

CARLOS FELIPE BARRERA SÁNCHEZ

**SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES DERIVADAS DE
ACASALAMENTO AO ACASO OU DE AUTOFECONDAÇÃO**

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

VIÇOSA
MINAS GERAIS – BRASIL
2013

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

B272s
2013

Barrera Sánchez, Carlos Felipe, 1978-

Seleção genômica ampla em populações derivadas de
acasalamento ao acaso ou de autofecundação / Carlos Felipe
Barrera Sánchez. – Viçosa, MG, 2013.
x, 76f. : il. (algumas col.) ; 29cm.

Orientador: Cosme Damião Cruz

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

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

CDD 22. ed. 576.5

CARLOS FELIPE BARRERA SÁNCHEZ

**SELEÇÃO GENÔMICA AMPLA EM POPULAÇÕES DERIVADAS DE
ACASALAMENTO AO ACASO OU DE AUTOFECUNDAÇÃO**

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

APROVADA: 05 de março de 2013.

Marcos Deon Vilela de Resende

Felipe Lopes da Silva

Leonardo Lopes Bhering

Marciane da Silva Oliveira

Cosme Damião Cruz

AGRADECIMENTOS

À Deus, por sempre se fazer presente me iluminando todos os dias.

Aos meus pais, Magnolia e Carlos e meus irmãos Alexander e Luis, pelo apoio e compreensão em todos os momentos das minhas decisões.

Ao professor Cosme Damião Cruz, pela amizade, paciência, confiança e incentivo, um exemplo a ser seguido.

À Universidade Federal de Viçosa, pela oportunidade de realização de meus estudos.

Aos grandes amigos Maria, Digner, Carolina, Geovanny, Katherine, Karen e Mercedes pela grande amizade, companheirismo, paciência e apoio.

Aos amigos de república Caio, Rafael, Gustavo, Dalcionei e Fábio pela convivência e companheirismo.

A Maria Fernanda pelo apoio e paciência na parte final do trabalho.

Aos amigos do laboratório de Bioinformática, Caio, Marciane, Gislayne, Moisés, Leonardo, Isís, Rafael, Danielle e David pelo convívio agradável durante a realização deste curso.

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

Aos professores de graduação e de pós-graduação, pela atenção, pela disponibilidade e pelos ensinamentos.

Muchas gracias!!

Meus sinceros agradecimentos!!

SUMÁRIO

RESUMO.....	vii	
ABSTRACT.....	ix	
INTRODUÇÃO GERAL.....	1	
CAPÍTULO 1		
DESEQUILÍBRIO DE LIGÃO (LINKAGE DISEQUILIBRIUM – LD) EM POPULAÇÕES DE ACASALAMENTO AO ACASO E AUTOFECUNDAÇÃO		
Resumo.....	5	
Abstract.....	6	
Introdução.....	7	
Material de Métodos.....	10	
<i>Estudo do desequilíbrio de ligação na população F_2 e suas descendências.....</i>	11	
<i>Estudo do desequilíbrio de ligação na população P, de origem desconhecida, e suas descendências.....</i>	12	
<i>Análise populacional.....</i>	13	
Resultados e Discussão.....	16	
<i>População F_2 – LD e mapa de ligação.....</i>	17	
<i>Populações avançadas por autofecundação da F_2 (F_{2s1} a F_{2s5}).....</i>	18	
<i>Populações avançadas por acasalamento ao acaso da F_2 (F_{2a1} a F_{2a5}).....</i>	22	
<i>População P de origem desconhecida.....</i>	25	
<i>Populações avançadas por autofecundação da P (P_{s1} a P_{s5}).....</i>	26	
<i>Populações avançadas por acasalamento ao acaso da P (P_{a1} a P_{a5}).....</i>	27	
Conclusões.....	29	
Referências.....	30	
CAPÍTULO II		
USO DE SELEÇÃO GENOMICA AMPLA EM POPULAÇÕES SIMULADAS E SUAS GERAÇÕES AVANÇADAS POR AUTOFECUNDAÇÃO E ACASALAMENTO AO ACASO.....		35
Resumo.....	36	
Abstract.....	38	
Introdução.....	40	
Material de Métodos.....	45	
<i>Origens das populações.....</i>	45	

<i>Simulação dos dados genotípicos e fenotípicos</i>	46
<i>Caracterização genética das populações simuladas</i>	47
<i>Metodologia de análise</i>	48
<i>Avaliação dos dados</i>	50
<i>Aplicativos computacionais para análise de dados</i>	51
Resultados e Discussão.....	51
<i>Dinâmica das populações avançadas por autofecundação e acasalamento ao acaso</i>	52
<i>Valor genético genômico (VGG)</i>	60
<i>Estimativas de confiabilidade e correlação</i>	66
Conclusões.....	73
Referências	74

RESUMO

BARRERA SANCHEZ, Carlos Felipe, D.Sc. Universidade Federal de Viçosa, Março de 2013. **Seleção genômica ampla em populações derivadas de acasalamento ao acaso ou de autofecundação.** Orientador: Cosme Damião Cruz. Coorientadores: Marcos Deon Vilela de Resende e Leonardo Lopes Bhering.

O homem depende das plantas para sua sobrevivência, a maneira mais econômica e sustentável de se aumentar a produtividade é através da obtenção de cultivares com maior potencial de produção. É neste ponto que o melhoramento de plantas atua, este é, em geral, uma atividade que pode levar muitos anos. Por isso, o melhorista deve tentar prever necessidades futuras e desenvolver metodologias para a redução desses longos períodos. A biologia molecular está disponibilizando ferramentas que podem eliminar essas dificuldades, e potencializar os processos seletivos, e obter grande rapidez na obtenção de ganhos genéticos com a seleção e baixo custo, em comparação com a tradicional seleção baseada em dados fenotípicos. A seleção genômica tornou-se uma ferramenta importante no melhoramento genético de animais e plantas o que pode permitir melhores acurácias e seleção precoce. O objetivo deste trabalho foi, em um primeiro momento, estudar o desequilíbrio de ligação (LD) em uma população F_2 com mapa genético previamente conhecido e a causa do LD determinada pela ligação fatorial. Em uma segunda etapa, estudou-se outra população em que o LD era determinado pelo acasalamento direcionado.

Foram simulados dados de indivíduos e de informações moleculares das populações F_2 , derivadas de genitores homozigotos contrastantes. As populações geradas foram de tamanho de 500 indivíduos e considerou-se cinco grupos de ligação com três cenários de saturação equivalentes a 11, 31 ou 51 marcas moleculares codominantes por grupo de ligação. As populações foram submetidas a cinco gerações de autofecundação e acasalamento ao acaso. Para populações conhecidas o genoma da espécie, o tipo de sistema de acasalamento afetaram de forma diferencial as taxas de desequilíbrio entre pares de locos, enquanto, para populações de origem desconhecidas, as causas do desequilíbrio não são diretamente relacionadas com a ligação fatorial e as taxas são preservadas de forma similar à F_2 com as sucessivas gerações de acasalamento ao acaso ou autofecundação.

Na segunda etapa do trabalho foram simuladas populações com estrutura populacional apresentando dados fenotípicos e genotípicos de cada indivíduo dentro da população, imitando alguns dos cenários em que a seleção genômica ampla (GWS) é aplicada. Se considerarem 1500 locos e tamanho da população de 1000 indivíduos para todas as gerações originadas por acasalamento ao acaso e autofecundação. Foram simuladas três características, em cada característica, o número de locos que controlava o caráter foi estabelecido em 500 e herdabilidades 20%, 40% e 60%, com base nos valores genotípicos e fenotípicos das populações em todas as gerações, foram estimados os parâmetros genéticos. Além da estimação de parâmetros, as populações foram caracterizadas quanto a distribuição de seus dados sendo feitos testes de normalidade e obtenção de coeficientes de simetria e curtose nas diversas gerações. Após estabelecidos todos os parâmetros genéticos se avalio a correlação dos valores fenotípicos observados com os valores fenotípicos preditos via informação de marcadores e a acuracia de seleção. A simulação utilizada foi eficaz em preservar a estrutura genética das populações e descrever a sua dinâmica ao longo de sucessivas gerações de acasalamento ao acaso ou autofecundação, os resultados amostram que o sistema de acasalamento afeta a eficiência do uso da utilização das estimativas dos efeitos dos marcadores em gerações avançadas.

ABSTRACT

BARRERA SANCHEZ, Carlos Felipe, D.Sc. Universidade Federal de Viçosa, March, 2013. **Genomic analysis for selection in populations derived from random mating or selfing.** Adviser: Cosme Damião Cruz. Co-Advisers: Marcos Deon Vilela de Resende e Leonardo Lopes Bhering

Man depends on plants for their survival, the most economical and sustainable way to increase productivity is by obtaining cultivars with higher yield potential. This is where the plant breeding works, this is usually an activity that can take many years. Therefore, the breeder should try to anticipate future needs and develop methodologies to reduce these long periods. Molecular biology is providing tools that can eliminate these problems and strengthen the selection process, very quickly in obtaining genetic gains with selection and low cost compared with traditional selection based on phenotypic data. The genomic selection has become an important tool in genetic improvement of animals and plants which can provide better accuracies and early selection. The objective of this work was, at first, to study the Linkage disequilibrium (LD) in an F_2 population with genetic map previously known and the cause of LD determined by the link factorial. In a second step, we studied another population where LD was determined by mating directed.

Were simulated data of individuals and molecular information of F_2 populations derived from contrasting homozygous parents. The populations were generated size of 500 individuals and it is considered five linkage groups with three scenarios saturation equivalent to 11, 31 or 51 codominant molecular markers by linkage group. The populations were subjected to five generations of selfing and random mating. For the genome known populations of the species, the type of mating system affected differentially rates disequilibrium between pairs of loci, whereas for populations of unknown origin, the causes of imbalance are not directly related to the link factor and rates are preserved similarly to F_2 with successive generations of random mating or selfing.

In the second stage of labor were simulated populations with population structure presenting phenotypic and genotypic data of each individual within the population, imitating some of the scenarios where genome wide selection (GWS) is applied. Were considered 1500 locus and population size of 1000 individuals for all generation originated by random mating and selfing. We simulated three features in

each feature, the number of loci that controlled the character was established in 500 and heritability 20%, 40% and 60%, based on genotypic and phenotypic values of the people in all generations, parameters were estimated genetic. Besides the parameter estimation, the populations were characterized as the distribution of your data being made normality tests and obtaining coefficients of symmetry and kurtosis in several generations. After all genetic parameters set, were evaluate the correlation of phenotypic values observed with the phenotypic values predicted using markers information and acuáracia selection. The simulation used was effective in preserving the genetic structure of populations and describe its dynamics over successive generations of random mating or selfing, results shows the mating system affects the efficiency of the use of the use of estimates of the effects of the markers in advanced generations.

INTRODUÇÃO GERAL

As plantas têm uma grande importância para a humanidade. Utilizamos as plantas diretamente na nossa alimentação e indiretamente para alimentação de animais que fornecerão alimento. Além disso, as plantas também são utilizadas como vestimenta, energia, habitação, ornamentação e medicamento. Podemos afirmar que o homem depende das plantas para sua sobrevivência. Projeções da ONU mostram que a população mundial deve alcançar 8,2 bilhões de pessoas em 2030 e 9,1 bilhões de pessoas em 2050. Esse crescimento populacional deve ser concentrado principalmente nos países menos desenvolvidos da Ásia, África e América Latina. Esses dados mostram um grande desafio que as nações, principalmente as menos desenvolvidas, terão de enfrentar nos próximos anos: aumentar a produção de alimentos para que não haja fome no mundo. Existem duas maneiras de aumentar a produção de alimentos: aumento da área cultivada pela incorporação de novas áreas ou aumento da produtividade.

A maneira mais econômica e sustentável de se aumentar a produtividade é através da obtenção de cultivares com maior potencial de produção/produtividade. É neste ponto que o melhoramento de plantas atua. Para Vavilov, o melhoramento de plantas é a “Evolução direcionada pela vontade do homem”, pois, o homem utiliza no melhoramento de plantas os mesmos mecanismos que a natureza utiliza para a evolução das espécies. O melhoramento de plantas é, em geral, uma atividade que pode levar muitos anos. A produção de novas cultivares em espécies anuais leva em média 12 anos, enquanto para cultivares perenes esse tempo pode ser de 15-20 anos. Por isso, o

melhorista deve tentar prever necessidades futuras e desenvolver metodologias para a redução desses longos períodos.

O melhoramento genético, ao selecionar variedades mais produtivas, de melhor qualidade industrial, mais tolerantes a estresses e com melhor adaptação ecológica, possibilita aumentar os rendimentos agrícolas e até mesmo reduzir o uso de insumos pelo agricultor, o que ajuda a preservar a saúde humana e o meio ambiente. Porém alguns caracteres agronômicos, especialmente os de herança quantitativa, apresentam dificuldades na seleção fenotípica, tanto na escolha dos pais como na seleção em populações segregantes. A biologia molecular está disponibilizando ferramentas que podem eliminar essas dificuldades. A partir de pesquisas de biologia molecular e genômica, foram identificados marcadores genéticos com potencial de aplicação na localização de regiões genômicas que controlam características de interesse (QTLs).

Estes marcadores podem ser utilizados para elucidar a arquitetura genética de caracteres complexos em plantas via mapeamento genético e detecção de QTLs. Sua aplicação foi vislumbrada para auxiliar os procedimentos de seleção no melhoramento convencional, o que foi chamado de melhoramento genômico. Seu aproveitamento foi vislumbrado para aperfeiçoar os processos seletivos, grande rapidez na obtenção de ganhos genéticos com a seleção e baixo custo, em comparação com a tradicional seleção baseada em dados fenotípicos. Visando a esses objetivos, Meuwissen et al. (2001) propuseram um novo método de seleção denominado seleção genômica (GS) ou seleção genômica ampla (*genome wide selection* – GWS), a qual pode ser aplicada em todas as famílias em avaliação nos programas de melhoramento genético de espécies alógamas e autógamias. A GWS, é definida como a seleção simultânea para centenas ou milhares de marcadores, os quais cobrem o genoma de maneira densa, de forma que todos os genes

de um caráter quantitativo estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores (Resende et al., 2008).

Uma vez gerado um grande número de marcadores espalhados por todo o genoma de um indivíduo, alguns destes marcadores estarão muito perto do QTL e em desequilíbrio de ligação (LD) com este (Hastbacka et. al., 1994). O conceito de desequilíbrio de ligação refere-se à associação não aleatória entre dois genes ou entre um QTL e um loco marcador. Quando as frequências alélicas e genotípicas de um ou mais locos autossômicos são constantes de uma geração para a outra e as frequências genotípicas são determinadas pelas frequências alélicas, diz-se que este loco se encontra em equilíbrio de ligação. Com a ligação gênica, dois genes ligados apresentam uma associação que não se dá ao acaso e estão em desequilíbrio de ligação. Com os eventos de recombinação, a cada nova geração, os locos tendem ao equilíbrio, e o tamanho de um dado segmento cromossômico que contém dois locos quaisquer e que não sofreu recombinação diminui, o que conseqüentemente reduz o LD.

O desenvolvimento teórico da GWS coincide com a tecnologia Single Nucleotide Polymorphisms (SNP), a qual é acurada e relativamente barata. A GWS usa associações de um grande número de marcadores SNPs em todo o genoma com os fenótipos, capitalizando no desequilíbrio de ligação entre os marcadores e QTLs proximamente ligados, sem uma prévia escolha de marcadores com base nas significâncias de suas associações com o fenótipo. Predições são então obtidas para os efeitos dos haplótipos marcadores ou dos alelos em cada marcador. Essas predições derivadas de dados fenotípicos e de genótipos SNPs em alta densidade em uma geração são então usadas para obtenção dos valores genéticos genômicos dos indivíduos de qualquer geração subsequente, tendo por base os seus próprios genótipos marcadores (Resende et al., 2008).

O objetivo deste trabalho foi, em um primeiro momento estudar o desequilíbrio de ligação (*linkage disequilibrium*) e os fatores que o influenciam. Em uma segunda etapa, fornecer subsídios para melhor entender a seleção genômica ampla, técnica que permite a seleção de indivíduos baseado apenas na informação dos seus marcadores.

CAPÍTULO 1

DESEQUILÍBRIO DE LIGAÇÃO (LINKAGE DISEQUILIBRIUM - LD) EM POPULAÇÕES DE ACASALAMENTO AO ACASO E AUTOFECUNDAÇÃO

Resumo - Os dois fenômenos envolvidos com a geração de polimorfismos de DNA, detectados por marcadores moleculares, são a mutação e a recombinação. Métodos são usados para estudar o fenômeno da ligação e recombinação, e o histórico de mutações da população. Um método é o desequilíbrio de ligação (LD) baseado na análise de associação, que tem recebido nos últimos anos uma atenção especial dos geneticistas de plantas. O objetivo deste trabalho foi, em um primeiro momento, estudar o LD em uma população F_2 com mapa genético previamente conhecido e a causa do LD determinada pela ligação fatorial. Num segundo momento, estudou-se outra população em que o LD era determinado pelo acasalamento direcionado. Foram simulados dados de indivíduos e de informações moleculares das populações F_2 , derivadas de genitores homocigotos contrastantes. As populações geradas foram de tamanho de 500 indivíduos e considerou-se cinco grupos de ligação com três cenários de saturação equivalentes a 11, 31 ou 51 marcas moleculares codominantes por grupo de ligação. As populações foram submetidas a cinco gerações de autofecundação e acasalamento ao acaso. Após a simulação foram construídos mapas genéticos de ligação fatorial e de desequilíbrio. Ademais foi gerada uma população P de origem desconhecida derivada de duas outras populações genitoras em equilíbrio de Hardy-Weinberg, estabelecida por 500 indivíduos de cada população e informações moleculares relativas a 55, 155 ou 255 locos. Para populações conhecidas o genoma da espécie, o tipo de sistema de acasalamento afetaram de forma diferencial as taxas de desequilíbrio entre pares de locos, enquanto, para populações de origem desconhecidas, as causas do desequilíbrio não são diretamente relacionadas com a ligação fatorial e as taxas são preservadas de forma similar à F_2 com as sucessivas gerações de acasalamento ao acaso ou autofecundação.

Termos para indexação: genômica; simulação; desequilíbrio de ligação; seleção genômica ampla; ligação fatorial.

Abstract- The two phenomena involved in the generation of DNA polymorphisms, detected by molecular markers are mutation and recombination. Methods are used to study the phenomenon of link and recombination, and mutation of population history. One method is the linkage disequilibrium (LD) - based association analysis, which in recent years has received special attention from plant breeders. The aim of this work was, at first, to study the LD in an F2 population with genetic map previously known and the cause of LD determined by the link factor. Secondly, were studied another population where LD was determined by mating directed. Were simulated data of individuals and molecular information of F2 populations derived from contrasting homozygous parents. The populations were generated size of 500 individuals and it is considered five linkage groups with three scenarios saturation equivalent to 11, 31 or 51 codominant molecular markers by linkage group. The populations were subjected to five generations of selfing and random mating. After the simulation were constructed genetic maps binding factor and imbalance. Moreover has been generated a population P of unknown origin derived from two other progenitor populations in Hardy-Weinberg equilibrium is established for each population of 500 individuals and molecular information regarding 55, 155 or 255 loci. For the genome known populations of the species, the type of mating system affected differentially rates disequilibrium between pairs of loci, whereas for populations of unknown origin, the causes of imbalance are not directly related to the binding factor and rates are preserved similarly to F2 with successive generations of random mating or selfing.

