

AUGUSTO TESSELE

**SELEÇÃO GENÔMICA AMPLA PARA ESCOLHA DE GENITORES DE SOJA
E PREDIÇÃO DO DESEMPENHO DE POPULAÇÕES HÍBRIDAS**

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

VIÇOSA
MINAS GERAIS – BRASIL
2017

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

T

T338s
2017
Tessele, Augusto, 1993-
Seleção genômica ampla para escolha de genitores de soja e
predição do desempenho de populações híbridas / Augusto
Tessele. – Viçosa, MG, 2017.
x, 51f. : il. (algumas color.) ; 29 cm.

Orientador: Felipe Lopes da Silva.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f.44-51.

1. Soja - Melhoramento genético. 2. *Genome Wide
Selection - GWS*. 3. *Glycine max* (L.) Merr. 4. Híbridos -
Desempenho. I. Universidade Federal de Viçosa. Departamento
de Fitotecnia. Programa de Pós-graduação em Fitotecnia.
II. Título.

CDD 22 ed. 633.343

AUGUSTO TESSELE

**SELEÇÃO GENÔMICA AMPLA PARA ESCOLHA DE GENITORES DE SOJA
E PREDIÇÃO DO DESEMPENHO DE POPULAÇÕES HÍBRIDAS**

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

APROVADA: 28 de abril de 2017.

Cosme Damiano Cruz
(Coorientador)

Moyses Nascimento
(Coorientador)

Thais Roseli Corrêa

Felipe Lopes da Silva
(Orientador)

Aos meus pais e ao meu avô Antônio

Dedico

AGRADECIMENTOS

À Deus, pelo dom da vida, por me abençoar com saúde e felicidade e por me permitir a busca diária por sabedoria.

Aos meus pais, Antônio Augusto Tessele Filho e Tania Regina Gralha Tessele, pelo amor, apoio, confiança, ensinamentos e exemplo de vida.

Ao meu avô, Antônio Augusto Tessele, pela amizade e por transmitir o amor pelo campo.

À Priscila Marchioro, pelo amor, companheirismo, compreensão, apoio e alegria.

À minha grande irmã, Rhaquel Tessele, pelo amor, paciência e apoio.

À minha família, por sempre compreender a minha ausência e me dar o suporte para trilhar este caminho.

Ao professor Felipe Lopes da Silva, pela orientação, respeito e grande amizade construído ao longo dessa jornada.

Ao professor Cosme Damião Cruz, pela orientação, ensinamentos, amizade e exemplo de profissional e ser humano.

Aos meus amigos do Laboratório de Oleaginosas e de Bioinformática, pela grande amizade, companheirismo e suporte.

Aos professores da Universidade Federal de Viçosa, pelo ensinamento científico transmitido.

Aos meus professores da Universidade Federal do Paraná, pela grande amizade construída ao longo dos anos e pela confiança depositada em mim.

A todos meus amigos, pelos inúmeros momentos que compartilhamos boas recordações.

BIOGRAFIA

Augusto Tessele, filho de Antônio Augusto Tessele Filho e Tania Regia Gralha Tessele, nasceu em 16 de agosto de 1993, na cidade de Palotina, Paraná.

Em dezembro de 2010, formou-se no curso de Técnico Agrícola pelo Colégio Agrícola Estadual Adroaldo Augusto Colombo, em Palotina.

Em janeiro de 2016, graduou-se em Agronomia pela Universidade Federal do Paraná, setor Palotina, Palotina.

Em fevereiro de 2016 ingressou no curso de Mestrado em Fitotecnia, na Universidade Federal de Viçosa, Minas Gerais.

SUMÁRIO

RESUMO.....	vii
ABSTRACT.....	ix
1. INTRODUÇÃO	1
2. REFERENCIAL TEÓRICO	4
2.1 Melhoramento genético	4
2.2 Escolha de genitores programa de melhoramento	5
- Importância.....	5
- Utilização de híbridos – heterose.....	6
- Utilização de populações segregantes e obtenção de transgressivos	7
2.3 Abordagem tradicional na escolha de genitores.....	8
-Cruzamento ‘bom x bom’	8
- Cruzamento entre bons e divergentes	8
- Estudo da diversidade genética (processos preditivos).....	9
- Análise dialélica	12
2.4. Abordagem molecular na escolha de genitores.....	14
2.5 Emprego de marcadores moleculares na predição de valores genéticos.....	15
3. MATERIAIS E MÉTODOS	19
3.1 Simulação dos dados.....	19
3.1.1 Simulação genotípica da população genitora.....	19
3.1.2 Simulação genotípica das populações híbridas	20
3.1.3 Simulação das características fenotípicas associada a genitores (RILs) e híbridos ...	20
3.2 Métodos para seleção de genitores.....	21
3.2.1 Métodos para escolha de genitores para formação de população de cruzamento	21
1. Escolha univariada com base em informação fenotípica da característica de interesse..	22
2. Escolha multivariada com base em informação fenotípica.....	22
3. Escolha univariada com base em valor genotípico predito dos híbridos	23
3.2.2 Estabelecimento das melhores combinações híbridas.....	25
- Predição com base em efeitos da capacidade combinatória	25
- Predição do valor genético de híbridos.....	26

-Análise fenotípica das melhores combinações híbridas	26
3.3 Predição de híbridos não realizados.....	26
3.4 Aplicativos computacionais para análise dos dados	27
4. RESULTADOS E DISCUSSÕES	27
- Estratégia 1: Escolha de genitores e performance de híbridos a partir de análise univariada com base em informação fenotípica da característica de interesse.	27
- Estratégia 2: Escolha de genitores e híbridos a partir de sistemas dialélicos estabelecidos por análise multivariada com base em informação fenotípica	30
- Estratégia 3: Escolha de genitores e híbridos baseada no valor genético genômico predito de linhagens	34
Predição de híbridos não realizados.....	36
5. CONSIDERAÇÕES FINAIS.....	43
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	44

RESUMO

TESSELE, Augusto, M.Sc., Universidade Federal de Viçosa, abril de 2017. **Seleção genômica ampla para escolha de genitores de soja e predição do desempenho de populações híbridas.** Orientador: Felipe Lopes da Silva. Coorientadores: Cosme Damião Cruz e Moysés Nascimento.

A seleção de genitores é a primeira etapa em um programa de melhoramento e define o potencial de sucesso no desenvolvimento de uma cultivar superior. A seleção genômica ampla (do inglês, *genome wide selection* – GWS) associa informações moleculares e fenotípicas e prediz o desempenho de progênes futuras (valor genético genômico predito) com informações moleculares. Neste cenário, o objetivo deste estudo foi avaliar o potencial da GWS na predição do desempenho de híbridos e, conseqüentemente, na seleção de genitores para cruzamentos, utilizando-se somente informações genotípicas dos híbridos. Ademais, almejou-se comparar os resultados com técnicas tradicionais de seleção de genitores (seleção univariada e multivariada), visando o estabelecimento de blocos de cruzamento, para, então, parametrizar o potencial da GWS. Para execução do experimento foram utilizados dados simulados tomando como referência o genoma da soja. Foram criadas 200 RILs (do inglês, *recombinant inbred line*) com informações moleculares de 5400 SNPs e quatro características fenotípicas (produtividade, altura de planta, acamamento e ramificações laterais). Além disso, foram gerados 19900 híbridos oriundos do inter cruzamento de todas as RILs. O conjunto de informações das RILs foram criadas fazendo alusão ao processo de fenotipagem e genotipagem de um conjunto de linhagens de soja com potencial para serem selecionadas como genitores em um programa de melhoramento. Primeiramente, foram selecionadas 10 linhagens seguindo critério univariado para a característica de interesse, ou seja, baseando-se no desempenho fenotípico *per se* para a produtividade. Em seguida, utilizou-se de um critério multivariado para seleção de genitores. Neste, foi empregado o método de agrupamento de Tocher, utilizando-se a distância euclidiana média, e a técnica de componentes principais para seleção de 10 linhagens geneticamente divergentes. Posteriormente, a seleção de genitores (10 linhagens) foi baseada no valor genético genômico estimado das RILs, cuja estimação foi obtida a partir da metodologia de seleção genômica ampla, considerando o carácter produtividade. Neste caso, foram selecionadas as linhagens com maior valor genético genômico estimado. Os três blocos de cruzamentos gerados foram avaliados pelo modelo de análise dialélica proposto por Griffing (1956). Paralelamente, os valores fenotípicos dos híbridos foram analisados visando a seleção das melhores

