILLO, E.; HE, X.; SINGH, P.; CICHY, K. Genomic bayesian prediction model for count data with genotype \times environment Interaction. **G3: Genes|Genomes|Genetics**, v. 6, n. 5, p. 1165-1177. 2016.

MUEHLBAUER, F. J.; BURNELL, D. G.; BOGYO, T. P.; BOGYO, M. T. Simulated comparisons of single seed descent and bulk population breeding methods. **Crop Science**, v. 21, n. 4, p. 572-577. 1981.

PEIXOTO, L. A.; FERREIRA, A.; SILVA, M. F. F.; CRUZ, C. D. Valuation of methodologies for mapping oligogenic trait loci in Recombinant Inbred Lines (RILs). **Acta Scientiarum**, v. 36, n. 1, p. 19-26. 2014.

R DEVELOPMENT CORE TEAM **R: A language and environment for statistical computing**. Vienna, Austria, R Foundation for Statistical Computing. 2016.

RAMALHO, M. A.; ABREU, A. D. F.; SANTOS, J.; NUNES, J. A. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. Lavras - MG - Brasil: UFLA, 2012.

RESENDE JR, M. F. R. **Seleção genômica ampla no melhoramento vegetal**. 2010. 78 f. Tese (Mestrado em Genética e Melhoramento de Plantas)- Universidade Federal de Viçosa, Viçosa MG, 2010.

RESENDE, M. D. V.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal**, v. 56, n. 1, p. 63-77. 2008.

RESENDE, M. D. V.; SILVA, F. F.; LOPES, P. S.; AZEVEDO, C. F. **Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada e estatística espacial**. Viçosa, MG: Editora UFV, 2012. 291 p.

ROCHA, R. B.; BARROS, W. S.; MURO-ABAD, J. I.; TOMAZ, R. S.; CRUZ, C. D.; BARROS, E. G.; ARAÚJO, E. F. Método para mapeamento de locos controladores de características oligogênicas. **Ciência Rural**, v. 40, n. 2, p. 302-308. 2010.

RUTKOSKI, J. E.; POLAND, J.; JANNINK, J.-L.; SORRELLS, M. E. Imputation of unordered markers and the impact on genomic selection accuracy. **G3: Genes|Genomes|Genetics**, v. 3, n. 3, p. 427-439. 2013.

SILVA, L. C. **Simulação do tamanho da população e da saturação do genoma para mapeamento genético de RILs**. 2005. 120 f. Tese (Magister Scientiae)- Universidad Federal de Viçosa, Viçosa MG, 2005.

SILVA, L. C.; CRUZ, C. D.; MOREIRA, M. A.; BARROS, E. G. Simulation of population size and genome saturation level for genetic mapping of recombinant inbred lines (RILs). **Genetics and molecular biology**, v. 30, p. 1101-1108. 2007.

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

VAN OEVEREN, A. J.; STAM, P. Comparative simulation studies on the effects of selection for quantitative traits in autogamous crop: early selection versus single seed descent. **Heredity**, v. 69, n. 4, p. 342-351. 1992.

VEIGA, R. D.; FERREIRA, D. F.; RAMALHO, M. A. P. Eficiência dos dialelos circulantes na escolha de genitores. **Pesquisa Agropecuária Brasileira**, v. 35, n. 7, p. 1395-1406. 2000.

WANG, J. Modelling and simulation of plant breeding strategies. In: ABDURAKHMONOV, I. Y. **Plant Breeding**. InTech, 2012. Cap. 2, p. 19-40.

WANG, J.; VAN GINKEL, M.; PODLICH, D.; YE, G.; TRETOWAN, R.; PFEIFFER, W.; DELACY, I. H.; COOPER, M.; RAJARAM, S. Comparison of two breeding strategies by computer simulation. **Crop Science**, v. 43, n. 6, p. 1764-1773. 2003.

XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, v. 163, n. 2, p. 789-801. 2003.

YU, J.; ARBELBIDE, M.; BERNARDO, R. Power of in silico QTL mapping from phenotypic, pedigree, and marker data in a hybrid breeding program. **Theoretical and Applied Genetics**, v. 100, n. 6, p. 1061-1067. 2005.

YU, J.; HOLLAND, J. B.; MCMULLEN, M. D.; BUCKLER, E. S. Genetic design and statistical power of nested association mapping in maize. **Genetics**, v. 178, p. 539-551. 2008.

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**, v. 182, n. 1, p. 355-364. 2009.

CAPÍTULO II

ANÁLISES DE CENÁRIOS COM GENOTIPAGEM DE BAIXA SATURAÇÃO
PARA A SELEÇÃO GENÔMICA DE CARACTERÍSTICAS OLIGOGÊNICAS EM
Coffea arabica L.

1. INTRODUÇÃO

Espécies perenes como *Coffea arabica* L., que tem longo período juvenil, necessidade de avaliações da produção por vários anos e altos custos nos testes de campo, o processo de melhoramento pode tardar mais de 20 anos. Portanto, é indispensável a procura de ferramentas que aumentem a eficácia e reduzam tempos e custos nos ciclos de seleção de genótipos (Sera, 2001; Ramalho *et al.*, 2013).

De forma geral, os programas de melhoramento têm como objetivo desenvolver novos cultivares geneticamente superiores aos atualmente disponíveis, adaptáveis a condições específicas e de comportamento previsível. Para atingir este objetivo são utilizadas estratégias que transferem, aumentam ou fixam a frequência de alelos favoráveis na população, o que demanda recursos físicos, humanos e de tempo. Para potencializar esses recursos a busca da maior eficiência é constante, visando aumentar o ganho genético com menor custo. Por isso, é necessária a reavaliação das estratégias de melhoramento e a análise das novas propostas que surgem com o avanço das pesquisas.

Métodos como a seleção genômica (GS) são ideais pela comprovada capacidade de inferência sobre os valores genéticos dos indivíduos, gerando alta eficiência preditiva, que a conduz à obtenção de rápidos ganhos genéticos (Resende *et al.*, 2012). Para alcançar esse resultado, a maioria da variação da característica deve ser explicada pelos marcadores, por isso é necessário elevado número deles, para aumentar a probabilidade de que alguns destes marcadores estejam fortemente ligados às regiões controladoras da característica (Meuwissen *et al.*, 2001; Resende, 2008).

No entanto, nem todos os programas de melhoramento tem ao alcance os recursos econômicos para a implementação deste método (Byrne, 2007; Zhang *et al.*, 2011; Wellmann *et al.*, 2013; Kainer *et al.*, 2015; Ma *et al.*, 2016). Um limitante adicional para *C. arabica* é o custo da manutenção de grandes populações no campo durante os longos períodos. Portanto, populações de estimacão grandes raramente poderiam ser alcançadas. Porém, todos os desenvolvimentos gerados ao redor da GS podem ser aproveitados, com a adaptação às necessidades e alcances de cada programa.

Muitas das características de interes agrônômico são poligênicas, que exigem inversões superiores para aplicar os métodos da GS, pela necessidade de usar altas densidades de marcação e tamanhos populacionais grandes. No entanto, este trabalho explorou as condições em que esses métodos poderiam ser utilizados com recursos limitados, que

seriam utilizados para a seleção de características menos complexas, como as oligogênicas, mas de alta importância para *C. arabica* e em qualquer outro programa de melhoramento.

Nesse contexto, os benefícios estariam no aumento da eficácia e eficiência do uso de recursos, especificamente na genotipagem e fenotipagem. Além disso, no ganho genético, por aumento na intensidade de seleção pelo aumento da acurácia.

Neste capítulo, usando os métodos de simulação descritos no capítulo anterior, são expostos os resultados de diferentes arranjos experimentais complementares, procurando as condições em que a *GS* permitiria auxiliar o melhoramento de características oligogênicas e identificando as circunstâncias que afetariam seu uso. Para isso foram contrastados cenários com diferentes densidades de marcadores e tamanhos populacionais, considerando o efeito de métodos preditivos, tipos de marcador e intensidades de seleção. Além disso, foi analisado o evento em que existe falta de associação entre os marcadores moleculares e o total das regiões codificadoras, fato que pode ocorrer em baixa densidade e, ou, baixo tamanho populacional.

2. MATERIAL E MÉTODOS

As análises se basearam na comparação de diferentes cenários, com avaliação de populações simuladas, variando níveis de saturação de marcadores e tamanhos populacionais. Além disso, foram comparados dois métodos de *GS*, aplicados em populações F_2 (segunda geração de autofecundação, descendentes de pais contrastantes), para gerar modelos preditivos dos valores genéticos genômicos (*VGG*). Destas populações, foram selecionados os indivíduos com maiores *VGG* para avançar às seguintes gerações por meio da autofecundação. Esta rotina foi repetida até a obtenção da geração F_6 , procurando a fixação dos alelos favoráveis (Figura 1).

2.1. Simulação de populações

Foram simulados 1040 locos equidistantes a um centimorgan (*cM*), que representaram genes com alelos codominantes, divididos em oito grupos de ligação de 130 *cM* cada. Foram simulados quatro locos controladores da característica, sendo um loco controlador por grupo de ligação, com efeito uniforme e ausência de dominância e epistasia.

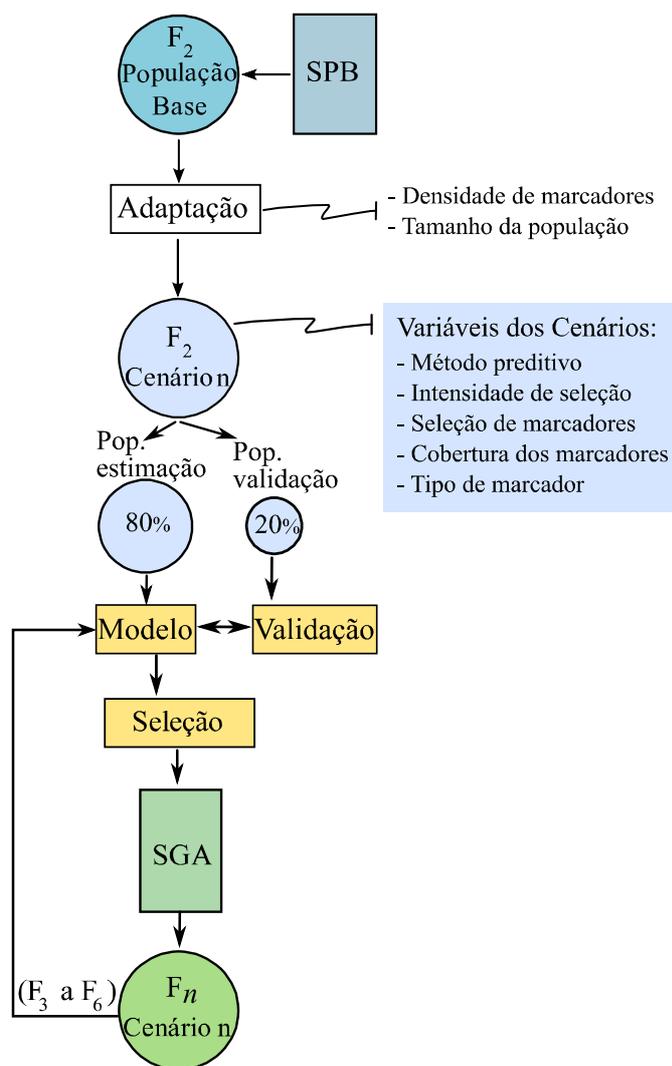


Figura1. Esquema de simulação e análise. A simulação da população base (SPB) foi feita uma única vez e logo foi adaptada para cada cenário. A simulação de gerações avançadas (SGA) foi feita com o genótipo dos indivíduos com maior valor genotípico predito pelo modelo, a cada geração.

A população base do estudo correspondeu a uma geração F_2 com 625 indivíduos, da qual foram extraídas as informações necessárias para as avaliações, segundo o cenário analisado. Pela sua estrutura, tanto em relação ao número de indivíduos como ao nível de saturação, esta população foi considerada adequada para a análise de GS e, em vista disso, usada como controle. Uma vez gerado o genoma base, foram simulados os respectivos valores genômicos, além de dois conjuntos de valores fenotípicos, com herdabilidade (h^2) de 40% e 80%.

Os valores genotípicos (gt_i) e fenotípicos (ft_i) foram determinados usando as expressões:

- $gt_i = \mu + a_i + d_i$, sendo μ : a média genotípica populacional, d_i : o efeito devido a dominância, considerado nulo ($d_i = 0$) e $a_i = \sum_{j=1}^4 p_j \alpha_j$, onde α_j é a contribuição do loco j (1, 0 ou -1 para as classes genotípicas AA, Aa e aa, respectivamente) e p_j o efeito do alelo favorável, com igual contribuição para todo loco j .
- $ft_i = gt_i + E_i$, em que E_i é o efeito ambiental gerado segundo uma distribuição normal com média e variância compatível com a herdabilidade do caráter, que correspondeu ao quadrado da correlação entre os valores fenotípicos e genotípicos.

O modelo preditivo dos VGG foi determinado e validado com indivíduos da segunda geração de autofecundação (F_2) e aplicado nas gerações posteriores (F_3 a F_6). Em cada geração foram selecionados os indivíduos com maiores valores genômicos estimados pelo modelo. As gerações F_3 a F_6 , de 200 indivíduos, foram simuladas a partir do genótipo dos indivíduos selecionados. Esses indivíduos aportaram um número proporcional de descendentes para a seguinte geração.

Como em uma população submetida a vários ciclos de autofecundação há aumento na frequência de genótipos homozigotos contra os heterozigotos, esta tendência foi monitorada em cada geração para verificar o adequado processo de simulação da autofecundação. Foi calculado o coeficiente de endogamia (F) a partir da expressão:

$$F = (H_e - H_o)/H_e,$$

sendo: $H_e = (\sum(1 - \sum p_i^2))/M$, $p_i = (n_{ii} + (n_{ij}/2))/N$ e $H_o = (\sum n_{ij})/N.M$,

em que n_{ii} é o número de homozigotos do alelo i , n_{ij} o número de heterozigotos, p_i a frequência estimada do alelo i , N o número total de indivíduos e M o número total de marcadores.

2.2. Análise de seleção genômica

Foi estabelecido que, em cada análise GS, a população de estimação fosse formada por 80% dos indivíduos da F_2 e a população de validação pelos 20% restante. A estimação dos VGG dos indivíduos foi feita com dois métodos: *RR-BLUP* (Meuwissen *et al.*, 2001) e *BLASSO* (Bayesian *LASSO*) (Park & Casella, 2008; De los Campos *et al.*, 2009). O método *RR-BLUP* pressupõe um número grande de *QTL* com pequeno efeito, assumindo que os efeitos dos *QTL* apresentam distribuição normal com uma variância

comum para os efeitos dos marcadores (Resende *et al.*, 2010). O método *BLASSO* pressupõe uma distribuição exponencial dupla para os efeitos de marcadores com variâncias heterogêneas, ou seja, considera que existam algumas marcas com maior efeito associado à característica e que o efeito da maioria dos marcadores seja próximo de zero (Resende *et al.*, 2012).

Para ambos os métodos, considerou-se o modelo:

$$y = \mu + Xm + e,$$

em que y é o vetor de observações fenotípicas, μ é a média geral, m o vetor de efeitos dos marcadores, assumidos como aleatórios, e e representa o vetor de efeitos aleatórios residuais. X é a matriz de incidência dos marcadores, construída a partir dos genótipos.

