

**EMPREGO DA SIMULAÇÃO NA ESCOLHA
DO NÚMERO DE INDIVÍDUOS NO MÉTODO
“*BULK*” NO MELHORAMENTO DE
PLANTAS AUTÓGAMAS**

JOSÉ WILSON DA SILVA

2009

JOSÉ WILSON DA SILVA

**EMPREGO DA SIMULAÇÃO NA ESCOLHA DO NÚMERO DE
INDIVÍDUOS NO METODO “BULK” NO MELHORAMENTO DE
PLANTAS AUTÓGAMAS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de “Doutor”

Orientador

Prof. Dr. Daniel Furtado Ferreira

LAVRAS
MINAS GERAIS - BRASIL
2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Silva, José Wilson da.

Emprego da simulação na escolha do número de indivíduos
no método “bulk” no melhoramento de plantas autógamas / José
Wilson da Silva. – Lavras : UFLA, 2009.

59 p. : il.

Tese (Doutorado) – Universidade Federal de Lavras, 2009.

Orientador: Daniel Furtado Ferreira.

Bibliografia.

1. Método Bulk. 2. Simulação computacional. 3. Feijoeiro. I.
Universidade Federal de Lavras. II. Título.

CDD – 635.6523

JOSÉ WILSON DA SILVA

**EMPREGO DA SIMULAÇÃO NA ESCOLHA DO NÚMERO DE
INDIVÍDUOS NO METODO “BULK” NO MELHORAMENTO DE
PLANTAS AUTÓGAMAS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de “Doutor”

APROVADA em 22 de junho de 2009

Prof. Dr. Magno Antônio Patto Ramalho	DBI
Prof. Dr. Ruben Delly Veiga	DEX
Dra. Ângela de Fátima Barbosa Abreu	EMBRAPA Arroz Feijão
Prof. Dr. Antônio Alves Soares	DAG

Prof. Dr. Daniel Furtado Ferreira
UFLA
(Orientador)

LAVRAS
MINAS GERAIS – BRASIL

Dedico o presente trabalho a toda minha família que me ensinou a superar as dificuldades com luta e perseverança e, em especial, a meus filhos, Wilson e William, que souberam superar minha ausência com uma dose multiplicada de amor a cada ida e volta e a Gláudia, a mulher da minha vida, por cultivar sempre em primeiro plano o nosso infinito amor.

O pensamento científico não deseja persuadir, mas inspirar; inspirar um outro pensamento, por em movimento o pensar (José Wilson da Silva, 2009)

AGRADECIMENTOS

A Deus, por me dar a vida e permitir mais uma benção em minha vida.

À Capes, pela concessão de bolsas de estudos.

À UFAL, pelo título de Engenheiro Agrônomo e Mestre em Agronomia.

Ao professor Daniel Furtado Ferreira, amigo, pela brilhante orientação e por muitos ensinamentos que levarei por toda a minha vida.

Aos professores do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, pelos conhecimentos transmitidos durante o curso.

Aos colegas do Programa de Genética e Melhoramento de Plantas, pela convivência.

Ao Dr. Paulo Vanderlei Ferreira, pelo incentivo na escolha do curso.

A meus pais e todos os meus irmãos, pelo amor, compreensão e apoio durante a realização deste curso.

E a todos que contribuíram para a elaboração deste trabalho, desejo-lhes felicidades e muito obrigado.

SUMÁRIO

	Página
RESUMO	i
ABSTRACT	ii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Melhoramento genético de espécies autógamas.....	3
2.2 Método bulk.....	5
2.3 Amostragem no método <i>bulk</i>	8
2.4 Simulação no melhoramento de plantas	9
3 MATERIAL E MÉTODOS	18
3.1 Simulação computacional	18
3.2 Experimentos simulados e parâmetros avaliados.....	23
4 RESULTADOS E DISCUSSÃO	25
4.1 Efeito do tamanho do <i>bulk</i>	25
4.2 Efeito da amostragem na seleção natural.....	38
4.3 Considerações gerais	41
5 CONCLUSÕES	44
REFERÊNCIAS BIBLIOGRÁFICAS	45
ANEXOS.....	49

RESUMO

SILVA, José Wilson da. **Emprego da simulação na escolha do número de indivíduos no método “bulk” no melhoramento de plantas autógamas**. 2009. 59 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras.¹

O método *bulk* de condução de populações segregantes apresenta como principal limitação perdas de progênies durante a amostragem. Estudos para entender ou minimizar o seu efeito são escassos. O objetivo deste estudo foi verificar o efeito da amostragem e da seleção natural no método *bulk* em plantas autógamas. Para isso, foram simuladas várias configurações considerando as gerações F_4 e F_{24} e as famílias $F_{4:5}$, $F_{4:6}$ e $F_{24:25}$. Os efeitos no número médio, mínimo e máximo de locos favoráveis, da variância genética e do número de alelos favoráveis foram mensurados ao longo das 1000 simulações de cada configuração. As simulações contemplaram diferentes herdabilidades, interações alélicas e números de locos. O efeito da amostragem foi avaliado com diferentes tamanhos de *bulk* os quais foram 50, 200, 1000, 2000 e 5000 indivíduos. O efeito da seleção natural foi incorporado na simulação por meio do valor adaptativo dos indivíduos. Duas situações foram contempladas uma em que o valor adaptativo mínimo variava de 20% a 100% o número de indivíduos e outra em que o número de indivíduos era igual. Conclui-se que na determinação do número de indivíduos no método *bulk* deve-se considerar a herdabilidade do caráter. O *bulk* de 2000 indivíduos nas condições simuladas mostrou-se igualmente eficiente ao de 5000 indivíduos. A seleção natural só resulta em ganho com *bulks* superiores a 1000 indivíduos.

¹ Orientador: Daniel Furtado Ferreira – UFLA

ABSTRACT

SILVA, José Wilson da. **Use of simulation in choosin the number of individuals in the “bulk” method of breeding self-pollinated.** 2009. 59 p. Thesis. (Doctorate in Genetics and Plant Breeding) – Federal University of Lavras, Lavras.²

The bulk method to conduct segregating populations presents as the main limitation the loss of progenies during sampling. Studies to undertand and minimize this limitation are scarce. This study aimed to verify the effect of sampling and natural selection on the bulk method in self-pollinated plants. Many configurations were simulated using the F_4 and F_{24} generations and families $F_{4:5}$, $F_{4:6}$ e $F_{24:25}$. The effects on the average, minimum and maximum numbers favorable of loci, the genetic variance and the number of favorables alleles were measurad through 1000 simulations for each configuration. The simulations considered different heritabilities, allelic interactions and number of loci. The sampling effect was evaluated with different bulk size of 50, 200, 1000, 2000 e 5000 individuals. The effect of natural selection was incorporated in to the simulation using individuals' adaptative value. Two situations were considered: first- the minimam adaptative value ranged from 20% to 100% of the number of individuals and second- the number of individuals was equal. It was concluded that to determine the number of individuals for the bulk method the heritability should be considered. The *bulk* size 2000 of individuals in the simulated conditions were equally efficient as 5000 individuals. Natural selection results in genetic gain only for *bulks* sizes over 1000 individuals.

² Major professor: Daniel Furtado Ferreira – UFLA

1 INTRODUÇÃO

A maioria dos caracteres de importância econômica, como a produção de grãos, é controlada por muitos genes, cujos alelos favoráveis estão em linhagens distintas. Nesse sentido, para combinar os alelos favoráveis que estão em duas ou mais linhagens em uma única, faz-se necessária a utilização do método de hibridação. No emprego deste método, há algumas decisões que devem ser tomadas, tais como quais genitores a serem utilizados na hibridação, como proceder aos cruzamentos e que método utilizar para conduzir as populações segregantes, sabendo que desta escolha depende a eficiência do programa.

Dentre as metodologias disponíveis para o avanço de populações segregantes está o método *bulk*. É um dos métodos mais utilizados no avanço de populações de plantas autógamas e se destaca pela facilidade de condução. Entre as vantagens atribuídas a este método, destaca-se a seleção natural que mantém os indivíduos mais adaptados, que de modo geral são os mais produtivos, como também permite adiar a abertura do *bulk*, tornando seu uso bastante flexível. Alguns trabalhos abordando esses aspectos foram realizados (Corte et al., 2002; Gonçalves et al., 2001; Pirola et al., 2002; Silva et al., 2004).

Contudo, na condução de populações segregantes pelo método *bulk*, o melhorista se depara com o questionamento de quantos indivíduos serão necessários em cada geração de avanço. Este aspecto deve ser realçado, pois a impossibilidade de avançar com todas as sementes colhidas a cada geração constitui a principal restrição do método *bulk*, provocando perdas substanciais de progênes promissoras de uma geração para outra.

As consequências da amostragem foram estimadas por Kervella & Fouilloux (1992), que observaram grandes perdas, especialmente nas gerações

mais avançadas e quando há grande variação no número de descendentes produzidos por planta. Apesar de informativos, os resultados simulados obtidos não refletem uma situação real; as limitações apontadas estão ligadas à escolha de modelos que não incluem aspectos como a seleção natural, a herdabilidade, a variância e os variados tamanhos de amostra do *bulk*. Não há muitos relatos sobre número de indivíduos em cada geração; os existentes pregam um número de 1.000 a 2.000 indivíduos por amostra.

Uma alternativa que pode corrigir estas limitações, com relação ao efeito da amostragem no *bulk*, é o emprego da simulação computacional. Por meio desta, pode-se imitar uma situação real do melhoramento, utilizando-se modelos genéticos diferentes, variando do mais simples, que contém somente os efeitos genéticos aditivos e dominantes, aos modelos complexos que contêm alelos múltiplos, epistasia e pleiotropia (Wang & Wolfgang, 2007).

Dessa forma, foi realizado o presente trabalho, com o objetivo de determinar o efeito da amostragem e da seleção natural na condução de populações segregantes no método *bulk*, em plantas autógamas.

2 REFERENCIAL TEÓRICO

2.1 Melhoramento genético de espécies autóгамas

O feijoeiro é uma planta autóгama e, nela, a ocorrência da cleistogamia torna a autofecundação seu sistema quase obrigatório de reprodução. Assim, os métodos de melhoramento aplicáveis ao feijoeiro são aqueles comuns às plantas autóгamas, como a introdução de linhagens, a seleção de linhas puras e a hibridação.

Neste contexto, o método de hibridação será abordado com mais detalhes por interesse deste trabalho. Para esclarecimentos sobre os outros métodos, consultar Borém & Miranda (2005), Fehr (1987) e Ramalho et al. (2001).

O método de hibridação tem a finalidade de reunir em uma única linhagem os alelos desejáveis que se encontram em linhagens distintas (Fehr, 1987; Ramalho, 1997). Portanto, na condução de um programa de melhoramento por hibridação, há algumas decisões a serem tomadas, tais como: a escolha dos genitores a serem cruzados (Baenziger & Peterson, 1991; Abreu et al., 2002), como combinar os genitores (Carneiro et al., 2002) e a escolha do método de condução da população segregante (Fehr, 1987; Ramalho et al., 2001).

A população segregante, a partir da qual será realizada a extração de linhagens, é obtida normalmente, mediante cruzamento entre genitores superiores (Rasmusson & Phillips, 1997). Deste cruzamento, obtém-se uma geração F_1 , homogênea, de indivíduos altamente heterozigóticos. Os indivíduos F_1 são em seguida autofecundados para a liberação da variabilidade, obtendo-se a geração F_2 , a qual é heterogênea, possuindo indivíduos homozigóticos e heterozigóticos. A população F_2 é considerada a população base em equilíbrio e

a endogamia só será levada em conta a partir desta geração (Ramalho et al., 1993; Wricke & Weber, 1986).

A população base formada tem grande potencial de melhoramento, pois se caracteriza por apresentar média e variabilidade elevadas, aumentando a probabilidade de selecionar linhagens de performance superior (Fouilloux & Bannerot, 1988).

Depois de obtida a população base, o próximo passo consiste em realizar o avanço da endogamia até um nível de homozigosidade desejado (Fehr, 1987). Existem vários métodos de avanço ou condução da população segregante destinados às espécies autógamas (Borém & Miranda, 2005; Fehr, 1987; Jensen, 1988; Ramalho et al., 2001). Esses métodos são classificados em duas categorias: a primeira, em que estão os métodos que não separam as fases de endogamia e seleção, na qual estão a seleção massal e o método genealógico, e a segunda, que abrange os métodos que separam as fases de endogamia e seleção, entre os quais estão os métodos *bulk* ou populacional e descendente de única semente ou SSD (Fouilloux & Bannerot, 1988). Constata-se, ainda, a existência de vários outros métodos que se constituem de modificações e combinações destes, tendo surgido em função das desvantagens ou limitações apresentadas pelos métodos originais.

Diante dos relatos, torna-se necessária a escolha do método de condução de população segregante durante as gerações de endogamia, de modo a manejá-la de forma adequada, melhorando a eficiência do processo seletivo. Ressalta-se que, para a tomada da decisão sobre que método de condução utilizar, o melhorista deve observar, cuidadosamente, qual é o mais adequado para os seus objetivos, além de avaliar outros fatores que interferem diretamente, como herdabilidade dos caracteres sob seleção, adequação dos ambientes para a seleção e, principalmente, a quantidade de recursos e a infraestrutura disponível (Fehr, 1987; Fouilloux & Bannerot, 1988). Como exemplo, Wang et al. (2003)

destacaram que o sucesso do programa de melhoramento do trigo desenvolvido pelo CIMMYT deve-se, dentre outros fatores, às estratégias de melhoramento adotadas por seus melhoristas.