populações híbridas de cada bloco de cruzamento. Além disto, estimou-se o valor genético genômico predito de todos os híbridos oriundos do inter cruzamento das 200 linhagens e o potencial preditivo foi verificado analisando-se o desempenho fenotípico dos melhores híbridos preditos. As metodologias de seleção de genitores univariada com base em valores fenotípicos e predição pela GWS apresentaram quatro populações híbridas promissoras de acordo com a análise dialélica, enquanto que o critério multivariado para seleção de genitores rendeu seis híbridos superiores. Entretanto, considerando-se a seleção dos 20% melhores híbridos baseados apenas em dados fenotípicos, foi observado que as populações mais promissoras foram encontradas no bloco de cruzamento baseado no valor genético genômico estimado dos genitores, seguido pelo critério multivariado e univariado. O desempenho fenotípico médio destes híbridos superiores foi 1,14, 1,11 e 0,93, respectivamente. A predição de performance empregada pela GWS para quatro características fenotípicas apresentou resultados promissores. O desempenho fenotípico dos melhores híbridos preditos para as características produtividade, altura de planta, acamamento e ramificações laterais apontou que 30%, 47%, 46% e 46% dos melhores híbridos preditos apresentaram excelente desempenho fenotípico, respectivamente. Além disso, observou-se que os genitores das populações híbridas com excelente desempenho fenotípico apresentavam elevados valores genético genômicos, destacando a importância de se considerar informações de genitores. Este resultado ressalta o potencial da GWS na predição do desempenho de híbridos e, conseqüentemente, na determinação dos genitores selecionados para cruzamentos.

ABSTRACT

TESSELE, Augusto, M.Sc., Universidade Federal de Viçosa, April, 2017. **Soybean parental selection with genome wide selection and prediction of hybrid populations performance.** Advisor: Felipe Lopes da Silva. Co-advisors: Comes Damiano Cruz and Moysés Nascimento.

Parental selection is the main stage in a breeding program, once it delimits the success in developing a new cultivar. The Genome Wide Selection (GWS) enables the association of molecular information with phenotypic data and predicts the performance of future progenies (estimated breeding value) using molecular information. In this scenario, the aim of this study was to evaluate the potential of GWS to predict hybrid performance and, consequently, support parental selection, only employing genotypic information of hybrid populations. Besides, we aimed to compare the results with traditional methods of parental selection (univariate and multivariate selection), in order to form crossing blocks, and, therefore, to parametrize the GWS potential. We ran this study based on simulated data from the soybean reference genome. We created 200 RILs (recombinant inbred line) associated to molecular information (5400 SNP markers) and four phenotypic traits (yield, plant height, lodging and number of branches). We created the 19900 hybrids from the intercross of all RILs as well. The group of RILs data was created aiming to allude the process of genotyping and phenotyping a set of soybean inbred lines with potential to yield promising hybrids. First, 10 inbred lines were selected according to *per se* performance criteria for yield, that is, the most yielding lines were selected. Then, a multivariate approach was employed for parental selection. In this case, the Tocher grouping technique, based on average Euclidean distance, and Principal Components analysis were employed to select the most genetically divergent inbred lines (10 inbred selected). Next, the parental selection (10 inbred lines selected) was based on estimated breeding value of RILs, whose estimation was made according to the GWS methodology for yield. All three crossing blocks were evaluated through diallele cross analysis following the method proposed by Griffing (1956). Alongside, the hybrids phenotypic performance was analyzed solely as well. Moreover, we estimated all hybrids breeding value from the intercross of the 200 RILs and its prediction capability was verified analyzing the phenotypic performance of the best predicted hybrids. The parental selection approaches based on univariate criteria and GWS prediction displayed four promising hybrid populations according to the diallele cross analysis, while the

multivariate criteria yielded six superior hybrids. However, considering selecting the 20% best hybrids based solely on phenotypic performance, we observed that the most promising ones were found in the crossing block based on estimated breeding value, follow by the multivariate approach and univariate criteria. The phenotypic average performance of these superior hybrids populations were 1.14, 1.11 and 0.93, respectively. The performance prediction employed for four agronomic traits by GWS delivered promising results. The phenotypic analysis of the best hybrids according to the GWS prediction model (ones with highest estimated breeding value) for yield, plant height, lodging and number of branches pointed out that 30%, 47%, 46% and 46% of these hybrids performed phenotypically greatly, respectively. In addition, the genitors of the hybrid populations with excellent phenotypic performance had great estimated breeding value, indicating parental information importance. These results highlight the potential of GWS in predicting the best hybrids and, therefore, establishing the best parents for crossing.

1. INTRODUÇÃO

A segurança alimentar é um estado atingido somente quando todas as pessoas têm condições físicas, sociais e econômicas para ter livre acesso a alimento nutritivo, seguro e em quantidade suficiente (FAO, 2003). Entretanto, para que a demanda mundial por alimento seja atendida, é necessário a produção de plantas em volume e com valores nutricionais adequados para consumo humano (BRUULSEMA *et al.*, 2012).

Apenas 30 espécies de plantas cultivadas são responsáveis por 95% da proteína e energia (calorias) consumidas no mundo, dentre estas figura a soja (SCHMIDT & WEI, 2006). A soja [*Glycine max* (L.) Merrill] é considerada a mais importante oleaginosa cultivada no mundo. A produção mundial esperada para a safra 2016/17 é de 337,85 milhões de toneladas, sendo os Estados Unidos da América o maior produtor com 118,69 milhões de toneladas (USDA, 2017). No Brasil, a área ocupada na safra 2016/17 foi estimada em 33,8 milhões de hectares, com produção recorde esperada de 104 milhões de toneladas (USDA, 2017).

Essa espécie oleaginosa atingiu esse patamar por apresentar características desejáveis, como o elevado teor de óleo (20%) e de proteína de qualidade (40%), junto com inúmeros nutrientes favoráveis e fatores bioativos (SINGH *et al.*, 2008). Assim, a soja e seus subprodutos são utilizados para consumo humano e animal e para fins industriais (HARTMAN *et al.*, 2011). Além de fatores alimentícios, a soja ainda apresenta características agrônômicas importantes em sistemas de rotação de culturas. A sua importância é decorrente da sua capacidade de fixar nitrogênio atmosférico, via associação simbiótica que, posteriormente, será incorporado ao solo e utilizado pela cultura subsequente (SINGH, 2010).

Em vistas das suas características desejáveis, a demanda por soja nas últimas décadas cresceu drasticamente, criando pressão no segmento agrícola para suprir esta necessidade, a qual pode ser atingida expandindo as áreas de cultivo ou aumentando a eficiência na produção. A expansão de fronteiras agrícolas enfrenta barreiras legais e sociais, sendo, portanto, evitada. Desta forma, torna-se necessário o aumento de produtividade, que pode ser atingido com a melhoria nas condições de cultivo (ambiente) ou pelo melhoramento da espécie (genótipo).

As melhorias nas condições de cultivo da soja incluem precisão no cultivo, adubação, irrigação, controle de pragas, doenças e plantas daninhas, rotação de culturas, conservação do solo, dentre outros.

Por sua vez, o melhoramento da espécie consiste em acumular alelos favoráveis

para as características agronômicas de interesse, de forma que a nova cultivar apresente elevado potencial produtivo, respondendo positivamente as condições favoráveis do ambiente, mas atenuando os efeitos adversos deste.