Para *RR-BLUP*: $m \sim N(0, G = I\sigma_m^2)$, $e \sim N(0, R = I\sigma_e^2)$, onde σ_m^2 e σ_e^2 são a variância dos efeitos dos marcadores e a variância residual, respectivamente. No método *BLASSO*: $m|\lambda \sim \prod \left(\frac{\lambda}{2}\right) \exp[-\lambda|m_i|]$, $e|\sigma_e^2 \sim MVN(0, I\sigma_e^2)$, em que λ representa o parâmetro de regularização (*sharpness*) e *MNV* a distribuição normal multivariada. Em termos do modelo hierárquico aumentado inclui um componente de variância extra τ_i^2 , em que: $p(m|\tau) \sim N(0, D)$, $diag(D) = (\tau_1^2 \dots \tau_n^2)$; $p(\tau|\lambda) = \prod \left(\frac{\lambda^2}{2}\right) \exp\left(\frac{-\lambda^2 \tau_i^2}{2}\right)$, assim, entende-se que a variância genética em cada loco marcador é dada por $\sigma_{m_i}^2 = \tau_i^2$ (Resende Jr *et al.*, 2012).

O problema de otimização a ser resolvido, para a obtenção dos componentes de regressão, pode ser generalizado para ambas as metodologias, segundo Lorenz *et al.* (2011), assim:

$$\min\{\sum(y - \sum Xm)^2 + \lambda \sum |m_i|^q\},$$

em que o primeiro termo da equação é a soma de quadrados dos resíduos da regressão e o segundo termo é a penalização. No método *RR-BLUP* $q = 2$ e $\lambda = \sigma_e^2 / \sigma_{g_i}^2 = \sigma_e^2 / (\sigma_m^2 / n)$, enquanto no método *BLASSO* $q = 1$ e o parâmetro λ pode ser estimado dos próprios dados pelos métodos *MCMC*. O termo $\sum |m_i|^q$ é a soma dos valores absolutos dos coeficientes de regressão contidos no vetor m , de modo que nas soluções nas quais os coeficientes de regressão se afastam do zero sofrem penalização.

Em cada avaliação, o modelo preditivo foi validado com a informação dos indivíduos não utilizados no seu desenvolvimento. Foi estimada a capacidade preditiva dos genótipos (r_{gg}^2), medida como o quadrado do coeficiente de correlação entre os valores genotípicos preditos (VGG) e os verdadeiros valores genéticos (VG), que informa o quanto o valor estimado é próximo do valor real. Também foi estimada a capacidade seletiva, medida como a proporção de indivíduos com real VG para ser selecionado entre os que foram selecionados pelo seu VGG . Estes indicadores também foram calculados em cada geração. Também foram determinadas as médias e variâncias genotípicas e fenotípicas e as frequências alélicas dos QTL .

As populações analisadas atingiram o valor de 11,0 na média genotípica e variância igual a zero, quando os alelos favoráveis foram conduzidos à homozigose (fixados). Para verificar esse evento, os resultados de cada cenário foram contrastados com os valores paramétricos, com teste de Wilcoxon. Se os alelos favoráveis foram fixados em uma das gerações avaliadas, o processo seletivo foi considerado *eficaz* e a seleção foi considerada *eficiente* na medida em que foram necessárias menos gerações para conduzir os QTL a homozigose.

As simulações e análises foram feitas nos programas *GENES* (Cruz, 2013), *RBIO* (Bhering, 2017) e *R* (R Development Core Team, 2016), usando os pacotes *BLR* (De los Campos *et al.*, 2009) e *rrBLUP* (Endelman, 2011).

2.3. Variação da distância de ligação entre loco marcador e loco controlador

Em cada cenário foram analisadas sete distâncias entre marcadores, para identificar a separação máxima na qual os marcadores permanecem eficientes no processo seletivo, no caso de características oligogênicas. Inicialmente, foram usados todos os locos simulados, portanto, em cada QTL estava inserto um marcador, sendo uma condição ideal que serviu como referência. Desse modo, a distância entre marcadores correspondeu a um cM e foram usados os 1040 locos gerados para o genoma base. Esse cenário foi denominado D1 (Figura 2). Foram selecionados marcadores equidistantes para gerar as variantes, de forma que os QTL ficaram no meio de dois marcadores, assim a distância QTL -*Marcador* foi a metade da distância *Marcador*-*Marcador*.

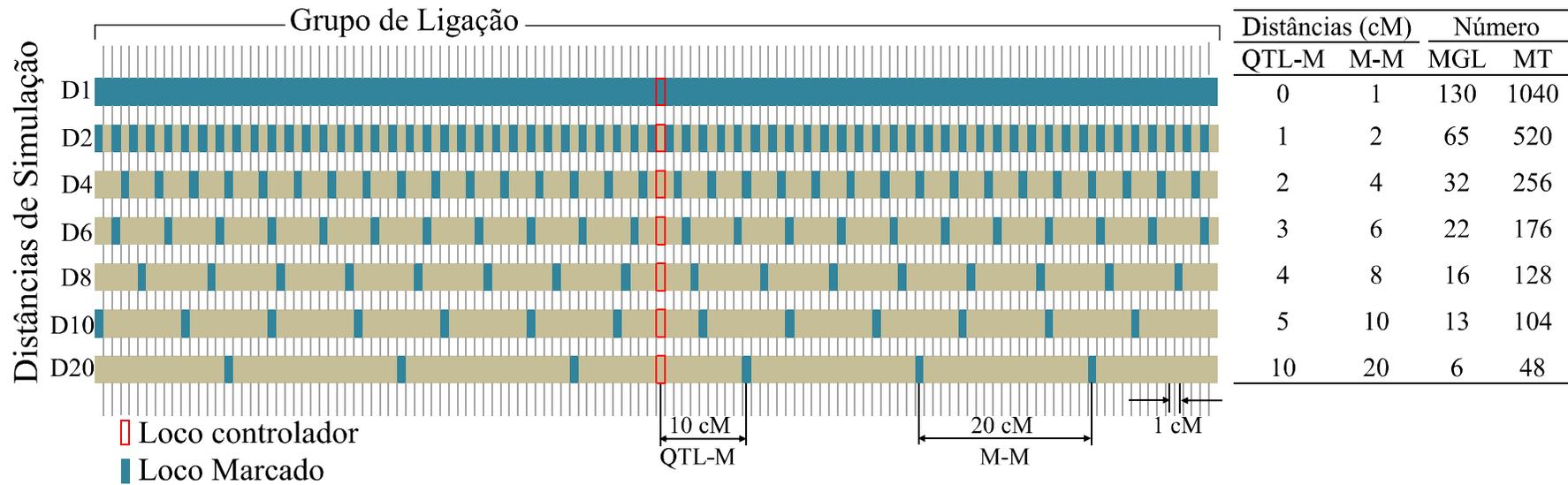


Figura 2. Esquema de distribuição de marcadores em um grupo de ligação segundo as distâncias analisadas. *QTL-M*: Distância entre *QTL* e loco Marcador; *M-M*: Distância entre locos Marcadores; *MGL*: Número de Marcadores por Grupo de Ligação; *MT*: Número de Marcadores no Total do genoma. Distâncias nomeadas segundo a separação entre marcadores. O D1 é o cenário controle, em que um dos marcadores corresponde ao mesmo *QTL*.

2.4. Cenários avaliados

Foram realizadas quatro análises principais, cada uma com diferentes arranjos experimentais complementares (Tabela 1), tendo em comum a análise de distâncias entre marcadores. Foram usadas populações de 625 indivíduos, com exceção da segunda análise. Cada cenário teve 100 repetições, com arranjos aleatórios nas populações de estimação e validação.

Tabela 1. Condições experimentais de cada cenário.

Análise	Tipo de marcador	Método GS	Porcentagem de seleção ^a	Quantidade de Marcadores ^b
Avaliações de métodos, intensidade de seleção, quantidade de marcadores				
1a	Codominante	<i>BLASSO</i>	20%	Total
1b	Codominante	<i>BLASSO</i>	10%	Total
2a	Codominante	<i>RR-BLUP</i>	20%	Total
2b	Codominante	<i>RR-BLUP</i>	10%	Total
3	Codominante	<i>RR-BLUP</i>	20%	Selecionados
Avaliações do tamanho populacional				
4a	Codominante	<i>RR-BLUP</i>	10%	Total
4b	Codominante	<i>RR-BLUP</i>	10%	Selecionados
Avaliação da não associação de marcadores a <i>QTL</i>				
5a	Codominante	<i>RR-BLUP</i>	10%	Total
5b	Codominante	<i>RR-BLUP</i>	10%	Selecionados
Avaliações do tipo de marcador				
6a	Dominante	<i>RR-BLUP</i>	10%	Total
6b	Dominante	<i>RR-BLUP</i>	10%	Selecionados

^a Porcentagem de seleção a partir da F_3 (na F_2 foram selecionados 50 indivíduos).

^b “Total” indica que foi usado o número de marcadores que fez cobertura total do genoma, segundo a densidade explicada na Figura 1. “Selecionados” indica que foram usados só os marcadores que maximizaram a acurácia preditiva, segundo a densidade do cenário.

O primeiro grupo de experimentos foi realizado para avaliar a acurácia dos métodos preditivos da GS, a resposta na fixação dos alelos favoráveis pela variação da intensidade de seleção e pela seleção de marcadores que maximizam a acurácia preditiva.

Para a seleção de marcadores em cada cenário, inicialmente foram pré-calculados os efeitos de todos os marcadores. Posteriormente, foram recalculados modelos e acurácias de grupos de marcadores. Primeiro foi analisado o marcador com maior valor absoluto

do efeito, o segundo grupo foi formado por os dois marcadores com maior efeito e assim por diante (Resende Jr *et al.*, 2012). Finalmente, para as análises foi usado o conjunto de marcadores que produziram o maior valor de acurácia preditiva, em cada cenário avaliado.

Na segunda análise, avaliou-se quatro tamanhos da F₂: 200, 300, 400 ou 500 indivíduos, com o total de marcadores ou com aqueles que maximizam a acurácia, para identificar o tamanho adequado nas diferentes densidades de marcadores.

Em condições reais, na análise genômica com baixa saturação de marcadores, regiões codificadoras podem não ser cobertas com marcadores, fator que pode afetar a eficácia da GS. Portanto, na terceira análise foram comparados genomas com um *QTL* não associado a marcadores, para quantificar as consequências desse evento.

Na quarta análise foram avaliadas populações transformando os marcadores codominantes em dominantes, para comparar a resposta à seleção usando esses marcadores, frequentemente empregados nos estudos de *C. arabica*.

3. RESULTADOS

3.1. Validação da autofecundação

Para verificar os processos de autofecundação foi calculado o coeficiente de endogamia de cada geração. Nas duas características simuladas, os valores observados foram próximos aos esperados (Tabela 2). Esse resultado confirma que os processos de simulação foram capazes de reproduzir o comportamento genético na autofecundação.

Tabela 2. Coeficiente de Endogamia médio em seis gerações de autofecundação de duas características.

Geração	Esperado	Observado	
		$h^2=0,8$	$h^2=0,4$
F2	0,000	-0,003	-0,003
F3	0,500	0,548	0,535
F4	0,750	0,787	0,778
F5	0,875	0,900	0,891
F6	0,937	0,948	0,944

3.2. Avaliação de métodos preditivos, intensidade de seleção e seleção de marcadores.

Foram simuladas populações com características de herdabilidade média e alta em sete cenários, constituídos de diferentes densidades de marcadores (Figura 1). Os cenários permitiram avaliar a resposta das populações à seleção, ao longo de cinco gerações analisadas, além de identificar os procedimentos ótimos para uma adequada estratégia de seleção (Tabela 1, análises 1 a 3).

Dependendo do cenário, a capacidade preditiva na geração F_2 apresentou valores entre 0,99 e 0,67, (Figura 3). Posteriormente, seus valores se tornaram decrescentes no avanço das gerações, em todos os cenários, até valores próximos a zero na geração F_6 , exceto nos cenários onde os alelos favoráveis foram fixados. Entretanto, os valores da capacidade seletiva se mantiveram quase constantes ao longo das gerações, mas em diferentes níveis segundo o cenário avaliado. Ao mesmo tempo foi observado o aumento geral das médias e a diminuição das variâncias a cada geração (Figura 4). Porém, cada cenário teve uma resposta diferente, resultando na fixação, ou não, dos alelos. O cálculo da capacidade seletiva se mostrou mais adequado para quantificar a eficácia na seleção.

Foram encontradas diferenças entre os métodos preditivos usados, e, em geral, o método *RR-BLUP* foi mais favorável que o *BLASSO*. Na característica com herdabilidade intermédia ($h^2=0,4$), as capacidades preditiva e seletiva foram maiores com o método *RR-BLUP* em todas as gerações e independentemente da intensidade de seleção (Figura 3), mas os alelos não foram fixados (Figura 4). Por sua vez, a característica com alta herdabilidade ($h^2=0,8$), usando as menores distâncias entre marcadores, só fixou os alelos com o método *RR-BLUP*, mesmo que as capacidades preditiva e seletiva foram similares nos dois métodos.

O aumento na porcentagem de seleção (*PS*) teve efeito tanto na eficácia (fixação dos alelos favoráveis, tendo como limite a geração F_6), quanto na eficiência, entendida como o uso de menos gerações e/ou menor densidade de marcadores para fixar os alelos. Isso foi mais claro na característica com herdabilidade alta, dado que a característica com herdabilidade intermédia não fixou os alelos os alelos favoráveis nos cenários um e dois (Figura 4).