Em razão do objetivo do trabalho, somente o método *bulk* será abordado mais detalhadamente. Para mais detalhes sobre os métodos de avanço da endogamia, consultar as referências citadas.

2.2 Método bulk

O método *bulk* foi desenvolvido, inicialmente, pelo melhorista sueco Nilsson-Ehle, em 1908, com a cultura do trigo, na Suécia, realizando trabalhos na *Swedish Seed Association* (SSA) (Fehr, 1987; Jensen, 1988).

A descrição e a forma de condução da população pelo método *bulk* são simples. Inicia-se a partir da geração F_2 , em que as plantas são colhidas massalmente e, então, uma amostra de sementes é retirada para a obtenção da população F_3 . Esse procedimento é repetido até a geração que apresente um nível adequado de homozigose, na qual será procedida a seleção de plantas individuais. A partir deste ponto, as melhores linhagens são identificadas e participarão de ensaios de competição de cultivares (Ramalho et al., 1993; Borém & Miranda, 2005).

É oportuno salientar três aspectos inerentes ao emprego do método *bulk* que devem ser esclarecidos: a ação da seleção natural durante as gerações de autofecundação, a determinação da época adequada para iniciar a seleção e a amostragem.

No método *bulk*, é geralmente aceita a premissa da ação da seleção natural durante as gerações de endogamia. Estudos sobre o método *bulk*, durante as gerações de endogamia, demonstraram que a seleção natural exerce influência, promovendo, dessa forma, o aumento da frequência de genótipos

agronomicamente desejáveis. No entanto, questiona-se se, realmente, a seleção ocorrerá no sentido desejado pelo melhorista (Ramalho et al., 2001).

O efeito da seleção natural foi estudado por 50 gerações em uma população de cevada. O referido estudo tomou como referência uma população de cevada sintetizada em 1928, constituída por sementes F_1 de todos os possíveis cruzamentos aos pares de 28 cultivares de cevada Allard (1988). Houve mudanças para vários caracteres considerando o efeito da seleção natural. Para o rendimento de grãos, houve um aumento de, aproximadamente, 1% a 2% ao ano, o que representa ganho comparável ao da seleção artificial (Ramalho et al., 2001). Segundo os autores, as alterações no rendimento de grãos de cevada resultaram do aumento no número de sementes e não propriamente de um aumento do tamanho da semente. Em outro trabalho, realizado por Hamblin (1977), envolvendo quatro populações segregantes avaliadas nas gerações F_2 e F_6 , ocorreu aumento expressivo na produtividade de grãos apenas no caso das populações segregantes com menor média. Deve-se ressaltar que isso ocorreu porque, nas populações menos produtivas, ocorrem indivíduos menos adaptados e a seleção natural pode atuar com maior intensidade.

Alguns trabalhos experimentais foram realizados com feijoeiro, conduzidos pelo método *bulk*. Constatou-se, pela avaliação de populações durante 15 gerações, que a ação da seleção natural contribui para um incremento médio do rendimento de grãos de 2,5% por geração (Corte et al., 2002). Os autores ressaltaram, ainda, que, se o objetivo for selecionar linhagens com determinados fenótipos desejáveis, tais como hábito de crescimento determinado, sementes grandes, semente padrão carioca e florescimento precoce, a seleção artificial deve ser praticada em F_3 . Esta prática tem por finalidade eliminar plantas com características indesejáveis e direcionar o *bulk* para o objetivo pretendido pelo melhorista (Fehr, 1987; Jensen, 1988).

Em outro estudo com populações de feijão avançadas pelo método *bulk* da geração F₂ a F₁₃, Gonçalves et al. (2001) observaram que os indivíduos com hábito de crescimento indeterminado e com menores sementes foram mantidos pela seleção natural.

Os resultados anteriormente citados evidenciam que a seleção natural atua no sentido de eliminar os indivíduos menos adaptados, que são aqueles com menor capacidade de sobreviver e reproduzir num dado conjunto de ambientes Allard (1988).

Outro aspecto que deve ser levado em consideração é a escolha dos ambientes, onde as gerações pelo método *bulk* serão avançadas. Estes ambientes deverão refletir as condições de cultivo e possibilitar a expressão das diferenças genéticas entre os indivíduos para os caracteres de interesse, de forma que a seleção natural atue no sentido desejado. Este aspecto foi estudado por Pirola et al. (2002), utilizando uma população de feijão avançada pelo método *bulk* em diferentes ambientes. Os autores mostraram que a seleção natural atuou mantendo indivíduos mais adaptados para aqueles ambientes onde as populações foram avançadas.

Na utilização do método *bulk*, a seleção é postergada até que um nível de homozigose seja atingido. Assim surge uma questão relevante: em que momento deverá ser aberto o *bulk* (Ramalho et al., 1993)? Com base em alterações ocorridas nas frequências genotípicas e por acréscimos cada vez menores nas frações da variância aditiva explorada ao longo das gerações de endogamia, não restam justificativas teóricas para adiar a seleção além das gerações F₄ e F₅. Contudo, estudos conduzidos por Silva et al. (2004) revelaram, por meio de avaliação experimental de progênies de feijão extraídas nas gerações F₂, F₈ e F₂₄, conduzidas em *bulk*, que a frequência de progênies superiores aumentou quando a extração de progênie foi adiada para gerações mais avançadas. Este resultado é um indicativo importante de que o método *bulk* é flexível em permitir o atraso

da extração das linhagens sem proporcionar perda para o melhorista. Vale salientar que o método *bulk* apresenta restrição, a qual será tratada no tópico seguinte.

2.3 Amostragem no método *bulk*

Um aspecto interessante, e que se enquadra como foco principal deste trabalho, refere-se ao número de indivíduos que deve ser utilizado a cada geração para que se tenha probabilidade de sucesso com a seleção. Este aspecto deve ser cuidadosamente observado, pois se enquadra como a principal desvantagem atribuída ao método *bulk*.

No caso do método *bulk*, a impossibilidade de avanço com todas as sementes colhidas a cada geração torna necessária a aplicação da amostragem, o que ocasiona a perda de alelos favoráveis, isto é, a perda de combinações genotípicas devido à amostragem deficiente durante o avanço subsequentes das gerações. Assim, Kervella & Fouilloux (1992) realizaram um trabalho de simulação conduzido para avaliar o efeito da amostragem no método *bulk*, na qual consideraram diferentes tamanhos de populações e percentagem de perdas por amostragem. Envolveram também diferentes taxas de multiplicação e verificaram que a taxa de extinção esperada de progênies, após uma geração, seja superior a 20%.

Dessa maneira, a fim de minimizar problemas com a amostragem, os autores sugeriram que, com o aumento da população ao longo das gerações, haveria uma redução na taxa de extinção das progênies. Segundo Ramalho et al. (2001), esta alternativa é exequível para o método *bulk*, em virtude da facilidade de condução. Porém, comparações neste contexto não foram realizadas para testar essas afirmações, por conta das conduções experimentais que são frequentemente onerosas e a experimentação convencional apresenta restrições devido à insuficiente amostragem.

De forma geral, são muitos os relatos sobre o número de indivíduos utilizados no avanço das gerações pelo método *bulk* nos programas de melhoramento do feijoeiro, recomendando o emprego de 1.000 indivíduos (Ramalho et al., 2001). No caso do feijoeiro, alguns trabalhos apresentaram diferentes números de indivíduos, como relatado a seguir: Corte et al. (2002) utilizaram 1.600 indivíduos para estudar o efeito da seleção natural em populações avançadas pelo método de *bulk* por dezessete gerações; Raposo et al. (2000) utilizaram uma amostra de 2.000 indivíduos para estudo de comparação de métodos de condução de populações segregantes e Gonçalves et al. (2001) utilizaram 1.280 indivíduos para avanço das gerações F₂ a F₁₃, em *bulk* visando estudar o efeito da seleção natural. Estes números são bastante inferiores ao utilizado em cevada, cerca de 15.000 (Allard, 1988; Solliman & Allard, 1991). Vale salientar que um número maior de indivíduos poderá inviabilizar o avanço das gerações em experimentos com repetição. Portanto, é de fundamental importância o uso de estratégias que possibilitem avaliar um número maior de indivíduos e mensurar o efeito da amostragem no *bulk* e a simulação computacional mostra-se adequada neste caso.

2.4 Simulação no melhoramento de plantas

No melhoramento de plantas, dispõe-se de várias estratégias que podem ser adotadas pelos melhoristas em seus programas. Para o caso de espécies autógamas, as diferentes estratégias de melhoramento distinguem-se, e principalmente, pela possibilidade de utilização de vários métodos de condução de populações segregantes. Dessa forma, tem sido grande o interesse em comparações dessas estratégias, em termos de eficiência, para, assim, decidir pela sua adoção. Segundo Fouilloux & Bannerot (1988), Van Oeveren & Stam (1992), comparações experimentais são, frequentemente, muito onerosas e podem ser feitas por meio de simulação Monte Carlo. Este método tipicamente

envolve a geração de observações de alguma distribuição de probabilidades e o uso da amostra obtida para aproximar a função de interesse.

A simulação, de acordo com Dachs (1988), consiste em imitar, por meio de programa de computador, o comportamento de um sistema real, para estudar seu funcionamento sob condições alternativas e viabilizar a extração de informações importantes, com baixo custo e rapidez, mediante alterações de interesse nos fatores envolvidos no fenômeno em estudo (Ferreira, 2001). Na realização de simulações, há a necessidade da formulação de modelos matemáticos que descrevam adequadamente o comportamento dos eventos reais em estudo (Mood et al., 1974). Por isso, a simulação computacional tem sido de grande utilidade em estudos genéticos sob vários contextos, como o de populações, do indivíduo ou do próprio genoma. Ela demanda dos geneticistas o desenvolvimento de modelos biológicos que retratem, da melhor forma possível, os fenômenos de interesse e, dos programadores, as rotinas para o processamento adequado, apesar de impor restrições, para que a influência de certos fatores possa ser avaliada (Cruz, 2001). O uso da simulação é muito amplo, podendo ser aplicada em áreas estratégicas, como medicina e agricultura.

No campo da genética e melhoramento de plantas, os modelos são inerentemente probabilísticos, por envolverem várias variáveis de natureza aleatória (Ferreira, 2001). Estes modelos necessitam ser suficientemente simples para permitir a operacionalização na forma de algoritmos e interpretação de forma adequada. Neste aspecto, surge, em muitos casos, uma desvantagem nos modelos simulados no que se refere à simplificação exagerada da realidade. Por outro lado, estes modelos permitem a variação nos valores assumidos pelos parâmetros, de tal forma que uma amplitude realística de fatores pode ser facilmente testada e, mais importante, o efeito relativo destes fatores sobre resultados da seleção pode ser estudado.

Como em todas as técnicas, cuidados devem ser tomados para evitar erros devidos a problemas amostrais, escolha inadequada das distribuições de probabilidade dos eventos de natureza aleatória, simplificação inadequada da realidade e erros de implementação do sistema simulado. Para isso, o programador deve utilizar os processos de validação, fazendo com que o sistema simulado possa operar nas condições do sistema real e verificar, por meio de testes de hipóteses e outras análises estatísticas, se os resultados observados na simulação são condizentes com os do sistema real (Ferreira, 2001). Conforme Ferreira (2001), no melhoramento de plantas, o uso da simulação é justificado quando: as soluções analíticas não existem ou o grau de dificuldade e o número de variáveis envolvidas não permitem a realização de inferências adequadas sobre o problema; pretende-se comparar a eficiência de um novo procedimento ou técnica em relação a outros já existentes e até mesmo consagrados e os métodos que são rotineiramente empregados em algumas etapas de um programa de melhoramento possam ser modificados para propiciar qualquer melhora ou aumento de sua eficiência.

O primeiro trabalho a utilizar simulação na genética foi realizado por Fraser (1957). A partir deste, a simulação computacional tem tido grande aplicabilidade nas pesquisas de genética e melhoramento de plantas, contribuindo de forma substancial para a resolução e a elucidação de problemas complexos. Vários exemplos podem ser citados, como estudos de determinação do número de famílias a serem avaliadas (Ferreira et al., 2000; Gurgel, 2004) e eficiência de dialelos circulantes na escolha de genitores (Veiga et al., 2000) dentre várias outras aplicações. Para o caso particular de espécies autógamas, as diferentes estratégias de melhoramento distinguem-se, principalmente, pela possibilidade de utilização de variados métodos de condução de populações segregantes. Dessa forma, um grande interesse tem sido focalizado para aperfeiçoar as estratégias em termos de eficiência, para, assim, decidir ou não

pela sua adoção. Para o melhoramento genético de plantas, é de fundamental importância a obtenção de informações sobre o potencial genético das cultivares, bem como sua capacidade de combinação, que resultem em híbridos produtores de populações segregantes promissoras.