A hibridação artificial é a fusão direcionada de gametas geneticamente diferentes para a obtenção de híbridos heterozigóticos (BORÉM & MIRANDA, 2013). Em autógamias, a autofecundação destes híbridos permite a exploração da máxima variabilidade genética existente dentro deste cruzamento, permitindo o estabelecimento de combinações alélicas não existentes anteriormente, possibilitando a ocorrência de indivíduos superiores.

Para que um programa de melhoramento desenvolva uma cultivar de soja superior, é necessário que o cruzamento entre genitores seja realizado de forma adequada. Caso a seleção envolva parentais com superioridade para as características de interesse e ampla variabilidade genética, haverá mais genótipos segregantes superiores (FALCONER, 1981). Portanto, a probabilidade de sucesso de um programa de melhoramento está intimamente ligada a uma criteriosa seleção de genitores (OLIVEIRA, 1999).

Uma das formas de determinar os melhores cruzamentos é por meio de dialelos, que consistem no cruzamento de todas as combinações possíveis de n genitores (HAYMAN, 1954). Neste esquema de cruzamentos, os híbridos que manifestam elevada heterose são aqueles com maior diversidade genética entre os genitores (FALCONER, 1981) e, portanto, estas populações segregantes tem maior potencial de apresentar indivíduos superiores e devem receber maior atenção.

Embora essa metodologia permita a determinação das melhores populações com precisão e ainda forneça informações genéticas, existem limitações. Em situação com um elevado número de genitores candidatos, a necessidade de se realizar vários cruzamentos artificiais e a ocupação de uma elevada área experimental inviabilizam a utilização de cruzamentos dialélicos (CARPENTIERI-PÍPOLO *et al.*, 2000; FREITA-JUNIOR *et al.*, 2006)

Por conseguinte, outra forma de determinar os melhores cruzamentos é por procedimentos preditivos. Nesse caso, busca-se estimar a variância genética da população segregante a partir de estimativas da distância genética entre genitores. A diversidade genética entre os genitores é mensurada sem a necessidade de combinações híbridas, sendo estabelecida com base em informações fenotípicas (diferenças agronômicas, morfológicas, fisiológicas, etc) ou genotípicas (marcadores moleculares) dos próprios genitores. Assim sendo, pode-se prever que os melhores híbridos serão aqueles oriundos do cruzamento de genitores com ampla diversidade e com características desejáveis. Este

procedimento permite ao melhorista concentrar esforços somente nos melhores cruzamentos preditos (CARPENTIERI-PÍPOLO *et al.*, 2000).

A determinação da diversidade genética baseada em caracteres fenotípicos sofre influência do ambiente, especialmente as características de interesse agrônomo, permitindo que linhagens geneticamente divergentes em um ambiente se apresentem similares em outras condições ambientais. Além disto, uma precisa análise de diversidade genética é dependente de uma caracterização fenotípica de qualidade, que despende mão de obra e tempo. Desta forma, para se aperfeiçoar e aumentar a eficácia da predição da diversidade genética entre genótipos, pode se utilizar informações moleculares. A principal vantagem do uso de informações moleculares é que estas não se alteram em diferentes condições ambientais e, devido ao nível tecnológico atual, podem ser coletadas de forma rápida e barata. Com o uso de marcadores moleculares almeja-se prever com precisão o desempenho das progênies baseando-se em estimativas de divergência genética (GRATTAPAGLIA, 2001).

O uso da seleção assistida por marcadores moleculares (do inglês, *molecular assisted selection* - MAS), utilizando simultaneamente dados fenotípicos e genotípicos, aumentou a eficiência da seleção de genitores e outros processos no melhoramento de plantas. Porém, o advento de novas plataformas de genotipagem capazes de saturar o genoma com marcadores e genotipar simultaneamente vários indivíduos permitiu o surgimento de um novo método de seleção denominada seleção genômica (GS) ou seleção genômica ampla (do inglês, *genome wide selection* - GWS) (MEUWISSENET *et al.*, 2001).

A seleção genômica (GS) preconiza o uso do polimorfismo de única base (*single nucleotide polymorphisms* – SNP) como marcadores moleculares. Estes marcadores estão amplamente dispersos e, aproximadamente, equidistantemente distribuídos no genoma (SMARAGDOV, 2009). Sendo assim, os milhares ou milhões de marcadores que cobrem o genoma maximizam o número de QTLs em desequilíbrio de ligação com pelo menos uma parte dos marcadores. Conseqüentemente, um maior número de QTLs com efeito no fenótipo serão capturados, inclusive aqueles de pequeno efeito (HEFFNER *et al.*, 2009).

O processo central da seleção genômica consiste na estimação de valores genéticos genômicos para todos os marcadores de cada genótipo para posteriormente prever os valores genético genômico estimado (do inglês, *genomic estimated breeding value* – GEBV) de progênies futuras utilizando somente suas informações genéticas (HEFFNER *et al.*, 2009). Esta predição é realizada utilizando os efeitos dos marcadores genéticos estimados em um modelo treinado a partir de dados fenotípicos e genotípicos

de outros indivíduos (MEUWISSEN et al., 2001).

A diferença fundamental entre a predição de valores genéticos tradicionais e valores genéticos genômicos reside no fato de que no primeiro são utilizadas informações fenotípicas para inferir sobre o genótipo dos indivíduos, enquanto que no segundo as informações genotípicas (marcadores genéticos) são utilizadas para prever valores fenotípicos futuros dos indivíduos (RESENDE *et al.*, 2008). Desta forma, ambas metodologias buscam selecionar indivíduos com maior valor genético (real)

Especialmente em espécies perenes, a GWS possibilita a predição do desempenho de indivíduos genotipados e, conseqüentemente, permite uma seleção precoce baseada no valor genético-genômico predito. Entretanto, tanto para espécies perenes como anuais, a GWS também representa uma nova e promissora ferramenta para a seleção de genitores, baseada em dois pontos cruciais. Primeiramente, em espécies de plantas cujo seleção de genitores se baseia em linhagens (maioria dos loci em homocigose), como a soja, é possível a obtenção da genotipagem das populações híbridas F_1 pela simples associação par a par das marcas moleculares dos genitores, previamente a existência do híbrido. Neste caso, não se despende recursos e tempo para hibridação, condução e genotipagem das populações híbridas. O segundo ponto crucial para o sucesso desta metodologia reside na capacidade da GWS em prever o desempenho de indivíduos genotipados. Nestas condições, somente com a genotipagem e criteriosa fenotipagem dos genitores é possível prever quais combinações de genitores renderão as melhores populações híbridas (a partir do valor genético-genômico predito). Desta forma, observando satisfatória eficiência no método proposto, almeja-se recomendar a realização de cruzamentos exclusivos. Assim, ao invés de selecionar um conjunto de genitores para formar o bloco de cruzamentos, selecionam-se especificamente os genitores que geram híbridos superiores, evitando a realização de cruzamentos pouco promissores.

2. REFERENCIAL TEÓRICO

2.1 Melhoramento genético

O melhoramento genético de plantas é a arte, ciência e negócio voltado a tornar as plantas mais adaptadas as necessidades humanas e as condições ambientais. Inicialmente, era uma arte, visto que nos primórdios da agricultura os lavradores salvavam as sementes das plantas mais desejáveis, sem critério científico (BORÉM & MIRANDA, 2013).

Por muito tempo esta filosofia predominou nos campos. Entretanto, foram as

descobertas de Mendel o ponto chave para desencadear uma série de avanços no melhoramento genético das espécies, que passou a ter um cunho mais científico. Juntamente, o aperfeiçoamento do conhecimento científico de plantas e o domínio da hibridação artificial foram fatores cruciais para a rápida e recente evolução do melhoramento genético de plantas (SÁNCHEZ, 2013).