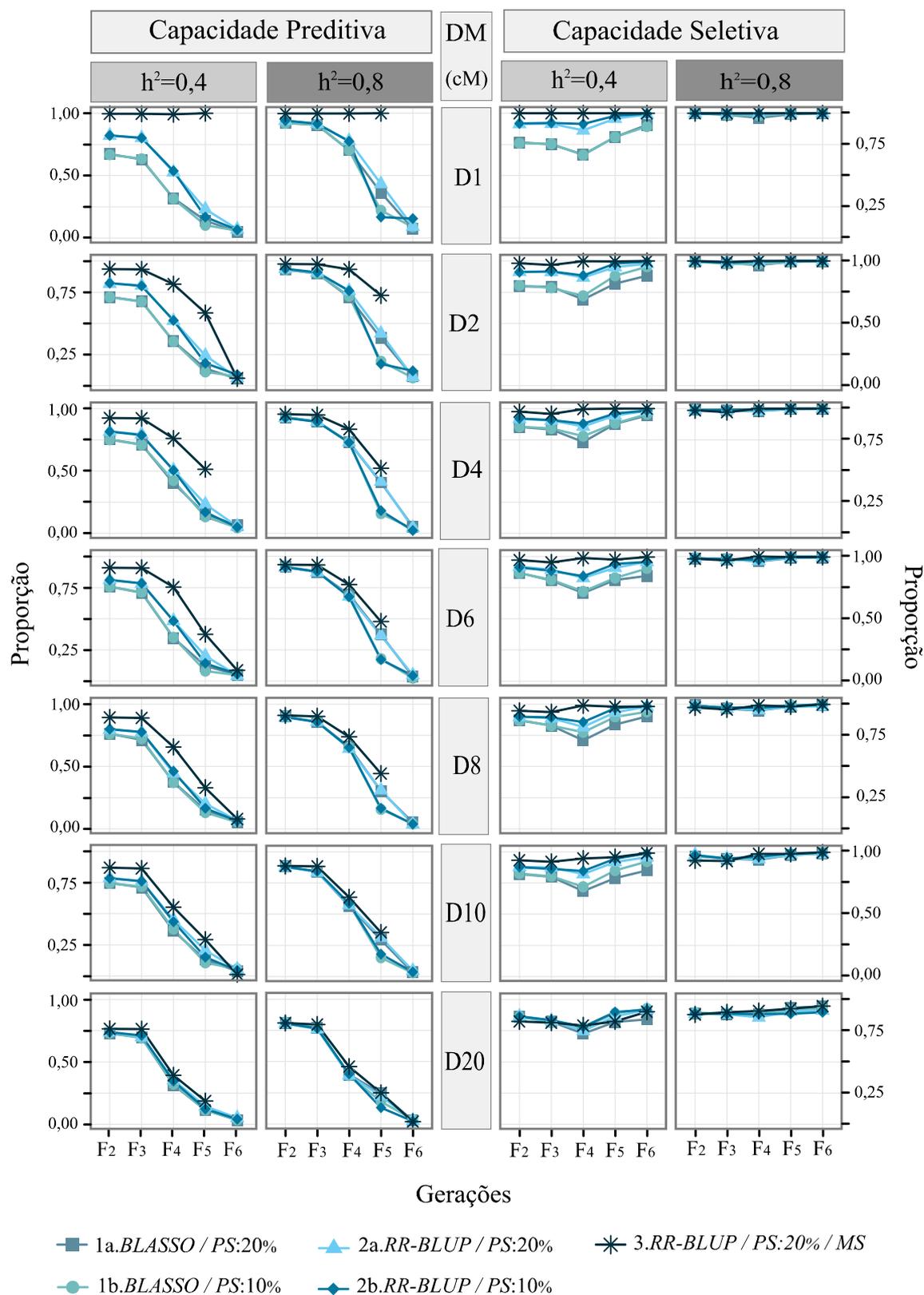


Figura 3. Capacidade preditiva dos modelos e capacidade seletiva em cinco cenários de simulação, cada um com sete distâncias entre marcadores (*DM*) em cM (Figura 1), e duas herdabilidades da característica ($h^2=0,4$ e $h^2=0,8$). Os cinco cenários correspondem à avaliação com o método *BLASSO* com duas porcentagens de seleção (*PS*): 20% e 10% (Tabela 1, Análise 1a e 1b); avaliação com o método *RR-BLUP* com as mesmas porcentagens de seleção (Tabela 1, Análise 2a e 2b); e a avaliação de marcadores selecionados (*MS*) com o método *RR-BLUP* com seleção de 20% (Tabela 1, Análise 3).

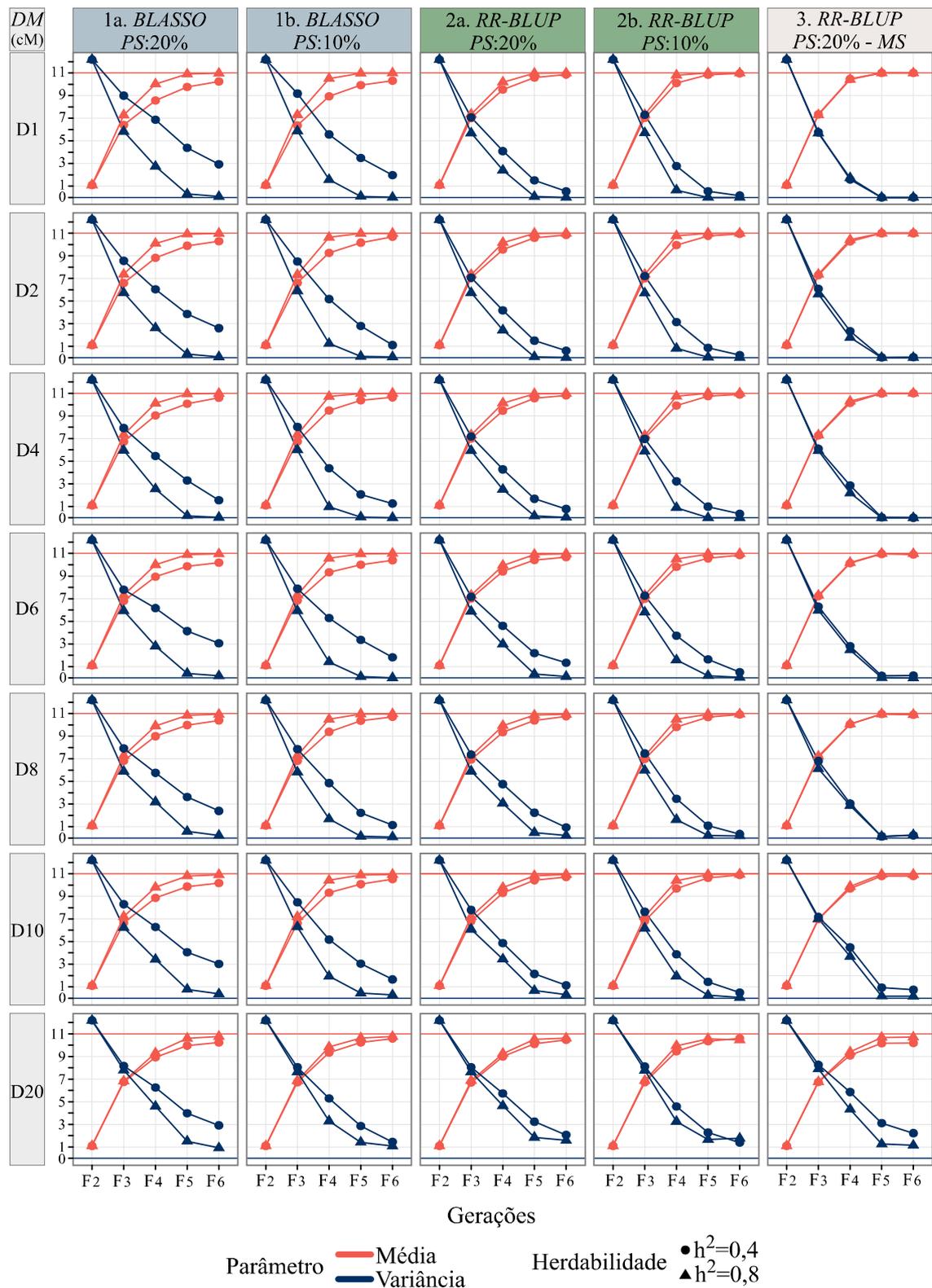


Figura 4. Médias e variâncias genóticas no avanço da seleção em cinco cenários de simulação, cada um com sete distâncias entre marcadores (*DM*), em cM, e duas herdabilidades da característica. Dois cenários usando o método *BLASSO* com duas porcentagens de seleção (*PS*): 20% e 10% (Tabela 1, Análises 1a e 1b); dois cenários com o método *RR-BLUP* com as mesmas *PS* (Tabela 1, Análises 2a e 2b); e a avaliação de marcadores selecionados (*MS*) com o método *RR-BLUP* e *PS* de 20% (Tabela 1, Análise 3).

Diminuir a proporção de indivíduos selecionados de 20% para 10%, usando o método *BLASSO*, passou de não fixar em nenhuma das características, a fixar os alelos em F_6 na característica de herdabilidade alta, usando distâncias entre marcadores inferiores a 6 cM. Por sua vez, com o método *RR-BLUP* e *PS* de 20% a fixação de alelos foi obtida com distâncias menores a 2 cM na geração F_6 , enquanto que com *PS* de 10% a fixação foi obtida na F_5 com distâncias menores de 4 cM, portanto melhorou a eficiência.

A seleção dos marcadores que maximizaram a acurácia foi o procedimento mais relevante para o aumento das capacidades preditiva e seletiva e em consequência para o aumento da eficácia e eficiência de fixação (Figuras 3 e 4). Esta análise foi feita unicamente com *PS* de 20% e o método *RR-BLUP* (Tabela 1, análise 3), superando a resposta do mesmo método quando foi usado o total de marcadores, do respectivo cenário. Esses resultados concordaram com os trabalhos de seleção de marcadores em *pinus*, milho e soja (Fritsche-Neto *et al.*, 2012; Resende Jr *et al.*, 2012; Ma *et al.*, 2016).

A melhoria na eficiência e na eficácia foi mais evidente na característica com herdabilidade intermédia, que teve um aumento significativo nas capacidades preditiva e seletiva, o que permitiu fixar os alelos favoráveis na geração F_5 , usando marcadores distanciados até 4 cM. A característica de herdabilidade alta também teve aumento na eficiência, permitindo fixar os alelos favoráveis na geração F_5 com marcadores com distância menor a 6 cM.

A seleção com marcadores distanciados por 20 cM aumentou a média populacional, mas, a diferença das distâncias menores a 10 cM, em nenhum dos cenários os alelos foram fixadas.

Por outro lado, comparando os cenários com grupo completo de marcadores ou com marcadores selecionados, a redução no desequilíbrio de ligação entre marcadores e o *QTL* foi similar na F_3 , mas na F_4 foi maior quando foram usados os marcadores selecionados (Figura 5). Os marcadores flanqueadores mostraram os maiores valores através das gerações. Os locos com distâncias maiores de 10 cM reduziram o *DL* rapidamente a valores abaixo de 0,2, com valores de 0,0 nas distâncias extremas.

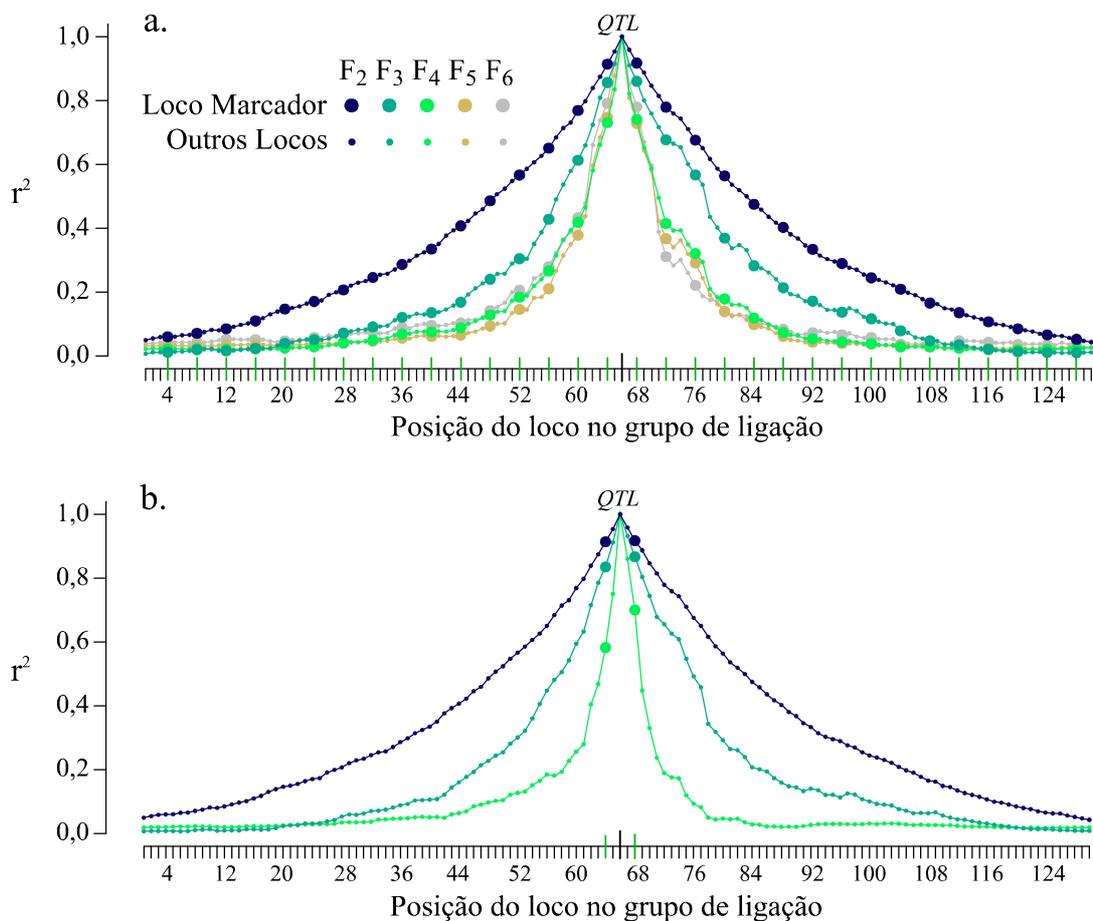


Figura 5. Desequilíbrio de ligação dos locos simulados no grupo de ligação e dos marcadores em relação a *QTL*, em cinco gerações de seleção e autofecundação. Informação da característica com $h^2=0,4$ nos cenários com distância entre marcadores de 4 cM, **a.** com marcadores igualmente espaçados (Tabela 1, análise 2a) e **b.** com marcadores selecionados (Tabela 1, análise 3), em que os alelos foram fixados na F_5 . Cada ponto é a média dos quatro grupos de ligação com *QTL*, de cada genoma, das 100 repetições.

3.3. Avaliações do tamanho populacional

Como o método *RR-BLUP* foi o que apresentou os melhores resultados em termos de eficácia e eficiência na fixação de alelos favoráveis, a partir da segunda etapa de avaliações, apenas esse método foi utilizado em todos os cenários para gerar o modelo preditivo, junto com alta intensidade de seleção (10%).

Inicialmente, foram feitas avaliações usando a totalidade de marcadores nos diferentes tamanhos populacionais da F_2 (Tabela 1, análise 4a). Para a característica com herdabilidade de 40% não houve fixação dos alelos em nenhum dos cenários, enquanto que a característica com 80% de herdabilidade fixou os alelos em F_6 , com distâncias entre marcadores com no máximo 6 cM, para populações de 500 indivíduos (Figura 6).