Dessa forma, Van Oeveren & Stam (1992) sugerem que o uso da simulação computacional pode contribuir efetivamente para otimização de programas de melhoramento por sobrepor estas limitações de ordem prática e por possibilitar que um grande número de repetições seja utilizado, de forma que informações mais seguras sejam obtidas por meio dos resultados médios desses experimentos simulados.

O primeiro trabalho a utilizar simulação computacional em plantas autógamas foi realizado por Casali & Tigchelaar (1975). Foram comparados os métodos de pedigree, *bulk*, SSD e métodos combinando pedigree ou *bulk*, até as gerações F₃ e F₄, seguidos pelo SSD. A população F₂ simulada constou de 400 indivíduos. A seleção foi baseada no valor fenotípico. Ao final, foram simuladas 32 progênies F₆ para os métodos pedigree e *bulk*, e 32, 64 e 320 para SSD e métodos combinados. Os autores utilizaram os valores genotípicos (VG) médios da melhores progênies F₆ para verificar o progresso com cada método.

Os valores genéticos foram gerados considerando um sistema genético, assumindo 20 locos com distribuição independente, com dois alelos cada, de efeitos iguais e aditivos, admitindo a ausência de dominância. Para a obtenção dos valores fenotípicos, foram acrescentados aos VG efeitos ambientais aleatórios normalmente distribuídos. As simulações foram realizadas para as herdabilidades variando de 10%, 25%, 50% e 75%. Dois modelos genéticos foram empregados: Modelo I, assumindo um só caráter controlado por 20 locos; Modelo II, assumindo cinco caracteres controlados por 2, 3, e 5 locos com herdabilidade de 100%, 75%, 50%, 25% e 10%, respectivamente.

Os autores constataram que a escolha do método variou com a herdabilidade do caráter. O método de pedigree foi preferido para os casos de herdabilidade alta (75%) e moderada (50%), enquanto que, para herdabilidades de 10% a 25%, os métodos *bulk* e SSD, respectivamente, resultaram nas melhores progênies F_6 . Concluíram, ainda, que a seleção em gerações iniciais para caracteres altamente herdáveis, seguida pelo SSD até a geração F_6 e posterior seleção das progênies para caracteres de baixa herdabilidade, seria mais efetiva.

Muehlbauer et al. (1981) compararam por simulação os métodos *bulk* e SSD, assumindo um modelo constando de sete cromossomos com seis locos cada e apresentando taxas de recombinação constante entre locos vizinhos de 0,0; 0,1; 0,2; 0,3 e 0,5. A taxa de multiplicação média (TM) por planta para o método *bulk* foi de 50 sementes, com desvios de 25, 50 e 100 sementes por planta, ao passo que, para o SSD, as taxas de sobrevivência (TS) das plantas de uma geração para outra foram de 0,707, 0,780, 0,841, 0,95 e 0,99. Toda a variação genética presente foi estritamente aditiva. Considerou-se na simulação uma população F_2 de 200 indivíduos para cada método. Os critérios utilizados para as comparações dos métodos foram a variância genética aditiva, os valores genéticos médios na geração F_6 e a porcentagem de progênies F_2 representadas em F_6 .

Os autores observaram que a variância genética aditiva na F_6 no método *bulk* foi inferior à obtida no SSD, relatando que a TM afetou consideravelmente a variabilidade genética no *bulk*. Considerando-se os desvios de TM de 25, cerca de 75% das progênies F_2 não estavam representadas na F_6 . Para o método SSD, a TS foi importante. A perda de progênies na F_6 no SSD foi equiparada ao *bulk* quando o TS foi inferior a 75% a cada geração de avanço. Relatam, ainda, que a ligação teve pouca influência na variância genética aditiva.

Tendo em vista que a seleção pode ser praticada em diferentes estágios de um processo de melhoramento e com variadas intensidades, Van Oeveren & Stam (1992) realizaram um estudo via simulação computacional com o objetivo de avaliar a eficiência de dois distintos métodos: um método com teste em geração precoce (TGP) no caso, em F_3 , estimando a probabilidade de obter linhagem na F_∞ que superem um determinado padrão e o método SSD, no qual a seleção é atrasada para gerações de endogamia mais avançada (F_6).

Os métodos de seleção foram comparados para os casos de um e cinco cruzamentos sob seleção, sendo somente os cruzamentos mais promissores avançados para o TPG. Cada cruzamento teve diferente número de locos segregantes (no máximo 30). Admitiu-se, nas simulações, um modelo genético para o caráter quantitativo com todos os locos de efeitos iguais e aditivos e sem presença de ligação. Consideraram os seguintes fatores: número de plantas F_2 (25 a 400), herdabilidade individual (0,02 a 1,0) e o efeito de dominância ($d=0$, $d=0,5$, $d=a$), considerando $a=1$. As dez progênies fenotipicamente superiores selecionadas de cada método foram avaliadas e comparadas. Os critérios de comparação utilizados foram a média, o máximo e o desvio padrão genotípico dessas progênies.

Os autores concluíram que a herdabilidade e o número de plantas na F_2 foram as duas variáveis mais importantes. Além disso, considerando somente um cruzamento, o TPG resultou em progênies superiores aos do SSD. Porém, quanto mais cruzamentos envolvidos, os métodos não diferiram, exceto quando a herdabilidade foi baixa, em que o SSD foi até melhor, devido à ineficiência do TPG em selecionar os melhores cruzamentos. Relatou-se, ainda, que a dominância teve pouca influência sobre ambos os métodos estudados.

Os métodos *bulk-pedigree* (BP) e *selected bulk method* (SB), largamente utilizados pelo CYMMYT, foram comparados via simulação computacional (Wang et al., 2003). As simulações foram realizadas utilizando-se a plataforma

de simulação para análise quantitativa de modelos genéticos QU-GENE e o módulo de aplicação QUCIM, especificamente desenvolvido para simular o procedimento de melhoramento para a cultura do trigo, adotado pelo CYMMYT. Os métodos foram comparados em termos de ganho genético ajustado, do número de cruzamentos mantidos após um ciclo de melhoramento e da alocação de recursos. O ciclo de melhoramento iniciou-se com o cruzamento e terminou na geração em que as novas linhagens obtidas retornam à fase de cruzamento como genitores. Para este estudo, a geração considerada foi a F_{10} .

Os métodos avaliados neste estudo distinguiram-se apenas pela forma de condução das populações nas gerações segregantes F_3 a F_5 . Após o início, com a seleção das plantas na F_2 , ambos seguiram por três gerações de endogamia pelo método *bulk*, com a diferença de que, no método SB, apenas as plantas selecionadas foram colhidas em *bulk*. Após isto, novamente praticou-se a seleção de plantas individuais na geração F_6 . Nas gerações seguintes (F_7 a F_{10}), as progênies foram conduzidas por meio de *bulk* dentro de progênies e avaliadas em experimentos. Para ambos os métodos, consideraram-se os mesmos 1.000 cruzamentos a partir de 200 genitores.

Foram considerados nove caracteres do trigo, com diferentes números de locos segregantes, baseados em estudos realizados. Para o caráter produtividade de grãos, devido ao desconhecimento do número de locos em situações reais, considerou-se a segregação de 20 e 40 locos na simulação. A dominância não foi considerada. Os efeitos dos genes pleiotrópicos foram assumidos presumivelmente causadores da correlação entre caracteres. A herdabilidade individual para o rendimento de grãos foi fixada 0,05. Foram simuladas doze configurações, considerando o caráter produtividade de grãos, sendo dois níveis para o número de locos com frequência 0,5 (20 e 40), três níveis de epistasia (ausente, digênica e trigênica) e dois níveis de pleiotropia (ausente e presente).

Os autores concluíram que o ganho genético padronizado com método SB foi de 3,3% superior ao alcançado com BP. Evidenciou-se que aumento do número de locos e a presença concomitante de epistasia e pleiotropia dificultam a identificação dos melhores genótipos, independente da estratégia utilizada, reduzindo o ganho genético. Uma maior proporção de cruzamentos foi escolhida pelo método SB, comparado ao método BP. Quanto aos recursos alocados da F₁ a F₈, o método SB ocupou cerca de 1/3 a menos de área do que o método BP, além de produzir 40% menos progênies durante estas gerações. Assim, Wang et al. (2003) concluíram que o método SB foi mais efetivo quanto ao tempo, ao trabalho e aos custos envolvidos, além de resultar em ganho superior.

Nunes et al. (2008) realizaram estudo com o objetivo de avaliar a eficiência do método pedigree pela inclusão da informação de parentesco entre as progênies a partir da genealogia e, ainda, compará-lo com o método *bulk* via simulação computacional e em experimentos de campo. Nas simulações, assumiu-se um modelo genético de 20 locos segregantes, não ligados, de efeitos iguais e aditivos e com frequência alélica 0,5. Para as configurações avaliadas, foram considerados valores de herdabilidade no sentido amplo na média das progênies de 10%, 25%, 50% e 75% e grau médio de dominância de 0,0, 0,5 e 1,0. Foram simulados 1.000 experimentos, para cada configuração e método de condução, referentes à avaliação de 256 progênies F_{4,5}. Nos experimentos de campo, com a cultura do feijoeiro, foram avaliadas 256 progênies F_{4,5} na safra das águas 2004/2005 e as correspondentes 256 progênies F_{4,6} na safra das secas 2005, sendo estas oriundas dos métodos genealógico e *bulk*.

Para o método *bulk*, a análise foi realizada considerando as progênies como não relacionadas. Os critérios de seleção avaliados foram a média das progênies e o melhor preditor linear não tendencioso (*BLUP*) com a inclusão do parentesco, denotado por *BLUPA* e na ausência de parentesco, denotado por *BLUP*. A dominância teve pouca influência sobre os resultados das simulações

para ambos os métodos de condução. O *BLUPA* foi mais eficiente que a média em todas as configurações. Nos experimentos de campo, a variação genética entre as progênies foi significativamente não nula para a produção de grãos e altura de plantas, em ambos o método de condução.

Os autores concluíram que, utilizando-se o método pedigree, considerando a genealogia entre as progênies por meio do procedimento *BLUPA*, resulta em ganhos seletivos superiores aos alcançados quando esta informação é ignorada, de forma a compensar o trabalho despendido na obtenção desses registros, sobretudo para caracteres de herdabilidade baixa.

Os autores afirmam, ainda, que outra observação importante diz respeito aos resultados obtidos para os valores médios das estimativas da variância genética entre progênies para os métodos genealógico e *bulk*. Constataram que as estimativas médias para ambos os métodos foram semelhantes. Neste ponto, convém mencionar que, para o método *bulk*, assumiu-se que as plantas simuladas tinham igual taxa de multiplicação (de cada planta foram gerados 40 descendentes). Isso implica que a ação da seleção natural não foi considerada durante o processo de condução.

Outro enfoque dessa simulação foi o de verificar a eficiência relativa do método pedigree com o *bulk*. Levando-se em consideração as ressalvas já mencionadas no que tange à condução de ambos os métodos, observou-se que o valor genético médio das progênies selecionadas, em ambos os métodos, foi semelhante, em todas as configurações estudadas. Concluiu-se, assim, com base nos resultados simulados, que os dois métodos, em termos de ganhos proporcionados, foram igualmente eficientes.

3 MATERIAL E MÉTODOS

Foram utilizados métodos de simulação computacional para avaliar o efeito da amostragem no método *bulk* de melhoramento de autógamias. As linhagens obtidas nas gerações $F_{4:5}$, $F_{4:6}$ e $F_{24:25}$ foram utilizadas para avaliar o efeito na presença ou na ausência da seleção natural. Foram considerados diferentes números de genes segregando e diferentes tamanhos de *bulks*. A seleção natural também foi incorporada na simulação. Para isso, os indivíduos com genótipos mais favoráveis, ou seja, com mais locos com alelos favoráveis tinham a maior probabilidade de deixar descendentes.

Nas situações subsequentes, os mecanismos de simulação utilizados e as configurações escolhidas para realizar cada simulação são descritos. Também são descritos os métodos utilizados para obter os resultados. Foram considerados diferentes valores de herdabilidade, tomando-se por base a geração F_2 .

3.1 Simulação computacional

Para a realização da simulação computacional, foi realizado um cruzamento entre duas linhagens P_1 e P_2 homocigotas e contrastantes. Foram obtidas as gerações filiais F_1, F_2, \dots, F_t , as quais foram conduzidas em *bulk*. Na ℓ -ésima geração, foram obtidas famílias $F_{t: (t+1)}$ e estimados diferentes parâmetros. A média, o mínimo, o máximo e a variabilidade dessas medidas foram computados em um total de 1.000 simulações Monte Carlo de cada configuração. Diferentes tamanhos das populações (n) conduzidas em *bulk* foram considerados para que fossem avaliados efeitos da amostragem no método. Além disso, foram considerados diferentes números de locos segregantes g independentes, ou seja, sem ligação e sem efeito epistáticos ou

pleotrópicos. Também foram consideradas diferentes interações alélicas, diferentes herdabilidades e ausência e presença da seleção natural.