A medida que o melhoramento de plantas foi se tornando um negócio, a otimização do uso de recursos físicos, humanos e financeiros, aliados a manutenção de um constante aumento produtivo e a melhoria de características de interesse das plantas se tornou uma necessidade. Desta forma, o melhoramento passou a empregar, cada vez mais, novas ferramentas tecnológicas que aumentam a eficiência e reduzem gastos e tempo para o lançamento de uma cultivar promissora.

Atualmente, uma das razões do sucesso de um programa de melhoramento de plantas reside na capacidade de predizer as melhores populações. Caso o melhorista tenha as ferramentas tecnológicas e o conhecimento científico adequado para predizer quais cruzamentos lhe renderão os melhores híbridos e/ou populações segregantes, a probabilidade de sucesso é elevada. A possibilidade de concentrar esforços somente em populações promissoras permite melhor explorar a variação genética e otimiza os recursos financeiros.

2.2 Escolha de genitores programa de melhoramento

- Importância

A seleção de genitores e, conseqüentemente, de híbridos e populações segregantes, constitui a primeira etapa de um programa de melhoramento (PIMENTEL *et al.*, 2013). Esta etapa é a mais importante para determinar o nível de sucesso do programa de melhoramento, uma vez que a variância genética nas populações iniciais é fundamental para se obter genótipos superiores (BERTAN *et al.*, 2007). Caso a variância genética nas populações iniciais seja reduzida, haverá menor variabilidade para ser explorada e selecionada.

Embora seja crucial a correta escolha dos genitores, possivelmente este seja um dos maiores dilemas dos melhoristas, pois muito pouco se foi elucidado sobre as bases científicas da seleção de genitores (BORÉM & MIRANDA, 2013). É verdade que o avanço genético da soja nos últimos anos demonstra um cenário oposto e de grande sucesso, visto que a cultura da soja apresentava uma média de 1089 kg/ha na década de 60 e chega à safra 2016/2017 com uma produtividade estimada de 3070 kg/ha (USDA, 2017). Entretanto, o que realmente tem se observado é que empresas públicas e privadas

vêm desenvolvendo e avaliando elevado número de populações segregantes anualmente, porém nem todas apresentam potencial genético suficiente para extração de genótipos com constituição genética superior (BARROSO & HOFFMANN, 2003). Desta forma, nota-se que a estratégia estabelecida é avaliar elevado número de populações segregantes esperando-se que um mínimo se apresente promissora. Todavia, até que se determine o real potencial de uma população produzir genótipos superiores, esforços, tempo e dinheiro foram dispendidos para obtenção e avaliação destas populações (BARROSO & HOFFMANN, 2003).

- Utilização de híbridos – heterose

A produção de híbridos superiores, em espécies alógamas e algumas espécies de cruzamento misto, é altamente dependente da diversidade genética entre os genitores, visto que o principal objetivo é explorar os efeitos da heterose. A heterose ou vigor híbrido é o aumento do vigor, produtividade, altura de planta e da intensidade de outros fenômenos fisiológicos decorrente do cruzamento entre indivíduos contrastantes (BORÉM & MIRANDA, 2013).

Segundo Falconer (1989), a heterose refere-se ao oposto da endogamia. A endogamia ocorre quando cruzamentos aparentados são realizados, aumentando a frequência de alelos idênticos por ascendência e provocando uma redução na média populacional (CRUZ, 2012). Entretanto, ao se cruzar linhagens endogâmicas, observa-se um aumento na média daquelas características que sofreram uma redução decorrente dos cruzamentos aparentados. Desta forma, busca-se o desenvolvimento de linhagens endogâmicas geneticamente divergentes as quais, ao serem cruzadas, resultem em híbridos vigorosos.

A primeira etapa para a obtenção de linhagens endogâmicas contrastantes é a identificação de populações geneticamente divergentes, de onde plantas com características favoráveis, sendo a principal delas a produtividade, serão autofecundadas (HALLAUER, 1990). Não somente a diversidade entre populações é levada em consideração. De fato, opta-se por autofecundar indivíduos de populações que apresentam elevada frequência de alelos favoráveis e com reduzida carga genética (alelos deletérios) (BERNINI *et al.*, 2013). Com o avanço das gerações são realizadas avaliações *per se* das linhagens e, às vezes, em combinações híbridas para eliminar linhagens pouco promissoras (JENKINS, 1935).

Em posse das linhagens superiores oriundas da autofecundação de populações geneticamente distantes, inicia-se os testes de cruzamento. Neste, o objetivo é encontrar

uma combinação híbrida de dois genitores com elevada frequência de alelos favoráveis e geneticamente complementares, cujo híbrido apresente vários loci em heterozigose e manifeste elevado vigor híbrido, preferencialmente com média superior ao melhor genitor.

- Utilização de populações segregantes e obtenção de transgressivos

O objetivo de se cruzarem genitores em espécies autógamas, como a soja, é a obtenção de genótipos superiores (transgressivos) em populações avançada e linhagens de alto desempenho ao final do programa. Para tanto, almeja-se obter uma população segregante com alta frequência de alelos favoráveis, mas com baixa variação genética para os loci contendo alelos favoráveis na população original (FUMES, 2013). Desta forma, a obtenção de segregantes transgressivos é dependente do número de alelos favoráveis contrastantes entre os genitores, da probabilidade de fixação dos alelos em um único bloco gênico, da contribuição relativa dos alelos desejáveis dos genitores e das diferenças genéticas necessárias para que a distinção e seleção da planta superior seja praticável (ISLEIB, 1999).

Em espécies autógamas, o bom desempenho *per se* do híbrido não necessariamente proporcionará uma população segregante promissora. O cruzamento de genitores com grande dissimilaridade genética tende a resultar em híbridos com muitos loci em heterozigose que, devido aos efeitos de desvio de dominância, proporciona elevado valor fenotípico. Entretanto, nas gerações subsequentes até a formação de linhagens, a segregação dos loci pode resultar no estabelecimento de poucas ou nenhuma combinação alélicas superiores às dos pais (transgressivos). Desta forma, visando a obtenção de linhagens superiores, os efeitos de desvio de dominância atuam como agente perturbador na seleção fenotípica, dificultando o estabelecimento das melhores populações segregantes ainda na geração híbrida F_1 .

Visto que a distância genética entre indivíduos reflete a quantidade de loci divergentes, podendo corresponder a alelos favoráveis ou não, este não é um bom parâmetro para ser utilizado isoladamente na seleção de genitores (RAMALHO & LAMBERT, 2004).

Destro (1991) reportou que o cruzamento de linhagens de soja com elevado D^2 (distância generalizada de Malhalanobis) não necessariamente proporcionavam maiores produtividades e que pequenos valores de D^2 estava geralmente associado a mal desempenho produtivo. Neste caso, o ideal seria o cruzamento de linhagens com D^2 intermediário, adicionando-se informações do desempenho destas linhagens.

Portanto, os genitores selecionados para cruzamento devem apresentar elevada frequência de alelos favoráveis e a variância genética entre si deve ser aquela observada entre os locos favoráveis que um genitor tem e o outro não (DUDLEY, 1982).

2.3 Abordagem tradicional na escolha de genitores

-Cruzamento ‘bom x bom’

Uma das primeiras e mais simplistas abordagens na seleção de genitores é baseada no desempenho *per se* dos indivíduos. Neste, a seleção de genitores é determinada pela avaliação simplesmente com base em dados sobre produtividade e seus componentes (RANGEL *et al.*, 1991)

Este método de seleção de genitores se fundamenta no conceito genético que propõe que a probabilidade de se obter linhagens superiores está em função da frequência de alelos favoráveis presente nas populações genitoras. Assim, esta metodologia prega o cruzamento ‘bom x bom’.