Index terms: genomics; simulation; linkage disequilibrium; wide selection genomics; binding factor

Introdução

O século XX foi marcado por grandes avanços na genética feitos na compreensão de como genes individuais controlam características simples. No entanto, os frutos da revolução em genética molecular provavelmente serão visto neste século, quando os genes e os alelos que controlam características complexas [locos de características quantitativas (QTL)] serão identificados e compreendidos (Flint-Garcia et al., 2003). O desenvolvimento e a utilização de marcadores moleculares para a detecção e exploração de polimorfismos de DNA em plantas e animais é um dos desenvolvimentos mais significativos na área da biologia molecular e biotecnologia (Gupta et al., 2005), mas o seu uso no entendimento da expressão dos genes e sua aplicação no melhoramento ainda faltam por ser pesquisada. A geração de polimorfismos de DNA envolve dois fenômenos detectados por marcadores moleculares, são a mutação e a recombinação. A detecção de ligação e o histórico do polimorfismo de DNA tem sido o centro de diversos estudos com marcadores moleculares (Terwilliger & Weiss, 1998; Nordborg & Tavaré, 2002; Gupta & Rustgi, 2004; Gupta et al., 2005; Abdurakhmonov & Abdurakarimov, 2008; Resende, 2008). Entretanto, para o estudo de ligação é preciso realizar cruzamentos apropriados antes do desenvolvimento do mapeamento das populações. Esta será, em alguns casos, a limitação ao uso de marcadores moleculares, pois os cruzamentos poderiam ser de difícil realização, como nas espécies florestais, e/ou a população de mapeamento a ser utilizada pode ser muito pequena, com apenas dois alelos por loco amostrados (Gupta et al., 2005).

Neste sentido, métodos alternativos têm sido desenvolvidos e usados para estudar o fenômeno da ligação e recombinação e o histórico de mutações da população.

Um método é o desequilíbrio de ligação (LD) baseado na análise de associação, que tem recebido nos últimos anos, atenção especial dos geneticistas de plantas (Meuwissen & Goddard, 2000). O mapeamento por associação, também conhecido como mapeamento por desequilíbrio de ligação, ou mapeamento por desequilíbrio de fase gamética, baseia-se em conceitos de genética de populações para identificar relações entre marcadores genéticos e caracteres fenotípicos (Wray & Visscher, 2008).

Historicamente, a análise de ligação foi usada para mensurar a proximidade genética entre um loco com outros, para mapas de genes controladores de caracteres qualitativos (Palaisa et al., 2003, 2004). Em plantas, a maioria destas análises de co-segregação tem sido conduzidas em populações altamente estruturadas com pedigree conhecido, como populações F_2 . Entretanto, estas populações têm duas grandes limitações; a primeira, o número limitado de eventos de recombinação, consequência da pobre resolução de caracteres quantitativos; a segunda é o fato de apenas dois alelos de qualquer loco pode ser estudado simultaneamente.

O desequilíbrio de ligação (LD) é a associação não ao acaso de alelos em diferentes locos. É a correlação entre polimorfismos que é causada pela sua história compartilhada de mutação e recombinação. Inicialmente, o desequilíbrio de ligação está presente na população em uma taxa determinada pela distância genética entre os dois locos e o número de gerações desde que ele surgiu (Flint-Garcia et al., 2003). Diferentes metodologias para estimar os níveis de LD têm sido amplamente descritas em plantas (Flint-Garcia et al., 2003; Gaut & Long, 2003; Gupta et al., 2005; Simko et al., 2004). Por este método, certos alelos de um loco marcador estão associados com alelos particulares em outro loco ligado, afetando uma característica de interesse. Em algumas revisões são descritos os métodos disponíveis, a estatística utilizada para se testar a significância das medidas obtidas e estimativas obtidas envolvendo locos multialélicos e

condições multilocos. As duas estatísticas mais usadas para mensurar o desequilíbrio de ligação são r^2 e D' (Gupta et al., 2005; Jorde, 2000; Liang et al., 2001; Gorelick & Laubichler, 2004; Flint garcia et al., 2003).

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

Uma das maiores utilizações atuais e futuras do LD em plantas, provavelmente será no estudo da associação marcador - característica o qual é geralmente feita por análise de ligação, utilizando análise de regressão simples e mapeamento de QTL por intervalo. Algumas limitações deste método têm sido superadas com o mapeamento de associação baseado no LD. Para o estudo de associação característica – marcador usando o LD, as metodologias diferem para caracteres quantitativos e caracteres discretos (qualitativos), embora ocasionalmente caracteres quantitativos possam ser tratados como caracteres qualitativos. Como exemplo de estudo envolvendo característica qualitativa em plantas foi conduzido um trabalho em milho (Palaisa et al.,

2003) para procurar associação entre polimorfismo do gene *y1* com coloração do endosperma. Outro estudo foi conduzido por Kumar et al., 2004, onde 200 famílias de irmãos completos foi usada para estudar associação marcador – característica. O uso do LD para mapeamento de QTL para caracteres quantitativos é mais laborioso, porém é também mais recompensador, por que ele permite localizar a posição do QTL que controla a característica de interesse de forma mais precisa. Mackay (2001) e Glazier et al. (2002) sugerem que se utilize análise de ligação para uma localização preliminar do QTL e depois usa-se o LD para uma localização mais precisa. Outras utilizações importantes são no estudo de diversidade genética em populações naturais, em coleções de germoplasma, o uso em genética de populações, e em programas de melhoramento de plantas implementando seleção genômica ampla (Gupta et al., 2005).

O objetivo deste trabalho foi estudar o LD, considerando vários locos gênicos, com dois tipos de populações. Para isso, foi feita abordagem sobre o LD em uma população F_2 com mapa genético previamente conhecido e a causa do LD determinada unicamente pela ligação fatorial. Numa segunda situação conforma-se outra população em que o LD era determinado pelo acasalamento direcionado entre duas outras populações em equilíbrio de Hardy – Weinberg. Também foi objetivo avaliar a mudança das taxas de LD ocorridas nas populações após sucessivas gerações de acasalamento ao acaso ou autofecundação.

Material e Métodos

Neste trabalho foi realizado um estudo sobre desequilíbrio de ligação em dois cenários relativos à população inicial. No primeiro considerou uma população F_2 , obtida de genitores homocigotos contrastantes, e suas gerações derivadas por autofecundação e acasalamento ao acaso. No segundo, considerou uma população referencial cujo

desequilíbrio fosse estabelecido em consequência de ser derivada da hibridação de duas outras populações, genitoras, simuladas em equilíbrio de Hardy-Weinberg. A partir desta população, aqui denominada de população desconhecida, foram geradas novas populações por acasalamento ao acaso e autofecundação. Dessa forma, todos os alelos de cada loco, na população F_2 terão frequências iguais a meio e a causa do desequilíbrio será determinada unicamente pela ligação fatorial. Para a população desconhecida e suas gerações derivadas o desequilíbrio de ligação será proporcionado pela variação nas frequências gaméticas e as frequências alélicas serão variadas para cada loco.

Estudo do desequilíbrio de ligação na população F_2 e suas descendências

Foram simulados dados de indivíduos e de informações moleculares das populações F_2 , derivadas de genitores homocigotos contrastantes. No processo de obtenção dos dados foi utilizado o módulo de simulação do programa para análise de dados moleculares e quantitativos GQMOL (Cruz, 2012). São encontrados trabalhos que reportam tanto para cruzamentos controlados quanto para populações exogâmicas, populações de 400 indivíduos recuperam as informações genômicas para fins de mapeamento genético (Bhering et al., 2008). Pelo mencionado, as populações geradas foram de tamanho de 500 indivíduos. Para avaliação do desequilíbrio, foram considerados cinco grupos de ligação com três cenários de saturação equivalentes a 11, 31 e 51 marcas moleculares codominantes, em cada grupo de ligação, espaçadas a 10 cM.

As populações geradas a partir dos genomas F_2 , foram submetidas a cinco gerações de autofecundações gerando as populações F_{2s1} a F_{2s5} , e a cinco gerações de acasalamento ao acaso, gerando as populações F_{2a1} a F_{2a5} . Medidas de desequilíbrio de ligação e de porcentagem de recombinação entre pares de locos foram obtidas e

comparadas. Se um mesmo número de plantas F_{2s1} for obtido da autofecundação de cada planta F_2 , ou um mesmo número de plantas F_{2s2} for obtido de cada planta F_{2s1} , e assim por diante, essas populações têm estrutura genética previsível e, portanto, podem ser utilizadas para mapeamento genético.

Assim, após a simulação das populações F_2 e F_{2s1} a F_{2s5} foram construídos mapas genéticos de ligação fatorial e de desequilíbrio empregando a propriedade transitiva, ou seja, se o loco A está ligado ao loco B, e o loco B está ligado ao loco C, logo o loco A está ligado ao loco C, independente da frequência de recombinação de A e C e, portanto, A, B e C pertencem ao mesmo grupo de ligação. Os critérios de agrupamento foram frequência máxima de recombinação (r_{\max}) e o LOD mínimo (LOD_{\min}), para inferir se dois locos estão ligados. Foram utilizados os valores 30% e 3, respectivamente, para r_{\max} e LOD_{\min} . A conversão da medida de desequilíbrio entre pares de marcadores ($U_{ii'}$) em distância genética ($d_{ii'}$) foi dada por $d_{ii'} = 50(1 - U_{ii'}/U_{\max})$.

Estudo do desequilíbrio de ligação na população P, de origem desconhecida, e suas descendências

Foi inicialmente gerada a população P, derivada de duas outras populações genitoras em equilíbrio de Hardy-Weinberg. Esta população foi submetida a cinco gerações de autofecundações gerando as populações P_{s1} a P_{s5} , e a cinco gerações de acasalamento ao acaso, gerando as populações P_{a1} a P_{a5} . Foram simulados dados de 500 indivíduos de cada população e estabelecidas informações moleculares relativas a 55, 155 ou 255 locos que expressam dois alelos co-dominantes.

Análise populacional

As frequências genotípicas são determinadas por forças evolutivas e por sistemas de acasalamentos. Acasalamento ao acaso conduz ao equilíbrio, enquanto a autofecundação provoca variação com acréscimo da frequência de homozigoto em detrimento dos heterozigotos. As frequências alélicas não devem mudar de uma geração para a outra em consequência do tipo de acasalamento, mas em virtude da ação de forças evolutivas sistemáticas e dispersivas, sendo as mais importantes a mutação, a migração, a seleção natural e a deriva genética. Considerando tais fatos, o número de locos em equilíbrio, em cada população analisada, foi avaliado pela estatística qui-quadrado confrontando os resultados observados com os esperados iguais a p^2 , $2pq$ e q^2 para AA, Aa e aa, respectivamente.

Utilizando o principio de Equilíbrio de Hardy-Weinberg (EHW), é possível prever a descendência resultante do acasalamento ao acaso considerando a população como um todo, em vez de particularizar os cruzamentos individuais. Os cruzamentos aleatórios de genótipos são equivalentes à união aleatória de gametas, onde as frequências genotípicas de AA, Aa e aa na geração parental são escritas como D , H , R .

Desequilíbrio de ligação: Para o cálculo do desequilíbrio de ligação considerou-se dois locos, com dois alelos cada, ou seja, no loco 1, A e a , e no loco 2, B e b . Os gametas produzidos pela população, na geração 1 tomada como referência, são dados por:

Gameta	Frequência
AB	P_{AB}
Ab	P_{Ab}
aB	P_{aB}
ab	P_{ab}

Estes gametas são de dois tipos, os gametas AB e ab são gametas não recombinantes, porque os alelos estão associados da mesma maneira que na geração

anterior, e os gametas Ab e aB são recombinantes, porque os alelos estão associados de modo diferente da geração anterior. O desequilíbrio de fase gamética é dado pela diferença destes dois tipos de gametas, ou seja, a frequência dos gametas não recombinantes menos a frequência dos gametas recombinantes, que nesta população, quantificado por meio de:

$$D = P_{AB} P_{ab} - P_{Ab} P_{aB}$$

Dessa forma a frequência alélica é influenciada pelo desequilíbrio de ligação da geração, ficando:

$$P_{AB(n)} = p_A p_B + D_n$$

$$P_{Ab(n)} = p_A q_b - D_n$$

$$P_{aB(n)} = q_a p_B - D_n$$

$$P_{ab(n)} = q_a q_b + D_n$$

$$\text{Sendo: } P_{AB} + P_{Ab} + P_{aB} + P_{ab} = 1$$

A taxa de aproximação do equilíbrio (D) é diretamente proporcional à frequência de recombinação (r) entre os genes, como mostra a tabela a seguir. Diante disto, para quantificar a aproximação do equilíbrio de ligação é necessário estimar as frequências alélicas na próxima geração.

Gametas	Frequência de ocorrência	
AB	$(1-r)/2$	p_{AB}
ab	$(1-r)/2$	p_{ab}
Ab	$r/2$	p_{Ab}
aB	$r/2$	p_{aB}

As frequências genótípicas esperadas, em relação a dois genes, foram dadas conforma apresentado a seguir:

Genótipos	Frequência esperadas	Nº observado
AABB	$p_1 = p_A^2 p_B^2 + p_A p_B D + D^2$	n_1
AABb	$p_2 = 2p_A^2 p_B q_b + 2p_A (q_b - p_B) D - 2D^2$	n_2
AAbb	$p_3 = p_A^2 q_b^2 - 2p_A q_b D + D^2$	n_3
AaBB	$p_4 = 2p_A q_a p_B^2 + 2(q_a - p_A) p_B D - 2D^2$	n_4
AaBb	$p_5 = 4p_A q_a p_B q_b - 2(q_a - p_A) (p_B - q_b) D + 4D^2$	n_5
Aabb	$p_6 = 2p_A q_a q_b^2 - 2(q_a - p_A) p_B D - 2D^2$	n_6
aaBB	$p_7 = q_a^2 p_B^2 - 2q_a p_B D + D^2$	n_7
aaBb	$p_8 = 2q_a p_B q_b - 2p_a (p_B - q_b) D - 2D^2$	n_8
aabb	$p_9 = q_a^2 q_b^2 + 2q_a q_b D + D^2$	n_9

Uma vez que os valores de p_A , q_a , p_B e q_b são obtidos é possível estimar o valor de D pelo método da máxima verossimilhança, admitindo que o número de ocorrência das classes genótípicas segue distribuição multinomial, a função de verossimilhança é utilizada é descrita a seguir:.

$$L(p_A, q_a, p_B, q_b, D; n_i) = \frac{N!}{n_1! n_2! \dots n_9!} p_1^{n_1} p_2^{n_2} \dots p_9^{n_9}$$

Em que p_1, p_2, \dots, p_9 são as frequências observadas das classes genótípicas.

O estimador de verossimilhança de D é dado pela derivada primeira em relação a D e igualada a zero da função $L(p_A, q_a, p_B, q_b, D; n_i)$, sendo:

$$\frac{\partial L(p_A, q_a, p_B, q_b, D; n_i)}{\partial D}$$

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

$$r^2 = \frac{(D)^2}{p_A q_a p_B q_b}$$

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

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

$$|D'| = \frac{D^2}{\min(p_A q_b, q_a p_B)} \quad D < 0$$

$$|D'| = \frac{D^2}{\min(p_A p_B, q_a q_b)} \quad D > 0$$

A estatística D' é baseada nas frequências observadas, e irá variar entre 0 e 1 se as frequências alélicas diferirem entre os locos. D' poderá ser menor do que 1 apenas se todos os quatro possíveis gametas foram observados, conseqüentemente assumindo que eventos de recombinação ocorreram entre os locos. Sempre que uma das quatro frequências haplotípicas for zero, D' será igual a 1, o que ocorre com frequência, quando se trabalha com populações pequenas (Wray & Visscher, 2008).

Foram apresentadas formas de visualização da extensão do desequilíbrio de ligação entre pares de locos. Os gráficos de declínio do desequilíbrio de ligação são usados para visualizar a razão na qual o desequilíbrio de ligação diminui em função da distância genética ou física, nas sucessivas gerações de acasalamento ou autofecundação. São construídos gráficos de dispersão (scattered plots) dos valores r^2 versus a distância genética ou física entre todos os pares de alelos.

Resultados e Discussão

Dois fenômenos tem sido de grande importância nos estudos genômicos aplicados ao melhoramento genético. O primeiro diz respeito à ligação fatorial e o

segundo ao desequilíbrio de ligação ou desequilíbrio de fase gamética. Eles podem estar intimamente relacionados, em certas condições populacionais. Entretanto, genes ligados que se encontram em grande desequilíbrio numa população F_2 podem perder a condição de desequilíbrio com o avanço de sucessivas gerações de acasalamento ao acaso. Por outro lado, genes localizados em cromossomos diferentes podem apresentar considerável desequilíbrio de fase gamética em certas populações avançadas.

Como os dois fenômenos são mais ou menos relacionados, em certas condições referentes ao tipo de população e o sistema de acasalamento, há o interesse de se estabelecer, para fins práticos, mapas de ligação e de desequilíbrio. Nos mapas de desequilíbrio realizados (Figuras 1, 2), deve-se ter em mente que o valor máximo irá variar para cada par de marcadores e será estabelecido em função de suas frequências alélicas. Nos mapas de ligação, a distância entre dois genes reflete a sua porcentagem de recombinação, cujo limite máximo será de 50%.

População F_2 – LD e mapa de ligação

Quando se considera uma população F_2 é possível calcular a distância entre dois genes por meio de equações de máxima verossimilhança baseado em distribuição multinomial e estabelecer o mapa genético que seria exatamente igual ao mapa estabelecido a partir das medidas de desequilíbrio entre dois pares de marcadores (Hill, 1975; Dempster et al., 1977; Excoffier & Slatkin, 1995; Long et al., 1995; Slatkin & Excoffier, 1996; Weir, 1996). Na população F_2 , as frequências dos alelos de cada loco seriam iguais a 0.5 e o valor máximo de desequilíbrio seria de 25%. Uma relação entre a distância e a taxa de desequilíbrio pode ser estabelecida por meio da expressão:

$$d_{ij} = 50(1 - U_{ij}/U_{\max})$$

Assim, o desequilíbrio máximo entre dois pares de locos corresponde à completa ligação fatorial e a falta de desequilíbrio se manifesta para locos com segregação independente em que a porcentagem de recombinação é igual a 50cM.

Neste trabalho, considerou inicialmente uma população F_2 e foram estabelecidos os mapas de ligação e desequilíbrio, conforme apresentados nas figuras 1_A e 1_B. Como era esperado, realizadas as comparações quanto aos números de grupos de ligação obtidos, o número de marcas por grupo, os tamanhos dos grupos de ligação, as distâncias médias entre marcadores adjacentes nos grupos de ligação, as variâncias das distâncias entre marcas adjacentes nos grupos de ligação, e se ocorrerá ou não inversão da ordem dos marcadores, os mapas são equivalentes, dado que nesta população toda causa de desequilíbrio de fase gamética é atribuída à ligação fatorial.