Os genótipos dos indivíduos foram constituídos de g locos independentes, sendo $g = 5, 50, 100$ e 200 . Inicialmente, foram gerados os genótipos de duas linhagens contrastantes para g locos e obtida a geração F_1 , cujos indivíduos eram heterozigotos para todos os locos. A partir daí eram geradas as populações F_2, F_3, \dots, F_t conduzidas em *bulk*. Foi considerado que cada indivíduo em uma dada geração, inicialmente, poderia gerar, em média, 80 sementes (descendentes). No entanto, o número de sementes de cada indivíduo foi gerado de uma Poisson, com média $\lambda = 80$. O valor 80 foi motivado por resultados experimentais, envolvendo plantas de feijoeiro comum.

Para simular a ação da seleção natural, foi considerado um valor de adaptação de cada indivíduo proporcional ao número de alelos favoráveis em sua constituição genotípica. Também foi considerado um valor mínimo desse valor adaptativo.

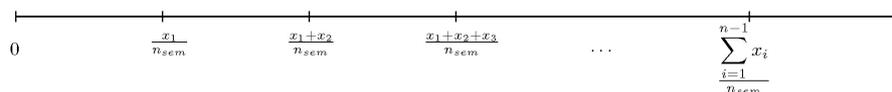
Seja F_i o valor adaptativo do i -ésimo indivíduo em uma dada geração; δ_i o número de alelos favoráveis desse mesmo indivíduo e $F_{\min} = 0,2$ e $1,0$ o valor adaptativo mínimo. Então:

$$F_i = 0,2 + \frac{\delta_i}{2g} (1 - F_{\min}) \quad (1)$$

seja z_i uma realização de variável aleatória Poisson com $\lambda = 80$, que representa o número de descendentes do i -ésimo indivíduo. O número de descendentes do i -ésimo indivíduo é, então, ajustado para seu valor adaptativo na seguinte forma:

$$x_i = z_i F_i \quad (2)$$

Se todos os indivíduos tivessem o mesmo número de sementes, eles deveriam ter probabilidade uniforme de serem selecionados para gerarem por autofecundação uma semente (descendente) para formar a próxima geração. No entanto, como os valores x_i variam potencialmente de indivíduo para indivíduo, considerando $i = 1, 2, \dots, n$. Então, o indivíduo candidato a formar a próxima geração era sorteado a partir de uma escala uniforme determinada pelo número de sementes (descendentes), obtido pela equação (2). Sendo n_{sem} o total de sementes dos descendentes, foi criada a seguinte escala uniforme:



Um número uniforme entre 0 e 1 era gerado e se o seu valor estivesse no primeiro segmento entre 0 e x_1/n_{sem} o primeiro indivíduo foi o elemento sorteado para ser autofecundado e deixar um descendente para a próxima geração; se o valor estivesse no segundo trecho, entre x_1/n_{sem} e $(x_1 + x_2)/n_{sem}$, o segundo indivíduo foi selecionado e assim por diante, até o n-ésimo intervalo.

Se o i-ésimo indivíduo selecionado para autofecundar tiver descendente a serem gerados, $x_i > 0$, então, é gerado esse descendente por autofecundação e o número de sementes é atualizado por $x_i = x_i - 1$. A escala anterior é recalculada com o novo valor x_i e o processo repetido n vezes. Se, em algum instante, o número de descendentes x_i do indivíduo selecionado for nulo, então, outro número uniforme deve ser gerado até que o valor do número de sementes desse indivíduo seja superior a zero. É importante salientar que, nesse processo, a probabilidade de sortear um indivíduo depende do seu valor adaptativo.

Cada indivíduo selecionado produziu um único descendente cada vez que foi sorteado, a menos que o valor de x_i seja nulo. Nesse caso, outro indivíduo é sorteado até que $x_i > 0$ e um descendente por autofecundação seja gerado. O processo foi repetido n vezes, após as devidas atualizações.

Com a geração filial obtida, o processo é repetido para se obter a próxima geração. Foram considerados tamanhos da população *bulk* de $n = 50, 200, 1000, 2000, 5000$.

Cada indivíduo, em cada geração, tem seu genótipo armazenado, sendo que o valor 0 representa *aa*, o valor 1 *Aa* e 2 *AA*, em que *a* e *A* são respectivamente os alelos desfavorável e favorável para um dado loco. Os valores genotípicos do *i*-ésimo indivíduo foi obtido por:

$$g_i = \sum_m^g a_m + \sum_m^g d_m \quad (3)$$

em que a_m representa o efeito aditivo do *m*-ésimo loco, sendo fixado como 1 qualquer que seja $k_2 = 1, 2, \dots$ e d_m é o efeito de dominância para o *m*-ésimo loco. Foram consideradas três situações para interações alélicas, as quais sejam, ausência de dominância ($d_m = 0$), dominância parcial ($d_m = 0,5$) e dominância completa ($d_m = a_m = 1$). Para cada situação, os valores de d_m , ao longo dos *g* locos foram todos iguais, ou seja, $d_1 = d_2 = \dots = d_g = d$.

Os descendentes de uma dada planta em uma dada geração foram obtidos por autofecundação, sendo sorteado um gameta masculino e outro feminino do conjunto total de gametas possíveis daquele indivíduo. Os gametas aleatórios, considerando os *g* locos independentes, foram combinados para formar um descendente. Se o genótipo do indivíduo no *m*-ésimo loco for igual a 2 (*AA*), seus gametas serão do tipo *A*, com 100% de probabilidade; se o genótipo do loco for 0 (*aa*), os gametas serão somente do tipo *a* e se o genótipo for 1(*Aa*), os gametas serão *A* com probabilidade 1/2 ou *a* com probabilidade 1/2. Nesse último caso, foi considerada a distribuição uniforme para sortear dois gametas desse loco. O indivíduo resultante foi obtido da junção desses dois gametas. Esse processo foi repetido para todos os locos e dependia do genótipo do indivíduo em questão para a obtenção de um descendente. Os valores foram

armazenados em uma matriz, com n linhas, em que cada uma correspondia a um indivíduo. As g primeiras colunas continham os valores genotípicos dos g locos com os valores (0, 1 ou 2) em cada célula. A coluna $g+1$ continha os valores genotípicos g_i dos indivíduos.

Neste trabalho, o *bulk* simulado foi conduzido até a geração F_4 , de onde foram obtidas famílias. Assim, o valor ℓ será fixado em 4, embora o programa de simulação permita variar esse número de acordo com a necessidade de seu usuário. O programa de simulação foi interpretado em linguagem R e foi disponibilizado no Anexo.

A simulação da geração $F_{4.5}$ foi realizada a partir de famílias obtidas de cada planta da geração F_4 . Como foram realizados experimentos com r repetições, sendo cada parcela constituída de k_1 plantas, então, as famílias $F_{4.5}$ foram formadas de rk_1 indivíduos cada.

Cada planta da geração F_4 originou rk_1 indivíduos, cuja autofecundação foi realizada da mesma forma como descrito para a condução do *bulk*. Cada grupo k_1 descendente dessa planta constituiu uma parcela experimental. Assim, r parcelas de cada família foram simuladas. Um total de nrk_1 plantas foi obtido no experimento da geração $F_{4.5}$.

A geração $F_{4.6}$ foi simulada a partir das famílias $F_{4.5}$ de forma um pouco diferente, uma vez que os descendentes são provenientes de rk_1 plantas da família $F_{4.5}$ e não apenas uma planta da família F_4 . Assim, de cada uma das famílias $F_{4.5}$ era sorteada uma planta entre as rk_1 plantas pertencentes à mesma e gerado um descendente por autofecundação para formar a família $F_{4.6}$ correspondente. O processo foi repetido, dentro de cada família, até a geração das rk_2 plantas da família $F_{4.6}$. O número de repetições dos experimentos $F_{4.5}$ e $F_{4.6}$ foram os mesmos, mas o número de plantas por parcela foi maior na geração $F_{4.6}$ ($k_2 > k_1$).

3.2 Experimentos simulados e parâmetros avaliados

O valor genotípico de cada planta, referente a i -ésima família, j -ésima planta da k -ésima repetição, foi calculado de acordo com a expressão (3), exceto pelo fato de que ele é representado nesse instante por g_{ijk} , sendo $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k_1$ ou k_2 e $k = 1, 2, \dots, r$. Os valores de k_1, k_2 e r foram fixados em 15, 60 e 3, respectivamente. O valor fenotípico de cada indivíduo foi obtido por:

$$F_{ijk} = \mu + g_{ijk} + \varepsilon_{ijk}, \quad (4)$$

em que μ foi fixado em 1800 kg/ha, tomando-se por referência o feijoeiro, sem perda de generalidade e ε_{ijk} é o erro experimental associado à i -ésima família, à j -ésima planta e à k -ésima repetição. O erro foi considerado normal independente e identicamente distribuído com média zero e variância σ^2 .

O valor de σ^2 foi determinado em função da h^2 tomando-se por referência a população F_2 . O valor de h^2 foi fixado em 1, 0,8, 0,5 e 0,1. Assim, considerando que a variância genética total σ_G^2 de F_2 é dada por:

$$\sigma^2 = 1/2 ga^2 + 1/4 gd^2, \quad (5)$$

pode-se determinar σ^2 da relação:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma^2}$$

resultando em:

$$\sigma^2 = \frac{1-h^2}{h^2} \sigma_G^2$$

Na expressão (5), considerou-se o fato de que os valores de a_m e d_m eram idênticos ao longo dos g locos, ou seja, $a_1 = a_2 = \dots = a_g = a = 1$ e $d = d_1 = d_2 = \dots = d_g = d$.

De cada indivíduo da geração $F_{4,5}$ e $F_{4,6}$, foi determinado o número total de alelos favoráveis (δ_i), cujo valor máximo é igual a $2g$. Em cada simulação, foram obtidos o valor médio, o mínimo e o máximo desse número de alelos

favoráveis, considerando as progênies $F_{4.5}$ e $F_{4.6}$. Essas quantidades foram representadas por δ_{med} , δ_{min} e δ_{max} . Também foram realizadas análises de variâncias em nível de famílias. Para isso, foram obtidas as médias dos valores fenotípicos para k_1 ou k_2 plantas das parcelas. Esse valor médio Y_{ik} foi modelado por:

$$Y_{ik} = \mu + f_i + \varepsilon_{i\delta} \quad (6)$$

em que f_i é o efeito da i -ésima família e $\varepsilon_{i\delta}$ é o erro médio entre parcelas, imposto normal com média 0 e variância σ^2/k_1 ou σ^2/k_2 .

O componente de variância entre famílias foi determinado utilizando o seguinte estimador:

$$\sigma_F^2 = \frac{QMF - QME}{r}$$

em que QMF e QME são os quadrados médios obtidos na análise de variância entre famílias e residual, respectivamente.

Como foram realizadas $N = 1000$ simulações Monte Carlo de cada configuração, pode-se determinar, para o número médio, mínimo e máximo de alelos favoráveis, os valores médio, mínimo e máximo na distribuição gerada. Da mesma forma, para variância genética estimada entre famílias, ao longo das $N = 1000$ simulações, foram obtidos os valores médio, mínimo e máximo. Essas quantidades foram comparadas entre si para diferentes tamanhos amostrais, o que permitiu que fosse mensurado o efeito da amostragem na variabilidade e nas perdas de linhagens no *bulk*.

4 RESULTADOS E DISCUSSÃO

4.1 Efeito do tamanho do *bulk*

Os resultados da Tabela 1 referem-se aos valores médios, mínimos e máximos, na distribuição gerada por simulação, a qual considerou 1.000 simulações Monte Carlo, do número de locos favoráveis (AA e Aa) no indivíduo do *bulk* de tamanho n que continha o maior valor fenotípico. Neste ponto, convém mencionar que se tomou por base uma planta de feijoeiro contendo 80 sementes por planta. É necessário observar que quando a herdabilidade é igual a 1, a relação fenotípica é idêntica à relação genotípica. Nesse caso, o indivíduo de maior número de locos favoráveis é selecionado. Essa situação foi utilizada como referência.

Numa geração $F_{4,5}$ ou $F_{4,6}$, espera-se, em média, $9/16$ (56,25%) de locos favoráveis na geração como um todo. Se considerarmos $n = 100$ locos e probabilidade de sucesso igual a $9/16$, então, a variável x que se refere à contagem do número de locos favoráveis, impondo locos independentes, tem distribuição binomial. Utilizando-se uma aproximação normal da binomial, pode-se ver que 99% dos indivíduos de uma dada geração são esperados com valores entre 43,48 e 69,02 locos favoráveis. Assim, apenas 0,5% superam 69,02 locos favoráveis. Isso quer dizer que, selecionando o melhor indivíduo com herdabilidade igual a 0,1, corresponderia a ter uma média próxima do limite de 69 locos favoráveis, se o *bulk* fosse de tamanho 200 indivíduos. Isso decore do fato de 0,5% dos indivíduos superarem o limite superior do intervalo e 0,5% corresponder, em média, a um indivíduo a cada 200 com essa característica. Na Tabela 1, as médias do número de locos favoráveis no melhor indivíduo com $h^2 = 1$ foram próximas desse limite teórico nas gerações $F_{4,5}$ ou $F_{4,6}$. É importante observar quando realiza-se simulação computacional verificar se o programa

implementado foi eficaz e, mais ainda, se os resultados refletem o que se espera que ocorra em condições reais (Ferreira, 2001). Os valores médios obtidos para as estimativas de locos favoráveis nas simulações estudadas foram próximas aos valores teóricos definidos.