Todavia, este método desconsidera informações da divergência genética entre genitores. Desta forma, os híbridos gerados do cruzamento de genitores com *background* genético aparentado apresentarão elevada média, mas pouca variância genotípica, limitando os ganhos genéticos. Isto é decorrente do fato de que genitores não distantes geneticamente tendem a compartilhar muitos alelos em comum (CRUZ *et al.*, 2011).

Com os sucessivos ciclos de seleção de genitores, o cruzamento de linhagens superiores resultará no estreitamento da base genética, aumentará o parentesco entre genitores e pouca variação genética será observada nas populações segregantes, limitando os ganhos com seleção.

- Cruzamento entre bons e divergentes

A avaliação da divergência genética anseia identificar genitores apropriados para a obtenção de híbridos com elevado efeito heterótico e que proporcionem maior segregação em recombinações, permitindo o aparecimento de transgressivos em populações avançadas (CRUZ *et al.*, 2011).

A diversidade genética tem sido inferida por meio de técnicas biométricas, baseadas na quantificação da heterose, ou por processos preditivos (CRUZ *et al.*, 2012). Entre as técnicas biométrica ou de natureza quantitativa, citam-se as análises dialélicas, que quantificam a capacidade de combinação específica quanto aos efeitos heteróticos manifestada nos híbridos. Entre os métodos preditivos da heterose, destacam-se aqueles baseados nas diferenças morfológicas, fisiológicas ou moleculares, em que a diversidade

genética entre os indivíduos é medida através de alguma medida de dissimilaridade e visa prever a heterose do híbrido.

- Estudo da diversidade genética (processos preditivos)

Os estudos de diversidade genética podem ser aplicados tanto a um conjunto de dados fenotípicos ou a informações moleculares. Muito embora marcadores moleculares provêm um grande volume de informação para análise da diversidade genética, tem se optado por continuar a utilizar dados fenotípicos, especialmente de natureza quantitativa (CRUZ *et al.*, 2011), em vista da importância agrônômica destas características.

Várias técnicas multivariadas podem ser empregadas na predição da diversidade genética, como métodos aglomerativos, análise de componentes principais e variáveis canônicas. A escolha do método mais apropriado deve ser de acordo com a precisão desejada, facilidade de análise e a forma com que os dados foram coletados (RANGEL *et al.*, 1991; CRUZ *et al.*, 2011)

As medidas de dissimilaridade estimam a distância genética par a par de um conjunto de indivíduos, gerando uma matriz de dissimilaridade. Dentre as medidas métricas disponíveis para este propósito, os que mais vem sendo utilizadas são a distância euclidiana média e a distância de Mahalanobis (DE CARVALHO *et al.*, 2003; CRUZ *et al.*, 2012). O segundo método é preferível, pois considera as variâncias e covariâncias residuais entre as características mensuradas, porém, por requerer a matriz de covariância residual estimada a partir de experimentos com repetição, apresenta uso limitado.

A distância Euclidiana média pode ser obtida por meio de observações individuais, sem a necessidade de repetições ou delineamentos experimentais. Inicialmente, foi desenvolvida para a análise restrita de variáveis quantitativas não correlacionadas (DIAS, 1998). Embora não adequada para uso no melhoramento de plantas, a distância Euclidiana vem sendo utilizada indiscriminadamente e tem mostrado grande utilidade mesmo na análise de características correlacionadas (CRUZ *et al.*, 2011).

Diversos autores têm optado pelo emprego da distância Euclidiana média para mensurar a dissimilaridade entre acessos de soja, como Val *et al* (2014), Teodoro *et al* (2015), Torres *et al* (2015) e Cantelli *et al* (2016).

Ao obter a matriz de dissimilaridade entre os genótipos, o melhorista busca estabelecer padrões de agrupamento para definir quais cruzamentos irão melhor explorar a diversidade entre genitores. Entretanto, ao se analisar um conjunto de dados de grande volume, torna-se impraticável o reconhecimento de grupos homogêneos pelo exame visual das estimativas (CRUZ *et al.*, 2011). Desta forma, é necessário o uso de técnicas

de agrupamentos, cuja finalidade é reunir os progenitores em vários grupos, de forma que haja homogeneidade dentro do grupo e heterogeneidade entre os grupos. Diferentes métodos de agrupamento dispõem de diferentes critérios de classificação dos indivíduos em grupos, como por exemplo o hierárquico e o de otimização.

O método hierárquico realiza a partição os dados sucessivamente até que seja estabelecido uma representação hierárquica dos agrupamentos (dendrograma ou diagrama de árvore). Neste caso, o número de grupos formados é dado simplesmente por inspeção gráfica visual ou estabelecidas entre os pontos de alta mudança de nível, tomando-os como delimitadores do número de progenitores para determinar o grupo (MILLIGAN, 1981; CRUZ *et al.*, 2012). Entretanto, a subjetividade na identificação do número ideal de grupos formados é uma problemática, visto que diferentes critérios de corte podem conduzir a resultados muito diferentes (MILLIGAN & COOPER, 1985).

Os métodos de otimização utilizam um algoritmo feito de modo direto, cuja solução é considerada ótima em algum sentido. Uma vez que a estrutura é obtida em certo nível de fusão, não há mudanças nos passos seguintes. Neste método, o objetivo é alcançar uma partição de indivíduos que otimize (maximize ou minimize) alguma medida predefinida (CRUZ *et al.*, 2011).

Um dos métodos de otimização mais usuais é o método de Tocher (CRUZ *et al.*, 2012). Neste, o primeiro agrupamento é realizado entre os indivíduos com maior similaridade, conforme a matriz de dissimilaridade. A inclusão de novos indivíduos será realizada se a distância média intragrupo for menor que a distância média intergrupo. Visto que a inclusão de um novo indivíduo sempre aumenta o valor médio da distância dentro do grupo, pode-se decidir pela inclusão ou não de um indivíduo em um grupo comparando o acréscimo no valor médio da distância intragrupo com um valor pré-definido. Este valor pode ser determinado arbitrariamente ou, como comumente é realizado, adotado o valor máximo do conjunto das menores distância envolvendo cada indivíduo.

De Almeida *et al* (2011) avaliaram oito características fenotípicas para aferir a divergência genética entre de 12 cultivares de soja. A dissimilaridade entre os indivíduos foi mensurada pela distância generalizada de Mahalanobis e o agrupamento foi realizado pelo método de otimização de Tocher e do vizinho mais próximo. Com base na diversidade genética e elevada média para os caracteres estudados, duas hibridações promissoras para a obtenção de populações segregantes com variabilidade superior foram detectadas. Os autores ainda relataram que os métodos de otimização de Tocher e vizinho mais próximo foram concordantes entre si.

Cantelli *et al* (2016) utilizaram a distância Euclidiana média para aferir a dissimilaridade entre 51 linhagens de soja e três testemunhas (cultivares). O agrupamento dos indivíduos foi realizado de acordo com o método de otimização de Tocher e diversos métodos hierárquicos. Os autores também relataram uma elevada similaridade entre o método de Tocher e o vizinho mais próximo.

Carpentieri-Pípolo *et al* (2000) estimaram a diversidade genética de 34 linhagens de soja com base em 12 características fenotípicas utilizando o método de agrupamento de Tocher, com base na distância generalizada de Mahalanobis. Os autores visavam uma apropriada seleção de genitores e, portanto, a recomendação dos cruzamentos também considerou o desempenho das linhagens. Ao final, somente 10 cruzamentos promissores foram recomendados, representado 2% do total de cruzamentos possíveis.

Uma outra forma de estudo da diversidade genética entre indivíduos é por Componentes Principais. Este método realiza a transformação de um conjunto de dados de variáveis dependentes e correlacionadas em um novo conjunto de dimensão equivalente, mas com propriedades importantes, que são de grande interesse em estudos de melhoramento. Este novo conjunto, os componentes principais, são a combinação linear das variáveis originais, dispostos agora de forma independente. Os componentes principais são estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos de variação total, contida nos dados iniciais. Além disso, estabelecem as variáveis de maior efeito na variação total.