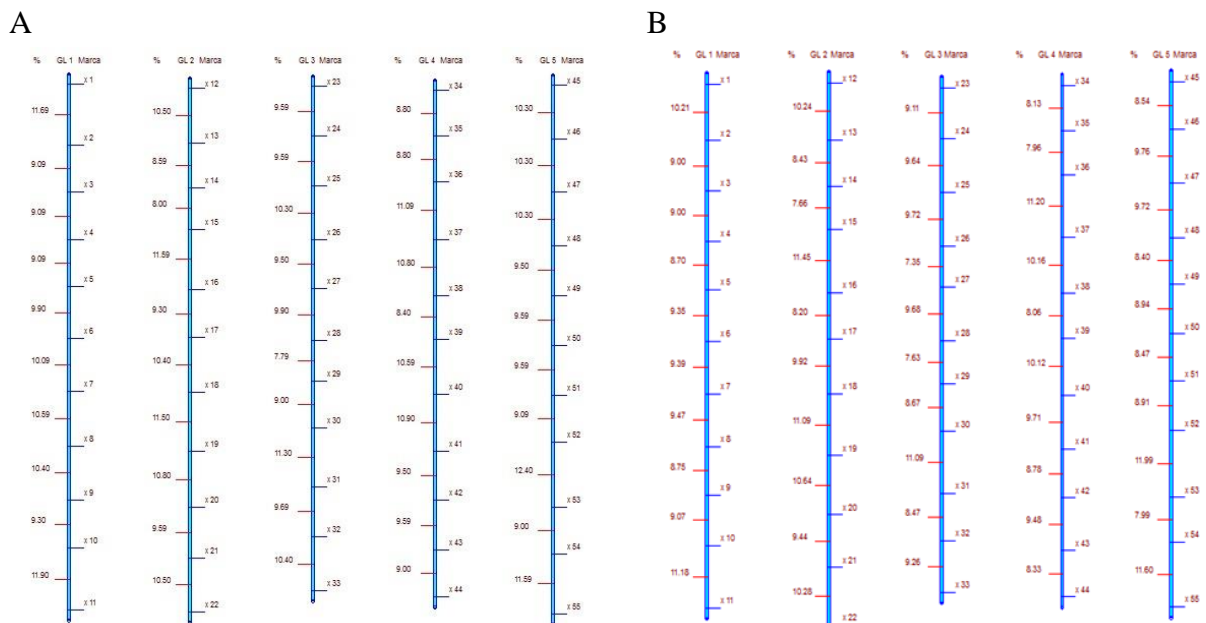


Figura 1. A) Mapa de ligação de uma população F_2

B) Mapa de desequilíbrio de uma população F_2 .

Populações avançadas por autofecundação da F_2 (F_{2s1} a F_{2s5})

Vários fatores podem afetar o desequilíbrio gamético entre pares de marcadores em uma população. Neste trabalho, procurou enfatizar a influência do sistema de acasalamento sobre o desequilíbrio tendo em vista futuras investigações de uso da

análise genômica no melhoramento vegetal e, portanto, processos de acasalamento ao acaso e de autofecundação são relevantes. A cada geração, a transferência de alelos fisicamente ligados é reduzida na proporção da fração de recombinação, até que seja adquirido o equilíbrio de ligação, onde os haplótipos recombinantes e não recombinantes são igualmente distribuídos na população (Gebhardt et al., 2004).

Um sistema de acasalamento a ser investigado é a autofecundação, próprio de plantas autógamas e, neste caso, por questões de aplicabilidade, as teorias de análise genômica relativas a mapeamento genético já estão bem estabelecidas. Assim, para a construção de mapas genéticos com populações F_{2s1} , a F_{2s5} já se tem funções de verossimilhança conhecidas e aplicáveis, de forma que a ligação fatorial possa ser acompanhada nas diferentes gerações e a associação entre desequilíbrio de fase gamética e ligação fatorial possa ser melhor visualizada.

Nas Figuras 2_A e 2_B são mostrados os mapas genéticos estabelecidos para o mesmo conjunto de genes (ou marcadores), estabelecidos a partir de indivíduos das populações F_{2s1} e F_{2s5} . O número de grupos de ligação esperado no processo de mapeamento nas populações depois das gerações de autofecundação permaneceu igual a 5. Porém, foi observado no grupo 4 que não houve a recuperação de todos os marcadores, apresentando diferenças nas variâncias das distâncias entre marcas adjacentes nos grupos de ligação (figura 2_B). Hagenblad & Nordborg, (2002) sequenciaram 14 pequenos fragmentos de 400 kb de *Arabidopsis*, eles encontraram que o LD decresce entre 250 kb, o que equivale a 1 cM. Segundo esses autores grande parte da informação da ligação fatorial é preservada, e as discrepâncias podem ser atribuídas a erros de amostragem, tamanho populacional e casualização gamética. Genes originalmente ligados, certamente continuam sendo visualizados como pertencentes ao mesmo grupo de ligação com os sucessivos avanços de autofecundação, como

demonstram as Figuras 2_A e 2_B, porém, apesar de suas relações de ligação fatorial não serem afetadas, considerável mudança na taxa de desequilíbrio está ocorrendo conforme pode ser visualizado nos valores de D' e r^2 da tabela 1.

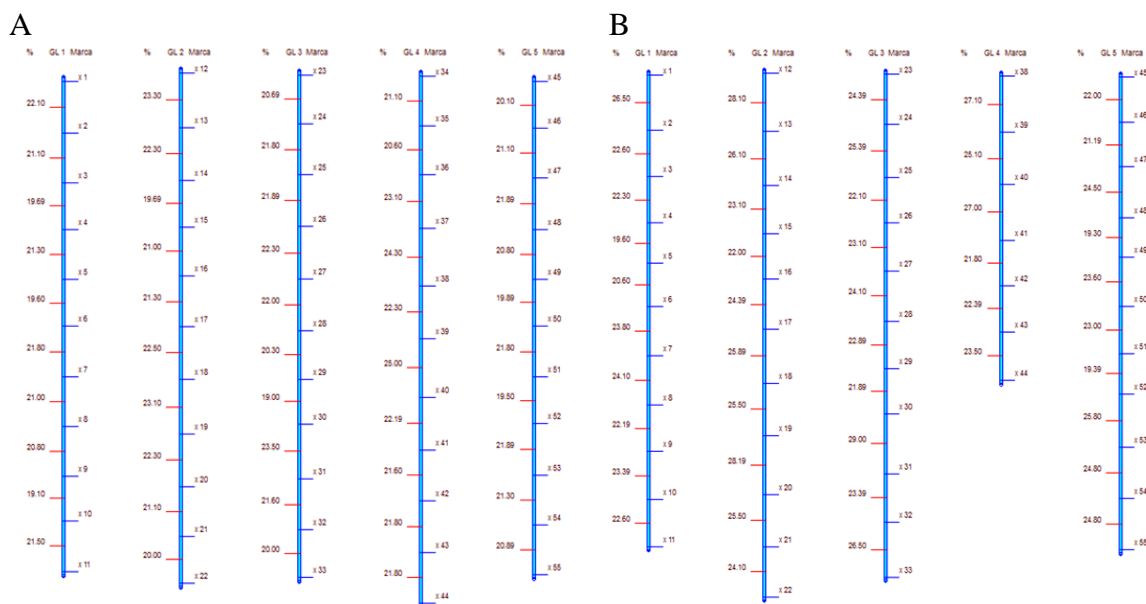


Figura 2. Mapas de ligação originados a partir de uma população conhecida. A) população F_{2s1}, B) população F_{2s5}.

O intuito da seleção genômica ampla (GWS) é obter um modelo que prevê o valor genético do indivíduo, mas que não necessariamente determina genes específicos envolvidos no controle do caráter. Cabe destacar que o tipo de população utilizada tem impacto relevante sobre os padrões de LD. Em função destas características a GWS tem chamado mais atenção de melhoristas recentemente, pela possibilidade real de sua operacionalização em programas de melhoramento (Hayes et al., 2009).

A Tabela 1 deve ser analisada considerando as particularidades do estudo realizado, em que foram considerados marcadores em 5 grupos de ligações com níveis diferenciados de saturação. Assim, no caso do trabalho para as pressuposições do estudo que o grupo de ligação era estabelecido por 11 marcadores, espera-se que a ligação fatorial, em 30cM de distância, possa produzir alguma taxa de desequilíbrio perceptível,

sendo predita pela razão 27/55 pares de marcadores em desequilíbrio em um determinado grupo de ligação. Para todo o genoma estudado, com $g=5$ grupos de ligações, a taxa de desequilíbrio perceptível, admitindo o caso específico e distância mínima de 30cM, seria de 9.1% resultante da relação 135/1485 pares de marcadores em desequilíbrio por pares de marcadores totais. Este valor poderia ser dissipado com o avanço dos acasalamentos. Assim, tendo em vista que o uso da GWS é fundamentada em LD, preocupa-se imaginar que haveria poucos pares de locos que proporcionariam considerável desequilíbrio e, ainda, que o acasalamento reduziria esta taxa de forma que o trabalho de genotipagem, das gerações básicas, necessitaria de reajuste onerando a técnica de seleção.

Tabela 1. Porcentagens de pares de locos em relação aos valores esperados D' e r^2 do desequilíbrio de ligação entre dois locos, em função do sistema de acasalamento de autofecundação para três conjuntos de genes (genomas com 55, 155 e 255 locos) durante 5 gerações.

D'	55				155				255			
	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9
Gerações												
F_2	79.1	9.6	7.8	3.6	79.2	7.3	7.6	6.0	77.2	8.9	7.5	6.4
F_{2s1}	82.0	13.1	5.0	0.0	80.8	12.5	6.7	0.0	79.7	13.1	7.2	0.0
F_{2s2}	82.0	15.4	2.6	0.0	81.0	14.3	4.8	0.0	79.7	15.3	5.0	0.0
F_{2s3}	82.2	15.5	2.3	0.0	80.8	15.0	4.2	0.0	79.4	16.0	4.6	0.0
F_{2s4}	81.8	16.4	1.9	0.0	80.5	15.5	4.0	0.0	79.1	16.5	4.5	0.0
F_{2s5}	81.4	16.8	1.8	0.0	80.3	15.7	4.0	0.0	78.9	16.7	4.5	0.0
r^2												
Gerações												
F_2	79.1	9.6	7.8	3.6	79.3	7.4	7.6	5.7	77.2	8.9	7.5	6.4
F_{2s1}	82.2	13.5	4.3	0.0	81.1	12.6	6.4	0.0	80.0	13.3	6.7	0.0
F_{2s2}	82.4	15.5	2.1	0.0	81.3	14.5	4.3	0.0	80.1	15.4	4.5	0.0
F_{2s3}	82.6	16.0	1.4	0.0	81.1	15.1	3.8	0.0	79.9	16.0	4.1	0.0
F_{2s4}	82.4	16.3	1.3	0.0	80.9	15.5	3.6	0.0	79.6	16.5	3.9	0.0
F_{2s5}	82.0	16.7	1.3	0.0	80.7	15.7	3.7	0.0	79.4	16.7	3.9	0.0

Eq- equilíbrio

As estatísticas D' e r^2 (tabela 1) nas populações derivadas de diferentes ciclos de autofecundação não apresentam diferença em relação aos três conjuntos de genes, amostrando um comportamento ascendente, com o incremento de pares de genes em equilíbrio após de diversas gerações de autofecundação, deduzindo que o número de

pares de marcas não afeta à identificação do LD nas populações, pode observasse que a porcentagem de pares de locos em equilíbrio vai aumentando como ocorreu nas gerações das colunas 1 e 2 da tabela 1. Resultados similares aos observados na Tabela 1 são ressaltados nos gráficos de declínio do LD (Figura 3) que permitem visualizar a extensão do LD entre pares de locos para as gerações F_{2s1} e F_{2s5} após diversos ciclos de autofecundação, se observa um comportamento similar tanto acima como abaixo da diagonal onde pode-se observar grande parte do gráfico em amarelo demonstrando um alto porcentagem de pares de locos em equilíbrio, e uma diferença na ocorrência de locos no LD dados em vermelho e azul. Assim, constata-se diminuição na taxa do LD com base à distância genética ou física. Observa-se um resultado similar para as estatísticas r^2 e D' , concordando com os resultados de Fearnhead & Donnelly (2001).

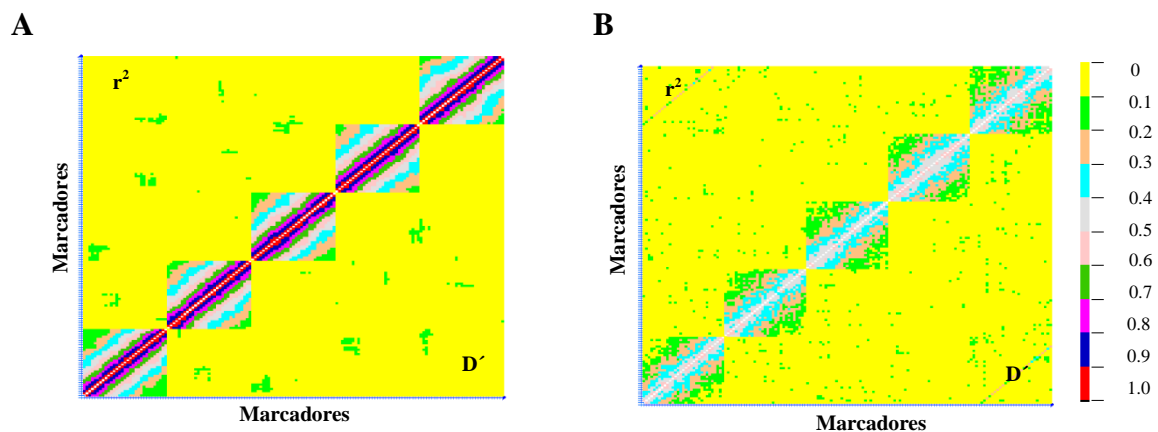


Figura 3. Padrões de pares de locos em LD para duas gerações de autofecundação A) F_2 e B) F_{2s11} , estimado por r^2 e D' plotados para cada comparação de pares com D' abaixo da diagonal e r^2 acima dela.

Populações avançadas por acasalamento ao acaso da F_2 (F_{2a1} a F_{2a5})

Apesar de já serem conhecidas as bases teóricas do efeito do acasalamento ao acaso sobre a taxa de desequilíbrio, considerou ser apropriado apresentar evidências de seus efeitos numa população de tamanho finito, mais compatível com a experimentação agrícola, e um conjunto de pares de locos cujo agrupamento, por meio de mapeamento

genético, já era conhecido. Neste caso, foi considerado que a população F_2 passou por vários ciclos sucessivos de acasalamento ao acaso e os valores da taxa de desequilíbrio foram calculados (Tabela 2). Trata-se de uma situação restritiva, porém de grande aplicação prática. Ou seja, nestas gerações, as frequências gênicas ainda permanecem inalteradas ($p=q=0.5$ para todos os locos), não há seleção, migração, deriva ou mutação e o equilíbrio de Hardy-Weinberg é atingido para cada loco. Os pares de genes certamente mantêm as suas mesmas posições físicas no mapa, pois não se considera ocorrência de nenhuma alteração estrutural ou cromossômica, porém os efeitos da recombinação passam ter influência marcante nas taxas de desequilíbrio.

Na Figura 4 é apresentada a tentativa de se gerar um mapa de ligação com as gerações avançadas por acasalamento ao acaso. Outra maneira de visualizar o efeito deste sistema de acasalamento é por meio da comparação dos valores das taxas de desequilíbrio, expressos em D' e r^2 , conforme apresentado na Tabela 2.

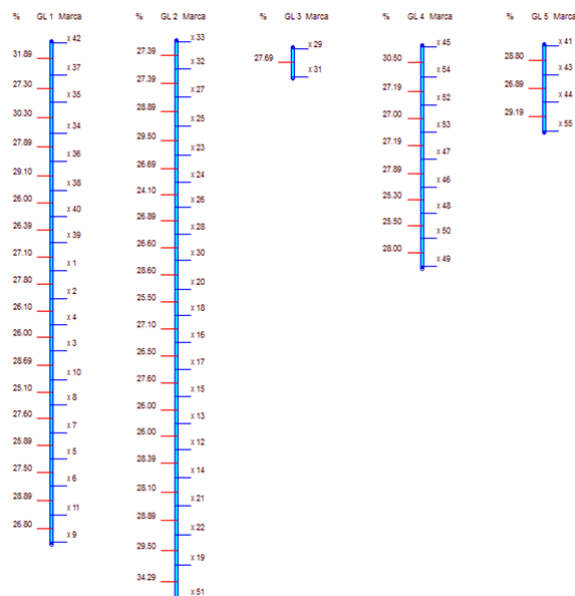


Figura 4. Mapa de ligação numa população $F_{2,a1}$ obtida por acasalamento ao acaso.

Pode-se observar na Tabela 2 que as estatísticas D' e r^2 revelam resultados mais expressivos, apresentando mais do que 96% dos pares de locos em equilíbrio após cinco gerações de acasalamento ao acaso, comparado com o valor obtido pela autofecundação que foi 82% após o mesmo número de gerações. O valor de LD perceptível (taxas acima de 0.4, somatório colunas 3 e 4 da tabela 2) esteve em torno do esperado nas gerações iniciais (11% a 14%), para as particularidades do estudo e que, após primeira geração de acasalamento ao acaso, este valor já era nulo. Assim, nesta situação a eficiência do uso de marcadores com a genotipagem realizada em gerações anteriores a da população de melhoramento poderá não ser tão consistente, independente do efeito do marcador sobre a expressão da característica.

Tabela 2. Porcentagens de pares de locos em relação aos valores esperados D' e r^2 do desequilíbrio de ligação entre dois locos, em função do sistema de acasalamento ao acaso para três conjuntos de genes (genomas com 55, 155 e 255 locos) durante 5 gerações.

D'	55				155				255			
	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9
Gerações												
F_2	79.1	9.6	7.8	3.6	79.2	7.3	7.6	6.0	77.2	8.9	7.5	6.4
F_{2a1}	86.3	13.7	0.0	0.0	83.9	15.9	0.2	0.0	83.5	16.1	0.4	0.0
F_{2a2}	96.6	3.4	0.0	0.0	95.3	4.5	0.0	0.2	95.0	4.7	0.0	0.4
F_{2a3}	93.3	6.7	0.0	0.0	95.7	4.0	0.0	0.2	96.2	3.4	0.0	0.4
F_{2a4}	95.0	5.0	0.0	0.0	95.4	4.4	0.0	0.2	95.6	4.0	0.0	0.4
F_{2a5}	96.4	3.6	0.0	0.0	95.4	4.4	0.0	0.2	95.8	3.8	0.0	0.4
r^2	55				155				255			
Gerações	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9
F_2	79.3	9.5	7.7	3.4	79.2	7.3	7.6	6.0	77.4	8.9	7.6	6.1
F_{2a1}	87.0	13.0	0.0	0.0	84.3	15.4	0.2	0.0	84.1	15.5	0.4	0.0
F_{2a2}	96.9	3.1	0.0	0.0	96.1	3.7	0.0	0.2	95.6	4.0	0.0	0.4
F_{2a3}	94.8	5.2	0.0	0.0	96.9	2.9	0.0	0.2	97.1	2.6	0.0	0.4
F_{2a4}	96.7	3.3	0.0	0.0	96.8	3.0	0.0	0.2	96.6	3.0	0.0	0.4
F_{2a5}	97.8	2.2	0.0	0.0	96.6	3.1	0.0	0.2	96.8	2.9	0.0	0.4

Eq- equilíbrio

Os resultados obtidos ratificam as conclusões de diversos estudos que demonstram que o LD diminui mais rapidamente em espécies alógamas, devido ao processo de recombinação ser mais efetivo em relação às espécies que se autofecundam, em que os indivíduos são mais semelhantes com redução da frequência de duplo

heterozigotos, que é onde ocorre a recombinação alterando o LD (Gupta et al., 2005; Flint-Garcia et al., 2003; Gaut & Long, 2003; Rafalski & Morgante, 2004). Em milho, vários estudos têm sido conduzidos para investigar o LD ao longo da população e tipos de marcadores. O LD varia substancialmente para cada população escolhida. Tenailon et al., (2001) investigaram a diversidades nas sequencias em 21 locos do cromossomo 1 no germoplasma do milho. O LD, mensurado como r^2 , decresce menos de 0.25 dentro de 200pb em média. Análises interlocos revelaram pequeno LD entre locos, apesar de que todos os locos foram localizados no mesmo cromossomo.