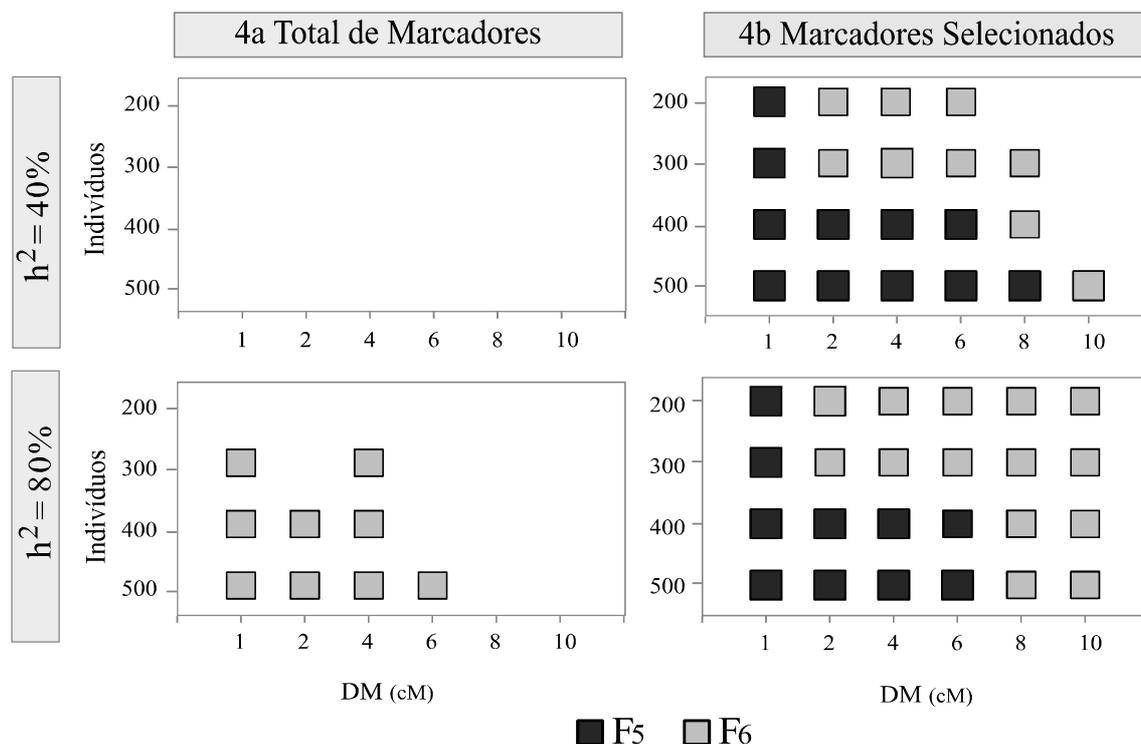


Figura 6. Gerações em que os alelos favoráveis foram fixados, considerando a distância entre marcadores (*DM*) e tamanhos populacionais (Indivíduos). Foram avaliadas duas características, uma com herdabilidade de 40% e outra de 80%, e duas versões do experimento, segundo a quantidade de marcadores usados. Análise 4a (Tabela 1): Uso do total de marcadores, de acordo com o cenário de distâncias da Figura 2; Análise 4b (Tabela 1): Uso de marcadores selecionados por maximizar a acurácia preditiva no respectivo cenário de *DM*. A porcentagem de seleção utilizada foi de 10%.

Em contraste, usando marcadores selecionados por aumentar a acurácia preditiva (Tabela 1, análise 4b.), a eficácia e a eficiência de seleção aumentaram significativamente em cada cenário. Houve fixação de alelos em ambas as características, tanto na geração F_6 quanto na F_5 , sendo relevante para a característica com herdabilidade intermédia (Figura 6).

Na característica com herdabilidade intermédia os alelos favoráveis foram fixados em todos os tamanhos populacionais avaliados. Para a população com 200 indivíduos, a maior distância entre marcadores que fixou em F_6 foi de 6 cM. A maior distância entre marcadores avaliada (10 cM) só conseguiu fixar os alelos favoráveis com a população de 500 indivíduos.

Por sua vez, na característica com herdabilidade alta, distâncias entre marcadores de 10 cM foram suficientes para fixar os alelos em F_6 , em qualquer dos tamanhos da população inicial. Para fixar em F_5 foram necessárias populações maiores de 400 indivíduos e distância entre marcadores igual ou menor a 6 cM.

A máxima distância entre marcadores selecionados e o mínimo tamanho populacional necessários para fixar em ambas as características em F_6 foram 8 cM e 300 indivíduos, respectivamente. Mas, a combinação mais eficiente, ao fixar os alelos em ambas as características na F_5 , foi de 6 cM entre marcadores e 400 indivíduos. Para o uso de maiores distâncias, seriam necessários maiores tamanhos da população F_2 .

3.4. Avaliação da não associação de marcadores a *QTL*

Com baixa cobertura do genoma é provável que regiões controladoras da característica não sejam cobertas com marcadores. Portanto, é fácil presumir que isso afete diretamente a eficácia da seleção, nesse caso a explicação da variação da característica seria parcial. Para quantificar esse problema foram avaliados genomas com uma das regiões controladoras sem marcação, ao longo de várias gerações (Tabela 1, análise 5).

Independentemente da densidade de marcadores, a capacidade preditiva foi afetada quando um dos *QTL* não foi associado a marcadores. Na geração F_2 , por exemplo, esta foi diminuída em 30%, em relação à capacidade preditiva quando teve associação com todos os *QTL*. Nas gerações posteriores, a capacidade preditiva caiu rapidamente, até valores perto de zero. Na avaliação com marcadores selecionados não foi possível calcular devido à fixação dos *QTL* marcados (variância igual a zero) (Figura 7).

A seleção ao longo das gerações aumentou o valor genotípico, mas, em nenhum dos cenários a média atendeu ao valor esperado para homozigotos dominantes (nesta simulação igual a 11), ao contrário, tenderam a estabilizar-se em um valor inferior (Figura 7). As variâncias, por sua vez, apresentaram valores distantes de zero, portanto, em qualquer dos casos a seleção tornou-se ineficiente.

As frequências alélicas indicaram que o incremento das médias foi resultado do aumento na frequência dos *QTL* associados a marcadores, devido à seleção. Em média a frequência do *QTL* sem associação teve um ligeiro aumento quando foram usados todos os marcadores. Em contraste, o uso de marcadores selecionados produziu uma resposta diferencial nas distâncias analisadas, mas não teve aumento da frequência.

Por outro lado, este cenário sugeriu que características com menos de quatro locos controladores podem ser fixados em poucas gerações de autofecundação e seleção.

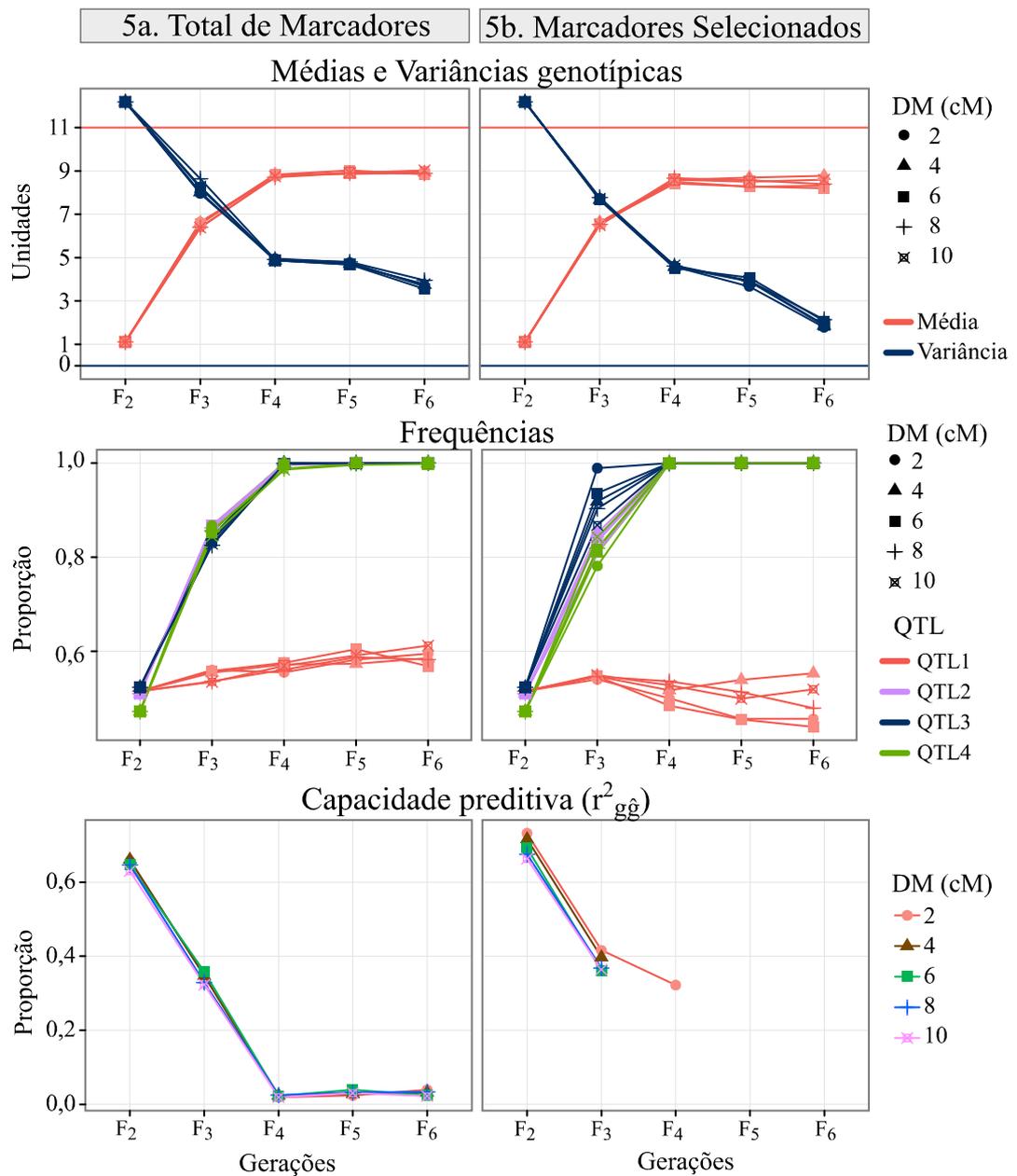


Figura 7. Parâmetros da característica com 80% de herdabilidade, analisada sem a associação de marcadores a um *QTL*, até a sexta geração. Na análise não foram colocados marcadores no grupo de ligação dois. Foram avaliadas cinco distâncias entre marcadores (*DM*). Resultados com referência às análises da Tabela 1, análise 5a, com o número total de marcadores e análise 5b, com marcadores selecionados que maximizaram a acurácia preditiva.

3.5. Avaliações do tipo de marcador

Os resultados da seleção com marcadores dominantes são apresentados na Figura 8, comparáveis com o cenário 2b-D1 das Figuras 3 e 4, com parâmetros de simulação equivalentes, mas com marcadores codominantes.

Com o uso de marcadores dominantes, a capacidade preditiva na F_2 foi 30% inferior, comparada com o uso de marcadores codominantes. Isso levou ao aumento da variância genotípica na F_3 , de forma similar como aconteceu na seleção fenotípica com $h^2=0,4$,

descrita no capítulo I. Contudo, a partir da F₄, a seleção com marcadores dominantes aumentou a média e a variância genotípica foi reduzida (Figura 8). O processo de seleção resultou no aumento progressivo das frequências alélicas. No entanto, a diferença do resultado com marcadores codominantes que fixou os alelos favoráveis na F₅, estas não atingiram o valor máximo e, por conseguinte, os alelos não foram fixados durante as gerações analisadas, embora a tendência indicasse que os alelos seriam fixados nas próximas gerações. Esses resultados foram similares para as duas herdabilidades.

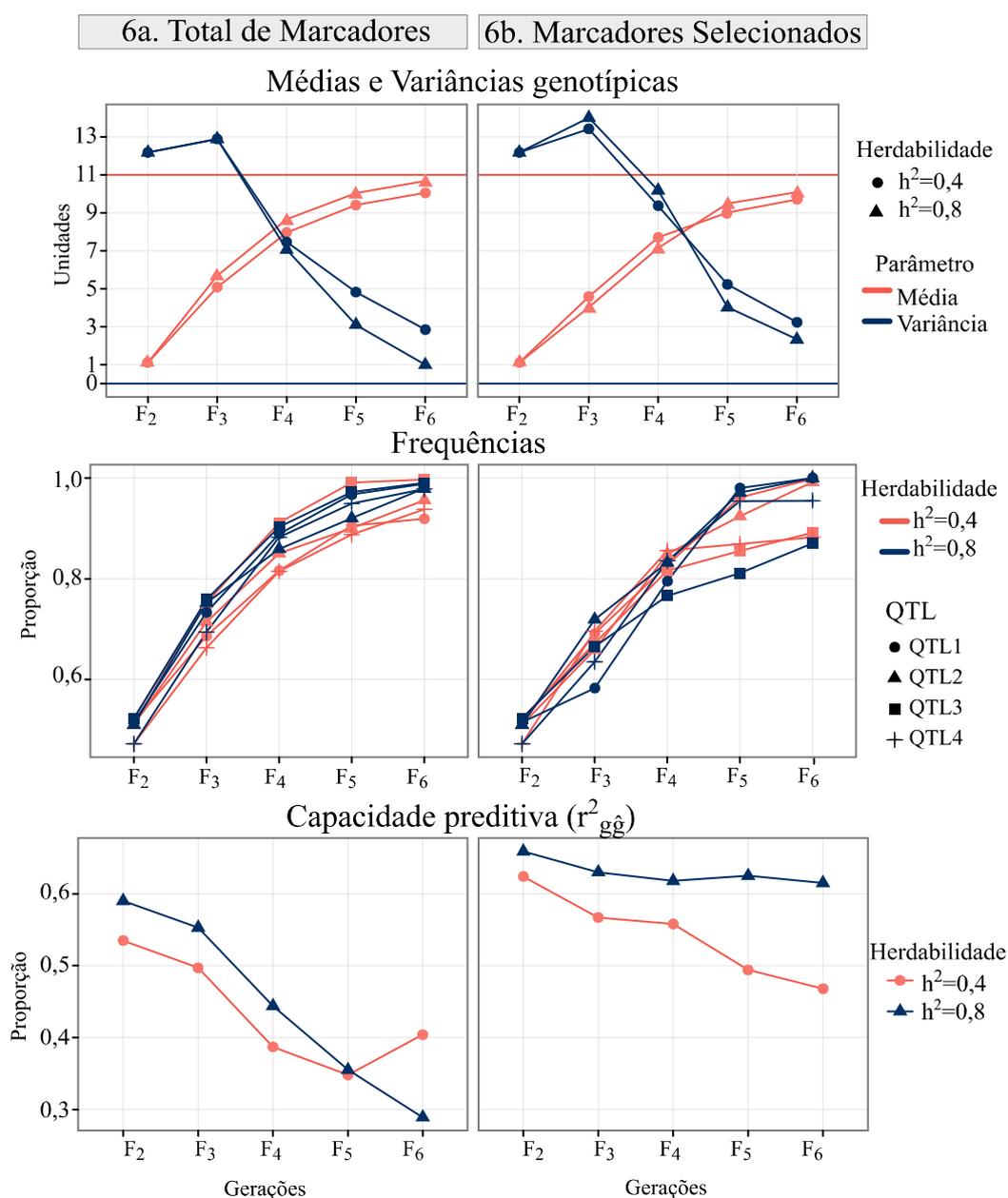


Figura 8. Parâmetros para características com 40% e 80% de herdabilidade, analisadas usando marcadores dominantes e com um cM entre marcadores, até a sexta geração. Resultados com referência às análises da Tabela 1, análise 6a, com o número total de marcadores e análise 6b, com marcadores selecionados que maximizaram a acurácia preditiva.

Como nas análises anteriores, o uso de marcadores selecionados melhorou a capacidade preditiva, mas isso não foi refletido em uma maior frequência de alelos favoráveis. Além disso, os valores de variância foram levemente superiores aos das variâncias registradas para o conjunto de marcadores completo.