A viabilidade dos componentes principais na análise de diversidade genética será dependente da capacidade de poucos componentes resumirem o conjunto de variáveis originais. Desta forma, um conjunto de dados v -dimensional poderá ser expresso em um espaço bi ou tridimensional, proporcionando uma simplificação considerável nos cálculos estatísticos e interpretação dos resultados com relação aos demais métodos alternativos. No melhoramento genético, os componentes principais têm sido utilizados para análise de diversidade genética quando os primeiros componentes principais explicam mais do que 80% da variação total, permitindo a análise de dispersão gráfica bi ou tridimensional.

Kein *et al* (1989) avaliaram a diversidade molecular de 58 acessos de soja utilizando marcadores RFLP. O método de componentes principais foi utilizado para plotar graficamente a dissimilaridade ente indivíduos. Os autores reportaram que a diversidade genética entre cultivares era reduzida e que havia elevada dissimilaridade entre diferentes espécies de soja, como *Glycine max* e *Glycine soja*. Ainda, ressaltaram que o método de componentes principais foi útil na redução no conjunto de dados genotípicos multidimensionais e na identificação de relações genéticas.

Maughan *et al* (1996) avaliaram a diversidade e relações genética entre *Glycine soja* e *Glycine max* utilizando marcadores moleculares AFLP. A análise de componentes principais foi realizada e permitiu uma clara distinção de *G. soja* e *G. max* em dois grupos.

É notória a existência de ampla gama de opções para o estabelecimento da diversidade genética entre um universo de indivíduos. O emprego de mais de um método de agrupamento (otimização ou hierárquico), associado ou não a análise de componentes principais ou variáveis canônicas se tornou uma prática corriqueira entre melhoristas. Desta forma, dispondo de maior conjunto de informações, o melhorista tem maior certeza na decisão a ser tomada e a probabilidade de insucesso baseada em suas decisões é reduzida.

Villela *et al* (2014) estimaram a diversidade genética de 74 cultivares de soja contendo a tecnologia RR oriundas de diferentes programas de melhoramento brasileiros. As análises foram baseadas em técnicas multivariadas de 10 características fenotípicas e 86 marcadores microsatélite. No que diz respeito a diversidade fenotípica, a matriz de dissimilaridade, obtida pela distância Euclidiana, foi utilizada para agrupamento das cultivares pelos métodos de Tocher e UPGMA. Além disso, o método de componentes principais também foi empregado. Os métodos de agrupamento foram concordantes quanto ao número de grupos formados (sete grupos), enquanto que o método de componentes principais destacou as variáveis de maior influência na diversidade genética entre as 74 cultivares. O estudo concluiu que há variabilidade genética entre as cultivares RR brasileiras e que alguns programas de melhoramento apresentam cultivares mais dissimilares que outros.

- Análise dialélica

As análises dialélicas consistem em um conjunto de $p(p-1)/2$ híbridos, resultante do cruzamento de p genitores, podendo ainda incluir os respectivos pais, híbridos recíprocos e/ou outras gerações relacionadas (F_2 , retrocruzamento, etc.).

A avaliação do comportamento das progênes permite a seleção das melhores combinações parentais, além de determinar a ação gênica que controla o caráter (CRUZ *et al.*, 2012; PIMENTEL *et al.*, 2013). Estas informações balizam o melhorista na escolha de um método de condução das populações segregantes e no critério de seleção a ser adotado.

Entre as metodologias mais comumente utilizadas em cruzamentos dialélico citam-se a metodologia proposta por Griffing (1956), pela qual são estimados os efeitos e as somas de quadrados de efeitos das capacidades geral e específica de combinação; a

metodologia proposta por Gardner & Eberhart (1966), na qual são avaliados os efeitos das variedades e heterose varietal; e a proposta por Hayman (1954), que fornece informações sobre o mecanismo básico de herança do caráter em estudo, dos valores genéticos dos genitores e do limite de seleção (CRUZ *et al.*, 2012).

Os diferentes tipos de dialelos são classificados de acordo com os esquemas de cruzamento utilizados. Os dialelos balanceados (completos ou meia-tabela) incluem os híbridos F1 entre todas as combinações dos progenitores, podendo alternativamente incluir os híbridos recíprocos e progenitores. Os dialelos parciais envolvem o cruzamento de dois grupos de genitores e seus respectivos híbridos, possibilitando reduzir o número de cruzamentos. Nos dialelos circulantes, os progenitores são representados por um mesmo número de cruzamentos, porém inferior a $p-1$, como ocorre nos balanceados; neste, informações sobre os progenitores são obtidas com um menor número de cruzamentos, porém com perda de informações de certos cruzamentos. Nos dialelos incompletos os progenitores apresentam número variável de cruzamentos em função de perdas de tratamento durante a condução dos ensaios ou quaisquer outros problemas.

Em todos os tipos de cruzamento dialélicos supracitados é possível estimar a capacidade geral (CGC) e específica (CEC) de combinação dos genitores e de seus cruzamentos, seguindo o método proposto por Griffing (1956). O primeiro termo é cunhado para designar o comportamento médio de um genitor em uma série de combinações híbridas, sendo este dependente da quantidade de alelos favoráveis que ele possui e transmite para a descendência. O segundo diz respeito ao desvio que um cruzamento apresenta em relação ao que seria esperado com base na CGC de seus genitores, em que suas estimativas apontam combinações híbridas com desempenho superior ou inferior ao esperado diante da capacidade geral de combinação (CRUZ & VENCOVSKY, 1989). A CGC, estimada por \hat{g}_i , decorre principalmente da variância genética aditiva, além de parte dos efeitos epistáticos do tipo aditivo x aditivo, enquanto a CEC, estimada por \hat{s}_{ij} , resulta da variância não aditiva, principalmente variância genética dominante e vários tipos de variância epistática (GRIFFING, 1956; MELO, 1987).

A quantificação da CGC associada a cada genitor permite o planejamento de futuros cruzamentos para a obtenção de linhagens superiores, usufruindo da herança transgressiva nas diferentes gerações segregantes (MIRANDA *et al.*, 1996; ASSUNÇÃO FILHO, 2016). Já, o conhecimento da ação gênica não-aditiva (CEC) pode ser aproveitada na avaliação de híbridos com melhores possibilidades de comercialização (MARTINEZ *et al.*, 1989).

Estudando a capacidade combinatória de populações de soja em F₂ e F₃ para resistência a ferrugem, Maphosa *et al* (2012) observaram que os efeitos aditivos e epistáticos do tipo aditivo x aditivo foram a forma mais comum da CGC e que a CEC nem sempre contribui com a resistência a ferrugem. A correlação positiva entre severidade parental, taxa de esporulação e CGC sugeriram que a seleção de parentais para o desenvolvimento de cultivares de soja resistentes a ferrugem pode ser baseada no desempenho *per se* dos pais.

Buscando estabelecer relações entre vigor de semente e produtividade na soja, Cho & Scott (2000) realizaram uma análise dialélica entre genótipos de soja com elevado e reduzido vigor de sementes. A CGC e CEC encontradas foram significativas, sendo a taxa de quadrados médios entre a capacidade geral e específica de combinação próximo a unidade para as duas características, indicando a predominância de efeitos aditivos atuando no vigor de sementes e produtividade. O vigor médio das sementes e a produtividade apresentaram valores superiores nos cruzamentos entre genitores com elevado vigor em comparação aos cruzamentos entre genitores com baixo vigor e com vigor contrastante. Os autores concluíram que o vigor das sementes pode ser melhorado geneticamente enquanto mantendo elevadas produtividades, em virtude da predominância de efeitos aditivos para ambas características.

Assunção Filho (2016) verificou o potencial de cruzamentos dialélicos de soja para precocidade, produtividade e resistência a ferrugem. Os autores encontraram elevada variabilidade entre os genitores, mas nenhum apresentou CGC satisfatórios para todas as características, mostrando comportamento independentes entre os caracteres avaliados. O cruzamento que proporcionou maior CEC para a produção foi entre um genitor com elevada e positiva CGC e outro como CGC negativa, indicando a relevância da complementariedade gênica dos genitores.