População P de origem desconhecida

Ao contrário do observado em populações conhecidas, em uma população qualquer cujo agrupamento de pares de locos por meio de mapeamento genético é desconhecido, não se identifica se os genes em LD se encontram, ou não, em ligação fatorial, impossibilitando a formação de mapas de ligação e seu análogo mapa de desequilíbrio (Figura 5), o número de grupos de ligação esperado no processo de mapeamento nas populações não é formado, foi observado que não houve recuperação de todos os marcadores. A quantidade esperada de pares de locos em LD para os conjuntos gênicos analisados, nesta situação, não é possível de ser predita, mas pelos resultados das Tabelas 3 e 4 em relação as tabelas 1 e 2, constata-se grande redução no valor observado ficando em torno de 5 a 6% (Valores de D' acima de 0.4).

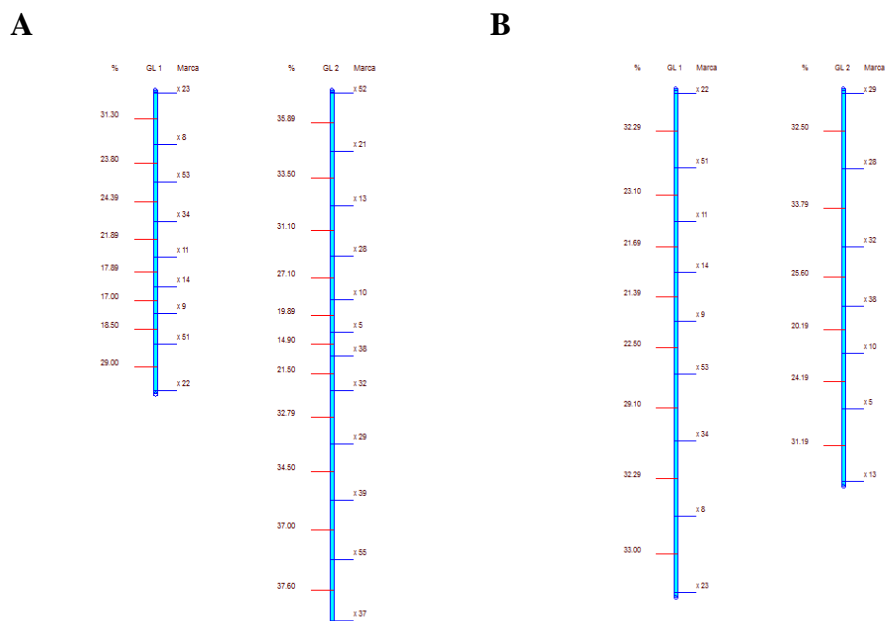


Figura 5. A) Mapa de ligação de uma população P de origem desconhecida B) Mapa análogo de desequilíbrio de uma população P.

Populações avançadas por autofecundação da P (P_{s1} a P_{s5})

Ciclos de autofecundação aumentam o número potencial de eventos de recombinação. Adicionalmente, o baixo número de alelos amostrados por loco em cada população dificulta examinar a totalidade extensão da diversidade genética disponível para várias espécies de plantas.

Na Tabelas 3 pode observar-se que não existem diferenças quanto ao comportamento dos pares de locos em equilíbrio em relação aos seus sistemas de acasalamento para uma população qualquer da qual desconhecesse sua estrutura genética. É possível perceber as estatísticas D' e r^2 descrevem o comportamento esperado neste tipo de população. Observa-se que a primeira coluna revela aumento na porcentagem de pares de locos em equilíbrio de ligação enquanto a terceira e quarta colunas são as que melhor demonstram o comportamento do LD perceptível,

apresentando redução expressiva com o passar das gerações. Em trabalhos Nordborg et al (2002) em populações isoladas de *Arabidopsis*, foi encontrado LD com 10 cM, produto do número limitado de eventos de recombinação.

Tabela 3. Porcentagens de pares de locos em relação aos valores esperados D' e r^2 do desequilíbrio de ligação entre dois locos, em função do sistema de acasalamento para três conjuntos de genes (genomas de 55, 155 e 255 locos) durante 5 gerações de autofecundação a partir de uma população qualquer, de origem desconhecida.

D'	55				155				255			
	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9
Gerações												
P_2	58.7	34.4	5.6	1.3	57.1	35.6	5.9	1.4	57.7	34.9	5.8	1.5
P_{s1}	78.4	21.3	0.3	0.0	76.7	22.1	0.8	0.4	77.7	21.1	0.7	0.5
P_{s2}	81.0	18.6	0.4	0.0	76.1	22.5	0.9	0.4	77.6	21.1	0.7	0.5
P_{s3}	78.1	21.1	0.7	0.1	75.1	23.6	0.9	0.5	77.8	21.0	0.7	0.5
P_{s4}	77.2	22.3	0.3	0.2	76.9	21.7	0.7	0.6	77.4	21.4	0.7	0.5
P_{s5}	79.9	19.5	0.6	0.1	76.7	21.8	0.8	0.6	78.3	20.6	0.6	0.5
r^2	55				155				255			
Gerações	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9
P_2	79.8	17.2	3.0	0.1	79.7	17.9	2.2	0.1	79.2	18.0	2.6	0.1
P_{s1}	97.2	2.8	0.0	0.0	97.2	2.6	0.2	0.0	96.9	2.8	0.2	0.0
P_{s2}	97.7	2.3	0.0	0.0	97.1	2.7	0.1	0.0	97.0	2.8	0.2	0.1
P_{s3}	96.4	3.6	0.0	0.0	96.8	3.0	0.1	0.1	96.9	2.8	0.1	0.1
P_{s4}	97.0	3.0	0.0	0.0	97.3	2.5	0.1	0.1	96.7	3.0	0.1	0.1
P_{s5}	97.6	2.4	0.0	0.0	97.2	2.6	0.1	0.1	97.1	2.6	0.1	0.1

Eq- equilíbrio

Populações avançadas por acasalamento ao acaso da P (P_{a1} a P_{a5})

A ausência de estrutura das populações, conduz a elevada redução do nível de desequilíbrio de ligação estabelecida pelos valores D' e r^2 , tabela 3 e 4 em comparação com as tabelas 1 e 2, detectando-se um grande porcentagem de lócus em equilíbrio (primeira coluna da tabela), para cada conjunto de genes. Comprometendo o uso da GWS no melhoramento genético pois dita metodologia é comprometida pela baixa quantidade de locos em LD em gerações avançadas de acasalamento. O resultado anterior faz ressaltar a importância do trabalho com populações, das quais se conheça sua estrutura genética. Segundo Grattapaglia (2007) o uso de elevadíssima densidade de marcadores e de populações não estruturados, permitem que os locos amostrados

potencialmente capturem toda a variabilidade genética da população em estudo e não apenas a variabilidade de dois genótipos parentais, como se verifica em estudos de gerações derivadas de cruzamentos controlados. Labate et al. (2000) examinaram o LD entre RFLP em duas populações sintéticas que tem tido cruzamentos aleatórios por algumas gerações. Estas populações originaram de 12 a 16 progenitores homocigóticos. Cada população original sofreu seleção recorrente por 12 gerações. É interessante notar que as populações responderam diferentemente a seleção. Uma população sofreu aumento substancial no LD nas 12 gerações, enquanto que outras tiveram um decréscimo.

Estes estudos revelam a necessidade de entendimento mais aprofundado das mudanças dos valores de LD em condições conhecidas do genoma das espécies e do sistema de acasalamento a que a população foi submetida.

Tabela 4. Porcentagens de pares de locos em relação aos valores esperados D' e r^2 do desequilíbrio de ligação entre dois locos, em função do sistema de acasalamento de autofecundação para três conjuntos de genes durante 5 gerações de uma população qualquer de origem desconhecida.

D'	55				155				255			
	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9	Eq	0.1-0.3	0.4-0.6	0.7-0.9
Gerações												
P ₂	60.3	33.7	4.8	1.2	57.8	35.3	5.4	1.4	82.3	16.9	0.7	0.1
P _{a1}	82.6	17.1	0.3	0.0	81.2	17.8	0.8	0.2	82.3	16.9	0.6	0.1
P _{a2}	81.1	18.4	0.5	0.0	79.9	18.9	0.9	0.2	80.8	18.4	0.6	0.2
P _{a3}	80.3	19.2	0.5	0.0	78.9	19.9	0.9	0.2	79.3	19.8	0.8	0.2
P _{a4}	78.9	20.5	0.5	0.1	77.6	21.1	1.0	0.2	78.3	20.7	0.8	0.2
P _{a5}	78.5	21.0	0.5	0.1	77.0	21.7	1.1	0.2	78.3	20.7	0.9	0.2
r ²												
		55			155				255			
Gerações												
P ₂	80.1	17.4	2.4	0.1	80.5	17.1	2.2	0.2	98.3	1.6	0.1	0.1
P _{a1}	98.9	1.1	0.0	0.0	98.3	1.6	0.1	0.0	98.4	1.5	0.1	0.0
P _{a2}	98.3	1.7	0.0	0.0	98.3	1.6	0.1	0.0	98.1	1.8	0.1	0.0
P _{a3}	97.8	2.2	0.0	0.0	97.8	2.1	0.1	0.0	97.6	2.3	0.1	0.0
P _{a4}	97.6	2.4	0.0	0.0	97.5	2.4	0.1	0.0	97.2	2.6	0.1	0.0
P _{a5}	97.2	2.8	0.0	0.0	97.1	2.9	0.1	0.0	97.2	2.7	0.1	0.0

Eq- equilíbrio

Conclusões

- O genoma da espécie, o tipo de população e os sistemas de acasalamento afetam de forma diferencial as taxas de desequilíbrio entre pares de locos.

- Nas populações F_2 os mapas de desequilíbrio e de ligação fatorial são equivalentes e menos preservados com o acasalamento ao acaso.

- Nas populações de origem desconhecidas, as causas do desequilíbrio não são diretamente relacionadas com a ligação fatorial e as taxas são preservadas de forma similar à F_2 com as sucessivas gerações de acasalamento ao acaso ou autofecundação.

- O tipo de população utilizada tem impacto relevante sobre os padrões de LD e conseqüentemente sobre o número de marcadores necessários para identificar genes que controlam características de interesse ao melhoramento e selecionar indivíduos superiores.

-A queda do desequilíbrio de ligação é mais lenta em autógamias de que alógamas.

Referências

ABDURAKHMONOV, I. Y.; ABDUKARIMOV, A. Application of association mapping to understanding the genetic diversity of plant germplasm resources. **International journal of plant genomics**, v. 2008, 2008.

BHERING, L. L.; CRUZ, C. D.; GOD, P. I. V. G. Estimativa de frequência de recombinação no mapeamento genético de famílias de irmãos completos. **Pesq. agropec. bras., Brasília**, v. 43, n. 3, p. 363-369, 2008.

CRUZ, C. D. Programa para análise de dados moleculares e quantitativos – GQMOL - Versão 2012.1. Viçosa: UFV, 2012.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 1-38, 1977.

EXCOFFIER, L.; SLATKIN, M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. **Molecular biology and evolution**, v. 12, n. 5, p. 921-927, 1995.

FEARNHEAD, P.; DONNELLY, P. Estimating recombination rates from population genetic data. **Genetics**, v. 159, n. 3, p. 1299-1318, 2001.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; IV, B. Structure of Linkage Disequilibrium in Plants*. **Annual Review of Plant Biology**, v. 54, n. 1, p. 357-374, 2003.

GAUT, B. S.; LONG, A. D. The lowdown on linkage disequilibrium. **The Plant Cell Online**, v. 15, n. 7, p. 1502-1506, 2003.

GEBHARDT, C.; BALLVORA, A.; WALKEMEIER, B.; OBERHAGEMANN, P.; SCHÜLER, K. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. **Molecular Breeding**, v. 13, n. 1, p. 93-102, 2004.

GLAZIER, A. M.; NADEAU, J. H.; AITMAN, T. J. Finding genes that underlie complex traits. **Science**, v. 298, n. 5602, p. 2345-2349, 2002.

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

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

GRATTAPAGLIA, D. Mapas genéticos e seleção assistida por marcadores moleculares, pp. 201-230 in *Biotechnology Florestal*, edited by A. BOREM. Editora UFV, Viçosa. 2007.

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

GUPTA, P.; RUSTGI, S. Molecular markers from the transcribed/expressed region of the genome in higher plants. **Functional & integrative genomics**, v. 4, n. 3, p. 139-162, 2004.

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

HAGENBLAD, J.; NORDBORG, M. Sequence variation and haplotype structure surrounding the flowering time locus FRI in *Arabidopsis thaliana*. **Genetics**, v. 161, n. 1, p. 289-298, 2002.

HAYES, B.; BOWMAN, P.; CHAMBERLAIN, A.; GODDARD, M. Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of dairy science**, v. 92, n. 2, p. 433, 2009.

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

HILL, W. G. Tests for association of gene frequencies at several loci in random mating diploid populations. **Biometrics**, p. 881-888, 1975.

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

KUMAR, S.; ECHT, C.; WILCOX, P.; RICHARDSON, T. Testing for linkage disequilibrium in the New Zealand radiata pine breeding population. **TAG Theoretical and Applied Genetics**, v. 108, n. 2, p. 292-298, 2004.

LABATE, J.; LAMKEY, K.; LEE, M.; WOODMAN, W. Hardy-Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. **Maydica**, v. 45, n. 3, p. 243-256, 2000.

LIANG, K.-Y.; HSU, F.-C.; BEATY, T. H.; BARNES, K. C. Multipoint Linkage-Disequilibrium-Mapping Approach Based on the Case-Parent Trio Design. **American journal of human genetics**, v. 68, n. 4, p. 937, 2001.

LONG, J. C.; WILLIAMS, R. C.; URBANEK, M. An EM algorithm and testing strategy for multiple-locus haplotypes. **American journal of human genetics**, v. 56, n. 3, p. 799, 1995.

MACKAY, T. F. The genetic architecture of quantitative traits. **Annual review of genetics**, v. 35, n. 1, p. 303-339, 2001.

MEUWISSEN, T.H.E. and GODDARD, M. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. **Genetics**, v. 155, n. 1, p. 421-430, 2000.

NORDBORG, M.; BOREVITZ, J. O.; BERGELSON, J.; BERRY, C. C.; CHORY, J.; HAGENBLAD, J.; KREITMAN, M.; MALOOF, J. N.; NOYES, T.; OEFNER, P. J. The extent of linkage disequilibrium in *Arabidopsis thaliana*. **Nature genetics**, v. 30, n. 2, p. 190-193, 2002.

NORDBORG, M.; TAVARÉ, S. Linkage disequilibrium: what history has to tell us. **TRENDS in Genetics**, v. 18, n. 2, p. 83-90, 2002.

PALAISSA, K.; MORGANTE, M.; TINGEY, S.; RAFALSKI, A. Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 26, p. 9885-9890, 2004.

PALAISSA, K. A.; MORGANTE, M.; WILLIAMS, M.; RAFALSKI, A. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. **The Plant Cell Online**, v. 15, n. 8, p. 1795-1806, 2003.

RAFALSKI, A.; MORGANTE, M. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. **TRENDS in Genetics**, v. 20, n. 2, p. 103-111, 2004.

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

SIMKO, I.; HAYNES, K.; EWING, E.; COSTANZO, S.; CHRIST, B.; JONES, R. Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. **Molecular Genetics and Genomics**, v. 271, n. 5, p. 522-531, 2004.

SLATKIN, M.; EXCOFFIER, L. Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. **Heredity**, v. 76, n. 4, p. 377-383, 1996.

TENAILLON, M. I.; SAWKINS, M. C.; LONG, A. D.; GAUT, R. L.; DOEBLEY, J. F.; GAUT, B. S. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). **Proceedings of the National Academy of Sciences**, v. 98, n. 16, p. 9161-9166, 2001.

TERWILLIGER, J. D.; WEISS, K. M. Linkage disequilibrium mapping of complex disease: fantasy or reality? **Current Opinion in Biotechnology**, v. 9, n. 6, p. 578-594, 1998.

WEIR, B. Genetic data analysis. Vol. II. **Sunderland, MA: Sinauer Associates**, 1996.

WEIR BS, HILL WG. Non-uniform recombination within the human β -globin gene cluster. **Am J Hum Genet** 38:776–778, 1986.

WRAY, N. R., VISSCHER, P. M. Population genetics and its relevance to gene mapping. In: NEALE, B. N. M.; FERREIRA, M. A. R.; MEDLAND, S. E.; POSTHUMA, D. (Ed.). **Statistical genetics: Gene mapping through linkage and association**. New York: Taylor & Francis Group. p. 87-112. 2008.

CAPÍTULO II
USO DE SELEÇÃO GENOMICA AMPLA EM POPULAÇÕES SIMULADAS E
SUAS GERAÇÕES AVANÇADAS POR AUTOFECONDAÇÃO E
ACASALAMENTO AO ACASO

Resumo- Um dos atrativos da genética molecular em benefício do melhoramento genético de plantas é a possibilidade de utilização direta das informações de DNA na seleção. O trabalho teve por objetivo avaliar a eficiência da seleção genômica ampla (GWS) na predição do valor genético em gerações avançadas por autofecundação e acasalamento ao acaso, com e sem avaliação fenotípica. Como todo o estudo se fundamenta em dados simulados, constitui-se também objetivo do trabalho constatar a própria eficácia do processo de simulação em gerar populações e gerações cujos princípios genéticos sejam preservados conforme princípios fundamentais da genética quantitativa e de populações. Foram simuladas cinco repetições de uma estrutura populacional. A estrutura de cada repetição foi criada para apresentar dois conjuntos de dados: fenotípicos e genotípicos de cada indivíduo dentro da população, imitando alguns dos cenários em que a GWS é aplicada. Em cada sequência de dados duplicados foram considerados 1500 locos, e um tamanho efetivo das populações de $N_e=1000$ para todas as geração originadas por acasalamento ao acaso e autofecundação. Foram simuladas três características, em cada característica, o número de locos que controlava o caráter foi estabelecido em 500 e as herdabilidades em 20%, 40% e 60%. Com base nos valores genotípicos e fenotípicos das populações em todas as gerações, foram estimados os parâmetros: média (μ), variância (σ^2), desvio padrão (σ), e a endogamia na população gerada pela autofecundação ou pelo acasalamento ao acaso. Além da estimação de parâmetros, as populações foram caracterizadas quanto a distribuição de seus dados sendo feitos testes de normalidade e obtenção de coeficientes de simetria (s) e curtose (k) nas diversas gerações. Após estabelecidos todos os parâmetros genéticos se avaliou a correlação dos valores fenotípicos observados com os valores fenotípicos preditos via informação de marcadores e a acurácia de seleção. A simulação utilizada foi eficaz em preservar a estrutura genética das populações e descrever a sua dinâmica

ao longo de sucessivas gerações de acasalamento ao acaso ou autofecundação. Os resultados mostram que o sistema de acasalamento afeta a eficiência do uso das estimativas dos efeitos dos marcadores em gerações avançadas. Ademais concluí-se que a agregação de informações moleculares e fenotípicas na predição do valor genético do indivíduo proporciona acréscimo no valor da acurácia.