Na Tabela 1 são apresentados os valores médios, mínimos e máximos de locos favoráveis ao longo das 1.000 simulações do melhor indivíduo selecionado fenotipicamente, as quais dependem da herdabilidade, do tamanho do *bulk* e das gerações consideradas, fixado o $a = d = 1$, ou seja, uma situação de dominância completa. Convém salientar que, em uma situação extrema em que a h^2 fosse nula, hipoteticamente falando, a relação do melhor fenótipo que teria associação nula com os valores genotípicos corresponderia a uma amostragem aleatória dos genótipos do *bulk*. Os 1.000 valores obtidos por simulação Monte Carlo corresponderiam a uma amostra aleatória de um *bulk* de tamanho 1000 e sua média deveria ser próxima de 9/16 de locos favoráveis nas gerações $F_{4,5}$ ou $F_{4,6}$.

Assim, o valor de 56,25% de locos favoráveis deverá ser utilizado nas discussões a seguir como referência para a média na comparação do efeito do tamanho de *bulks*. Os limites de 99% de confiança podem ser usados como valores de referência para mínimo e máximo; neste caso, o tamanho de *bulks* correspondente seria o de 200 (valor de referência).

O efeito da herdabilidade é nítido, como é esperado pela teoria. Para um tamanho fixo do *bulk* e da geração, observam-se ganhos maiores à medida que o *bulk* aumenta, quando se compara a média do número de locos favoráveis entre a herdabilidade 0,1 a 1,0. A partir de *bulks* de tamanho 1000, os ganhos no número de locos favoráveis de $h^2 = 1$ para $h^2 = 0,1$ ficam praticamente estáveis, próximos de 10 locos favoráveis por tamanho de *bulk*.

TABELA 1 Número médio, mínimo e máximo de locos favoráveis do melhor indivíduo dentro das famílias $F_{4:5}$ e $F_{4:6}$, nos 100 locos segregando com interação alélica de dominância completa $a = d = 1$.

Valores	n	h^2							
		0,1		0,5		0,8		1,0	
		$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$
Méd	50	56,67	56,68	63,50	62,63	66,16	64,42	66,52	65,98
	200	60,02	58,90	65,64	64,86	67,52	66,80	68,10	68,52
	1000	61,10	60,00	66,18	66,12	69,68	67,69	71,56	69,86
	2000	62,15	61,00	68,39	67,15	69,89	68,71	73,13	72,00
	5000	63,31	62,20	68,41	67,54	71,38	70,03	73,49	73,20
Mín	50	41,00	41,00	53,00	53,00	58,00	57,00	60,00	60,00
	200	45,00	44,00	56,00	55,00	60,00	60,00	64,00	62,50
	1000	45,00	44,00	57,00	56,12	60,00	60,12	68,00	65,00
	2000	46,91	46,00	58,61	57,23	61,09	61,10	69,13	68,09
	5000	47,88	47,00	58,80	57,78	62,10	64,08	69,36	68,26
Máx	50	69,00	68,34	75,00	74,00	76,00	75,00	76,00	75,00
	200	75,00	74,21	76,00	75,36	77,00	76,00	79,00	77,00
	1000	76,00	75,00	77,00	76,00	79,00	78,00	80,00	79,00
	2000	77,00	76,00	77,00	76,00	78,00	78,00	80,00	80,00
	5000	78,00	77,16	78,00	77,00	79,00	79,00	80,00	80,00

Para o valor mínimo do número de locos favoráveis do melhor fenótipo ao longo das 1.000 simulações Monte Carlo, o efeito da herdabilidade foi muito pronunciado, tanto em pequenos *bulks* quanto em grandes *bulks*. Com *bulk* de tamanho $n = 50$ e $h^2 = 0,1$, o valor mínimo do número de locos do melhor fenótipo é igual a 41, que está muito aquém da média da geração $F_{4:5}$. Isso indica que, na pior situação, com *bulks* muito pequenos e herdabilidade baixa, o indivíduo selecionado é pior que a média da geração. Para $h^2 = 1$ com *bulk* de tamanho $n = 5000$, o valor de número de locos favoráveis mínimo do melhor fenótipo é de 68,26, que é bem inferior a média da geração. No entanto, nenhum

risco o melhorista terá de obter um indivíduo com número de alelos inferior à média da geração. A importância deste trabalho fica justificada por essa simples constatação.

Para o valor máximo, a herdabilidade tem grande efeito se o *bulk* for pequeno e efeito inexpressivo se o *bulk* for grande. Esse resultado é semelhante para o ocorrido para o mínimo. Na situação mais drástica, $h^2 = 0,1$ e $n = 50$, o valor máximo do número de locos favoráveis foi de 69,00 na $F_{4,5}$ e de 68,34 na $F_{4,6}$. Isso indica que o melhor cenário na pior configuração vislumbra situações de ganhos favoráveis, pois esse valor é maior que a média da geração (56,25).

Nas Tabelas 2 e 3, estão apresentados os resultados dos valores médios, mínimos e máximos obtidos nas 1.000 simulações Monte Carlo do número de locos favoráveis do melhor indivíduo identificado pelo fenótipo, ou genótipo se $h^2 = 1$, para um *bulk* de tamanho n , considerando dominância parcial e ausência de dominância, respectivamente. Como o valor fenotípico é resultante do valor genotípico e do ambiente, espera-se que haja mudanças no fenotípico com as mudanças no tipo de interação alélica. Assim, a seleção do fenótipo do melhor indivíduo terá efeito esperado nas medidas do número de locos favoráveis avaliados nesse indivíduo. Entretanto, considerando outro modelo de interação alélica, o padrão de resposta permaneceu o mesmo, uma vez que, na geração F_4 , ainda se tem 1/8 de locos em heterozigose.

Na Tabela 2 estão os resultados para $d = 0,5$ e, na Tabela 3, para $d = 0$. Em ambos os casos, o aumento da $h^2 = 0,1$ para 1 provoca acréscimo expressivos nas médias dos números de locos favoráveis. O que se pode depreender é que as alterações observadas nos resultados da Tabelas 1, 2 e 3 são consequência da frequência dos locos em heterozigose. A média do número de locos favoráveis com $n = 5000$ e com $d = 0$ (Tabela 3) é de 66,79 para $h^2 = 1$ de 73,49, com $h^2 = 1$ e $d = 1$ (Tabela 1). Essa diferença de praticamente sete locos favoráveis pode ser potencialmente atribuída à presença dessa dominância. Isso tenderá a

dissipar, se os *bulks* forem conduzidos a uma geração mais avançada que F_4 para se obter famílias. Antes de qualquer comentário, pode-se afirmar que a dominância não exerceu qualquer influência sobre as estimativas do número de locos favoráveis. Este resultado está de acordo com os obtidos por Van Oeveren & Stam (1992) e Nunes et al. (2008) que, por meio de simulação computacional, também destacaram que a dominância tem pouca importância no melhoramento de autógamias.

TABELA 2 Número médio, mínimo e máximo de locos favoráveis do melhor indivíduo dentro das famílias $F_{4:5}$ e $F_{4:6}$ nos 100 locos segregando com interação alélica dominância parcial $d = 0,5$

Valores	n	h^2							
		0,1		0,5		0,8		1,0	
		$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$
Méd	50	59,76	59,08	63,60	62,15	64,96	64,20	66,38	65,13
	200	60,62	59,96	64,90	63,01	65,38	64,80	68,86	68,08
	1000	61,21	61,00	64,12	63,00	66,28	65,15	69,33	68,19
	2000	62,03	61,00	65,13	64,00	66,73	66,00	70,19	69,00
	5000	63,49	62,31	66,28	65,00	68,19	67,00	71,11	70,00
Mín	50	43,00	42,00	45,00	44,00	48,00	46,00	46,00	45,00
	200	45,00	43,65	47,00	46,09	49,00	48,00	47,20	46,00
	1000	45,00	44,00	48,00	47,00	50,80	49,49	49,43	47,00
	2000	45,67	45,00	48,93	47,05	51,20	50,00	51,34	49,21
	5000	47,00	46,00	49,00	47,45	51,40	50,00	51,40	50,05
Máx	50	72,00	71,00	74,00	72,00	75,00	74,00	77,00	76,00
	200	72,00	71,00	75,00	75,00	76,10	75,06	77,00	76,05
	1000	75,00	74,02	77,00	76,03	78,00	77,30	78,00	77,00
	2000	75,89	74,34	77,29	76,45	78,59	77,46	79,06	78,00
	5000	78,00	77,00	78,00	77,39	79,00	78,00	80,00	79,00

TABELA 3 Número médio, mínimo e máximo de locos favoráveis do melhor indivíduo dentro das famílias $F_{4:5}$ e $F_{4:6}$ nos 100 locos segregando com interação alélica ausência de dominância $d = 0$

Valores	n	h^2							
		0,1		0,5		0,8		1,0	
		$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$	$F_{4:5}$	$F_{4:6}$
Méd	50	59,02	58,13	61,49	60,24	62,62	61,64	63,25	62,50
	200	60,95	59,15	62,79	61,12	63,17	62,59	64,08	63,92
	1000	61,76	60,12	63,82	62,30	63,93	62,91	64,89	63,26
	2000	62,12	61,05	64,23	63,37	64,98	64,03	65,38	63,77
	5000	62,48	61,45	65,29	64,20	65,63	65,20	66,79	65,69
Mín	50	45,00	44,00	48,00	47,00	50,00	49,00	51,00	50,00
	200	49,00	47,00	49,00	47,50	51,00	50,00	52,00	51,00
	1000	49,00	48,00	51,00	48,70	51,20	50,00	53,00	52,00
	2000	50,18	49,20	52,34	51,00	52,33	51,36	53,58	52,00
	5000	51,00	50,00	53,00	52,02	53,00	52,12	54,00	53,22
Máx	50	75,00	74,00	76,00	75,00	77,00	76,00	77,00	76,00
	200	76,00	75,00	77,00	76,00	78,00	77,00	78,00	77,00
	1000	76,00	75,00	77,00	76,18	79,00	78,00	79,00	78,03
	2000	76,89	75,40	77,89	76,90	79,89	78,60	80,06	79,01
	5000	78,00	77,00	79,00	78,00	80,00	79,00	81,00	80,00

A discussão mais importante refere-se ao efeito do tamanho da amostra na variabilidade e no potencial do *bulk*. Como o padrão de resposta é o mesmo, independente da interação alélica, foram considerados os resultados da Tabela 1 para discutir esse efeito. Fixado um valor de h^2 , verifica-se que existem ganhos acentuados com o aumento do *bulk*. Se a herdabilidade é grande, os ganhos com o aumento de n são menores. Este fato é esperado, pois, o número de locos favoráveis com herdabilidade crescente está em um patamar superior. Por exemplo, de $n = 2000$ para 5000 com $h^2 = 0,1$, o número médio de locos

favoráveis aumenta de 62,15 para 63,31, enquanto com $h^2 = 1$, os valores aumentaram de 73,13 para 73,49, na $F_{4,5}$, respectivamente.

Outro fato que merece ser destacado é que, em condições de herdabilidade baixas exigem-se *bulks* muito grandes em comparação a um *bulk* bem menor se a herdabilidade for alta. Tomando-se como exemplo $h^2 = 0,5$, um *bulk* de aproximadamente $n = 2000$ indivíduos seria necessário para equivaler a um *bulk* de tamanho $n = 200$ com $h^2 = 1$.

Este fato realça a importância da herdabilidade. Como comumente relatado na literatura, o sucesso do melhorista cresce com o incremento da h^2 do caráter (Falconer, 1960). O que se percebe também é que o aumento no tamanho da amostragem provoca acréscimos que vão diminuindo de magnitude na medida em que aumenta o tamanho de *bulk*.

Uma inspeção nas Tabelas 1, 2 e 3, com pouquíssimas exceções, permite inferir que um *bulk* de tamanho 2000 indivíduos, fixada a h^2 em qualquer valor, e se o *bulk* for além deste tamanho os ganhos são inexpressivos. Isto pode ser verificado no gráfico da Figura 1 para $h^2 = 0,1$ e $h^2 = 1$, situações extremas escolhidas para ilustrar o fato.