2.4. Abordagem molecular na escolha de genitores

O desenvolvimento de novas tecnologia no campo da biologia molecular, especialmente a reação em cadeia da polimerase, tornou possível o acesso de informações genótípicas de vários indivíduos simultaneamente e de detectar polimorfismos a nível de DNA, suficientes para discriminar a variação genética entre indivíduos e dentro de populações (CAIXETA *et al.*, 2009; CRUZ *et al.*, 2011). Assim sendo, o emprego de informações moleculares tem sido uma das principais ferramentas utilizadas no estudo da diversidade genética.

Os marcadores moleculares são estáveis em relação a influência ambiental e

estádio de desenvolvimento da planta e podem ser coletados de forma rápida e prática, sendo, em geral, muito polimórficos em relação aos caracteres fenotípicos (VILLELA, 2013). Neste sentido, os marcadores moleculares têm dado valiosas contribuições na abordagem molecular para escolha de genitores, como na detecção de parentesco genético entre acessos e indivíduos superiores, na predição da heterose e no estabelecimento de grupos heteróticos promissores para constituição de híbridos (CRUZ *et al.*, 2011).

Em estudo conduzido por Mulato (2009), buscou-se estimar a divergência genética entre acessos de soja de diferentes regiões do mundo para serem utilizados em programas de melhoramentos. Para tanto, foram utilizados dados fenotípicos e moleculares na análise da distância genética. O autor observou que houve reduzida concordância entre os métodos de agrupamento baseados em dados fenotípicos e genotípicos, mostrando que cada tipo de abordagem acessa diferentemente a variabilidade genética existente.

Em contrapartida, Ramalho & Lambert (2004) destacaram as consequências negativas do uso de medidas de dissimilaridade, a partir informações moleculares, na seleção de genitores para obtenção de linhagens superiores. Os autores ressaltaram que a escolha de genótipos com elevada divergência genética pode resultar em um híbrido heterozigótico para a maioria dos locos de interesse. Desta forma, a probabilidade do estabelecimento de uma linhagem com todos alelos favoráveis é extremamente baixa. Em contrapartida, o cruzamento de genitores superiores e com elevada frequência de alelos favoráveis resultará em um híbrido com muitos desses alelos favoráveis já fixados. Assim, a segregação de um reduzido número de locos favoráveis heterozigóticos aumenta a probabilidade de se encontrar uma linhagem superior com todos os locos favoráveis.

Outro aspecto negativo do uso isolado de dados moleculares para estimar a distância genética é a região do genoma acessada. Enquanto a análise fenotípica é realizada em caracteres de interesse, a abordagem molecular pode utilizar de regiões do genoma não transcritas, cuja diversidade entre genótipos resultante não é expressa fenotipicamente. Desta forma, a diversidade existente, embora real, não é observada fenotipicamente, impossibilitando ganhos genéticos com a seleção.

2.5 Emprego de marcadores moleculares na predição de valores genéticos

Com o advento dos marcadores moleculares e os avanços na área da biologia molecular tornou-se possível o aperfeiçoamento da determinação de valores genotípicos dos indivíduos, anteriormente realizada com informações fenotípicas por meio de procedimento BLUP (do inglês, *Best Linear Unbiased Prediction*) (RESENDE *et al.*,

2013). A primeira proposta para melhorar a eficiência deste procedimento foi a de integrar genética molecular com os métodos tradicionais de seleção artificial fenotípica, elaborada por Lande & Thompson (1990), por meio da técnica que ficou conhecida como seleção auxiliada por marcadores moleculares (MAS).

A MAS utiliza de informações fenotípicas e dados de marcadores moleculares em ligação próxima com alguns locos controladores de características quantitativas (QTL) (RESENDE *et al.*, 2008). Usualmente, os marcadores atuam como covariáveis na explicação dos valores fenotípicos dos indivíduos ou como efeitos aleatórios incorporados no modelo para o fenótipo (FERNANDO & GROSSMAN, 1989). Assim, após modelagem estatística sujeita a erro tipo II, ou seja, de não rejeitar uma hipótese falsa, os marcadores são eleitos ou não como determinantes dos efeitos de QTLs (RESENDE *et al.*, 2013).

O emprego da MAS em estudos de genética molecular é extenso. Fuganti *et al* (2004) encontraram sete QTLs no genoma da soja relacionados à resistência de *Meloidogyne javanica* via MAS. Entretanto, os autores ressaltaram o potencial de dois marcadores serem utilizados para seleção assistida, pois cada explicava quase 50% da variação fenotípica.

Silva *et al* (2001) também buscaram identificar QTLs associados a resistência a *M. javanica*. Os autores encontraram três locos apresentando associação com o número de galhas por planta, mas pouco explicavam a variação fenotípica observada.

Costa *et al* (2008) encontraram três fragmentos polimórficos em desequilíbrio de ligação com o loco de resistência a ferrugem asiática da soja, cuja segregação observada foi de 3:1 (herança monogênica, com dominância completa). Desta forma, estes marcadores foram considerados promissores na seleção assistida para resistência a ferrugem asiática da soja.

Embora uma ferramenta pioneira e promissora, a MAS apresenta gargalos que dificultam o seu amplo uso e aceitação em programas de melhoramento. Dentre esses, a necessidade de um reduzido número de populações segregantes para estabelecer uma correlação marcador-QTL (análise de ligação), pois, embora permita identificar o desequilíbrio de ligação entre marcador e QTL, restringe ao melhorista o uso de uma reduzida variabilidade genética, comprometendo os ganhos genéticos (RESENDE *et al.*, 2012a). Associado a isto, a MAS detecta somente QTLs de grande efeito na variação genética, o que não explica suficientemente toda a variação fenotípica (DEKKERS, 2004; GUIMARAES *et al.*, 2007), visto que são omitidos os genes de pequeno efeito (CROSBIE *et al.*, 2003). Desta forma, a MAS pouco contribui no melhoramento de

características de interesse agrônomo, visto a natureza poligênica e a grande influência do ambiente nos caracteres quantitativos. Segundo Resende *et al* (2013), a MAS só apresenta superioridade considerável frente a seleção fenotípica quando o tamanho da família empregado para fenotipagem e genotipagem é muito grande (na ordem de 500 ou mais).

Neste cenário, a MAS tem se mostrado mais adequada a manipulação genes de maior efeito e, tanto no setor público quanto privado, o seu uso tem se restringido a característica monogênicas ou oligogênicas de pouca influência ambiental (XU & CROUCH, 2008; HEFFNER *et al.*, 2009).

O atrativo da genética molecular em benefício do melhoramento genético aplicado é a utilização de informações diretamente do DNA para seleção, de forma a permitir elevada eficiência seletiva, rapidez na obtenção de ganhos genéticos com a seleção e baixo custo, em comparação com a tradicional seleção fenotípica (RESENDE *et al.*, 2008). Os avanços das tecnologias de genotipagem de larga escala e a descoberta de marcadores com densa cobertura no genoma permitiram que Meuwissen *et al* (2011) propusessem um novo método de seleção denominado seleção genômica (GS) ou seleção genômica ampla (GWS). Esta técnica pode ser aplicada em várias famílias de um programa de melhoramento em avaliação, apresenta alta acurácia seletiva para a seleção baseada unicamente em marcadores, não exige prévio conhecimento da posição dos QTLs e não está sujeita a erros do tipo II associados a seleção de marcadores ligados a QTLs (RESENDE *et al.*, 2013).