Termos para indexação: simulação; genômica; parâmetros genéticos; melhoramento; seleção genômica.

Abstract- One of the great attractions of molecular genetics for the benefit of plant breeding is the possibility of direct use of DNA information in the selection. The study aimed to evaluate the efficiency of genome-wide selection (GWS) in the prediction of genetic value in advanced generations by selfing and random mating, with and without phenotypic evaluation. As the whole study is based on simulated data, is also objective of the study noted the very effectiveness of the simulation process to generate generations and populations whose genetic principles are preserved as principles foundations of quantitative genetics and population. There five replicates of a simulated population structure. The structure of each repetition was designed to present two sets of data: phenotypic and genotypic of each individual within the population, mimicking some of the scenarios where GWS is applied. In each sequence of duplicate data, forum considered 1500 locus, and an effective population size of $N_e = 1000$ for all generation originated by random mating and selfing. Were simulated three features in each feature, the number of loci that controlled the character was established in 500 and heritability 20%, 40% and 60%, based on genotypic and phenotypic values of the populace in all generations, were estimated parameters: mean (μ), variance (σ^2), standard deviation (σ), and the inbreeding in the population generated by selfing or random mating. Besides the parameter estimation, the populations were characterized as the distribution of your data being made normality tests and obtaining coefficients of symmetry (s) and kurtosis (k) in several generations. After all genetic parameters set, were evaluate the correlation of phenotypic values observed with the phenotypic values predicted using markers information and acúaracia selection. The simulation used was effective in preserving the genetic structure of populations and describe its dynamics over successive generations of random mating or selfing, results shows the mating system

affects the efficiency of the use of the use of estimates of the effects of the markers in advanced generations.

Index terms: simulation, genomics, genetic parameters; improvement; genomic selection.

Introdução

O homem depende, para sua alimentação, direta ou indiretamente das plantas. Ademais, outros produtos necessários ao bem-estar da humanidade, como fibras, materiais de construção, etc., são igualmente fornecidos pelos vegetais. Com o crescente aumento da população na superfície terrestre, necessário se faz a correspondente elevação da produção agrícola. Isso pode ser conseguido pelo melhoramento das condições ambientais onde as plantas cultivadas se desenvolvem ou pelo melhoramento dessas próprias plantas. O ambiente pode ser modificado mediante o emprego das modernas técnicas agrônomicas de produção, incluindo adubação, irrigação, drenagem, controle de insetos, doenças e ervas daninhas, rotação de culturas, preparo e conservação de solo.

Altas produções são atingidas pelo emprego de melhores práticas culturais associadas ao plantio de variedades melhoradas. Considerando a importância que as plantas representam, não é de surpreender a preocupação do homem, em obter tipos mais adequados para a satisfação de suas necessidades. Inicialmente, o melhoramento de plantas era mais uma arte, pois os melhoristas eram pessoas práticas que tinham a habilidade de selecionar as plantas que apresentavam diferenças que podiam ter interesse econômico ou pessoal. Mas, à medida que o melhorador progredia nos conhecimentos sobre as plantas, a seleção passou a ser feita de maneira mais científica, proporcionando muito sucesso, a otimização de recursos físicos, pessoais e financeiros, o aumento de produtividade e a melhoria de várias outras características de interesse na agricultura. Embora muitos métodos tenham surgido ao longo dos anos, a estratégia básica utilizada foi a de predizer o valor genético do indivíduo, baseado em informações fenotípicas e em alguns casos em genealogia.

A seleção consiste em escolher os indivíduos que, pela união de seus gametas, formarão a próxima geração. A seleção, por permitir taxas reprodutivas diferenciadas, é uma das mais poderosas forças de alteração da frequência dos genes nas populações. Mas como escolher os indivíduos e as populações de maneira adequada a determinados objetivos, é a pergunta que está sempre intrigando os melhoristas. Infelizmente ainda não é possível conhecer, com total acurácia, o valor genético dos indivíduos para características poligênicas ou oligogênicas cuja expressão fenotípica é consideravelmente afetada pelo ambiente. Assim, a complexidade da seleção reside no fato do desempenho dos indivíduos, também denominado de fenótipo, ser resultado do patrimônio genético que a planta possui, o chamado genótipo e, ainda, dos efeitos de meio ambiente, existindo ainda uma interação entre os efeitos de genótipo e de meio ambiente, já que algumas plantas são superiores a outras em alguns ambientes, mas podem se tornar inferiores àqueles em outros ambientes.

O processo pelo qual são estimados os valores genéticos aditivos dos indivíduos conhecido como avaliação genética. Por envolver processos estocásticos, jamais será conhecido o valor genético verdadeiro que um indivíduo tem, para fins de sua utilização como genitor, mas, através de metodologias biométricas, é possível estimar esse valor de forma que tenha boa aproximação com o valor real. É, entretanto, necessário que a estimativa seja livre dos efeitos de meio ambiente e da interação genótipo x ambiente. A acurada predição do valor genético dos indivíduos depende da ação dos genes envolvidos na determinação das características, do número de informações ao respeito das plantas avaliadas (quanto maior este número, melhor a estimativa do valor genético), do parentesco entre os indivíduos avaliados além dos chamados efeitos permanentes de ambiente e da precisão com que os efeitos de ambiente são identificados. Até agora os processos seletivos se fundamentaram numa seleção com

base no desempenho fenotípico como indicador do potencial ou da predisposição genética em se expressar sob influência de determinado ambiente. Em um novo paradigma considera-se ser ideal ler o potencial genético dos indivíduos diretamente das informações hereditárias, agregando ou omitindo as informações do fenótipo, por meio de um índice indireto do fenótipo. A predição do valor genético e a prática da seleção genômica abordam essa nova estratégia.

Com o desenvolvimento dos marcadores moleculares e o avanço em técnicas de biologia molecular, criou-se a expectativa de que as informações genotípicas dos marcadores moleculares, uma vez correlacionados com características fenotípicas de interesse, pudessem ser amplamente utilizadas na obtenção e seleção de indivíduos com maior valor genético. Esta técnica ficou conhecida como seleção assistida por marcadores moleculares (MAS - Marker Assisted Selection) (Resende JR, Munoz et al., 2012; Smaragdov, 2009). Uma primeira proposição realizada para aumentar a eficiência desse procedimento baseado em dados fenotípicos foi descrita por Lande e Thompson (1990), por meio da seleção auxiliada por marcadores (MAS). A MAS utiliza simultaneamente dados fenotípicos e dados de marcadores moleculares em ligação gênica próxima com alguns locos controladores de características quantitativas (QTL) (Pérez et al., 2010).

A integração de métodos clássicos de melhoramento genético com as estratégias e tecnologias da genômica levará ao estabelecimento de novos paradigmas para o desenvolvimento de cultivares superiores de plantas. Com a perspectiva de um aumento nos ganhos de seleção e redução nos ciclos de melhoramento via seleção assistida por marcadores, muitas pesquisas foram feitas e QTLs foram detectados e mapeados nas mais variadas culturas (Frary et al., 2000; Yano et al., 2000; Liu et al., 2002). Não obstante das vantagens da MAS, uma das principais causas de insucesso foi a

necessidade do estabelecimento de associações entre os marcadores e os QTLs para cada família avaliada e o fato de serem feitas apenas a detecção de um pequeno número de QTLs de grande efeito, os quais, devido à natureza poligênica e à alta influência ambiental dos caracteres quantitativos, não explicam suficientemente toda a variação genética (Dekkers, 2004).

O atrativo da genética molecular em benefício do melhoramento genético aplicado é a utilização direta das informações de DNA na seleção, de forma a permitir alta eficiência seletiva, grande rapidez na obtenção de ganhos genéticos com a seleção e baixo custo, em comparação com a tradicional seleção baseada em dados fenotípicos (Resende et al., 2008; Meuwissen & Goddard, 2010). Os avanços de tecnologias de genotipagem em larga escala, a descoberta de novos marcadores como os SNPs (Single Nucleotide Polymorphisms) e a automação do processo de genotipagem de marcadores (Jenkins & Gibson, 2002), permitiram que Meuwissen et al. (2001) propusessem um novo método de seleção denominado seleção genômica (GS) ou seleção genômica ampla (genome wide selection – GWS), o qual é definido como a seleção simultânea para centenas ou milhares de marcadores, os quais cobrem o genoma de maneira densa, de forma que todos os genes de um caráter quantitativo estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores. A técnica seletiva se denomina ampla porque atua em todo o genoma, capturando todos os genes que afetam um caráter quantitativo sem a necessidade de identificar previamente os marcadores com efeitos significativos e de mapear QTLs, como no caso da MAS. Valores genéticos genômicos associados a cada marcador ou alelo são usados para fornecer o valor genético global de cada indivíduo (Crossa et al., 2006; Crossa et al., 2007; Oakey et al., 2006; Piepho, 2009).

A GWS usa associações de um grande número de marcadores SNPs em todo o genoma com os fenótipos, capitalizando no desequilíbrio de ligação entre os marcadores e QTLs proximamente ligados, sem uma prévia escolha de marcadores com base nas significâncias de suas associações com o fenótipo. Predições são, então, obtidas para os efeitos dos haplótipos marcadores ou dos alelos em cada marcador. Essas predições derivadas de dados fenotípicos e de genótipos SNPs em alta densidade em uma geração são usadas para a obtenção dos valores genéticos genômicos dos indivíduos de qualquer geração subsequente, tendo por base os seus próprios genótipos marcadores. Há uma diferença básica na predição de valores genéticos tradicionais e na predição de valores genéticos genômicos. Nos primeiros, informações fenotípicas são utilizadas visando à inferências sobre efeitos dos genótipos dos indivíduos e, nos últimos, informações genotípicas (genótipos para os alelos marcadores) são usadas visando à inferência sobre valores fenotípicos futuros (ou valores genéticos genômicos preditos) dos indivíduos. (Resende et al., 2008)

O estudo da herança e da variação nos caracteres qualitativos se baseia na análise de gerações, separando os indivíduos em classes e avaliando suas proporções nos resultados de certos cruzamentos. Entretanto, nos estudos genéticos, caracteres quantitativos que são, em geral, regulados por vários genes com pequena magnitude de efeitos, as estratégias de melhoramento genético aplicada a características quantitativas dependem fundamentalmente dos sistemas de acasalamento, que regulam como gametas se unem na fertilização. Apesar de diferentes estratégias de melhoramento e de particularidades de plantas autógamas e alógamas as questões de predição de valores genéticos são, essencialmente, as mesmas em especial quando se agregam informações moleculares. Desta forma, o objetivo deste trabalho é avaliar a eficiência da seleção genômica ampla (GWS) na predição do valor genético em gerações avançadas por

autofecundação e acasalamento ao acaso, com e sem avaliação fenotípica. Como todo estudo se fundamenta em dados simulados, constitui-se também objetivo do trabalho constatar a própria eficácia do processo de simulação em gerar populações e gerações cujos princípios genéticos sejam preservados conforme princípios fundamentos da genética quantitativa e de populações.

Material de Métodos

Origens das populações

Para estudar o impacto da seleção genômica ampla (GWS) em gerações avançadas por autofecundação e acasalamento ao acaso, foram simuladas cinco repetições de uma estrutura populacional. A estrutura de cada repetição foi criada para apresentar dois conjuntos de dados: fenotípicos e genotípicos de cada indivíduo dentro da população, imitando alguns dos cenários em que a GWS é aplicada. Em cada sequência de dados duplicados foram considerados 1500 locos, e um tamanho efetivo (N_e) das populações de 1000 indivíduos para todas as gerações originadas. Os dados de simulação foram obtidos por meio do aplicativo GENES 2013.1.1 (Cruz, 2012) onde se considerou uma população F_1 , derivada do cruzamento entre duas populações genitoras P_1 e P_2 , em equilíbrio de Hardy-Weinberg. Todas estas gerações foram obtidas por simulação. Inicialmente foram simulados 1000 indivíduos de cada população os quais foram genotipados em relação a 1500 locos que expressam dois alelos codominantes em cada loco. Posteriormente a população F_1 foi submetida a 10 ciclos de acasalamento ao acaso e 10 autofecundações para obter as gerações avançadas. Para fins práticos e didáticos as populações analisadas foram denominadas de A_j ($j = 1, 2, \dots, 10$) para as gerações derivadas por acasalamento ao acaso e representativas de populações alógamas

e S_j ($j=1,2...10$) para as gerações derivadas por autofecundação e representativas de populações autógamas.

Simulação dos dados genotípicos e fenotípicos

Para proceder às análises de GWS, foram simulados dados genotípicos e fenotípicos considerando ausência de dominância e ação aditiva entre os genes, diferentes herdabilidades (h^2), número de locos controlando a característica e tipo de marcador. Foram simulados três caracteres quantitativos com herdabilidades iguais a 0,20, 0,40, 0,60 controlados cada um por 500 locos, considerando dois alelos por loco.

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

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

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

Em que:

$$a_i = \sum_{j=1}^{500} p_j \alpha_j \qquad d_i = 0$$

Sendo α_j é o efeito do alelo favorável no loco j , considerado igual a 1, 0 ou -1 para as classes genotípicas AA, Aa e aa, respectivamente, e p_j é a contribuição do loco j para a manifestação da característica considerada, no trabalho, como tendo distribuição binomial (Figura 1). Foi estabelecido que os 500 primeiros locos genotipados foram os controladores da característica.

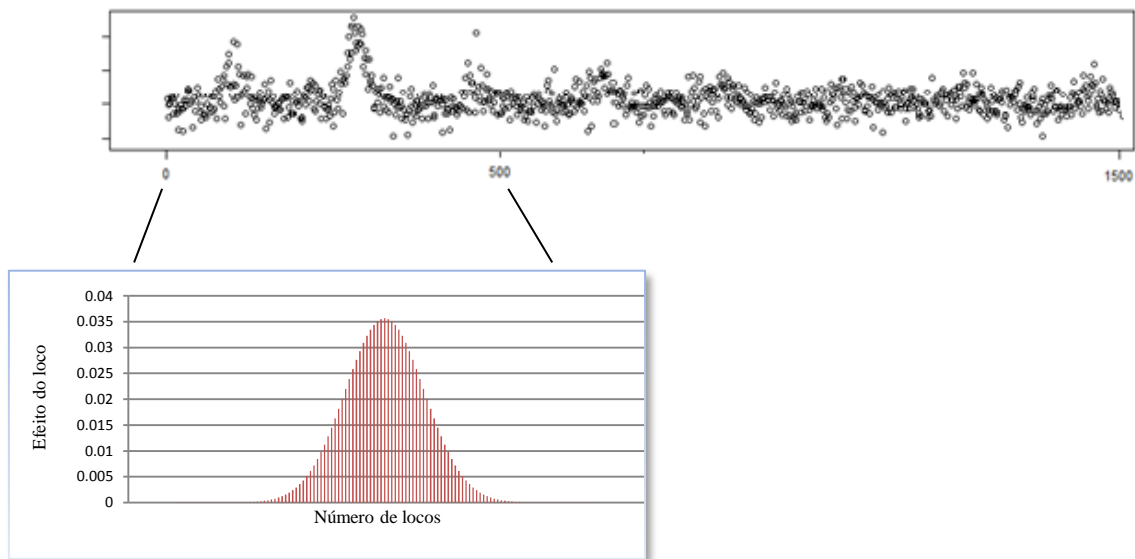


Figura 1. Distribuição binomial do efeito do loco sobre uma característica quantitativa.

Caracterização genética das populações simuladas

Para entender o comportamento das populações é importante o conhecimento dos parâmetros genéticos e estatísticos como consequências do uso de métodos de melhoramento genético. Sendo assim, foram utilizados alguns conceitos de genética que estão relacionados com a constituição genética dos indivíduos isoladamente ou dentro das populações, para assim discorrer sobre os principais fatores que modificam a estrutura genética de uma população tendo em vista os acasalamentos ao acaso e as sucessivas gerações de autofecundação nos dois cenários estudados.

Nos processos biológicos, é praticamente impossível medir todos os indivíduos da população. Por essa razão, com base nos valores genotípicos e fenotípicos das populações em todas as gerações, foram estimados os parâmetros: média (μ), variância (σ^2), desvio padrão (σ), e a endogamia na população gerada pela autofecundação ou pelo acasalamento ao acaso, em consequência do tamanho finito da população. Além da estimação de parâmetros, as populações foram caracterizadas quanto a distribuição de

seus dados sendo feitos testes de normalidade e obtenção de coeficientes de simetria (s) e curtose (k) nas diversas gerações.

Metodologia de análise

A GWS depende das informações sobre os dados fenotípicos e genotípicos dos indivíduos de uma população. Essas duas informações preliminares são usadas, em conjunto, para atingir o objetivo final da GWS que é a geração de um modelo de predição capaz de prever o valor genômico dos indivíduos da população-base e tal modelo ser utilizado, com base apenas nas informações genotípicas, para prever valores genéticos de indivíduos em gerações futuras. As variáveis preditoras são o conjunto de marcadores, o que requer a estimação, em uma etapa inicial, da contribuição (efeito) de cada marcador em explicar o fenótipo. Essas estimativas são utilizadas no modelo de predição, e em conjunto, compõem o Valor Genético Genômico (VGG) do indivíduo. Tomando como origem os dados fenotípicos, foram estimados os efeitos de cada um dos locos marcadores que somados, compõem o valor genético genômico predito de cada indivíduo. Os efeitos foram preditos por meio do procedimento BLUP/GWS que permite ajustar todos os efeitos alélicos simultaneamente.

Aplicou-se para estimar os efeitos dos marcadores, o modelo linear misto geral conforme Resende et al., (2008):

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

em que y é o vetor de observações fenotípicas, b é o vetor de efeitos fixos (média geral), h é o vetor dos efeitos aleatórios dos marcadores e e refere-se ao vetor de resíduos aleatórios. X e Z são as matrizes de incidência para b e h . A estrutura de médias e variâncias no modelo em questão é definida como:

$$\begin{aligned}
h &\sim N(0, G) & E(y) &= Xb \\
e &\sim N(0, R = I\sigma_e^2) & \text{Var}(y) &= V = ZGZ' + R \\
G &= \sum_i^N I\sigma_{gi}^2 = I\sigma_g^2
\end{aligned}$$

As equações de modelo misto genômicas para a predição de h via o método BLUP/GWS equivalem a:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I\frac{\sigma_e^2}{\sigma_g^2/n} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{h} \end{bmatrix} = \begin{bmatrix} X' & y \\ Z' & y \end{bmatrix} \text{ em que } \sigma_e^2 \text{ é a variância residual e } \sigma_g^2 \text{ é a variância}$$

genética. O valor genético genômico global do indivíduo j é dado por:

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

em que n é o número de marcadores dispostos no genoma, Z_i é a linha da matriz de incidência que aloca o genótipo do i -ésimo marcador para cada indivíduo, 0, 1 ou -1 para os genótipos A_1A_1 , A_1A_2 e A_2A_2 , respectivamente, para marcadores bialélicos e codominantes, e \hat{h}_i é o efeito estimado do i -ésimo marcador.