4. DISCUSSÃO

Os programas de melhoramento dependem da criação de estratégias que potencializem os recursos físicos, humanos e de tempo, para aumentar o ganho genético com o menor custo. Neste trabalho, as análises de métodos de seleção genômica foram restritas a características oligogênicas (controladas por quatro locos) em uma espécie autógama, procurando os cenários com as mínimas densidades de marcadores e tamanhos populacionais que permitam fixar os alelos favoráveis.

No capítulo anterior foi visto que, em populações de 200 indivíduos e com um método similar ao massal, a seleção fenotípica não conseguiu fixar os alelos favoráveis até a geração F_6 . Com o mesmo tamanho populacional, a seleção genômica fixou os alelos favoráveis, nas características com herdabilidade intermédia ou alta, usando pelo menos 6 cM entre marcadores. Portanto, esse método foi mais eficaz do que a seleção fenotípica, sendo a primeira evidência da viabilidade da seleção genômica com restrição do número de marcadores e tamanhos populacionais.

No entanto, foi observado que os alelos favoráveis foram fixados quando a capacidade preditiva na F_2 foi superior a 0,9, visto que a confiabilidade da estimativa dos efeitos dos marcadores foi reduzida ao longo das gerações, como também foi observado por Michel *et al.* (2016). Os valores adequados de capacidade preditiva só foram obtidos com a seleção do método preditivo adequado e com o uso de marcadores codominantes, selecionados por maximizarem a acurácia preditiva. Também foi necessário o aumento da intensidade de seleção.

O método *BLASSO* tinha a expectativa de ser mais eficiente para características oligogênicas, por considerar que poucos marcadores tem efeito direto sobre a característica, penalizando a maioria deles com efeitos perto de zero (De los Campos *et al.*, 2009). Mas, nas condições deste trabalho, o método *RR-BLUP* foi mais acurado do que o *BLASSO*, ainda assumindo que a característica depende de um número grande de *QTL* com pequeno efeito (Resende *et al.*, 2010).

O *RR-BLUP* por ser menos complexo que o *BLASSO*, possui menor exigência computacional. Essa vantagem foi usada para identificar os marcadores que maximizaram a acurácia preditiva, já que para essa análise são necessários múltiplos ciclos de reestimação da acurácia, usando consecutivos conjuntos de marcadores (Resende Jr *et al.*, 2012). A identificação e seleção destes marcadores fez o trabalho de penalização de forma direta, o que aumentou significativamente a acurácia. Esse foi o principal fator para alcançar alta eficácia em seleção e em consequência em fixação.

Para características oligogênicas, a eficácia de seleção foi influenciada pela distância entre marcadores e a eficiência pelo número de marcadores usado no modelo preditivo. Marcadores distanciados dos *QTL* por mais de 10 cM não contribuíram para a correta seleção de genótipos. Os alelos não foram fixados na população, dado que estes locos perdem rapidamente o *DL* com o *QTL*, prejudicando a predição de valores genotípicos. Quando a característica depende de poucos genes é desvantajoso o uso de grande número de marcadores, porque os efeitos estimados dos marcadores não associados perturbam a correta predição dos *VGG*, sendo recomendável a seleção de marcadores que aumentem a acurácia. Essa seleção descarta os locos que não estão associados com a característica, melhorando a estimação de efeitos no modelo (Fritsche-Neto *et al.*, 2012; Resende Jr *et al.*, 2012; Ma *et al.*, 2016).

Além disso, o uso de grande número de marcadores incrementa os custos. Assim, do ponto de vista prático, é ideal procurar o menor número de marcadores que explique a maioria da variância genética. Por isso, a seleção de marcadores é uma excelente opção que contribui para uma adequada relação custo-benefício (Ma *et al.*, 2016).

Os marcadores que maximizaram a acurácia preditiva, em todos os casos corresponderam aos marcadores flanqueadores do *QTL*. Isso diminuiu a probabilidade de perder o alelo favorável devido à recombinação. Ma *et al.* (2016) avaliaram três estratégias de amostragem de marcadores em soja: uniforme, aleatória e em blocos de haplótipos, sendo esta última a mais adequada, coincidindo com nosso resultado.

O outro elemento que contribuiu na eficácia seletiva foi o aumento na intensidade de seleção, importante componente na estimação do ganho por seleção. McClosky *et al.* (2013) observaram que a causa principal do aumento dos ganhos com seleção genômica em autógamias, não se deve a melhores previsões dos *VGG* do que em alógamas, e sim pelo aumento de genótipos superiores (homozigotos), a causa da autofecundação. Então, com o aumento do viés na acurácia de predição a cada geração selecionada (Michel *et*

al., 2016), o mecanismo que produz o progresso genético estaria condicionado ao descarte de indivíduos com menor valor genético (recessivos e heterozigotos), como um tipo de seleção negativa (Ma *et al.*, 2016). Neste contexto, foi observado que quanto maior a intensidade de seleção, maior a exclusão dos indivíduos com menor valor genético, o que foi refletido na fixação de alelos em menos gerações.

Neste trabalho, a simulação seguiu um delineamento simples para o avanço das gerações, semelhante ao método massal. É possível que com o uso de métodos mais estruturados, com seleção entre e dentro de progênies (como o genealógico), a seleção por valores genéticos preditos poderia ser ainda mais eficiente.

Por outro lado, em *C. canephora* foi estimado que 1 cM equivale a 460 Kb (Paillard *et al.*, 1996). Usando esse valor e considerando um genoma de *C. arabica* de 1.28 Gb (Clarindo & Carvalho, 2010), o tamanho estimado dessa espécie seria de 2780 cM. Como para fixar alelos favoráveis em populações de estimação de 200 indivíduos, foram necessários pelo menos 6 cM entre marcadores, assim, assumindo uma cobertura homogênea, seriam necessários aproximadamente 464 marcadores para cobrir o genoma de *C. arabica*.

No entanto, essa condição nem sempre pode ser satisfeita. Por isso foi analisada a situação em que, devido à baixa densidade, alguns dos locos controladores da característica não apresentavam marcadores ligados. Nesse cenário, foi verificado que a ausência de ligação de marcadores com um dos *QTL* levou ao aumento no valor da média, mas impediu a fixação de alelos. Este seria o maior risco de se utilizar baixas densidades de marcadores, dado que em espécies como *C. arabica* a cobertura do genoma é difícil, principalmente se são considerados unicamente marcadores codominantes.

Na literatura tem sido amplamente reportado o uso de marcadores dominantes nesta espécie (Chaparro *et al.*, 2004; Pearl *et al.*, 2004; Teixeira-Cabral *et al.*, 2004; Gichuru *et al.*, 2008; Brito *et al.*, 2010; Setotaw *et al.*, 2010; Diola *et al.*, 2011; Pestana *et al.*, 2015). No entanto, no presente estudo foi observado que com marcadores dominantes a capacidade preditiva da *GS* foi 30% menor na F_2 , em comparação com o marcador codominante, resultando em avanços menos eficazes, o que deve ser levado em conta caso esse tipo de marcador seja utilizado.

4.1. Aplicação em programas de melhoramento de café

A comprovada capacidade da *GS* sobre a inferência dos valores genéticos gera alta eficiência preditiva, o que permite a obtenção de rápidos ganhos genéticos (Resende *et al.*, 2012). Essa vantagem pode ser aproveitada, se os desenvolvimentos gerados ao redor da *GS* são adaptados às necessidades e alcances de cada programa de melhoramento. Este trabalho explorou as condições em que esse método poderia ser usado com recursos limitados, que seria utilizado para a seleção de características menos complexas, como as oligogênicas, mas de alta importância em qualquer programa de melhoramento.

Se as condições para aumentar a acurácia preditiva forem satisfeitas em *C. arabica*, junto com a combinação adequada de densidade de marcadores e tamanho populacional, a seleção genômica permitiria incrementar a eficácia e eficiência dos programas, reduzindo custos por genotipagem e fenotipagem. Além disso, o ganho genético seria aumentado com a seleção intensa em um número grande de mudas, antes de serem transplantadas. Assim, seriam transferidos ao campo apenas os melhores indivíduos, economizando espaço e custos de manutenção. No entanto, não seria possível obter ganho através da redução no tempo, dado que no programa de melhoramento é necessário verificar a produtividade das populações e para isso é necessário completar o ciclo de cultura.

No caso do melhoramento visando a incorporação de resistência a doenças, a seleção efetiva é feita sob alta pressão da doença. Essa situação nem sempre acontece devido a que as condições ambientais, no tempo e/ou no espaço, não são sempre favoráveis ao patógeno ou pela inviabilidade da inoculação artificial no campo. Assim, a seleção genômica, que uma vez gerado o modelo não precisa da fenotipagem, facilitaria a seleção na ausência da doença nas populações de seleção.

No melhoramento genético visando a resistência a nematoides, que é considerado o método mais econômico e eficaz para o controle destes em cafeeiro, a avaliação da resistência é feita em casa de vegetação, sendo necessária a clonagem de indivíduos e amostragem destrutiva, além de outras particularidades que fazem a fenotipagem complicada (Gonçalves & Silvarolla, 2007; Pereira *et al.*, 2012; Fatobene, 2014). Assim, de forma geral, a caracterização da resistência a doenças é um processo laborioso e demorado, que requer uma série de inoculações individuais (Prakash *et al.*, 2011). A seleção genômica possibilita a estimação de *VGG* dos indivíduos, reduzindo

os esforços em testes de campo de características difíceis, laboriosas ou custosas, como essas (Rutkoski *et al.*, 2011).

Por outro lado, no caso da resistência à ferrugem (*Hemileia vastatrix*), em que diferentes genótipos têm formas alélicas dominantes e/ou recessivas que atuam segundo a raça do patógeno (Pereira, 1995; Capucho *et al.*, 2009; Pestana, 2010; Prakash *et al.*, 2011; Pestana *et al.*, 2015), a seleção genômica seria eficiente pela transferência simultânea dos vários genes controladores, com a vantagem do avanço na seleção ainda sem a caracterização desses genes. Isso aumentaria a probabilidade de altos níveis de resistência, gerando resistência duradoura.

CONCLUSÕES

Segundo a avaliação da viabilidade da seleção genômica de características oligogênicas, considerando a genotipagem de baixa densidade e limitado tamanho populacional, pode-se concluir que:

Nas condições acima mencionadas, os métodos da seleção genômica podem ser utilizados para a seleção de características oligogênicas em *C. arabica*, porém, considerando que se não se conhecem as possíveis regiões controladoras da característica, a marcação no genoma deve ser uniforme. Além disso, altos níveis de acurácia do modelo devem ser alcançados com a seleção de marcadores e com a configuração adequada de densidades de marcadores e tamanhos populacionais.

Para fixar os alelos favoráveis até a F₆, com pequenas populações (~200 indivíduos), são necessárias distâncias entre marcadores no máximo de 6 cM. No entanto, 400 indivíduos e distâncias de 6 cM entre marcadores foi a combinação avaliada mais conveniente, pois nessas condições os alelos favoráveis foram fixados na geração F₅, nas características com herdabilidades intermédias e altas.

Em *C. arabica* a dificuldade de conseguir marcadores codominantes e/ou o uso de densidades em que regiões codificadoras não sejam marcadas, são os principais inconvenientes para a utilização da seleção genômica nos cenários avaliados.

5. REFERÊNCIAS

- BHERING, L. L. Rbio: A tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology**, v. 17, p. 187-190. 2017.
- BRITO, G. G.; CAIXETA, E. T.; GALLINA, A. P.; ZAMBOLIM, E. M.; ZAMBOLIM, L.; DIOLA, V.; LOUREIRO, M. E. Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. **Euphytica**, v. 173, n. 1, p. 255-264. 2010.
- BYRNE, D. H. Molecular marker use in perennial plant breeding. **Acta Horticulturae**, v. 751, p. 163-167. 2007.
- CAPUCHO, A. S.; CAIXETA, E. T.; ZAMBOLIM, E. M.; ZAMBOLIM, L. Herança da resistência do Híbrido de Timor UFV 443-03 à ferrugem-do-cafeeiro. **Pesquisa Agropecuária Brasileira**, v. 44, n. 3, p. 276-282. 2009.
- CLARINDO, W. R.; CARVALHO, C. R. Flow cytometric analysis using SYBR Green I for genome size estimation in coffee. **Acta Histochemica**, v. 113, n. 2, p. 221-225. 2010.
- CRUZ, C. D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v. 35, n. 3, p. 271-276. 2013.
- CHAPARRO, A. P.; CRISTANCHO, M. A.; CORTINA, H. A.; GAITAN, A. L. Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. **Genetic Resources and Crop Evolution** v. 51, n. 1, p. 291-297. 2004.
- DE LOS CAMPOS, G.; NAYA, H.; GIANOLA, D.; CROSSA, J.; LEGARRA, A.; MANFREDI, E.; WEIGEL, K.; COTES, J. M. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, v. 182, n. 1, p. 375-385. 2009.
- DIOLA, V.; DE BRITO, G. G.; CAIXETA, E. T.; MACIEL-ZAMBOLIM, E.; SAKIYAMA, N. S.; LOUREIRO, M. E. High-density genetic mapping for coffee leaf rust resistance. **Tree Genetics & Genomes**, v. 7, n. 6, p. 1199-1208. 2011.
- ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome** v. 4, n. 1, p. 250-255. 2011.
- FATOBENE, B. J. **Seleção de cafeeiros com resistência múltipla a nematoides do gênero *Meloidogyne***. 2014. 71 f. Tese (Doutorado)- Instituto Agronômico. Curso de pós-graduação em agricultura tropical e subtropical, Campinas, SP, 2014.
- FRITSCHÉ-NETO, R.; DO VALE, J. C.; LANE, E.; RESENDE, M. D. V.; MIRANDA, G. Genome wide selection for root traits in tropical maize under stress conditions of nitrogen and phosphorus. **Acta Scientiarum Agronomy**, v. 34, n. 4, p. 389-395. 2012.
- GICHURU, E. K.; AGWANDA, C. O.; COMBES, M. C.; MUTITU, E. W.; NGUGI, E. C. K.; BERTRAND, B.; LASHERMES, P. Identification of molecular markers linked

to a gene conferring resistance to coffee berry disease (*Colletotrichum kahawae*) in *Coffea arabica* L. **Plant Pathology**, v. 57, n. 6, p. 1117-1124. 2008.

GONÇALVES, W.; SILVAROLLA, M. B. A. Luta contra a doença causada pelos nematóides parasitos do cafeeiro. **O Agrônômico**, v. 59, n. 1, p. 54-56. 2007.

KAINER, D.; LANFEAR, R.; FOLEY, W. J.; KÜLHEIM, C. Genomic approaches to selection in outcrossing perennials: focus on essential oil crops. **Theoretical and Applied Genetics**, v. 128, n. 12, p. 2351-2365. 2015.

MA, Y.; REIF, J. C.; JIANG, Y.; WEN, Z.; WANG, D.; LIU, Z.; GUO, Y.; WEI, S.; WANG, S.; YANG, C.; WANG, H.; YANG, C.; LU, W.; XU, R.; ZHOU, R.; WANG, R.; SUN, Z.; CHEN, H.; ZHANG, W.; WU, J.; HU, G.; LIU, C.; LUAN, X.; FU, Y.; GUO, T.; HAN, T.; ZHANG, M.; SUN, B.; ZHANG, L.; CHEN, W.; WU, C.; SUN, S.; YUAN, B.; ZHOU, X.; HAN, D.; YAN, H.; LI, W.; QIU, L. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.). **Molecular Breeding**, v. 36, n. 8, p. 113. 2016.