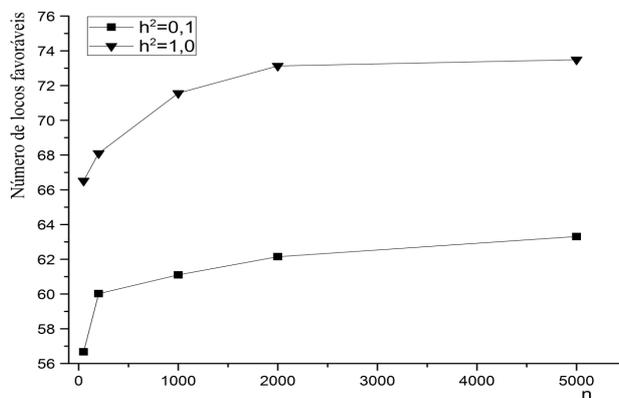


FIGURA 1 Número de locos favoráveis versus tamanho de *bulk* com $h^2 = 0,1$ e 1 para 100 locos segregando $d = a = 1$ em famílias $F_{4;5}$

A variabilidade do número de locos favoráveis pode ser indiretamente estudada pelos valores mínimos e máximos da distribuição, observando, no melhor fenótipo ou genótipo, se $h^2 = 1$. O que se verifica é que o mínimo tem seu menor valor em *bulks* pequenos ($n = 50$) e seu maior valor em *bulk* de maior tamanho ($n = 5000$). Da mesma forma, o menor valor do valor máximo é de 69, que foi observado no *bulk* de menor tamanho ($n = 50$) e maior valor do máximo ocorreu para $n = 5000$ (Tabela 1). Para os demais casos de interações alélicas, o mesmo padrão foi detectado (Tabelas 2 e 3). Para resumir, pode-se inferir que a distribuição do número de locos favoráveis é transladada para a direita no eixo real, à medida que n aumenta. Assim, espera-se que os mínimos e os máximos aumentem com o tamanho do *bulk*. Dessa forma, é possível afirmar que, para valores acima de 2.000 indivíduos, os ganhos nessa translação são inexpressivos.

Na Tabela 4 estão apresentados resultados idênticos aos das situações da Tabela 1 para a média, exceto pelo fato de que a geração considerada foi a $F_{24;25}$. Assim, a dominância é praticamente inexpressiva nessa geração. Os tamanhos de *bulk* e a h^2 tiveram o mesmo padrão de resposta daquele observado na Tabela 1.

A ressalva é a de que o número de locos favoráveis, na melhor situação, $n = 5000$ e $h^2 = 1$ (65,12), está bem aquém do valor observado nessa mesma situação em $F_{4:5}$ (73,49). Nesse caso, a média da geração é de praticamente 50% dos locos favoráveis (AA), com queda de 6,25 pontos percentuais em relação à média da geração $F_{4:5}$. O tamanho do *bulk* de 2.000 indivíduos com $h^2 = 0,1$ praticamente reproduz o potencial igual ou superior a 2000 com $h^2 = 1$. O aumento do tamanho do *bulk* de 2.000 para 5.000, praticamente não teve efeito no número médio de locos favoráveis do melhor indivíduo avaliado fenotipicamente.

TABELA 4 Número médio de locos favoráveis do melhor indivíduo nos 100 locos segregando com interação alélica de dominância completa $a = d = 1$ nas famílias $F_{24:25}$

Valores	n	h^2			
		0,1	0,5	0,8	1,0
	50	56,87	57,86	58,17	58,32
	200	61,16	61,43	61,62	61,86
Méd	1000	63,45	63,45	63,78	64,50
	2000	65,00	65,00	65,00	65,00
	5000	65,00	65,00	65,02	65,12

Nas Tabelas 5, 6 e 7 são apresentados os valores da média, do mínimo e do máximo da estimativa de variância genética entre famílias para a geração $F_{4:5}$, considerando $g = 100$ locos e diferentes interações alélicas em função da herdabilidade e do tamanho do *bulk*. Pode-se observar que, independentemente da interação alélica, a variância genética não se altera com o aumento da h^2 fixado o tamanho do *bulk*. Uma única exceção ocorreu para $n = 50$ e $h^2 = 0,1$ com $a = d = 1$ (Tabela 5).

Para o tamanho do *bulk*, o efeito foi pronunciado na média e na amplitude de variação, conforme pode ser observado pelas variações no mínimo e no máximo. Novamente, valores entre 1.000 e 2.000 parecem ser adequados, pois há uma aproximação dos valores para o limite estabelecidos com $n = 5000$, que se considera um tamanho adequado o suficiente para servir de referência. Há um maior indício, pelos resultados, de que o tamanho adequado $n = 2000$ seja o melhor.

TABELA 5 Variância genética média, mínima e máxima entre famílias $F_{4:5}$ com $g = 100$ locos segregando com interação alélica de dominância completa $d = a = 1$, em função de n e h^2

Valores	n	h^2			
		0,1	0,5	0,8	1,0
Méd	50	42,96	46,78	47,89	47,89
	200	70,22	70,17	70,15	70,06
	1000	89,86	89,58	89,65	89,98
	2000	89,86	89,63	89,85	89,98
	5000	89,90	89,73	89,75	89,89
Mín	50	31,27	34,37	32,90	30,15
	200	47,39	54,23	55,04	46,41
	1000	70,43	70,69	71,14	72,51
	2000	72,83	72,59	72,44	72,61
	5000	72,84	73,60	72,54	72,91
Máx	50	168,07	191,94	175,25	198,58
	200	140,69	134,28	134,69	146,24
	1000	111,72	112,75	123,25	109,75
	2000	100,34	112,95	113,65	110,15
	5000	101,73	113,09	114,18	100,45

TABELA 6 Variância genética média, mínima e máxima entre famílias $F_{4,5}$ com $g = 100$ locos segregando com interação alélica de dominância parcial $d = 0,5$, em função de n e h^2

Valores	n	h^2			
		0,1	0,5	0,8	1,0
Méd	50	42,20	45,90	46,80	47,64
	200	69,39	69,47	69,65	70,01
	1000	89,56	89,58	89,65	89,68
	2000	89,56	89,61	89,65	89,78
	5000	89,60	89,63	89,65	89,71
Mín	50	31,62	28,48	29,80	34,47
	200	52,61	55,30	55,15	52,45
	1000	62,36	67,64	47,74	67,03
	2000	72,65	71,88	72,58	71,23
	5000	73,00	72,38	72,20	71,73
Máx	50	200,77	202,86	199,40	196,04
	200	131,68	136,98	136,54	138,98
	1000	142,39	109,45	107,68	107,61
	2000	112,32	114,25	110,24	104,06
	5000	114,21	115,38	111,00	103,27

TABELA 7 Variância genética média, mínima e máxima entre famílias $F_{4:5}$ com $g = 100$ locos segregando com interação alélica ausência de dominância $d = 0,0$ em função de n e h^2 .

Valores	n	h^2			
		0,1	0,5	0,8	1,0
Méd	50	41,95	44,72	45,10	46,48
	200	68,59	68,62	68,64	69,71
	1000	88,75	89,18	89,22	89,28
	2000	89,11	89,17	89,20	89,30
	5000	89,20	89,21	89,25	89,31
Mín	50	24,90	26,14	31,56	32,60
	200	51,14	54,99	54,88	57,15
	1000	68,28	71,05	72,04	72,23
	2000	71,17	69,12	67,78	66,34
	5000	62,36	67,64	67,74	67,03
Máx	50	197,06	180,79	191,59	173,69
	200	141,13	138,22	137,62	134,78
	1000	107,92	103,61	106,94	107,52
	2000	112,08	114,05	112,56	111,35
	5000	142,39	109,45	147,68	107,61

Para uma geração mais avançada de endogamia $F_{24:25}$, os mesmos resultados apresentados na Tabela 5 foram reproduzidos com essa população (Tabela 8). Para essa situação, houve um pronunciado efeito do tamanho do *bulk* na média das estimativas da variância genética entre famílias $F_{24:25}$, para todos os valores de h^2 . Observando-se o tamanho de *bulk* de 2.000 indivíduos, nota-se que houve estabilização dos valores, comparado com os do *bulk* de tamanho 5.000. Para $h^2 = 0,1$, observou-se que, com $n = 50$, o valor da variância genética média foi de 43,70 e, para $n = 5000$, de 98. A diferença é muito grande quando se comparam os valores para $n = 50$ e $n = 5000$. É oportuno salientar que os resultados dos componentes de variância são provenientes de experimentos com

três repetições com quinze plantas por parcela em uma estrutura experimental casualizada. Assim, os resultados estão sujeitos ao erro experimental. Outra importante constatação é que, mesmo com herdabilidades altas, mas com amostras pequenas $n \leq 200$, há uma perda muito grande na variabilidade média. Assim, se as herdabilidades forem pequenas, $h^2 = 0,1$, que é comum em várias situações em autógamias (Carneiro et al., 2002 e Moreto et al., 2007), com amostras do *bulk* iguais ou superiores a 1.000, a variabilidade média supera o caso extremo $n = 200$ com $h^2 = 1,0$ em relação ao tamanho de amostra do *bulk*, considerando uma herdabilidade 0,1.

TABELA 8 Variância genética média entre famílias $F_{24:25}$ com $g = 100$ locos segregando com interação alélica dominância completa $d = a = 1$, em função de n e $h^2 = 0,1$

Valores	n	Variância
	50	43,70
	200	70,00
Média	1000	98,00
	2000	98,02
	5000	99,08

A validade da simulação pode ser checada observando-se os resultados dos valores médios, mínimos e máximos do número de alelos favoráveis da geração $F_{4:5}$, com $g = 100$ locos segregando (Tabelas 1A, 2A, e 3A). O número de alelos favoráveis é 100, uma vez que a geração estudada é resultante do cruzamento de duas linhagens puras e contrastantes. Os valores médios observados destoam ligeiramente desse valor médio. Uma explicação para esse desvio positivo do valor médio esperado pode ser dada pela incorporação do efeito da adaptação nos mecanismos de simulação. Os indivíduos mais

adaptados tendem a deixar mais descendentes do que os menos adaptados, o que pode aumentar a média do número de alelos favoráveis, conforme Allard (1988), Corte et al. (2002), Gonçalves et al. (2001), Pirola et al. (2002) e Silva et al. (2004).

Nenhum efeito da herdabilidade é percebido, uma vez que a avaliação é feita diretamente nos genótipos dos indivíduos da geração considerada. As pequenas diferenças observadas são atribuídas ao erro Monte Carlo da simulação. Também se verifica, como era esperado, uma não interferência da dominância. Isso corrobora a qualidade da simulação realizada, conforme Tabelas 1A, 2A, e 3A, no Anexo.

Importantes resultados, que confirmam todo o padrão de resposta encontrado para $g = 100$ locos, estão apresentados na Tabela 4A. Nesse caso, várias outras situações foram investigadas em função de g e de n . O efeito do *bulk* é consistentemente para $g = 5, 50$ e 200 . Um importante fato, que merece ser destacado, é que com $g = 200$ locos, os ganhos no número de locos favoráveis não estabilizaram com aumento no tamanho do *bulk*, mesmo para *bulk* de tamanho $n = 5000$. Isso indica que, se o caráter é controlado por mais de 100 genes, *bulks* de tamanhos iguais ou maiores que 5.000 indivíduos devem ser utilizados, se houver viabilidade prática e econômica para isso.

4.2 Efeito da amostragem na seleção natural

Para verificar o efeito da amostragem na seleção natural, simulações foram feitas com adaptabilidade mínima igual a 1. Nesse caso, cada indivíduo, ao deixar descendentes para a próxima geração, entrava em igualdade de condições com seu concorrente. As simulações da seção anterior foram realizadas com adaptabilidade mínima igual a 0,2. Isso implicava numa variação de 20% a 100% na probabilidade de o indivíduo deixar descendentes para a próxima geração. Esse valor era dependente do número de alelos favoráveis na

sua constituição genotípica. O valor da adaptação foi usado para modificar a média da Poisson, que foi o modelo usado para especificar o número de descendentes de um indivíduo para a próxima geração.

Na Tabela 9, estão apresentados os números médios de locos favoráveis no melhor indivíduo, considerando $h^2 = 1$ e também para a comodidade do leitor, as mesmas situações foram simuladas, considerando adaptabilidade mínima de 0,2. Convém salientar que o *bulk* de tamanho 5.000 indivíduos ficou inviabilizado de ser simulado em virtude de o laboratório de computação científica da universidade, onde eram realizadas as simulações de maior demanda de processamento e memória, ter entrado em colapso. Pode-se observar (Tabela 9) que, nos *bulks* pequenos ($n \leq 200$), o número de locos favoráveis do *bulk* simulado com adaptabilidades iguais de todos os indivíduos foi superior ao do *bulk* com adaptabilidade de 20% a 100% entre os indivíduos. Para *bulks* maiores ($n \geq 1000$), houve uma inversão desses desempenhos médios. Assim, o efeito do tamanho do *bulk* só irá capitalizar o efeito da seleção natural em grandes tamanhos de amostra ($n \geq 1000$). É uma afirmativa consistente, pois, com um número maior de descendentes, a seleção natural pode atuar com maior intensidade (Hamblin, 1977).

TABELA 9 Número médio de locos favoráveis no melhor indivíduo com $h^2 = 1$ em diferentes adaptabilidades e com $g = 100$, $d = a = 1$.