As equações de predição apresentadas acima assumiram, a priori, que todos os locos explicam iguais quantidades da variação genética. Assim, a variação genética explicada por cada loco é dada por σ_g^2/n , em que σ_g^2 é a variação genética total e n é o número de marcadores utilizados. [Meuwissen et al. 2001; Zhang et al., 2011; Bernardo & Yu, 2007 implementarem esta estratégia]. Na predição dos efeitos aleatórios via BLUP/GWS, não há necessidade do uso da matriz de parentesco (Schaeffer, 2006), pois a matriz de parentesco baseada em pedigree usada no BLUP tradicional é substituída pela própria matriz $Z'Z$ que é uma matriz de parentesco estimada pelos marcadores. Os resultados dos VGG, juntamente com os valores fenotípicos e genotípicos foram

apresentados em gráficos empregando o programa SigmaPlot 10.0 e tomando como referência 200 indivíduos.

A seleção genômica ampla requer o uso de uma população de estimação para estimar os efeitos dos marcadores e uma população de validação, para analisar a eficiência da estimação destes efeitos na recuperação do valor genômico em uma população independente. O trabalho tem como população de estimação os 1000 indivíduos de cada geração e considera como população de validação a fenotipagem e genotipagem dos indivíduos na mesma geração, e o fenotipagem das gerações anteriores de tal maneira que os 1000 indivíduos da geração anterior eram utilizados na estimação dos valores genômicos preditos. Uma vez estimados todos os efeitos, estes eram aplicados na população de validação para prever o valor genômico.

Avaliação dos dados

A confiabilidade é uma medida do quadrado da correlação entre o valor estimado e os valores verdadeiros, ou seja, mede o quanto a estimativa obtida é relacionada com o valor real do parâmetro. Ela informa o quanto o valor estimado é bom, ou o quanto o valor estimado é próximo do valor real e dá a confiabilidade daquela estimativa ou valor.

A eficácia da GWS para cada situação testada, foi avaliada calculando a correlação do valor genético predito com o genótipo conhecido dos 1000 indivíduos a cada geração de acasalamento ao acaso e autofecundação. Esta correlação é conhecida como a capacidade preditiva ($r_{y\hat{y}}$) da seleção genômica em estimar os fenótipos e ela é dada teoricamente pela acurácia de seleção ($r_{q\hat{q}}$) multiplicada pela raiz quadrada da herdabilidade individual (h) ou, em outras palavras, $r_{y\hat{y}} = r_{q\hat{q}}h$ (Resende et al., 2008). Assim, as confiabilidades obtidas pela GWS foram comparadas nas gerações avançadas

de acasalamento empregando o genotipagem e o fenotipagem da mesma geração, e a genotipagem de uma geração com o fenotipagem das anteriores gerações.

Aplicativos computacionais para análise de dados

As análises de simulação e parâmetros genéticos foram desenvolvidos no programa GENES 2013.1.1, amplamente utilizado em análises de modelos aplicados ao melhoramento de plantas e animais. É um software destinado à análise e processamento de dados por meio de diferentes modelos biométricos, contando com procedimentos uni e multivariados, enfatizando estimação de parâmetros genéticos.

Quanto às análises de estimação relacionados com a GWS foram desenvolvidas como um pacote rrBLUP, desenvolvido por Endelman, 2011, do software R.

Resultados e Discussão

As características genéticas a serem melhoradas em uma espécie agrícola, podem ser de dois tipos: caracteres qualitativos ou caracteres quantitativos. Os caracteres qualitativos são aqueles governados por um ou poucos genes. Contudo, grande parte das características agrônômicas de interesse dos melhoristas, apresentam herança quantitativa. Os caracteres quantitativos ou poligênicos são aqueles governados por múltiplos genes, sendo que cada gene apresenta segregação conforme as “Leis de Mendel”. Além disso, quando se analisa uma população segregante, observa-se que os caracteres de herança quantitativa apresentam distribuição contínua de fenótipos. Outra característica dos caracteres poligênicos, refere-se ao fato de serem influenciados pela variação do ambiente, dificultando ainda mais a identificação dos genótipos com base apenas no fenótipo observado.

O conhecimento da estrutura genética de uma população é indispensável ao melhorista para realizar sobre ela mudanças em magnitude e sentido desejado. O conhecimento da estrutura da população inclui o conhecimento da frequência dos alelos que compõem os diferentes genótipos das famílias, da frequência genotípica do sistema de acasalamento. Cada um dos sistemas reprodutivos, resulta em dinâmicas gênicas e fenotípicas contrastantes, o qual pode ser visto nos resultados apresentados a seguir. O conhecimento acumulado destas dinâmicas permitiu desenvolver e entender metodologias que fundamentam os diferentes métodos de melhoramento vegetal, utilizados atualmente, seja para plantas autógamas ou alógamas.

Com a finalidade de averiguar o que aconteceria no sistema real de populações se alterações de interesse fossem efetuadas em seu funcionamento, foi construído com simulação um sistema que imita o funcionamento de uma realidade. Já que, informações valiosas podem ser extraídas desse sistema simulado, com menor custo e maior rapidez. No sistema real, muitas opções de alteração são inviáveis de serem avaliadas, seja pelos custos que podem ser elevados, seja pelos longos períodos de resposta.

Dinâmica das populações avançadas por autofecundação e acasalamento ao acaso

A distinção do sistema reprodutivo é primordial para o melhoramento de uma espécie, pois todos os métodos destinados à condução de populações segregantes e destinados à seleção de plantas superiores, dependem da base genética de cada espécie e caracterização de estrutura. O conhecimento do sistema reprodutivo de uma população é fundamental para a condução adequada de um programa de melhoramento, já que os métodos aplicados para esse fim são diferentes e específicos, em função do sistema de reprodução prevalecente na população. Sabe-se que a ocorrência de cruzamentos e de autofecundações, levam a estruturas genéticas populacionais bem distintas em gerações

avançadas e isto pode ser comprovado por meio da média, da variância, da distribuição dos valores e da endogamia.

Verifica-se com base aos valores fenotípicos, que as medias permanecem similares com o passar dos ciclos de autofecundação (Tabela 1), em consequência da ausência de dominância ente os alelos. Tal fato também ocorreu com os ciclos de acasalamento ao acaso (Tabela 2) demonstrando que os processos de simulação foram capazes de retratar o comportamento genético esperado para a média populacional. Pode ser observado que a média , para a característica com h^2 0,20 da geração S_1 (20.077) é igual á media da geração F_1 (20.074) e, mesmo após dez gerações de autofecundação (S_{10}), continua com valor próximo igual a 20.078. Com dominância populações derivadas de acasalamento entre aparentados exibem depressão endogâmica, condicionada à redução da contribuição de locos em heterozigose que devem contribuir para o aumento da média do caráter em razão da heterose ou vigor híbrido. Neste estudo a depressão não é esperada, tendo em vista que foi assumido ausência de dominância entre os alelos controladores do caráter e, portanto, as médias obtidas traduzem apropriadamente este fenômeno, mesmo com os dados simulados de uma característica governada por 500 locos. Para as gerações obtidas de acasalamento ao acaso, para o caráter equivalente de herdabilidade igual a 0,20, se observa valores da média de A_1 de 20.080 e, após 10 gerações de acasalamento ao acaso, a média se manteve inalterada apresentando em A_{10} , valor igual a 20.096. O fato relatado também foi observado em relação às demais características simuladas neste estudo.

Tabela 1. Médias populacional das gerações avançadas de uma população F_1 por autofecundação, para três caracteres de herdabilidade 20, 40 e 60%.

h^2	F_1	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
h20	20.0748	20.0774	20.079	20.0782	20.0774	20.0771	20.0781	20.0787	20.0787	20.0788	20.0788
h40	40.0748	40.0774	40.079	40.0782	40.0774	40.0771	40.0781	40.0787	40.0787	40.0788	40.0788
h60	60.0748	60.0774	60.079	60.0782	60.0774	60.0774	60.0781	60.0787	60.0787	60.0788	60.0788

Tabela 2. Médias populacional das gerações avançadas de uma população F_1 , por acasalamento ao acaso, para os três caracteres de herdabilidade 20, 40 e 60%.

h^2	F_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
h20	20.0748	20.0801	20.0757	20.0758	20.0852	20.0928	20.093	20.096	20.0981	20.0995	20.0967
h40	40.0748	40.0801	40.0757	40.0758	40.0852	40.0928	40.093	40.096	40.0981	40.0995	40.0967
h60	60.0748	60.0801	60.0757	60.0758	60.0852	60.0928	60.093	60.096	60.0981	60.0995	60.0967

Em todas as amostras ou populações, ocorre variabilidade dos indivíduos que as constituem. Além disso amostras com mesma média como nesse caso, podem apresentar distribuições diferentes. portanto, só a média não nos dá uma idéia clara de como os dados se distribuem. Outro padrão que deve ser observado, nas gerações avançadas por autofecundação e acasalamento ao acaso, é a distribuição dos dados de características simuladas considerando ação poligênica e distribuição normal. Assim, mediante o teste de normalidade (Figura 2), pode ser observado um padrão de distribuição consistente normal, com curtose (2.6218 ± 0.6) e simetria (-0.1223 ± 0.1) (Tabela 3 e 4) e frequências observadas e esperadas próximas, em todas as gerações de autofecundação e acasalamento ao acaso. É visto na Figura 2A que o padrão de distribuição da população A_1 é conservado até a população A_{10} independente da característica ser de alta ou baixa herdabilidade. Na Figura 2B são apresentadas figuras ilustrativas da distribuição para as mesmas gerações, porém considerando a característica de herdabilidade igual a 60%, onde se percebe a curvatura e simetria conservada e próxima da distribuição normal. Resultados similares ocorreram para as gerações de populações derivadas por autofecundação (Figuras 2C e 2D).

Tabela 3. Valores de Curtose e simetria populacional das gerações avançadas de uma população F₁ por acasalamento ao acaso, para três caracteres de herdabilidade 20, 40 e 60%.

	Curtose				simetria		
	h20	h40	h60		h20	h40	h60
F ₁	2.9921	2.8518	3.2250	F ₁	0.055	0.1024	-0.0743
A ₁	2.8230	3.0222	3.1534	A ₁	-0.0527	-0.157	0.0298
A ₂	2.9519	2.9037	2.9071	A ₂	-0.0716	0.0567	0.0761
A ₃	3.0938	2.8929	2.9777	A ₃	0.0475	0.0567	-0.0146
A ₄	3.1567	2.8409	3.2240	A ₄	0.0052	0.0173	0.1029
A ₅	2.8266	3.4562	3.1772	A ₅	0.0588	-0.1223	0.0151
A ₆	2.6218	2.9684	2.8561	A ₆	-0.1237	0.0273	0.0419
A ₇	2.9398	2.9423	2.9562	A ₇	0.0123	0.0845	-0.0573
A ₈	3.2176	3.0513	3.0068	A ₈	0.0219	-0.0619	0.0163
A ₉	3.2144	2.9491	3.1664	A ₉	-0.0712	-0.0322	0.1301
A ₁₀	2.8242	2.8983	3.0635	A ₁₀	-0.0417	0.083	-0.063
Media	2.969264	2.979736	3.064855	Media	-0.01456	0.004955	0.018455

Tabela 4. Valores de Curtose e simetria populacional das gerações avançadas de uma população F₁ por autofecundação, para três caracteres de herdabilidade 20, 40 e 60%.

	Curtose				simetria		
	h20	h40	h60		h20	h40	h60
F ₁	2.9921	2.8518	3.2250	F ₁	0.055	0.1024	-0.0743
A ₁	2.8780	3.0532	2.9801	A ₁	0.1159	0.0107	0.1067
A ₂	3.1311	2.9780	3.4037	A ₂	0.0705	-0.0164	-0.1699
A ₃	2.9155	2.9847	3.1812	A ₃	0.0243	-0.02	-0.1507
A ₄	2.8098	2.8456	3.0185	A ₄	-0.0046	-0.1217	0.0327
A ₅	3.4106	3.2204	3.0185	A ₅	0.0493	0.0744	0.0327
A ₆	2.9654	2.8034	3.0810	A ₆	0.0402	0.0385	-0.0716
A ₇	2.9126	2.9553	2.8715	A ₇	-0.0722	0.0584	-0.0354
A ₈	2.9029	3.3854	2.9948	A ₈	-0.026	0.041	0.0349
A ₉	2.8477	2.8151	3.0847	A ₉	-0.0803	0.0461	0.1105
A ₁₀	2.9638	2.7587	2.9250	A ₁₀	0.1130	0.0995	0.0695
Media	2.975409	2.968327	3.071273	Media	0.025918	0.028445	-0.01045

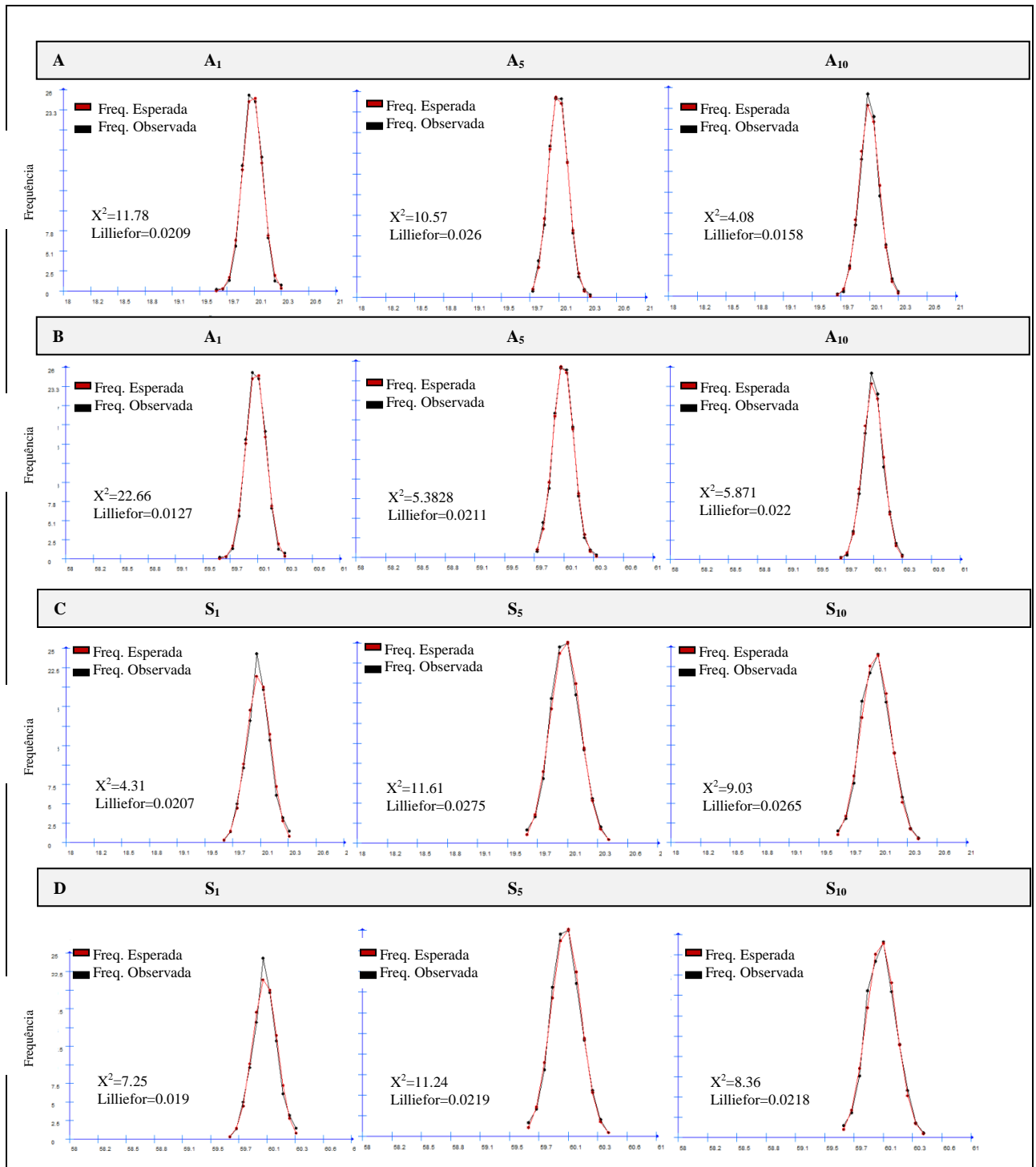


Figura 2. Distribuição de frequência do número total de indivíduos avaliados nos diferentes cenários simulados, apresentando as frequências observadas e esperadas, A. Populações A₁, A₅ e A₁₀ para característica h² 0,20; B. Populações A₁, A₅ e A₁₀ para h² 0,6; C. Populações S₁, S₅, S₁₀ para característica h² 0,20, D. Populações S₁, S₅, S₁₀ para h² 0,60.

Nas tabelas 5 e 6 são apresentados os coeficientes de endogamia (F), para as populações avançadas por autofecundação e acasalamento ao acaso, respectivamente. Nas populações, ou gerações, que caracterizam espécies autógamas em que o processo de acasalamento é por autofecundação constata-se incremento no grau de endogamia, sendo observado valor igual a -0.1608, que traduz ligeiro excesso de heterozigotos, na geração F₁ original, ou referencial, até valor de 0.9987 na geração S₁₀. Deve ser destacado que a endogamia medida no presente trabalho refere-se à comparação entre valores observados (h_o) e esperados (h_e) de heterozigotos de 1500 locos gênicos, obtidos mediante um processo de simulação, que ressalta o fato de a simulação ser de grande utilidade em estudos genéticos sobre vários contextos, incluindo estudos de populações, do indivíduo ou do próprio genoma. Nas plantas autógamas é conhecido que a medida que ocorram as autofecundações os locos controladores dos caracteres tem a sua heterozigosidade reduzida e, com base em modelos preditivos, geralmente fundamentados em apenas um loco gênico, sabe-se que as sucessivas autofecundações reduzem a proporção de heterozigotos à metade em cada geração aumentando o coeficiente de endogamia (Tabela 5). Sendo assim teoricamente era esperado valores de endogamia da geração F₁ até S₁₀ como sendo 0, 0.5, 0.75, 0.875, 0.937, 0.968, 0.984, 0.992, 0.996, 0.998, 0.999, valores estes muito próximos aos observados (Tabela 5).

Tabela 5. Coeficientes de endogamia (F) para geração avançadas por autofecundação.

Geração	he	Ho	F
F ₁	0.423511	0.491627	-0.16084
S ₁	0.423344	0.244445	0.422586
S ₂	0.423355	0.12219	0.711377
S ₃	0.423339	6.09E-02	0.856251
S ₄	0.423335	3.07E-02	0.92752
S ₅	0.423287	1.49E-02	0.96481
S ₆	0.423307	7.47E-03	0.982356
S ₇	0.423289	3.74E-03	0.991158
S ₈	0.423278	1.88E-03	0.995554
S ₉	0.423275	9.83E-04	0.997677
S ₁₀	0.423273	5.37E-04	0.998731

ho: heterozigose observada; he: heterozigose esperada sobre a hipótese de equilíbrio de Hardy wemberg; F: coeficiente endogamia.

Tabela 6. Coeficientes de endogamia (F) para geração avançadas por acasalamento ao acaso

Geração	he	ho	F
F ₁	0.423511	0.491627	-0.16084
A ₁	0.423592	0.423334	6.08E-04
A ₂	0.422999	0.424319	-3.12E-03
A ₃	0.422386	0.422525	-3.28E-04
A ₄	0.422677	0.42124	3.40E-03
A ₅	0.42249	0.422694	-4.84E-04
A ₆	0.422077	0.423461	-3.28E-03
A ₇	0.42204	0.421934	2.52E-04
A ₈	0.421964	0.421153	1.92E-03
A ₉	0.421528	0.422773	-2.95E-03
A ₁₀	0.421022	0.421476	-1.08E-03

ho: heterozigose observada; he: heterozigose esperada sobre a hipótese de equilíbrio de Hardy wemberg; F: coeficiente endogamia.