MCCLOSKEY, B.; LACOMBE, J.; TANKSLEY, S. D. Selfing for the design of genomic selection experiments in biparental plant populations. **Theoretical and Applied Genetics**, v. 126, p. 2907-2920. 2013.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 1, p. 1819-1829. 2001.

MICHEL, S.; AMETZ, C.; GUNGOR, H.; EPURE, D.; GRAUSGRUBER, H.; LÖSCHENBERGER, F.; BUERSTMAYR, H. Genomic selection across multiple breeding cycles in applied bread wheat breeding. **Theoretical and Applied Genetics**, v. 129, p. 1179–1189. 2016.

PAILLARD, M.; LASHERMES, P.; PETIARD, V. Construction of amolecular linkage map in coffee. **Theoretical and Applied Genetics**, v. 93, n. 1-2, p. 41–47. 1996.

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

PEARL, H. M.; NAGAI, C.; MOORE, P. H.; STEIGER, D. L.; OSGOOD, R. V.; MING, R. Construction of a genetic map for arabica coffee. **Theoretical and Applied Genetics**, v. 108, n. 5, p. 829-835. 2004.

PEREIRA, A. A. **Herança da resistência a Hemileia vastatrix Berk. et Br. Em cafeeiros derivados do Híbrido de Timor.** . 1995. 66 f. Tese (Doctor Scientiae)-Universidade Federal de Viçosa, Viçosa-MG, 1995.

PEREIRA, T. B.; SALGADO, S. M.; CARVALHO, G. R.; PEREIRA, A. A.; FERREIRA, A. D.; OLIVEIRA, L. P. Relação de genótipos de cafeeiro (*Coffea arabica* L.) a *Meloidogyne exigua* população sul de Minas. **Coffee Science**, v. 7, n. 1, p. 84-90. 2012.

PESTANA, K. **Caracterização fenotípica e molecular da resistência do cafeeiro híbrido de timor A *Hemileia vastatrix***. 2010. 60 f. Tese (MSc. Fitotecnia)-Universidade Federal de Viçosa, Viçosa, 2010.

PESTANA, N. K.; CAPUCHO, A. S.; CAIXETA, E. T.; ZAMBOLIM, E. M.; ALMEIDA, D.; CRUZ, C. D.; ZAMBOLIM, L.; PEREIRA, A. A.; OLIVEIRA, A. C. B.; SALIYAMA, N. S. Inheritance study and linkage mapping of resistance loci to *Hemileia vastatrix* in Híbrido de Timor UFV 443-03. **Tree Genetics & Genomes**, v. 11. 2015.

PRAKASH, N. S.; MUNISWAMY, B.; HANUMANTHA, B. T.; SREENATH, H. L.; SUNDARESHA, K. D.; SURESH, N.; SANTHOSH, P.; SOUMYA, P.; ASHA, B. M.; BHAT, S. S.; JAYARAMA. Marker assisted selection and breeding for leaf rust resistance in coffee (*Coffea arabica* L.). **Indian Journal of Genetics & Plant Breeding** v. 71, n. 2, p. 185-189. 2011.

R DEVELOPMENT CORE TEAM **R: A language and environment for statistical computing**. Vienna, Austria, R Foundation for Statistical Computing. 2016.

RAMALHO, M. A. P.; CARVALHO, B. L.; NUNES, J. A. R. Perspectives for the use of quantitative genetics in breeding of autogamous plants. **ISRN Genetics**, v. 2013, 6 p. 2013. Disponível em: <http://dx.doi.org/10.5402/2013/718127>. Acesso em: 04/10/2016.

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

RESENDE, M. D. V.; SILVA, F. F.; LOPES, P. S.; AZEVEDO, C. F. **Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada e estatística espacial**. Viçosa, MG: Editora UFV, 2012. 291 p.

RESENDE, M. D. V.; RESENDE JÚNIOR, M. F. R.; AGUIAR, A. M.; ABAD, J. I. M.; MISSIAGGIA, A. A.; SANSALONI, C.; PETROLI, C.; GRATTAPAGLIA, D. **Computação da Seleção Genômica Ampla (GWS)**. Colombo, PR: Embrapa Florestas, 2010. 79 p.

RUTKOSKI, J. E.; HEFFNER, E. L.; SORRELLS, M. E. Genomic selection for durable stem rust resistance in wheat. **Euphytica**, v. 179, n. 1, p. 161-173. 2011.

SERA, T. Coffee genetic breeding at IAPAR. **Crop Breeding and Applied Biotechnology**, v. 1, n. 2, p. 179-199. 2001.

SETOTAW, T. A.; CAIXETA, E. T.; PENA, G. F.; ZAMBOLIM, E. M.; PEREIRA, A. A.; SAKIYAMA, N. S. Breeding potential and genetic diversity of "Híbrido do Timor" coffee evaluated by molecular markers. **Crop Breeding and Applied Biotechnology**, v. 10, p. 298-304. 2010.

TEIXEIRA-CABRAL, T. A.; SAKIYAMA, N. S.; ZAMBOLIM, L.; PEREIRA, A. A.; SCHUSTER, I. Single-locus inheritance and partial linkage map of *Coffea arabica* L. **Crop Breeding and Applied Biotechnology**, v. 4, n. 1, p. 416-421. 2004.

WELLMANN, R.; PREUß, S.; THOLEN, E.; HEINKEL, J.; WIMMERS, K.; BENNEWITZ, J. Genomic selection using low density marker panels with application to a sire line in pigs. **Genetics Selection Evolution**, v. 45, n. 1, 11 p. 2013. Disponível em: <http://www.gsejournal.org/content/45/1/28>. Acesso em: 14 feb 2015.

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. 1, p. 3642-3650. 2011.

APÊNDICE A. Validação de rotinas para simulação de genótipos e valores genotípicos e fenotípicos executáveis no programa *R*.

Para o cumprimento dos objetivos, as análises precisaram de ciclos de simulação de populações, estimação e avaliação de parâmetros. Para otimizá-las, viu-se a necessidade de integrar estes ciclos em um único ambiente e, portanto, foram desenvolvidas rotinas executáveis no programa *R*. Esse incluíram duas rotinas de simulação: uma para o desenvolvimento dos genótipos de cada geração e outra para a simulação dos seus respectivos valores genotípicos e fenotípicos.

Neste anexo é apresentada a validação específica das rotinas, por meio da geração e avaliação de uma população F_2 de 500 indivíduos, procedente de um cruzamento inicial de progenitores homozigotos contrastantes. O genótipo foi gerado com 520 locos equidistantes a um cM, com quatro grupos de ligação de 130 cM e cada um deles com um *QTL* com efeito uniforme e ausência de dominância. Foram avaliadas duas características: uma correspondente aos valores genotípicos (*VGen*) e outra com herdabilidades de 80% (*VFen80*). Essa população não faz parte do núcleo de análises do trabalho, só foi usada para esta validação.

A validação foi feita seguindo os procedimentos descritos nos métodos do capítulo I. A continuação os resultados obtidos:

O mapa pode ser detalhado na Figura A1 e a comparação das suas variáveis na Tabela A1.

Tabela A1. Comparação do número esperado e simulado das principais variáveis do mapa.

Variável	Número esperado	Número simulado
Grupos de ligação	4	4
Número de marcas por grupo	130	130
Tamanho médio dos grupos	130	129,38
Distância média entre marcadores	1	0,99
Variância distância entre marcadores	0	0,09

Na segregação de locos individuais, realizado testes de qui-quadrado (5%) com a hipótese de segregação de 1:2:1, o resultado indicou que todos os locos apresentaram a razão esperada. Na ordem do genoma, segundo a correlação de Spearman, não houve inversão da posição de marcas nos grupos de ligação. O valor do estresse foi baixo (0,3).

As posições dos *QTL* em cada grupo de ligação corresponderam com aqueles preestabelecidos ($M66 = QTL1$; $M196 = QTL2$; $M326 = QTL3$; $M456 = QTL4$), localizados na região mediana dos respectivos grupos de ligação, segundo foi definido (Figura A2).

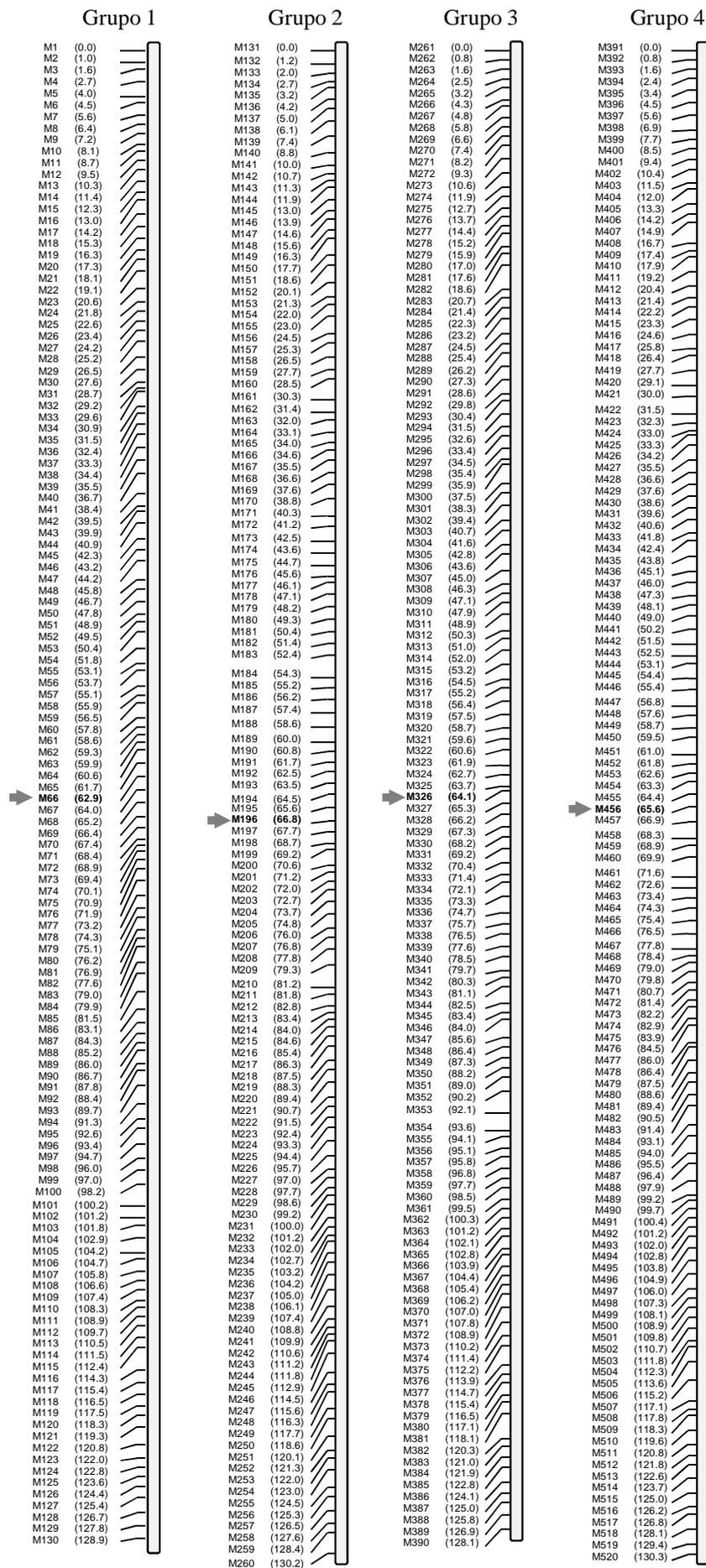


Figura A1. Mapa de ligação de uma população F₂ de 500 indivíduos e 520 marcadores

A proporção esperada da variação fenotípica explicada por cada *QTL* para *VGen* era de 25% e para *VFen80* de 20%, Foram obtidos os seguintes valores para cada *QTL* das respectivas características: 23,5, 22,7, 25,6, 30,0 e 19,0, 17,2, 19,9, 26,1.

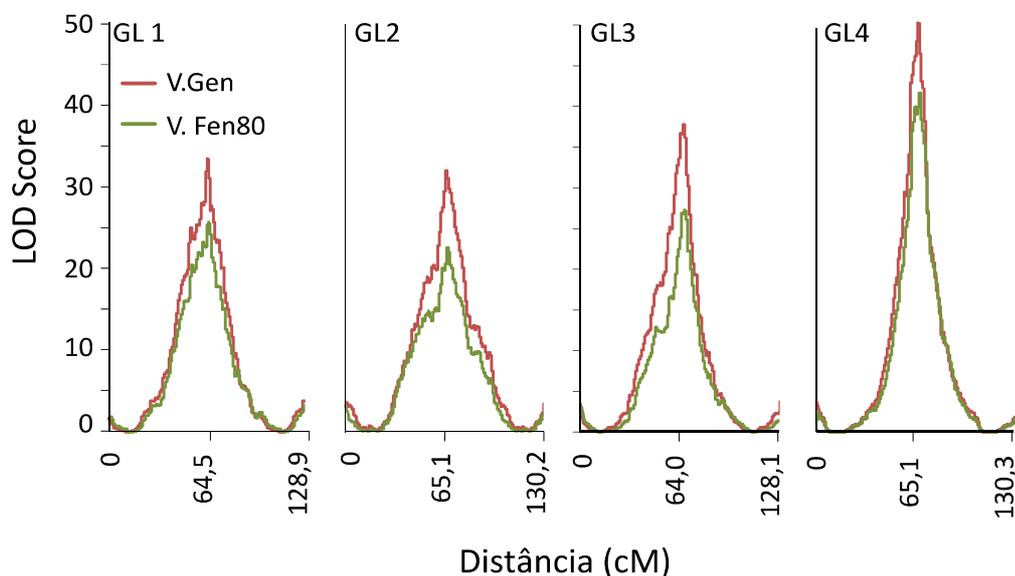


Figura A2. Localização dos *QTL* pelo método de Intervalo Simples para as duas características simuladas. GL: grupos de ligação.

Validação da simulação dos valores genotípico e fenotípico

Os valores genotípicos e fenotípicos gerados pelas rotinas reproduziram adequadamente os parâmetros de simulação indicados: médias em torno de um (1,0), variâncias correspondentes à herdabilidade solicitada e distribuições normais dos erros. Esses resultados podem ser conferidos na Tabela A2 e ao análise de resíduos na Figura A3.

Tabela A2. Caracterização dos valores genotípicos e fenotípicos simulados

	Valor Genotípico	V. Fenotípico $h^2=0,8$
Herdabilidade	1,0	0,803
Média	1,04	1,05
Variância	12,72	15,84
Máximo	11,3	13,64
Mínimo	-8,7	-8,80
Curtose	2,743	2,796
Coef. Simetria	-0,070	-0,013
Lilliefors ¹	0,030 ^{ns}	0,019 ^{ns}

¹ Mostra-se o valor *D* e teste de hipóteses (1%) de que os dados seguem distribuição normal, (**) rejeita *H₀*; (ns): não rejeita *H₀*.

Os resultados da validação demonstram que os procedimentos de simulação utilizados seguem os princípios genéticos, reproduzindo adequadamente os eventos biológicos e portanto confiáveis para sua utilização nas análises do trabalho.