Valores	n	$F_{\min} = 0,2$	$F_{\min} = 1,0$
Médios	50	66,52	68,70
	200	68,10	68,80
	1000	71,56	69,70
	2000	73,13	69,90

Na Tabela 10, só para a média $F_{\min} = 0,2$ e $F_{\min} = 1,0$ estão apresentados os números de locos favoráveis na geração $F_{4,5}$, em função do tamanho do *bulk* n para uma situação de h^2 igual a 1×10^{-8} . Esse valor foi escolhido para ilustrar uma situação de herdabilidade nula, que é impossível de ser simulada diretamente. Isso equivale, ao se realizar a seleção do melhor fenótipo, a um sorteio aleatório entre indivíduos daquela geração. Assim, a média do melhor fenótipo nas 1.000 simulações realizadas corresponderá a uma estimativa da média do número de locos favoráveis de uma geração F_5 com tamanho de amostra igual a 1.000. Essa média é teórica e facilmente determinada, sendo igual a 53,125 locos para $g = 100$. Como a seleção, nesse caso de melhor fenótipo, é totalmente ao acaso, o efeito do tamanho do *bulk* é praticamente inexistente, pois qualquer que seja o tamanho do *bulk*, o tamanho final conterà 1.000 indivíduos aleatório,s obtidos em cada uma das 1.000 simulações. Mesmo com *bulk* de tamanho $n = 50$, ao final das simulações, 50.000 indivíduos foram simulados e uma amostra de 1.000 foi obtida. Esse número de indivíduos simulados é expressivamente grande. Se a h^2 for extremamente pequena,

nenhum efeito da amostragem interferirá no resultado e na probabilidade de sucesso do método *bulk* resultar em uma linhagem promissora.

Comparando-se os valores médios dos locos favoráveis para $F_{\min} = 0,2$ e para $F_{\min} = 1,0$, ainda com h^2 igual a 1×10^{-8} , verifica-se que não há efeito da seleção natural, pelas mesmas razões comentadas anteriormente (Tabela 10)

TABELA 10 Número médio de locos favoráveis no melhor indivíduo com $h^2 = 1 \times 10^{-8}$ em diferentes adaptabilidades e com $g = 100$, $d = a = 1$

Valores	n	$F_{\min} = 0,2$	$F_{\min} = 1,0$
Médios	50	52,99	56,30
	200	53,74	58,76
	1000	69,71	68,50
	2000	69,92	69,68

4.3 Considerações gerais

Os efeitos da amostragem e da seleção natural foram investigados neste estudo. Uma das principais medidas utilizadas foi a média do número de locos favoráveis no melhor indivíduo identificado fenotipicamente. Foram consideradas diferentes herdabilidades e diferentes números de locos. Quando a h^2 foi considerada igual a 1, a seleção fenotípica era equivalente à seleção genotípica por razões óbvias.

O efeito da amostragem teve grande magnitude no desempenho do *bulk*. Esse efeito era mais pronunciado nas situações de baixa herdabilidade. Verificou-se que o ganho no número de locos favoráveis do melhor indivíduo, praticamente não apresentou acréscimo expressivo para *bulks* de tamanhos

iguais ou superiores a 2.000. Quando o *bulk* era conduzido até uma geração muito avançada, $F_{24:25}$, as médias reduziram-se, entretanto, o comportamento geral continua o mesmo, indicando que o tamanho ideal do *bulk* deveria ser de 2.000 indivíduos. Vários resultados foram utilizados para checar a consistência e a confiabilidade das simulações realizadas.

Finalmente, o efeito da seleção natural associado aos tamanhos de *bulks* foi avaliado. O que se constatou é que somente para *bulks* maiores ($n \geq 1000$) é que o efeito da seleção natural era capitalizado. Por exemplo, com $n = 2000$, a média do número de locos favoráveis foi de, aproximadamente 70 para $F_{\min} = 1,0$ e de 73 para $F_{\min} = 0,2$. Isso indica que um ganho médio de 3 locos favoráveis seria atribuído à seleção natural, considerando $h^2 = 1$. Resultados para herdabilidade de zero comprovam novamente a consistência dos resultados obtidos por simulação. O efeito do tamanho do *bulk* é o mesmo, independentemente de a adaptabilidade ter sido ou não considerada na simulação.

É preponderante ressaltar que o resultado da seleção natural apresentado neste trabalho seja visto com muitas ressalvas. Inicialmente, o mecanismo de simulação utilizado é limitado e dificilmente retrata a realidade com fidelidade, embora seja uma tentativa de simplificá-la e modelá-la.

Ademais, a situação retratada utiliza o feijoeiro comum como referência, sendo utilizado um número médio de 80 sementes por planta. Mesmo um genótipo com todos os alelos desfavoráveis terá um valor de adaptabilidade de 20%, o que corresponderia a um valor médio de 16 descendentes desse indivíduo. A não expressividade do efeito da seleção natural, possivelmente ocorreu por terem sido tomadas como referência 80 sementes, o que representa um feijoeiro adaptado, ressaltando que, nas populações menos adaptadas, a seleção natural pode atuar mais intensamente. Isso pode não ter refletido uma maior vantagem a favor da seleção natural se o resultado fosse comparado com

outra situação em que a média fosse bem inferior a 80 sementes do feijoeiro. Aperfeiçoamentos podem ser realizados em futuros trabalhos, uma vez que os códigos utilizados na simulação estão disponibilizados neste trabalho (Anexo B).

5 CONCLUSÕES

A utilização da simulação computacional demonstrou eficiência e coerência e foi informativa no estudo da amostragem no método *bulk*.

Na determinação do tamanho do *bulk* em autógamias, a herdabilidade deve ser considerada

O tamanho de *bulk* de 2.000 indivíduos foi equivalente ao *bulk* de 5.000 indivíduos.

O efeito da seleção natural só resultou em ganho quando os tamanhos de *bulks* foram superiores a 1.000 indivíduos.

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, A. de F. B.; RAMALHO, M. A. P.; SANTOS, J. B. Prediction of seed-yield potential of common bean populations. **Genetics and Molecular Biology**, Ribeirão Preto, v. 25, n. 3, p. 323-327, Sept. 2002.

ALLARD, R. W. Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. **Jornal of Heredity**, Cary, v. 79, n.4, p. 225-238, July/Aug 1988.

BAENZIGER, P. S.; PETERSON, E. J. Genetic variation: its origin and use for breeding self-pollinated species. In: STALKER, H. T.; MULTRIPHY, J. P. **Plant breeding in the 1990's**. Wallingford: North Carolina State University, 1991. p.69-100.

BORÉM, A.; MIRANDA, G. V. **Melhoramento de Plantas**. 4. ed. rev. amp. Viçosa: UFV, 2005. 525 p.

CARNEIRO, J. E. de S.; RAMALHO, M. A. P.; ABREU, A. de F. B.; GONÇALVES, F. M. A. Breeding potential of single double and multiple crosses in common bean. **Crop Breeding and Applied Biotechnology**, Londrina, v. 2, n. 4, p. 515-524, out./dez. 2002.

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

CORTE, H. R.; RAMALHO, M. A. P.; GONÇALVES, F. M. A.; ABREU, A. de F. B. A. Natural selection for grain yield in dry bean populations bred by the bulk method. **Euphytica**, Wageningen, v. 123, n. 3, p. 287-393, 2002.

CRUZ, C. D. A informática no melhoramento genético. In: NASS, L L.; VALOIS, A. C.; MELO, I. S. de.; VALADARES-INGLIS, M. C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, 2001. p. 1085-1118.

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

FALCONER, D. S. **Introduction to quantitative genetics**. New York: Ronald, 1960. 365 p.

FEHR, W. R. **Principles of cultivar development: theory and technique**. New York: Mamillann, 1987. v. 1, 525 p.

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

FERREIRA, W. D.; RAMALHO, M. A. P.; FERREIRA, D. F.; SOUZA, M. A. de. Family number in common bean selection. **Genetics and Molecular Biology**, Ribeirão Preto, v. 23, n. 2, p. 403-409, June 2000.

FOIULLOUX, G.; BANNEROT, H. Selection methods in common bean (*Phaseolus vulgaris* L.). In: GEPS, P. **Genetics resources of phaseolus beans**. Drecht: Kluwer Academic, 1988. p. 503-541.

FRASER, A. S. Simulation of genetics systems by automatic digital computers. II: introduction. **Australian Journal of Biological Science**, Melbourne, v. 10, p. 484-491, 1957.

GONÇALVES, F. M. A.; RAMALHO, M. A. P.; ABREU, A. de F. B. Natural selection in four common bean traits. **Crop Breeding and Applied Biotechnology**, Londrina, v. 1, n. 3, p. 213-220, jul./set.2001.

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

HAMBLIN, J. Plant breeding interpretations of the effects of bulk breeding on four populations of beans (*Phaseolus vulgaris* L.) **Euphytica**, Wageningen, v. 26, n.1, p. 157-168, Feb. 1977.

JENSEN, N. F. **Plant breeding methodology**. New York: J. Wiley & Sons, 1988. 676p.

KERVELLA, J.; FOUILLOUX, G. A theoretical study of bulk breeding method: importance and consequences of losses due to sampling. **Euphytica**, Wageningen, v. 60, n. 3, p. 185-195, Apr. 1992.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. Tokyo: McGraw-Hill, 1974. 556 p.

MORETO, A. L. ; RAMALHO, M. A. P. ; NUNES, J. A. R. ; ABREU, A. de F. B. Estimação dos componentes da variância fenotípica em feijoeiro utilizando o método genealógico. **Ciência e Agrotecnologia**, Lavras, v. 31, n. 4, p. 1035-1042, jul./ago. 2007.

MUEHLBAUER, F. J.; BURNELL, D. G.; BOGYO, M. T. Simulated comparisons of single descent and bulk populations breeding methods. **Crop Science**, Madison, v. 21, n. 4, p. 572-577, Jul./Aug. 2004.

NUNES, J. A. R. ; RAMALHO, M. A. P. ; FERREIRA, D. F. Inclusion of genetic relationship information in the pedigree selection method using mixed models. **Genetics and Molecular Biology**, v. 31, n. 3, p. 73-78, 2008.

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

PIROLA, L. H.; RAMALHO, M. A. P.; CARNEIRO, J. E. de. S.; ABREU, A. de F. B. Natural selection and family x location interaction in the common (dry) bean plant. **Genetics and Molecular Biology**, Ribeirão Preto, v. 25, n. 3, p. 343-347, set. 2002.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2006. Disponível em: ><http://www.r-project.org>>. Acesso em: 10 mar. 2008.

RAMALHO, M. A. P.; SANTOS, J. B. dos; ZIMMERMANN, M. J. de O. **Genética quantitativa em plantas autógamas**: aplicações ao melhoramento do feijoeiro. Goiânia: UFG, 1993. 271 p.

RAMALHO, M. A. P. Melhoramento do feijoeiro. In: SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 1997, Lavras. **Anais...** Lavras: UFLA, 1997. p. 167-196.

RAMALHO, M. A. P.; ABREU, A. F. B.; SANTOS, J. B. dos. Melhoramento de espécies autógamas. In: NASS, L. L.; VALOIS, A. C.; MELO, I. S. de.; VALADARES-INGLIS, M. C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, 2001. p. 201-230.

RAPOSO, F. C.; RAMALHO, M. A. P.; ABREU, A. de F. B. Comparação de métodos de condução de populações segregantes de feijoeiro. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 10, p. 1991-1997, out. 2000.

RASMUSSEN, D. C.; PHILLIPS, R. L. Plant breeding progress and genetic diversity from de novo variation and elevated epistasis. **Crop science**, Madison, v. 37, n. 2, p. 303-310, Mar./Apr. 1997.

SILVA, N. O.; RAMALHO, M. A. P.; ABREU, A. de F. B. CARNEIRO, J. E. de S. Performance of common bean families after different generations under natural selection. **Genetics and Molecular Biology**, Ribeirão Preto, v. 27, n. 4, p. 574-578, dez. 2004.

SOLIMAN, K. M.; ALLARD, R. W. Grain yield of composite cross populations of barley: effects of natural selection. **Crop Science**, Madison, v. 31, n. 3, p. 705-708, May./June 1991.

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

WANG, J.; WOLFGANG, H. P. Simulation modeling in plant breeding: Principles and applications. **Agricultural Sciences in China**, Beijing, v. 6, n. 8, p. 101-105, Aug. 2007.

WANG, J.; VANG 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**, Madison, v. 43, n. 6, p. 1764-1773, Nov./Dec. 2003.

WRICKE, G.; WEBER, W. E. **Quantitative genetics and selection in plant breeding**. Berlin: Walter de Gruyter, 1986. 395 p.