A endogamia medida nas gerações derivadas de acasalamento ao acaso de uma população referencial F₁ foi quantificada e apresentada na Tabela 6. Estes valores retratam o que seria esperado em uma população de plantas alógamas e mostram valores muito baixos de F indicando que o tamanho populacional empregado não foi suficiente para gerar endogamia significativa. Assim, como exemplo, pode ser analisada a geração A₁ com um valor F de -0.0006 e, após 10 sucessivas gerações de acasalamento ao acaso, apresentou valor F de -0.001. De maneira geral, verifica-se que não houve, como esperado, endogamia na população, nem fixação de formas genotípicas homozigotas e nem alteração nos valores de variância genotípica (Tabela 8). Quando as populações de espécies alógamas estão se reproduzindo segundo seu sistema reprodutivo natural, e na ausência de endogamia, desconsiderando-se eventos de mutação, seleção, migração e deriva genética, isto é, as frequências genotípicas e gênicas, permanecem inalteradas em consequência do encontro inteiramente ao acaso entre os gametas que se unem. É sabido que em tais condições as populações encontram-se no estado de “Equilíbrio perfeito” (Equilíbrio de Hardy-Weinberg); neste sentido, uma população de cruzamentos aleatórios encontra-se nas frequências $[(p + q)^2 = p^2 + 2pq + q^2 = 1]$, para cada loco

gênico. Tal fato foi preservado no processo de simulação e constatado para um conjunto relativamente grande de locos gênicos (1500 locos).

A endogamia leva à fixação de formas homozigotas e, conseqüentemente, à ampliação da variabilidade genética total na população como pode ser comprovado nos valores obtidos e descritos na Tabela 7. Em estudos biométricos comprova-se que a variância genética total em uma população derivada de n gerações de autofecundação é expressa por meio de:

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

Sendo assim na geração em que o coeficiente de endogamia é igual a 0, a variância aditiva total é $\sigma_A^2 + \sigma_D^2$, porém como foi considerado no processo de simulação ação aditiva ou seja variância de dominância igual a zero, a variância genética total era de σ_A^2 . Como o aumento do coeficiente de endogamia com o passar das gerações de autofecundações na S₁₀ o coeficiente de endogamia era de 0.999, próximo a unidade, onde seria esperado que a variância total seria de $2 \sigma_A^2$.

Desta forma os valores apresentados na Tabela 7 ratifica a expectativa, tendo em vista que, no presente estudo foi pressuposto ausência de dominância entre alelos dos diferentes locos gênicos que controlam o caráter, e, portanto, a variância esperada em S₁₀ deveria atingir o dobro do valor manifestado na geração referencial F₁ com endogamia nula. Para as gerações derivadas de acasalamento ao acaso a variabilidade, função das frequências genotípicas, deve ser preservada com valores apresentados em A_i (i=1,2..10) próximos dos manifestados na F₁ (Tabela 8).

Tabela 7. Valores de variância genotípica obtidos para dez gerações de autofecundação.

h^2	F_1	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
h20	0.0434	0.0853	0.0913	0.1076	0.1105	0.1184	0.1135	0.1096	0.1182	0.1198	0.112
h40	0.0225	0.0403	0.0459	0.0492	0.0557	0.0575	0.0565	0.0564	0.0576	0.0574	0.058
h60	0.0155	0.0276	0.0312	0.0333	0.0372	0.0372	0.0401	0.0388	0.0367	0.0376	0.0399

Tabela 8. Valores de variância genotípica obtidos para dez gerações de acasalamento ao acaso

h^2	F_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
h20	0.0434	0.053	0.0556	0.0532	0.0554	0.0507	0.0545	0.0537	0.0556	0.0576	0.0564
h40	0.0225	0.0289	0.0268	0.0256	0.0269	0.0257	0.026	0.0255	0.0276	0.0296	0.0271
h60	0.0155	0.0172	0.0188	0.0168	0.0187	0.0172	0.0174	0.0189	0.018	0.0195	0.0167

Os resultados de todos os parâmetros e informações genéticas foram similares nas cinco repetições efetuadas, permitindo inferir que a utilização dos dados obtidos por simulação para a formação das populações sob análise é um processo confiável. Tal fato permitirá inferir e extrapolar os resultados das análises efetuadas, considerando as particularidades da seleção genômica ampla para espécies de plantas autógamas e alógamas cujos processos reprodutivos foram considerados e suas consequências genéticas foram preservadas e adequadamente observadas nas gerações obtidas e estudadas.

Valor genético genômico (VGG)

Foram referenciados três critérios de seleção para fins de utilização em diferentes estratégias de melhoramento. O primeiro consiste em tomar como base os valores fenotípicos dos indivíduos avaliados cujas confiabilidades, medidas pelo quadrado da correlação entre o valor fenotípico e genotípico, são expressas pelas herdabilidades das características, estabelecidas em 20, 40 e 60%. Nestas situações, em especial na herdabilidade mais baixa, pode-se prever que o melhorista terá certa dificuldade em selecionar genótipos que, de fato, apresentam desempenho superior ou,

ainda, de descartar aqueles de desempenho genético não favorável uma vez que a associação entre o fenótipo e genótipo é de baixa magnitude em razão da ação dos efeitos perturbadores do ambiente. O segundo critério faz uso do valor genômico em que a seleção é praticada sobre valores genéticos preditos que levam em consideração os valores fenotípicos, mas agregam consideráveis informações genéticas obtidas a partir da genotipagem realizada. E, por fim, o critério utilizado nas gerações avançadas em que a seleção praticada sobre valores genômicos preditos que levam em consideração os efeitos dos marcadores, sem informação fenotípica da própria geração, mas com os pesos estimados por meio da análise genômica ampla considerando informações sobre a genotipagem e fenotipagem de uma determinada geração anterior a da geração analisada.

Assim, na geração inicial o melhorista dispõe, para fins de seleção, o valor fenotípico (V_{fen} - valor mensurado do genótipo sob influência do meio), o valor genômico (V_{GG} - valor predito pela GWS) e, neste estudo que é fundamentado em simulação, também é conhecido o valor genotípico verdadeiro (V_{gen}). Nas demais gerações (ou geração t), além destas três informações, também eram disponíveis os $t-1$ conjuntos de valores genômicos preditos indiretos obtido pelas informações da genotipagem da geração cujos efeitos dos marcadores era estimado nas gerações $i = 1, 2, \dots, t-1$.

Para fins de melhoramento convencional, o pesquisador teria a sua disposição, em condições de campo, apenas informações fenotípicas resultantes de sua mensuração que é afetada, em maior ou menor intensidade, pelas causas não genéticas. Entretanto, com os avanços em biologia molecular é possível que os melhoristas possam realizar a genotipagem dos indivíduos e agregar valiosa informação a ser utilizada como critério de seleção. Neste trabalho houve a comprovação de que o quadrado da correlação entre

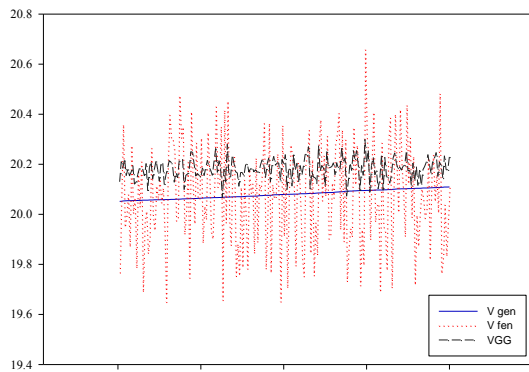
os valores fenotípicos e genotípicos dos indivíduos, para todas as populações (ou gerações), foi próximo aos valores das herdabilidades paramétricas estabelecidas para as características analisadas.

Nas Figuras 3 e 4 é apresentado o comportamento do valor genético genômico (VGG) para 200 indivíduos em relação aos valores fenotípicos e genotípicos dos mesmos para três características diferentes quanto ao grau de herdabilidades associada (0,20, 0,40 e 0,60). Pode ser observado que, para as populações representativas das espécies alógamas (Figura 3), o VGG encontra-se dispersos acompanhando os valores genotípicos verdadeiros, para características de herdabilidade 20 e 40 % e, com a vantagem, de apresentar menor amplitude de variação em torno do valor genotípico tanto nas gerações iniciais quanto nas gerações avançadas. Para a característica com herdabilidade de 60 % o VGG apresenta maior relação com o valor fenotípico, em especial na geração inicial (A_1) podendo prever que as informações moleculares pode não acrescentar grandes vantagens em casos de herdabilidade elevada. Entretanto na geração A_{10} , mesmo com herdabilidade igual a 0,60 a tendência foi a mesma, ou seja, o valor genômico acompanha o valor genotípico com menor variação em torno destes valores verdadeiros. A Figura 3 também possibilita visualizar a ação do ambiente como agente perturbador do processo seletivo de considerável magnitude, reduzindo os ganhos, permitindo que genótipos não tão superiores favorecidos pelo ambiente contribuíssem para a próxima geração e que genótipos superiores, com desempenho prejudicado pelo ambiente fossem descartados.

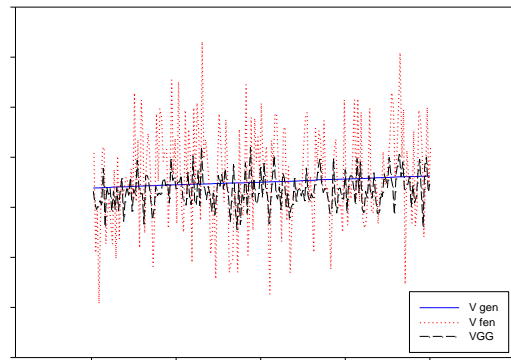
Com relação às populações representativas das espécies autógamas os resultados dos valores genotípicos, fenotípicos e genômicos são apresentados na Figura 4. Pode-se observar padrão similar para os três caracteres estudados na primeira geração de autofecundação. É de destacar que o VGG com o transcorrer das gerações apresenta

valor de frequência inferiores aos valores genotípicos apresentando um viés conforme pode ser visto na figura 4, assim a metodologia validada para alógamas não é adequada para autógamias, sendo necessária outra parametrização na matriz de parentesco genômico, precisando considerar mudança de escala de variação genética quando o coeficiente de endogamia muda de 0 para 1.

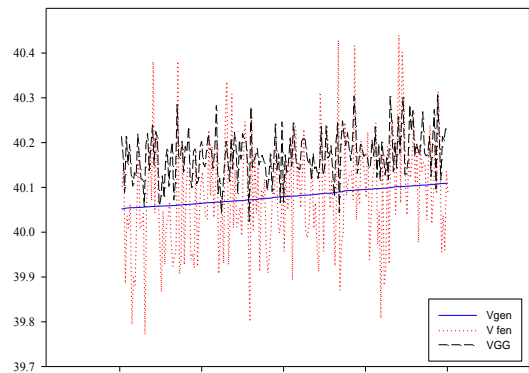
A₁- característica de $h^2 = 0.20$



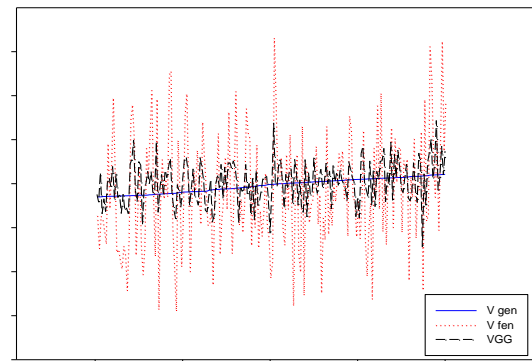
A₁₀ - característica de $h^2 = 0.20$



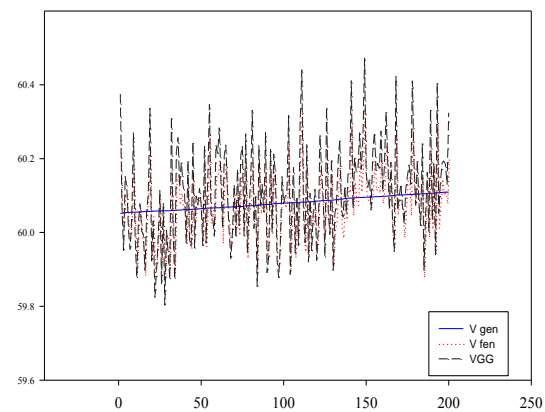
A₁- característica de $h^2 = 0.40$



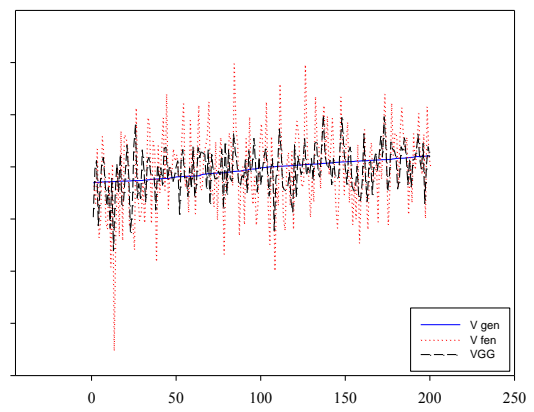
A₁₀ - característica de $h^2 = 0.40$



A₁- característica de $h^2 = 0.60$



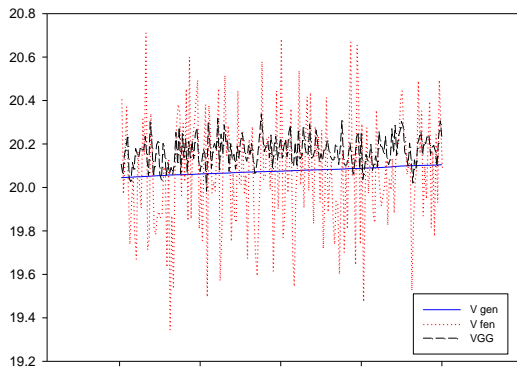
A₁₀ - característica de $h^2 = 0.60$



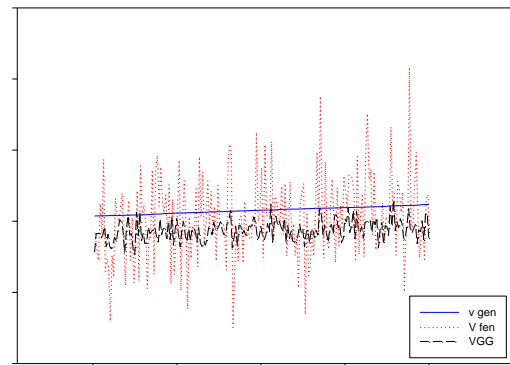
Indivíduos

Figura 3. Relação dos Valores genéticos genômicos de estimação (VGG) com valores genéticos (Vgen) e com valores fenotípicos (Vfen) para acasalamento ao acaso em gerações A₁ e A₁₀ para três caracteres de herdabilidade (20%, 40%, 60%).

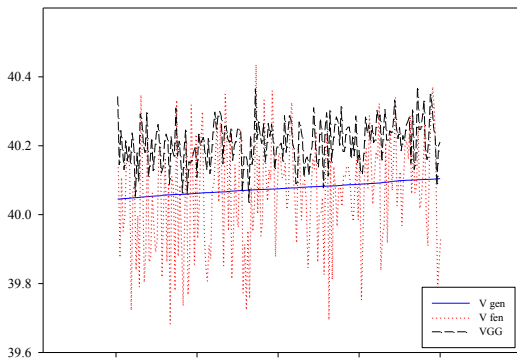
S₁- característica de h² = 0.20



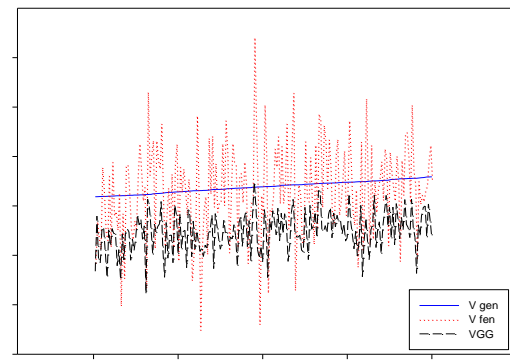
S₁₀ - característica de h² = 0.20



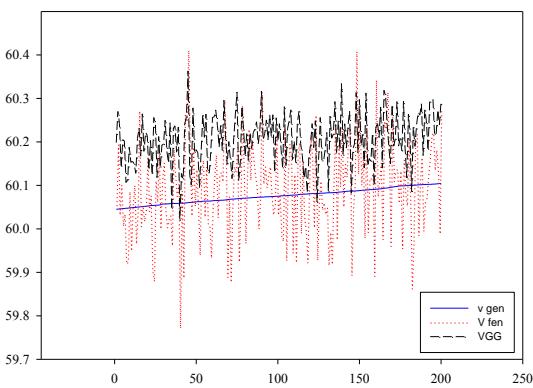
S₁- característica de h² = 0.40



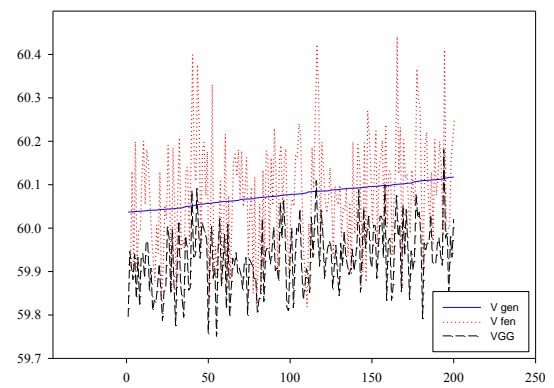
S₁₀ - característica de h² = 0.40



S₁- característica de h² = 0.60



S₁₀ - característica de h² = 0.60



Indivíduos

Figura 4. Relação dos Valores genéticos genômicos de estimação (VGG) com valores genéticos (Vgen) e com valores fenotípicos (Vfen) para autofecundação em dois gerações F_{2a1} e F_{2a10} para três caracteres de herdabilidade (20%, 40%, 60%).

Estimativas de confiabilidade e correlação

No melhoramento clássico a seleção é praticada tomando como critério os valores fenotípicos dos indivíduos que deve representar apropriadamente o desempenho genético de um indivíduo, permitindo distinguir a sua superioridade em relação a outros sob seleção. Entretanto, a ação do ambiente reduz a associação entre fenótipo e genótipo tornando a seleção menos efetiva. Neste contexto, é apresentada a GWS como forma de melhorar o critério de seleção agregando, aos valores fenotípicos, informações moleculares advinda de um processo de genotipagem. Neste trabalho pode-se observar (Tabelas 9 e 10) o quanto este procedimento proporciona benefícios à prática da seleção.

Como já preconizado, após ser feita a genotipagem na população e obtenção dos valores genômicos, para fins de seleção, os valores de confiabilidade calculados para as três características simuladas em todos os cenários testados (gerações e sistemas de acasalamento), apresentarem substancial aumento (Tabelas 9 e 10).