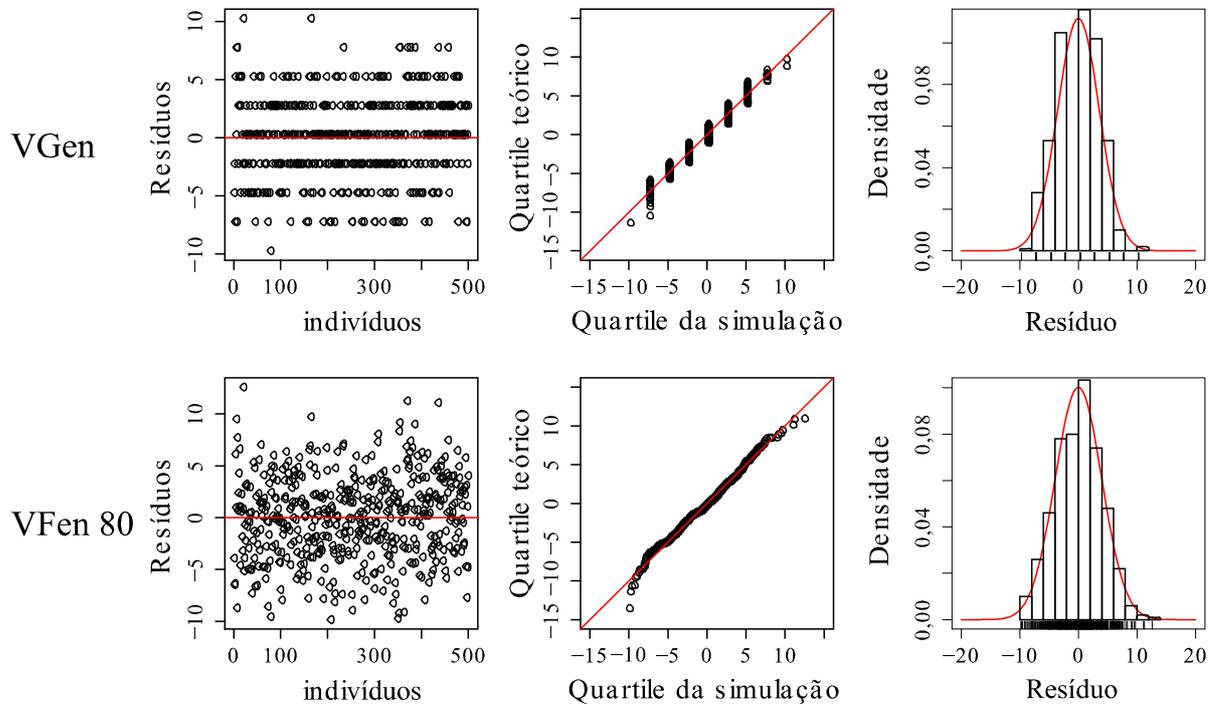


Figura A3. Evidências gráficas da distribuição dos resíduos genotípicos e fenotípicos simulados (Valores dos Resíduos, Normal Q-Q plot e Histogramas)

APÊNDICE B. Rotinas executáveis no R para simulação e seleção de genótipos.

Para um cenário particular, segundo as informações fornecidas, a rotina adapta as informações da população base, simulada no programa GENES, cria o modelo preditivo da GS, classifica e seleciona indivíduos superiores. Posteriormente, simula genótipos, valores genotípicos e fenotípicos dos descendentes de indivíduos selecionados. Ao mesmo tempo calcula e reúne as estimativas dos parâmetros, para posterior análise.

```
#####
#                               ADAPTAÇÃO DA POPULAÇÃO BASE (F2)  SIMULADA NO GENES                               #
#####

# As informações de cada repetição, de cada herdabilidade e cada densidade são salvas em arquivos
# diferentes. Isso permite voltar as informações básica, se é necessário, e em caso de falhas na
# simulação pode se continuar a partir do ponto de falha e no é necessário repetir toda a simulação.
# Logo essas informações são resumidas em um arquivo, em forma de tabelas, para posterior análise
# ("RESULTADOS.RData").

# INFORMAÇÕES INICIAIS #

library(rrBLUP)

setwd("C:\\Simulações") # Definição da trilha de dados                                <-----###Adaptar###

# Leitura do arquivo de marcadores moleculares e fenótipos
# colunas do arquivo: id Vgen  Vfen80  Vfen40  M1      M2      M3      M4      M5...M1040
# codificado 1 0 -1
FG=read.csv(file = "F2_Base.csv", row.names = 1)                                     # <-----###Adaptar###

caract<-3 # Herdabilidades                                                            <-----###Adaptar###

tipo.marcador<- 2 # DOMINATE= 1; CODOMINATE= 2                                       <-----###Adaptar###

genomF2<-data.frame(FG[-1:-(caract+1)])

# Recodifica 1 1 -1 se é marcador Dominante
if (tipo.marcador == 1) {genoF2<-replace (genomF2, genomF2==0, 1)} else {genoF2<-genomF2}
feno<-data.frame(FG[-(caract+2):-ncol(FG)]); FENOTIPOS<-as.matrix(feno[, -1])
rm(FG)

GL<- 8 # Grupos de ligação                                                            <-----###Adaptar###
GLF<- c() # Lista dos GL fora de análise                                             <-----###Adaptar###
#(se todos os GL estão incluídos, não colocar nada)
MGL<- ncol(genomF2)/GL # Marcadores por grupo de lig.
QTL<-c(196, 456, 716, 976) # número do marcador correspondente ao QTL                <-----###Adaptar###

nis<- 20 # número de indiv a selecionar                                              <-----###Adaptar###
n.des<- 10 # descendentes por individuo                                              <-----###Adaptar###

d<-read.table("densidad.dat") # Tabela que relaciona os marcadores que serão
# usados, neste caso com 7 colunas, uma para
# cada densidade analisado
for(R in 1:100) #Repetições                                                           <-----###Adaptar###
{
  for(s in 1:7) #<---##### Indica a saturação de marcadores segundo a lista do objeto "d"
  {D<- (switch(s,0,1,2,3,4,5,10)) #<---##### Indica a distância QTL-Marcador.

  #Prepara os marcadores que serão incluídos na análise
  #(depende da saturação de marcadores -objeto "d"- e GL excluídos)
  #
  if (length(GLF)>0) {MARC<-d[,s]; for (j in GLF) {LI<-((MGL*j)-MGL+1); LS<-(MGL*j); A<-numeric();
  B<- c(0)
  for (i in 1:(nrow(d))) {A[ (MARC[i]<LI || MARC[i]>LS) ] <- c(MARC[i])
  if (A!= B[length(B)]) {B<- c(B,A)}; MARC<- B}
  MARC<- MARC[-1]; rm(LI,LS,A,B,i,j) } else {MARC<- d[,s]}

  GENOTIPO<-as.matrix(genomF2[,MARC])

  #Divisão de matrizes para Validação
  #
  IVal = sample(1:(nrow(GENOTIPO)), (nrow(GENOTIPO)*0.2)) #corresponde ao 20% do total
  GENOTIPOE <- (GENOTIPO[-IVal,]) #Indivíduos Analisados
  GENOTIPOV <- (GENOTIPO[IVal,]) #Indivíduos para Validação

  #Análise por característica
  #

```

```

for(h in 1:caract) # Herdabilidades = 3
{ H<-(switch(h,100,80,40))
G<-2 # Valor para identificar a F2

VGeno<-(FENOTIPOS[,1]); VGeno<- as.data.frame(VGeno) # Acumula os valores genotípicos das gerações
VFeno<-(FENOTIPOS[,h]); VFeno<- as.data.frame(VFeno) # Acumula os valores fenotípicos das gerações
Here<- var(FENOTIPOS[,1])/var(FENOTIPOS[,h])
H2<-as.numeric(); H2<-c(H2,Here) # Acumula os valores de herdabilidade

variable<-(FENOTIPOS[-(IVal),h]) # fenótipo da pob de estimação
variableval<-(FENOTIPOS[IVal,h]) # fenótipo da pob de validação

# Análise das variáveis
#-----

fitRRB=mixed.solve(variable,Z=GENOTIPOE,K=diag(ncol(GENOTIPOE)))

# Estimação VGG
#-----
Ef.Marca<- fitRRB$u # Efeitos de Marcadores
VGG<-GENOTIPOE%*%Ef.Marca # Valores Genômicos

# Calculo de acurácias
#-----
VGS<- FENOTIPOS[-(IVal), 1] # valor Genômico Simulado
acuracia<- (cor(VGG, VGS))^2
confia<- (cor(VGG, variable))^2

# Validação
#-----
VGG_V<- GENOTIPOV%*%Ef.Marca
VGS_V<- FENOTIPOS[IVal, 1] # valor Genômico Simulado de Validação

acuracia_V <- cor(VGG_V,VGS_V)^2
confia_V<- cor(VGG_V,variableval)^2

A<-(rep("na", (nrow(VGG)-nrow(VGG_V)))); B<-as.numeric(c(VGG_V,A))
VGGs<- cbind(VGG, B) # Acumula os VGG estimados por geração
rm(A,B)

Acuracias<- c(acuracia, acuracia_V) # Acumula as acurácias por geração
Confiabil<- c(confia, confia_V) # Acumula as confiabilidades por geração

# Ranking de indivíduos
#-----

#classificação esperada segundo Valores genômicos simulados
cla.sim<-as.matrix(FENOTIPOS[,1]); rownames(cla.sim)<- rownames(FENOTIPOS)
cla.sim.s=sort(as.matrix(cla.sim)[,],decreasing=TRUE)
cla.sim<- as.numeric(names(cla.sim.s)); rm(cla.sim, cla.sim.s)
#classificação segundo Valores genômicos estimados
VGe= GENOTIPO%*%Ef.Marca + mean(FENOTIPOS[,h])
VGel=sort((VGe)[,],decreasing=TRUE)
cla<- as.numeric(names(VGel))

Ranking<- cbind(cla.sim, cla) # Acumula as listas de classificação de ind por valor genômico

# Frequência alélica e Coef. endogamia (F)
#-----
# frequência média segundo 1040 alelos, independente da densidade da variável

##FREQ. POPULACIONAL
freq.p<-
(0.5*length(which(genomF2[,]==0))+length(which(genomF2[,]==1)))/(nrow(genomF2)*ncol(genomF2))
He<-(1-((freq.p)^2 + (1-freq.p)^2))
Ho<-(length(which(genomF2[,]==0)))/(nrow(genomF2)*ncol(genomF2))
coef.endo.F<-(He-Ho)/(He)
Frec.ale<- freq.p; CoefEndo<- coef.endo.F #acumulas os valores por geração

##FREQ. QTL
freq.p<-
(0.5*length(which(genomF2[,QTL]==0))+length(which(genomF2[,QTL]==1)))/(nrow(genomF2)*length(QTL))
Ho<-(length(which(genomF2[,QTL]==0)))/(nrow(genomF2)*length(QTL))
coef.endo.F<-(He-Ho)/(He)
Frec.ale.QTL<- freq.p; CoefEndo.QTL<- coef.endo.F #acumulas los valores por geração

##FREQ. QTL INDIVIDUALES
Fp.QTL.ind<- numeric()
CE.QTL.ind<- numeric()
for (q in 1:4)
{freq.p<-
(0.5*length(which(genomF2[,QTL[q]]==0))+length(which(genomF2[,QTL[q]]==1)))/(nrow(genomF2)*1)
Ho<-(length(which(genomF2[,QTL[q]]==0)))/(nrow(genomF2)*1)
coef.endo.F<-(He-Ho)/(He)
Fp.QTL.ind<- c(Fp.QTL.ind, freq.p) #acumulas os valores por geração
CE.QTL.ind<- c(CE.QTL.ind, coef.endo.F) #acumulas os valores por geração
}

# Medias e Variâncias

```

```

#-----
mediaG<- mean(FENOTIPOS[,1])
varianzaG <- var(FENOTIPOS[,1])
mediaF<- mean(FENOTIPOS[,h])
varianzaF <- var(FENOTIPOS[,h])

#####
#          -> F3 a F6 <-          SIMULAÇÃO, CLASIFICAÇÃO E SELEÇÃO
#
#####

GENO=matrix(, 625, 1040) # ind F2 x Marcad #
GENO[genomF2== 0]<- c(2)
GENO[genomF2== 1]<- c(4)
GENO[genomF2==-1]<- c(1)

for (G in 3:6) # Gerações F3 a F6
{
##### SELEÇÃO #####
#-----

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

##### SIMULAÇÃO GERAÇÕES AVANÇADAS (SGA) #####
#-----

# Geração do genótipo dos descendentes

genoS<-matrix(, (nis*n.des), ncol(GENO))
MGL<- ncol(GENO)/GL # Marcadores por grupo de lig.
cont<- 0

for (ind in 1:nis)
{
  GI<- (GENERACION[ind,]) #genoma de um ind.
  for (des in 1:n.des)
  {res.ind<-NULL
  for (gl in 0:(GL-1))
  {
    fd<- (MGL*gl)
    Fn<- GI[:(1+fd):(MGL+fd)]

    # Diferencia homologos
    cromol=matrix(); cromos2=matrix()
    cromol[Fn==1]<- c(0); cromos2[Fn==1]<- c(0)
    cromol[Fn==2]<- c(0); cromos2[Fn==2]<- c(1)
    cromol[Fn==3]<- c(1); cromos2[Fn==3]<- c(0)
    cromol[Fn==4]<- c(1); cromos2[Fn==4]<- c(1)

    # recombina
    s1<-sample(c(1,2),1); s2<-sample(c(16:115), 1)
    if (s1==1) {col<- c(cromol[1:s2], cromos2[(s2+1): MGL])}
    if (s1==2) {col<- c(cromos2[1:s2], cromol[(s2+1): MGL])}
    if (s1==1 & s2> 100) {col<- c(cromos2[1:(s2-100)], col[(s2-99): MGL])}
    if (s1==2 & s2> 100) {col<- c(cromol[1:(s2-100)], col[(s2-99): MGL])}
    if (s1==1 & s2< 31) {col<- c(col[1: (99+s2)], cromol[(100+s2): MGL])}
    if (s1==2 & s2< 31) {col<- c(col[1: (99+s2)], cromos2[(100+s2): MGL])}

    s1<-sample(c(1,2),1); s2<-sample(c(16:115), 1)
    if (s1==1) {co2<- c(cromol[1:s2], cromos2[(s2+1): MGL])}
    if (s1==2) {co2<- c(cromos2[1:s2], cromol[(s2+1): MGL])}
    if (s1==1 & s2> 100) {co2<- c(cromos2[1:(s2-100)], co2[(s2-99): MGL])}
    if (s1==2 & s2> 100) {co2<- c(cromol[1:(s2-100)], co2[(s2-99): MGL])}
    if (s1==1 & s2< 31) {co2<- c(co2[1: (99+s2)], cromol[(100+s2): MGL])}
    if (s1==2 & s2< 31) {co2<- c(co2[1: (99+s2)], cromos2[(100+s2): MGL])}
    #co<- cbind(col,co2); co # resultado da recombinação

    res<- (c((3^col)+co2)) #codifica como 1 2 3 4
    res.ind<- c(res.ind,res)
  }
  cont<- (cont+1)
  genoS[cont,]<- res.ind
}
}

genoA=matrix(); # codifica como 1 0 -1 (Codominante)
genoA[genoS==1]<- c(-1)
genoA[genoS==2]<- c(0)
genoA[genoS==3]<- c(0)
genoA[genoS==4]<- c(1)
genoA<-matrix(genoA, (nis*n.des),)

if (G==3) {genomF3<-genoA; }# Reúne os genomas simulados
if (G==4) {genomF4<-genoA; }