ANEXOS

ANEXO A		Página
TABELA 1A	Média, mínimo e máximo do número de alelos favoráveis em 1.000 simulações em famílias F _{4:5} com g = 100 locos d = a = 1, em função de n e h ²	50
TABELA 2A	Média, mínimo e máximo do número de alelos favoráveis em 1.000 simulações em famílias F _{4:5} com g = 100 locos d = 0,5, em função de n e h ²	51
TABELA 3A	Média, mínimo e máximo do número de alelos favoráveis em 1.,000 simulações em famílias F _{4:5} com g = 100 locos d = 0,0 em função de n e h ²	52
TABELA 4A	Número de locos favoráveis médios em função de g e n para h ² = 1,0 e d = a = 1 em famílias F _{4:5}	53

TABELA 1A Média, mínimo e máximo do número de alelos favoráveis em 1.000 simulações em famílias $F_{4;5}$ com $g=100$ locos $d= a = 1$ em função de n e h^2

Valores	n	h^2			
		0,1	0,5	0,8	1,0
Méd	50	100,47	100,88	100,91	100,72
	200	100,82	100,83	100,83	100,81
	1000	100,81	100,69	100,78	100,80
	2000	100,81	100,76	100,79	100,80
	5000	100,83	100,82	100,84	100,84
Mín	50	73,47	75,18	75,21	74,88
	200	70,30	70,36	70,36	70,16
	1000	65,42	65,49	65,50	65,52
	2000	64,85	65,52	65,61	66,04
	5000	64,66	65,19	65,29	65,34
Máx	50	116,28	126,18	126,18	125,96
	200	131,11	131,02	131,02	131,11
	1000	136,08	136,00	135,21	135,90
	2000	138,09	138,11	139,08	139,10
	5000	139,17	139,22	139,25	139,30

TABELA 2A Média, mínimo e máximo do número de alelos favoráveis em 1.000 simulações em famílias $F_{4;5}$ com $g=100$ locos $d=0,5$

Valores	n	h^2			
		0,1	0,5	0,8	1,0
Méd	50	100,82	100,81	100,81	100,80
	200	100,84	100,85	100,85	100,81
	1000	100,82	100,78	100,86	100,82
	2000	100,82	100,78	100,80	100,81
	5000	100,83	100,79	100,80	100,80
Mín	50	75,01	75,17	75,17	75,15
	200	70,48	70,35	70,35	70,38
	1000	65,38	65,50	72,72	65,72
	2000	64,82	64,65	64,68	64,65
	5000	64,11	64,09	64,49	64,66
Máx	50	126,42	126,08	126,08	126,13
	200	131,19	131,02	131,02	131,11
	1000	136,45	135,71	128,58	136,12
	2000	137,01	136,34	136,23	136,20
	5000	137,45	137,23	136,95	136,40

TABELA 3A Média, mínimo e máximo do número de alelos favoráveis em 1.000 simulações em famílias $F_{4;5}$ com $g=100$ locos $d = 0,0$

Valores	n	h^2			
		0,1	0,5	0,8	1,0
Méd	50	100,80	100,80	100,71	100,78
	200	100,88	100,88	100,88	100,81
	1000	100,83	100,80	100,82	100,03
	2000	100,85	100,81	100,84	100,76
	5000	100,85	100,82	100,83	100,84
Mín	50	75,36	75,13	75,21	75,39
	200	70,42	70,42	70,42	70,26
	1000	65,52	65,38	65,75	65,85
	2000	64,42	64,34	64,89	65,65
	5000	64,28	64,15	64,16	64,50
Máx	50	126,16	126,17	126,05	126,31
	200	131,27	131,27	131,27	131,09
	1000	136,08	135,86	135,03	137,28
	2000	137,95	136,56	136,87	137,29
	5000	138,54	138,43	137,96	137,77

TABELA 4A Número de locos favoráveis médios em função de g e n para $h^2 = 1,0$ e $d = a = 1$ em famílias $F_{4:5}$

Valores	n	g		
		5	50	200
Méd	50	1,98	23,60	97,40
	200	2,03	24,62	98,50
	1000	2,98	27,58	106,50
	2000	3,47	33,15	116,18
	5000	3,50	34,33	120,20
Min	50	1,00	15,00	66,00
	200	1,00	16,00	67,00
	1000	1,00	19,00	74,00
	2000	2,00	21,00	80,00
	5000	2,00	22,00	85,00
Max	50	3,00	31,00	120,00
	200	3,00	31,00	128,00
	1000	4,00	38,00	164,00
	2000	5,00	41,00	164,00
	5000	5,00	41,00	165,00

ANEXO B

Página

Programa utilizado para simulação de uma das situações estudadas,
linguagem R. 55

Programa para gerar simular Bulks, com g genes independentes, n indivíduos por geração, com efeito aditivo a=1,0 e d=a, d=0,5 e d=0 # sem epistasia e considerando fitness proporcional ao número de alelos favoráveis na sua constituição genotípica. Cada indivíduo deixará# em média 80 descendentes por geração, mas iremos considerar uma Poisson com esta média para isso

função auxiliar que recebe um indivíduo de dimensão g x 1 com valores 2, 1, 0,...,2 por exemplo

```
# retorna um descende aleatório a partir dele g x 1
gera filho = function(GFj li)
{
  g = length(GFj li)
  descendente = GFj li
  for (k in 1:g)
  if (GFj li[k]==1)
  {
    u = runif(1)
    if (u<=0.25) descendente[k]=2 else
    if ((u>0.25) & (u<=0.75)) descendente[k]=1 else
    descendente[k]=0
  }
  return(descendente)
}
```

Função GFj: gera indivíduos de um Bulk da geração Fj, considerando os indivíduos da geração j-1, n e g, j>= 2
o valor da Poisson gerado com média 80 será alterado pelo Fitness de forma proporcional, Fitness mínimo de 20% (default)

```
GeraFj = function(GFj_1,n,g,fitmin=0.2)
{
  GFj = matrix(0,n,g)
  numsem = matrix(rpois(n,80),n,1)
  fitness = fitmin + apply(GFj_1,1,sum)/(g*2/(1-fitmin))
  numsem = round(numsem*fitness)
  atualiza_nsac = function(numsem)
  {
    nsac = matrix(numsem[1],n,1)
    for (i in 2:n) nsac[i]=nsac[i-1]+numsem[i]
    nsac=nsac/nsac[n]
    return(nsac)
  }
}
```

```

}
nsac = atualiza_nsac(numsem)
for (ii in 1:n)
{
repeat
{
ind = length(nsac[nsac<runif(1)]) + 1
if (numsem[ind]>0) break
}
GFj[ii,]=gerafilho(GFj_1[ind,])
if (numsem[ind]>0) numsem[ind]=numsem[ind]-1
nsac = atualiza_nsac(numsem)
}
return(GFj)
}
# Função para gerar um experimento a partir de n plantas Fj, recebe GFj n x g
# retorna GFj_jplus1 de dimensão nrk x g, em que para cada planta GFj
# k linhas de cada repetição serão geradas e de cada planta GFj serão geradas r rep.
# Assim, a primeira progênie terá k linhas da primeira rept, as k seguintes da segunda e
# assim por diante até a r-ésima repetição (totalizando rk plantas), as rk seguintes serão
da progênie
# 2 e assim por diante até a n-ésima progênie (planta da GFj)

GeraEFj_jplus1 = function(GFj,n,g,r,k,a,d,h2)
{
if (r*k>80) stop ("Número de sementes insuficiente para realizar o experimento")
GFEjplus1 = matrix(0,n*r*k,g+3)
cont = 0
VGF2 = 0.5*g*a^2+0.25*g*d^2
VE = (1-h2)/h2*VGF2
for (ii in 1:n)
{
repeat
{
xx = gerafilho(GFj[ii,])
GFEjplus1[cont+1,1:g] = xx
GFEjplus1[cont+1,g+1] = length(xx[xx==0])*(-a)+length(xx[xx==2])*a+
length(xx[xx==1])*d # valor genotípico
GFEjplus1[cont+1,g+2] = rnorm(1,0,VE^0.5) # erro
GFEjplus1[cont+1,g+3] = 1800 + GFEjplus1[cont+1,g+1] +
GFEjplus1[cont+1,g+2] # valor fenotípico
cont = cont + 1
if (cont %% (r*k)==0) break
}
}
return(list(GFEjplus1=GFEjplus1, h2=h2,VGF2=VGF2, VE=VE))

```

```

}

# Função para gerar um experimento a partir de n plantas Fjplus1, recebe GFEj_jplus1
n*r*k x (g+3)
# retorna GFEj_jplus2 de dimensão nrk x (g+3), em que das r*k plantas da GFEj_jplus1
# k linhas de cada repetição serão geradas e de cada progenie GFEjplus1 serão geradas r
rep.
# Assim, a primeira progênie terá k linhas da primeira rept, as k seguintes da segunda e
# assim por diante até a r-ésima repetição (totalizando rk plantas), as rk seguintes serão
da progênie
# 2 e assim por diante até a n-ésima progênie (planta da GFEj_jplus1)

GeraEFj_jplus2 = function(GFEj_jplus1,n,g,r,k,a,d,h2,r_a,k_a)
{
  GFEjplus2 = matrix(0,n*r*k,g+3)
  cont = 0
  VGF2 = 0.5*g*a^2+0.25*g*d^2
  VE = (1-h2)/h2*VGF2
  for (ii in 1:n)
  {
    repeat
    {
      iii=(ii-1)*r_a*k_a + trunc(runif(1)*r_a*k_a)+1
      xx = gerafilho(GFEj_jplus1[iii,1:g])
      GFEjplus2[cont+1,1:g] = xx
      GFEjplus2[cont+1,g+1] = length(xx[xx==0])*(-a)+length(xx[xx==2])*a+
      length(xx[xx==1])*d # valor genotípico
      GFEjplus2[cont+1,g+2] = rnorm(1,0,VE^0.5) # erro
      GFEjplus2[cont+1,g+3] = 1800 + GFEjplus2[cont+1,g+1] +
      GFEjplus2[cont+1,g+2] # valor fenotípico
      cont = cont + 1
      if (cont %% (r*k)==0) break
    }
  }
  return(GFEjplus2)
}

# recebe qualquer resultado experimental Fj_jplus1 ou Fj_jplus2
# monta estrutura para obter as estatísticas descritivas
Calc_EstDesc = function(Fj_jplusk,n,g,r,k)
{
  prog = 1:n
  prog = matrix(rep(prog,each = r*k),n*r*k,1)
  re = 1:r
  re = rep(re,each = k)
  re = matrix(rep(re,times = n),n*r*k,1)
  NAF = apply(Fj_jplusk[,1:g],1,sum)
}

```

```

nlf <- function(xx)
{
return(length(xx)-length(xx[xx==0]))
}
NLF = apply(Fj_jplusk[,1:g],1,nlf)
dat.1 = data.frame(prog=prog,re=re,F = Fj_jplusk[,g+3],NAF=NAF,NLF=NLF)
return(dat.1)
}
# obtém estatísticas descritivas-componentes de variância, alelos favoráveis
Est_Desc = function(data,r,n)
{
y = matrix(t(tapply(data$F,list(data$prog,data$re),
mean)),n*r,1)
prog = 1:n
prog = matrix(rep(prog,each = r),n*r,1)
re = 1:r
re = matrix(rep(re,each = n),n*r,1)
prog = as.factor(prog)
re = as.factor(re)
anova = aov(y ~prog)
anova = anova(anova)
sig2p = (anova$"Mean Sq"[1]-anova$"Mean Sq"[2])/r
sig2 = anova$"Mean Sq"[2]
maxg = max(data$NAF)
ming = min(data$NAF)
medg = mean(data$NAF)
maxNLF= max(data$NLF) # maior número de locos favoráveis observado diretamente
data <- data[order(-data$F),]
maxFNLF <- data$NLF[1] # pega o # locos do maior fenótipo selecionado
return(list(sig2p=sig2p,sig2=sig2,maxg=maxg,
medg=medg, ming=ming,maxNLF=maxNLF,maxFNLF=maxFNLF))
}

n=50; g=100; a=1; h2=0.01 ; d = 1.0
k_a = 15; r_a = 3
k = 60; r = 3
ng = 4 # geração final do bulk, onde serão abertas famílias
NSM = 1000
fitmin=1.0
ii = 1
Result = matrix(0, NSM, 14) # vg, ve,ming,medg,maxg,maxNLF,maxFNLF
# primeiras 7 colunas para geração ng:ng+1
# últimas 7 colunas para geração ng:ng+2 (na mesma sequência)
while (ii<=NSM)
{

```

```

GFj=matrix(1,n,g) # geração F1
ger = 2
while (ger <= ng)
{
GFj=GeraFj(GFj,n,g,fitmin) # geração ger + 1
ger = ger +1
}

GFEger_gerplus1 = GeraEFj_jplus1(GFj,n,g,r_a,k_a,a,d,h2)
#exemplo se ng=4, GFE4_5
data1 = Calc_EstDesc(GFEger_gerplus1$GFEjplus1,n,g,r_a,k_a)
estdescFj_jplus1 = Est_Desc(data1,r_a,n)
Result[ii,1] = estdescFj_jplus1$sig2p
Result[ii,2] = estdescFj_jplus1$sig2
Result[ii,3] = estdescFj_jplus1$ming
Result[ii,4] = estdescFj_jplus1$medg
Result[ii,5] = estdescFj_jplus1$maxg
Result[ii,6] = estdescFj_jplus1$maxNLF
Result[ii,7] = estdescFj_jplus1$maxFNLF

GFEger_gerplus2 = GeraEFj_jplus2(GFEger_gerplus1$GFEjplus1,n,g,
r,k,a,d,h2,r_a,k_a)
#exemplo para ng=4, GFE4_6
data2 = Calc_EstDesc(GFEger_gerplus2,n,g,r,k)
estdescFj_jplus2 = Est_Desc(data2,r,n)
Result[ii,8] = estdescFj_jplus2$sig2p
Result[ii,9] = estdescFj_jplus2$sig2
Result[ii,10] = estdescFj_jplus2$ming
Result[ii,11] = estdescFj_jplus2$medg
Result[ii,12] = estdescFj_jplus2$maxg
Result[ii,13] = estdescFj_jplus2$maxNLF
Result[ii,14] = estdescFj_jplus2$maxFNLF
ii = ii + 1
}
#Result
apply(Result,2,mean)
apply(Result,2,max)
apply(Result,2,min)
save.image("config1jws.RData")

```