Para os cenários em que foram consideradas gerações representativas de espécies alógamas (Tabela 9), os valores de confiabilidade de seleção por meio do valor genômico sempre é superior, com pequenas variações e algumas vezes acréscimos à medida que as gerações avançam, não sendo tão afetada pela redução do desequilíbrio de ligação da própria geração. Em muitas situações o acréscimo na acurácia é substancial indicando que a inclusão de informações sobre a genotipagem possa ser valiosa mesmo tendo em vista os custos adicionais que esta informação possa trazer para a pesquisa. Assim, deve ser ressaltado os resultados obtidos para as diversas gerações e a característica de baixa herdabilidade ($h^2 = 0,20$) onde se verificaram acréscimos na eficiência preditiva pois as confiabilidades variaram de 0.3381 na geração A_1 para 0.4791 na geração A_{10} . Deve-se ter em mente que dobrar o valor de

uma herdabilidade por técnicas convencionais pode não ser tão fácil ou exigir também custos elevados em ampliação da variabilidade genética, por inclusão de maior número de genótipos na avaliação, ou reduzir a influencia ambiental por meio de delineamentos, controle de fatores bióticos e abióticos e dimensionamento do experimento.

Para as gerações de acasalamento ao acaso e características com maiores herdabilidade ($h^2 = 0,40$ e $h^2 = 0,60$) foi verificado acréscimos atingindo valores de confiabilidade de 0.6066 e 0.7298 na geração A_1 , respectivamente, e valores de 0.6458 e 0.7292 na geração A_{10} , respectivamente. Bernardo & Yu (2007) demonstraram, via simulação, o potencial e as perspectivas do uso da GWS para características quantitativas em milho. Ao simular a performance test-cross de duplos haplóides em três ciclos de seleção, baseada em informações dos marcadores, para situações onde 20, 40 ou 100 QTLs estavam envolvidos no controle genético de características quantitativas (de diferentes herdabilidades), esses autores verificaram que a resposta a seleção foi 18-43% maior via GWS que a resposta via seleção recorrente assistida por marcadores moleculares.

Segundo Bernardo & Yu (2007), o esquema da GWS aplicados à gerações posteriores que minimiza fenotipagem e maximiza genotipagem é bastante favorável ao melhoramento da espécie, principalmente se o custo da genotipagem for bastante reduzido. Pode ser observado (Tabela 9), nos estudos por simulações, que as acurácias obtidas numa geração t , com apenas informações sobre os marcadores moleculares, cujos pesos foram obtidos a partir das informações sobre a genotipagem e fenotipagem das gerações anteriores (F_1 ou A_i , sendo $i=1,2,\dots,t-1$), que as acurácias permanecem altas, mesmo com a provável redução no número de locos em desequilíbrio de ligação proporcionada pelos rearranjos gaméticos durante o processo de acasalamento ao acaso e encontro não preferencial dos gametas produzidos pela população. Neste contexto,

deve ser ressaltado os resultados da geração A_{10} para as características analisadas utilizando a fenotipagem e genotipagem das gerações anteriores. Assim, se os efeitos dos marcadores são estimados em A_9 o valor de confiabilidade em A_{10} será reduzido (passando de 0.47910 para 0.39532 para característica de herdabilidade igual a 0,20) porém menos drasticamente se estes efeitos tiverem sido estimados em A_1 ou F_1 (passando de 0.47910 para 0.28389 ou 0.22045 respectivamente para característica de herdabilidade igual a 0,20).

Wong & Bernardo (2008) demonstraram que a GWS é também aplicável a espécies alógamas perenes, como é o caso da palma de óleo (dendê). Esses autores demonstraram, via simulações, com tamanho populacionais de 50 a 70, repostas a GWS foram 4 a 25% superiores as respostas correspondentes com a seleção fenotípica, dependendo da herdabilidade da característica e do número de QTLs. Segundo esses autores o custo por unidade de ganho foi 35 a 65% inferior com a GWS quando comparada a seleção fenotípica, quando o custo por data point foi considerado como US\$ 0,15. Os autores demonstraram ainda que a GWS pode viabilizar quatro ciclos de seleção no mesmo tempo requerido normalmente para dois ciclos de seleção com base em dados fenotípicos.

De maneira geral verifica-se que para sair de um patamar de herdabilidade igual a 0,20 para 0,30, como ilustração, é possível por meio da redução da fenotipagem em até oito gerações, como tomando como base a geração A_{10} (Tabela 9). Para ter acréscimos na herdabilidade de uma característica originalmente de 0,40 e mantê-la num patamar igual a 0,50, eliminando a fenotipagem, seria possível apenas com informações prévias da fenotipagem feita a, no máximo, duas gerações anteriores. E, para característica de herdabilidade igual a 0,60 a prática da seleção, via GWS, não seria vantajosa sem a utilização da fenotipagem da própria geração sob seleção.

Tabela 9. Valores de confiabilidade da seleção das gerações avançadas por acasalamento ao acaso obtidas a partir do fenotipagem e genotipagem das mesmas gerações (**diagonal**) ou apenas da genotipagem e fenotipagem anteriores (**fora da diagonal, leitura na horizontal**) para três caracteres de herdabilidade iguais a 0.20, 0.40, e 0.60.

Característica 1 - $h^2 = 0,20$

	F1	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈	A ₉	A ₁₀
F ₁	0.39523										
F _{2a1}	0.22953	0.33816									
F _{2a2}	0.21206	0.25517	0.42005								
F _{2a3}	0.19802	0.23212	0.31343	0.40936							
F _{2a4}	0.15520	0.25659	0.35884	0.36138	0.41134						
F _{2a5}	0.18756	0.25125	0.33913	0.31468	0.30306	0.37075					
F _{2a6}	0.20306	0.28807	0.35138	0.33892	0.32270	0.34149	0.46832				
F _{2a7}	0.19888	0.27263	0.33554	0.36921	0.29309	0.32109	0.39253	0.38609			
F _{2a8}	0.23374	0.29250	0.3339	0.38783	0.31797	0.36045	0.39896	0.30718	0.44041		
F _{2a9}	0.22334	0.29312	0.34678	0.36344	0.32958	0.35535	0.36621	0.29559	0.32162	0.43059	
F _{2a10}	0.22045	0.28389	0.36854	0.36423	0.36556	0.35782	0.36603	0.31701	0.27208	0.39532	0.4791

Característica 2 - $h^2 = 0,40$

	F1	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈	A ₉	A ₁₀
F ₁	0.59239										
F _{2a1}	0.3707	0.60663									
F _{2a2}	0.36590	0.51790	0.60276								
F _{2a3}	0.31637	0.42234	0.45270	0.58968							
F _{2a4}	0.30453	0.44488	0.46548	0.52680	0.60097						
F _{2a5}	0.32164	0.44871	0.47405	0.43240	0.47885	0.57797					
F _{2a6}	0.33394	0.46895	0.44623	0.47630	0.48674	0.47002	0.63487				
F _{2a7}	0.36706	0.44793	0.43357	0.44887	0.48344	0.46629	0.52130	0.62260			
F _{2a8}	0.35342	0.44824	0.43309	0.51072	0.54247	0.45018	0.51378	0.51560	0.61568		
F _{2a9}	0.32169	0.49138	0.44255	0.48744	0.51476	0.48009	0.50056	0.48167	0.53319	0.63486	
F _{2a10}	0.32919	0.47014	0.46279	0.45863	0.47046	0.46566	0.47937	0.45513	0.51809	0.58378	0.645

Característica 3 - $h^2 = 0,60$

	F1	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈	A ₉	A ₁₀
F ₁	0.74744										
F _{2a1}	0.52874	0.72986									
F _{2a2}	0.51941	0.56282	0.75534								
F _{2a3}	0.48134	0.54622	0.64035	0.73153							
F _{2a4}	0.50127	0.55236	0.62081	0.61262	0.75180						
F _{2a5}	0.50059	0.55472	0.62157	0.56374	0.58588	0.73412					
F _{2a6}	0.47056	0.55732	0.62428	0.60090	0.58922	0.65025	0.76138				
F _{2a7}	0.50045	0.56002	0.60220	0.56372	0.56364	0.62673	0.64212	0.78032			
F _{2a8}	0.48781	0.55853	0.61833	0.58520	0.57726	0.62403	0.61123	0.65559	0.73638		
F _{2a9}	0.50191	0.61125	0.60391	0.57928	0.60067	0.63450	0.60239	0.63053	0.63707	0.77112	
F _{2a10}	0.47067	0.57706	0.59341	0.56135	0.57460	0.63401	0.62828	0.62761	0.57883	0.68394	0.729

O aumento da confiabilidade, ou seja, na capacidade do pesquisador inferir sobre o valor genético do indivíduo a partir de valores mensurados, passa a ser indispensável, maximizando os ganhos e reduzindo o custo, tempo e mão de obra despendidos na experimentação. É possível observar na tabela 10 que, de maneira geral para as espécies autógamas, à medida que as gerações avançam, o valor de confiabilidade apresenta-se constante em todos os cenários simulados, sendo observados valores de confiabilidade nas gerações S_1 e S_{10} (0.3685 ± 0.02 ; 0.6019 ± 0.02 ; 0.7471 ± 0.02) para os três caracteres analisados respectivamente. Para espécies autógamas existe enorme necessidade de efetuar trabalhos de pesquisa no campo da seleção genômica, pois até a atualidade não existem muitos trabalhos ao respeito. Pode-se verificar, por meio desta pesquisa, que a inclusão das informações moleculares associadas ao uso de técnicas biométricas apropriadas propiciará uma informação ao pesquisador de grande valia.

De maneira similar ao verificado nas populações representativas de espécies alógamas, ao reaproveitar as informações genotípicas e fenotípicas das gerações anteriores, observassem valores de confiabilidade com valores superiores em relação a suas herdabilidades (Tabela 10). Entretanto, uma análise mais pontual indica que a GWS parece proporcionar resultados mais vantajosos nestas espécies talvez em razão da redução mais lenta do desequilíbrio de ligação entre os locos controladores das características quantitativas. De maneira geral verifica-se que para sair de um patamar de herdabilidade igual a 0,20 para 0,30, como ilustração, é possível por meio da redução da fenotipagem em até dez gerações, como tomando como base a geração A_{10} (Tabela 10). Resultado semelhante também é apresentado para a situação em que se deseja ter acréscimos na herdabilidade de uma característica originalmente de 0,40 e mantê-la num patamar igual a 0,50. E, para característica de herdabilidade igual a 0,60 ser elevada a 0,70 seria possível reduzir a fenotipagem em até oito gerações anteriores.

Segundo Calus et al. (2008), a GWS pode basear-se no uso de: (i) apenas dos marcadores; (ii) de haplótipos ou intervalos definidos por 2 marcadores; (iii) haplótipos definidos por mais de 2 marcadores, incluindo a covariância entre haplótipos devida à ligação, ademais determinarem que para caracteres de baixa herdabilidade (10%) não existem diferenças significativas entre essas 3 abordagens. Solberg et al. (2006) mostraram que é possível praticar a GWS eficientemente com o uso apenas dos marcadores, ou seja, com a predição direta dos efeitos dos marcadores. Relatam também que isso é vantajoso porque não há necessidade de estimar as fases de ligação entre os marcadores, as quais são estimadas com algum erro.

Tanto para populações representativas de espécies autóгамas ou alógamas a inclusão das informações a partir da genotipagem e fenotipagem na própria geração propicia acréscimo no valor da confiabilidade e, conseqüentemente, do ganho a ser obtido pela seleção. Isso pode ser explicado desde o ponto de vista, da superioridade da GWS sobre a seleção convencional, pode ser atribuída ao fator do uso da matriz de parentesco real própria de cada caráter, fato que aumenta a acurácia seletiva.

A metodologia GWS se adequou bem ao número de indivíduos analisados e se comportou efetiva em todos os cenários simulados para efetuar as análises de seleção genômica demonstrando um grande potencial em sua aplicação. Outras metodologias foram relatadas como superiores (Meuwissen, et al., 2001, Meuwissen et al., 2009; Hayes et al., 2009). Em trabalhos similares foi verificado que, em condições de herdabilidade (20 % e 20 locos), 500 ou mais indivíduos são necessários para se ter uma acurácia adequada. Com 1.000 indivíduos a acurácia esperada é de 95 % e para obter acurácia superior a 98 %, 2 mil indivíduos são necessários (Resende et al., 2008).

Tabela 10. Valores de confiabilidade da seleção das gerações avançadas por autofecundação obtidas a partir do fenotipagem e genotipagem das mesmas gerações (**diagonal**) ou apenas da genotipagem e fenotipagem anteriores (**fora da diagonal, leitura na horizontal**) para três caracteres de herdabilidade iguais a 0.20, 0.40, e 0.60.

Característica 1 - $h^2 = 0.20$

	F1	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀
F ₁	0.3952										
F _{2s1}	0.31885	0.3685									
F _{2s2}	0.31067	0.3621	0.3654								
F _{2s3}	0.29961	0.3419	0.3516	0.4221							
F _{2s4}	0.30353	0.3387	0.35114	0.4268	0.40081						
F _{2s5}	0.29808	0.3395	0.3507	0.4256	0.40153	0.45733					
F _{2s6}	0.30185	0.34141	0.35254	0.42756	0.40312	0.4586	0.39215				
F _{2s7}	0.29882	0.34192	0.35148	0.42177	0.40547	0.45790	0.38967	0.38724			
F _{2s8}	0.295052	0.33963	0.35055	0.42239	0.40342	0.45635	0.38790	0.38603	0.41968		
F _{2s9}	0.29436	0.33983	0.35089	0.42420	0.40493	0.45418	0.38741	0.38536	0.42078	0.44912	
F _{2s10}	0.29434	0.34048	0.35054	0.42512	0.40476	0.45354	0.38590	0.38503	0.42180	0.44893	0.385

Característica 2 - $h^2 = 0,40$

	F1	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀
F ₁	0.59329										
F _{2s1}	0.49276	0.60198									
F _{2s2}	0.47949	0.53791	0.57502								
F _{2s3}	0.46335	0.52557	0.56701	0.55824							
F _{2s4}	0.46544	0.52969	0.56185	0.55517	0.60081						
F _{2s5}	0.46551	0.52775	0.55893	0.54680	0.59632	0.59842					
F _{2s6}	0.46612	0.53055	0.55706	0.54382	0.59580	0.60172	0.57134				
F _{2s7}	0.46641	0.52803	0.55338	0.54123	0.59590	0.59637	0.57065	0.56090			
F _{2s8}	0.46557	0.52523	0.55124	0.53845	0.59403	0.59611	0.56859	0.56093	0.57510		
F _{2s9}	0.46568	0.52529	0.55028	0.53603	0.59379	0.59497	0.56883	0.56161	0.57443	0.59432	
F _{2s10}	0.46518	0.52570	0.55131	0.53578	0.59364	0.59513	0.56918	0.56130	0.57498	0.59405	0.5901

Característica 3 - $h^2 = 0.60$

	F1	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀
F ₁	0.74744										
F _{2s1}	0.62551	0.74712									
F _{2s2}	0.59258	0.70367	0.72808								
F _{2s3}	0.57575	0.67783	0.71163	0.72384							
F _{2s4}	0.57883	0.67609	0.70486	0.71357	0.73424						
F _{2s5}	0.58154	0.67714	0.70178	0.71111	0.73114	0.73776					
F _{2s6}	0.57782	0.67764	0.69719	0.70932	0.72655	0.73144	0.74205				
F _{2s7}	0.57494	0.67375	0.69444	0.70679	0.72581	0.72947	0.73692	0.72847			
F _{2s8}	0.57238	0.67269	0.69325	0.70589	0.72475	0.72813	0.73673	0.72532	0.71863		
F _{2s9}	0.57088	0.67186	0.69270	0.70550	0.72465	0.72767	0.73584	0.72424	0.71756	0.73250	
F _{2s10}	0.57047	0.67249	0.69373	0.70593	0.72398	0.72724	0.73591	0.72437	0.71782	0.73244	0.728

Conclusão

- A simulação utilizada foi eficaz em preservar a estrutura genética das populações e descrever a sua dinâmica ao longo de sucessivas gerações de acasalamento ao acaso ou autofecundação.

- A agregação de informações moleculares e fenotípica na predição do valor genético do indivíduo proporciona acréscimo no valor da confiabilidade.

- O sistema de acasalamento (autofecundação ou acasalamento ao acaso) afeta a eficiência do uso da utilização das estimativas dos efeitos dos marcadores em gerações avançadas.

- A redução da fenotipagem em populações representativas de plantas autógamas mostrou ser mais vantajosa do que a verificada em populações representativas de plantas alógamas.

Referências

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

CROSSA, J.; BURGUEÑO, J.; CORNELIUS, P. L.; MCLAREN, G.; TRETOWAN, R.; KRISHNAMACHARI, A. Modeling genotype× environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop science**, v. 46, n. 4, p. 1722-1733, 2006.

CROSSA, J.; BURGUEÑO, J.; DREISIGACKER, S.; VARGAS, M.; HERRERA-FOESEL, S. A.; LILLEMOM, M.; SINGH, R. P.; TRETOWAN, R.; WARBURTON, M.; FRANCO, J. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. **Genetics**, v. 177, n. 3, p. 1889-1913, 2007.

DEKKERS, J. C. M. Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, v. 82, n. 13 suppl, p. E313-E328, 2004.

ENDELMAN, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome** 4:250-255. doi: 10.3835/plantgenome2011.08.0024. 2011.

FRARY, A.; NESBITT, T. C.; GRANDILLO, S.; VAN DER KNAAP, E.; CONG, B.; LIU, J.; MELLER, J.; ELBER, R.; ALPERT, K. B. fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. **Science**, v. 289, n. 5476, p. 85-88, 2000.

HAYES, B.; BOWMAN, P.; CHAMBERLAIN, A.; GODDARD, M. Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of dairy science**, v. 92, n. 2, p. 433, 2009.

JENKINS, S.; GIBSON, N. High-throughput SNP genotyping. **Comparative and functional genomics**, v. 3, n. 1, p. 57-66, 2002.

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

LIU, J.; VAN ECK, J.; CONG, B.; TANKSLEY, S. D. A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. **Proceedings of the National Academy of Sciences**, v. 99, n. 20, p. 13302, 2002.

MEUWISSEN, T.; GODDARD, M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. **Genetics**, v. 185, n. 2, p. 623-631, 2010.

MEUWISSEN, T.; SOLBERG, T. R.; SHEPHERD, R.; WOOLLIAMS, J. A. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. **Genetics Selection Evolution**, v. 41, n. 1, p. 2, 2009.

OAKEY, H.; VERBYLA, A.; PITCHFORD, W.; CULLIS, B.; KUCHEL, H. Joint modeling of additive and non-additive genetic line effects in single field trials. **TAG Theoretical and Applied Genetics**, v. 113, n. 5, p. 809-819, 2006.

PÉREZ, P.; DE LOS CAMPOS, G.; CROSSA, J.; GIANOLA, D. Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. **The Plant Genome**, v. 3, n. 2, p. 106, 2010.

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

RESENDE JR, M.; MUNOZ, P.; ACOSTA, J.; PETER, G.; DAVIS, J.; GRATTAPAGLIA, D.; RESENDE, M.; KIRST, M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. **New Phytologist**, 2012.

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

SCHAEFFER, L. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, v. 123, n. 4, p. 218-223, 2006.

SMARAGDOV, M. Genomic selection as a possible accelerator of traditional selection. **Russian Journal of Genetics**, v. 45, n. 6, p. 633-636, 2009.

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

YANO, M.; KATAYOSE, Y.; ASHIKARI, M.; YAMANOUCHI, U.; MONNA, L.; FUSE, T.; BABA, T.; YAMAMOTO, K.; UMEHARA, Y.; NAGAMURA, Y. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. **The Plant Cell Online**, v. 12, n. 12, p. 2473-2483, 2000.

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