```

```

if (G==5) {genomF5<-genoA; }
if (G==6) {genomF6<-genoA; }

# Frequência alélica e Coef. endogamia (F)
# _____

# frequência média segundo 1040 alelos independente da densidade da variável ###

##FREQ. POPULACIONAL
freq.p<-(0.5*length(which(genoA[,]==0))+length(which(genoA[,]==1)))/(nrow(genoA)*ncol(genoA))
Ho<-(length(which(genoA[,]==0)))/(nrow(genoA)*ncol(genoA))
coef.endo.F<-(He-Ho)/(He)
Frec.ale<- c(Frec.ale,freq.p); CoefEndo<- c(CoefEndo, coef.endo.F) # acumula os valores por
geração

##FREQ. QTL
freq.p<-
(0.5*length(which(genoA[,QTL]==0))+length(which(genoA[,QTL]==1)))/(nrow(genoA)*length(QTL))
Ho<-(length(which(genoA[,QTL]==0)))/(nrow(genoA)*length(QTL))
coef.endo.F<-(He-Ho)/(He)
Frec.ale.QTL<- c(Frec.ale.QTL,freq.p); CoefEndo.QTL<- c(CoefEndo.QTL,coef.endo.F)
# acumula valores por geração

##FREQ. QTL INDIVIDUALES
Fp.QTL.Fn<- numeric()
CE.QTL.Fn<- numeric()
for (q in 1:4)
{freq.p<-(0.5*length(which(genoA[,QTL[q]]==0))+length(which(genoA[,QTL[q]]==1)))/(nrow(genoA)*1)
Ho<-(length(which(genoA[,QTL[q]]==0)))/(nrow(genoA)*1)
coef.endo.F<-(He-Ho)/(He)

Fp.QTL.Fn<- c(Fp.QTL.Fn, freq.p)
CE.QTL.Fn<- c(CE.QTL.Fn, coef.endo.F)
}

Fp.QTL.ind<- rbind (Fp.QTL.ind, Fp.QTL.Fn) # acumula os valores por geração
CE.QTL.ind<- rbind (CE.QTL.ind, CE.QTL.Fn) # acumula os valores por geração

# Simulação Valores Genotípicos
# _____

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

# Simulação Valores Fenotípicos
# _____

if (H== 100) {
  if (var(VG[,1])>0) {VF<-VG[,1]; corrf<- (cor(VG[,1], VF)); Here<- (var(VG[,1])/ var(VF))
  } else {VF<-VG[,1]; corrf<- "NA"; Here<- "NA"}}

if (H== 80) {
  if (var(VG[,1])>0) { sdv<- (sqrt((var(VG[,1])/0.80)-var(VG[,1])))
  repeat
  { VF<-(VG[,1] + rnorm(nrow(VG), mean= 0, sd= sdv))
  sdv<-sdv+0.001
  corrf<- cor(VG[,1], VF)
  Here<- (var(VG[,1])/ var(VF))
  if (Here > 0.794 & Here < 0.806) break
  } #corrf; Here; sdv
  } else {VF<-VG[,1]; corrf<- "NA"; Here<- "NA"}}

if (H== 40) {
  if (var(VG[,1])>0) { sdv<- (sqrt((var(VG[,1])/0.40)-var(VG[,1])))
  repeat
  { VF<-(VG[,1] + rnorm(nrow(VG), mean= 0, sd= sdv))
  sdv<-sdv+0.001
  corrf<- cor(VG[,1], VF)
  Here<- (var(VG[,1])/ var(VF))
  if (Here > 0.394 & Here < 0.406) break
  }
  } else {VF<-VG[,1]; corrf<- "NA"; Here<- "NA"}}

VF<- as.matrix(VF)

A<-rep("na", (nrow(VGGeno)-nrow(VG))); B<-as.numeric(c(VG,A)); C<-as.numeric(c(VF,A))
VGeno<- cbind(VGGeno,B) # Acumula os valores genotípicos das gerações
VFeno<- cbind(VFeno,C) # Acumula os valores fenotípicos das gerações
rm(A,B,C)
H2<-c(H2,Here) # Acumula os valores de herdabilidade

# Estimacão VGG
# _____

if (tipo.marcador == 1)
{genoA=numeric(); # Recodifica como 1 1 -1 (Dominante)# É ativado se foi indicado
genoA[genoS==1]<- c(-1) # que ia-se trabalhar com
genoA[genoS==2]<- c(1) # marcador Dominante
genoA[genoS==3]<- c(1) #
genoA[genoS==4]<- c(1) #
genoA<-matrix(genoA, (nis*n.des),) #

```

```

VGG<- (genoA[,MARC])%*%Ef.Marca #MARC: Filtra os GL a avaliar

#VGGs<- cbind(VGGs, VGG) # Acumula VGG estimados em cada geração
VGGs<- cbind(VGGs, rbind(VGG, matrix(rep(NA,(nrow(VGGs)-nrow(VGG))), ,1)))

# Calculo de acurácias
#-----
if (var(VG[,1])>0 && var(VGG)>0) {acuracia<- (cor(VG[,1], VGG))^2} else {acuracia<- "NA"}
if (var(VF[,1])>0 && var(VGG)>0) {confia<- (cor(VF[,1], VGG))^2} else {confia<- "NA"}

Acuracias<- c(Acuracias, acuracia) # Acumula as acurácias por geração
Confiabil<- c(Confiabil, confia) # Acumula as confiabilidades por geração

# Medias e Variâncias
#-----
mediaG<- c(mediaG, mean(VG[,1]))
varianzaG<- c(varianzaG, var(VG[,1]))
mediaF<- c(mediaF, mean(VF[,1]))
varianzaF<- c(varianzaF, var(VF[,1]))

# Ranking de indivíduos
#-----
#classificação esperada segundo Valores genômicos simulados
cla.sim_<-VG; rownames(cla.sim_)<- (seq(1,nrow(VG),1))
cla.sim.s=sort(as.matrix(cla.sim_),,decreasing=TRUE)
cla.sim<- as.numeric(rownames(cla.sim.s)); rm(cla.sim_, cla.sim.s)
#classificação segundo Valores genômicos estimados
VGe= VGG + mean(FENOTIPOS[,h]) #(beta=média do fenótipo pob estimacão(F2))
VGe<- as.matrix(VGe); rownames(VGe)<- seq(1, nrow(VGe),1)
VGe1=sort(as.matrix(VGe)[,],decreasing=TRUE); VGe1<- as.matrix(VGe1)
cla<- as.numeric(rownames(VGe1)); rm(VGe, VGe1)

A<- (rep("na", (nrow(Ranking)-length(cla)))); B<-as.numeric(c(cla.sim,A)); C<-as.numeric(c(cla,A))
Ranking<- cbind(Ranking,B,C) # Acumula as listas de classificação de ind por valor genômico
rm(A,B,C)

# Preparação arquivos para a seguinte geração
#-----
GENO<- genoS
}

# Salvando resultados
#-----
colnames(Ranking)<- c("F2s", "F2e", "F3s", "F3e", "F4s", "F4e", "F5s", "F5e", "F6s", "F6e")

colnames(VGeno)<- c("F2", "F3", "F4", "F5", "F6"); colnames(VFeno)<- c("F2", "F3", "F4", "F5", "F6")
colnames(VGGs)<- c("F2", "F2V", "F3", "F4", "F5", "F6")

resultados<- paste("D",D,"-H",H,"-R",R,".RData", sep = "")
save(genomF2, genomF3, genomF4, genomF5, genomF6, fitRRB, VGeno, VFeno, mediaG, varianzaG, mediaF,
varianzaF, Ranking, VGGs, Acuracias, Confiabil, Frec.ale, Frec.ale.QTL, Fp.QTL.ind,
CoefEndo, CoefEndo.QTL, CE.QTL.ind, He, H2, file = resultados)
}
}
}

#####
### Rotinas para reunir as informações principais, em tabelas, para posterior análise ###
#####

# Agrupando Resultados
#-----

# Objetos que receberão os parâmetros e resultados principais
x.He<- data.frame()
x.Frec.ale<- data.frame()
x.Frec.ale.QTL<- data.frame()
x.Fp.QTL.ind<- data.frame()
x.CoefEndo<- data.frame()
x.CoefEndo.QTL<- data.frame()
x.CE.QTL.ind<- data.frame()
x.H2<- data.frame()
x.Acuracias<- data.frame()
x.Confiabil<- data.frame()
x.mediaG<- data.frame()
x.mediaF<- data.frame()
x.varianzaG<- data.frame()
x.varianzaF<- data.frame()

for (r in 1:100) {for (d in c(0:5,10)) {for (h in c(100, 80, 40))
{archivo<- paste("C:/Simulações/D", d,"-H", h, "-R", r, ".RData", sep = "")
load(archivo)

Acuracias<-as.numeric(Acuracias)
Confiabil<-as.numeric(Confiabil)

```

```

H2<-as.numeric(H2)

x.He<- rbind(x.He, He)

x.Frec.ale<- rbind(x.Frec.ale, Frec.ale)
x.Frec.ale.QTL<- rbind(x.Frec.ale.QTL, Frec.ale.QTL)
x.CofEndo<- rbind(x.CofEndo, CofEndo)
x.CofEndo.QTL<- rbind(x.CofEndo.QTL, CofEndo.QTL)
x.H2<- rbind(x.H2, H2)
x.mediaG<- rbind(x.mediaG, mediaG)
x.mediaF<- rbind(x.mediaF, mediaF)
x.varianzaG<- rbind(x.varianzaG, varianzaG)
x.varianzaF<- rbind(x.varianzaF, varianzaF)

x.Acuracias<- rbind(x.Acuracias, Acuracias)
x.Confiabil<- rbind(x.Confiabil, Confiabil)

x.CE.QTL.ind<- rbind(x.CE.QTL.ind, CE.QTL.ind)
x.Fp.QTL.ind<- rbind(x.Fp.QTL.ind, Fp.QTL.ind)
}}

save(x.He, x.Frec.ale, x.Frec.ale.QTL, x.CofEndo, x.CofEndo.QTL, x.H2, x.mediaG, x.mediaF,
x.varianzaG,
x.varianzaF, x.Acuracias, x.Confiabil, x.CE.QTL.ind, x.Fp.QTL.ind, file = "RESULTADOS.RData")

# Agrega nveis e nomes
#
# ***** Deve ser ajustado caso as variveis cambiem (Distancia, Herdabilidades, repeties, ...)*****

niveles<- matrix(c(rep(rep(c(seq(0,5,1),10), rep(3,7)),1), rep(c(100,80,40),7)), ,2)
niveles2<- matrix(c(rep(c(seq(0,5,1),10), rep(15,7)), rep(rep(c(100,80,40),rep(5,3)),7),
rep(seq(2,6,1),21)), ,3)

geracoes1<- c("D","H","F2","F3","F4","F5","F6")
geracoes2<- c("D","H","F2","F2V","F3","F4","F5","F6")
nqtl<- c("D","H","G","QTL1","QTL2","QTL3","QTL4")

x.He<- cbind(niveles, x.He)

x.Frec.ale<- cbind(niveles, x.Frec.ale); colnames(x.Frec.ale)<- (geracoes1)
x.Frec.ale.QTL<- cbind(niveles, x.Frec.ale.QTL); colnames(x.Frec.ale.QTL)<- (geracoes1)
x.CofEndo<- cbind(niveles, x.CofEndo); colnames(x.CofEndo)<- (geracoes1)
x.CofEndo.QTL<- cbind(niveles, x.CofEndo.QTL); colnames(x.CofEndo.QTL)<- (geracoes1)
x.H2<- cbind(niveles, x.H2); colnames(x.H2)<- (geracoes1)
x.mediaG<- cbind(niveles, x.mediaG); colnames(x.mediaG)<- (geracoes1)
x.mediaF<- cbind(niveles, x.mediaF); colnames(x.mediaF)<- (geracoes1)
x.varianzaG <- cbind(niveles, x.varianzaG ); colnames(x.varianzaG )<- (geracoes1)
x.varianzaF <- cbind(niveles, x.varianzaF ); colnames(x.varianzaF )<- (geracoes1)

x.Acuracias <- cbind(niveles, x.Acuracias ); colnames(x.Acuracias )<- (geracoes2)
x.Confiabil <- cbind(niveles, x.Confiabil ); colnames(x.Confiabil )<- (geracoes2)

x.CE.QTL.ind<- cbind(niveles2, x.CE.QTL.ind)
colnames(x.CE.QTL.ind)<- (nqtl); row.names(x.CE.QTL.ind)<-NULL
x.Fp.QTL.ind<- cbind(niveles2, x.Fp.QTL.ind)
colnames(x.Fp.QTL.ind)<- (nqtl); row.names(x.Fp.QTL.ind)<-NULL

save(x.He, x.Frec.ale, x.Frec.ale.QTL, x.CofEndo, x.CofEndo.QTL, x.H2, x.mediaG, x.mediaF,
x.varianzaG,
x.varianzaF, x.Acuracias, x.Confiabil, x.CE.QTL.ind, x.Fp.QTL.ind, file = "RESULTADOS.RData")

# Ordena segundo os nveis
#
x.He<-x.He[order(x.He[,2], decreasing=T),]

x.Frec.ale<-x.Frec.ale[order(x.Frec.ale[,2], decreasing=T),]
x.Frec.ale.QTL<-x.Frec.ale.QTL[order(x.Frec.ale.QTL[,2], decreasing=T),]
x.CofEndo<-x.CofEndo[order(x.CofEndo[,2], decreasing=T),]
x.CofEndo.QTL<-x.CofEndo.QTL[order(x.CofEndo.QTL[,2], decreasing=T),]
x.H2<-x.H2[order(x.H2[,2], decreasing=T),]
x.mediaG<-x.mediaG[order(x.mediaG[,2], decreasing=T),]
x.mediaF<-x.mediaF[order(x.mediaF[,2], decreasing=T),]
x.varianzaG<-x.varianzaG[order(x.varianzaG[,2], decreasing=T),]
x.varianzaF<-x.varianzaF[order(x.varianzaF[,2], decreasing=T),]
x.Acuracias<-x.Acuracias[order(x.Acuracias[,2], decreasing=T),]
x.Confiabil<-x.Confiabil[order(x.Confiabil[,2], decreasing=T),]

x.CE.QTL.ind<-x.CE.QTL.ind[order(x.CE.QTL.ind[,3], decreasing = F),]
x.Fp.QTL.ind<-x.Fp.QTL.ind[order(x.Fp.QTL.ind[,3], decreasing = F),]

save(x.He, x.Frec.ale, x.Frec.ale.QTL, x.CofEndo, x.CofEndo.QTL, x.H2, x.mediaG, x.mediaF,
x.varianzaG,
x.varianzaF, x.Acuracias, x.Confiabil, x.CE.QTL.ind, x.Fp.QTL.ind, file = "RESULTADOS.RData")

```