A GS preconiza o uso do polimorfismo de única base única (SNPs) como marcadores, os quais estão amplamente e, aproximadamente, equidistantemente distribuídos no genoma (SMARAGDOV, 2009). Com uma densa cobertura do genoma, realiza-se a seleção simultânea para centenas ou milhares de marcadores, de forma que todos os QTLs de uma característica estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores. Conseqüentemente, um maior número de QTLs com efeito no fenótipo serão capturados, inclusive aqueles de pequeno efeito, e explicarão quase a totalidade da variação genética de um caráter quantitativo (HEFFNER *et al.*, 2009). Desta forma, a GS torna-se uma ferramenta molecular adequada para características de baixa herdabilidade.

Segundo Resende *et al* (2013), a GWS pode basear-se no uso de: (i) apenas dos marcadores; (ii) de haplótipos ou intervalos definidos por dois marcadores; (iii) haplótipos definidos por mais de dois marcadores, incluindo a covariância entre haplótipos devido a ligação. Entretanto, Callus *et al* (2008) ressaltam que, para

características de baixa herdabilidade, não existem diferenças significativas entre essas três abordagens.

Os efeitos dos marcadores são estimados a partir de uma amostra de indivíduos de várias famílias, minimizando o impacto de determinadas famílias no peso dos marcadores. Os marcadores terão seus efeitos genéticos estimados a partir de uma amostra populacional grande, permitindo várias repetições por marcador. Desta forma, embora a herdabilidade de cada marcador efetivo seja muito baixa, o elevado número de repetições torna a herdabilidade alta a nível populacional.

Os valores genéticos genômicos associados a cada marcador, estimados em um modelo treinado a partir de dados fenotípicos e genotípicos, são utilizados para fornecer o valor genético genômico global de cada indivíduo (MEUWISSEN *et al.*, 2001). Por conseguinte, é realizado a predição do valor genético genômico estimado (do inglês, *genomic estimated breeding value* – GEBV) de progênes futuras utilizando somente suas informações genéticas.

Assim sendo, na predição de valores genéticos genômicos, as informações genotípicas (marcadores genéticos) são utilizadas para predizer valores fenotípicos futuros dos indivíduos, enquanto que na abordagem tradicional são utilizadas informações fenotípicas para inferir sobre o genótipo dos indivíduos

Trabalhos estudando a aplicabilidade da GWS na cultura da soja são escassos e o seu emprego comumente é realizado em etapas mais avançadas em um programa de melhoramento, especialmente para espécies perenes.

Em espécies como o eucalipto e dendê, o emprego da GWS permite aumentar os ganhos por seleção através da seleção precoce de genótipos, utilizando somente dados genotípicos. Nestas situações, a estimativa dos valores genético-genômicos permite selecionar indivíduos antes destes expressarem a característica sob seleção, reduzindo os intervalos entre os ciclos de seleção (RESENDE *et al.*, 2013).

Em estudo apresentado por Resende *et al.* (2012b), os autores testaram a acurácia da GWS para crescimento e qualidade de madeira de eucalipto, a partir de informações de populações de melhoramento oriundas de duas empresas privadas. Os autores destacaram que o método auxilia no entendimento de características quantitativas e é uma ferramenta revolucionária no melhoramento de espécies perenes.

Em outro estudo Zhao *et al.* (2012) utilizaram a GWS em seis populações segregantes de milho para estimar os efeitos de marcadores e verificar a acurácia preditiva das características umidade de grãos e produtividade. Os autores atestaram que a GWS tem potencial de incrementar o ganho de seleção por tempo.

Nos dois estudos supracitados, a GWS foi empregada em populações segregantes, tanto em espécie perene como anual, para seleção de genótipos superiores. Entretanto, a GWS pode ainda ser utilizada na seleção de genitores, visando obter populações híbridas e segregantes com elevada frequência de alelos favoráveis.

Em estudo apresentado por Zhao *et al* (2013), os autores testaram diferentes metodologias estatísticas para determinar os efeitos dos marcadores, visando a predição da performance de populações híbridas. De fato, neste estudo os autores utilizaram linhagens de trigo na população de treinamento do modelo estatístico para então validar e selecionar na população híbrida. Os autores encontraram uma superioridade e similaridade entre as metodologias RR-BLUP e BayesA.

3. MATERIAIS E MÉTODOS

Neste estudo foi avaliada a eficiência na predição da performance de híbridos F₁ a partir de informações fenotípicas associadas a dados moleculares, pelo método de seleção genômica ampla, considerando cruzamentos entre genitores de uma população RILs. Todos dados moleculares e fenotípicos empregados neste estudo foram gerados por simulação, tomando como referência a cultura da soja, conforme descrito a seguir:

3.1 Simulação dos dados

3.1.1 Simulação genotípica da população genitora

Para a geração dos dados foi utilizado o módulo de simulação do aplicativo computacional GENES (CRUZ, 2013).

Neste estudo, o genoma da soja foi tomado como referência, sendo uma espécie diploide $2n=2x=40$ cromossomos. As informações do genoma foram obtidas junto ao banco de dados online da USDA-ARS, o SoyBase (<http://soybase.org>).

O genoma foi gerado com um nível de saturação de 270 marcas moleculares espaçadas variavelmente em cada grupo de ligação, resultando em um total de 5400 marcas moleculares no genoma. As marcas moleculares geradas eram codominantes, permitindo a identificação de indivíduos heterozigotos (marcadores SNPs).

Em seguida, foi simulado uma população de linhagens endogâmicas recombinantes (*Recombinant Inbred Line* – RILs) com 200 indivíduos, oriundas do cruzamento de dois genitores homozigotos contrastantes em fase de aproximação.

3.1.2 Simulação genotípica das populações híbridas

Todas as linhagens que constituíam a população RILs foram cruzadas entre si, desconsiderando os cruzamentos recíprocos, para a formação da população híbrida F₁. A informação genotípica de cada genitor permitiu a obtenção das informações genotípicas dos híbridos.

3.1.3 Simulação das características fenotípicas associada a genitores (RILs) e híbridos

A partir dos dados genotípicos, foram simuladas quatro características fenotípicas simultaneamente para genitores e híbridos, considerando relação intraalélica de ausência de dominância ($gmd=d/a=0$) para os locos controladores das mesmas. A relação intraalélica de ausência de dominância foi considerada a partir de trabalhos apresentados por Kamikoga (1989) e Brim & Cockerham (1961), cujo valores apresentados para a característica produtividade foram inferiores a unidade e próximos a zero. Desta forma, para fins de padronização, a relação de ausência de dominância foi aplicada também para as demais características.

As herdabilidades das características simuladas foram estabelecidas como sendo igual a 0,20, 0,50, 0,70 e 0,80, fazendo alusão a produtividade, acamamento, altura de planta e ramificações laterais, respectivamente. A determinação de tais valores de herdabilidade foi baseada nos trabalhos de Know & Torrie (1964), Aditya *et al* (2011), Wiggins (2012), Dilnesaw *et al* (2013), Ghodrati (2013), Baraskar *et al* (2014) e Hakim *et al* (2014).

O número de locos controladores de cada característica foi determinado com base nas informações contidas no banco de dados da SoyBase (<http://soybase.org>), porém mantendo-se um limite máximo de 100 locos (Tabela 1). Além disso, determinou-se o posicionamento de cada loco em relação ao grupo de ligação para cada característica.

Tabela 1. Número de locos controladores das características produtividade, acamamento, altura de planta e ramificações laterais (diagonal) e correlação fenotípica entre as características (acima da diagonal).

	Produtividade	Altura	Acamamento	Ramif. Lat.
Produtividade	100	-0,040	-0,140	0,150
Altura		73	0,469	0,562
Acamamento			32	0,313
Ramif. Lat.				21

As estimativas das correlações fenotípicas entre as características, necessárias

para a simulação dos valores fenotípicos adotando locos com efeitos pleiotrópicos para diferentes características, foram baseadas nas informações apresentadas por Bekele & Alemahu (2011).

Os fenótipos dos indivíduos (i) foram gerados segundo o modelo $F_i = G_i + M_i$, em que G_i é o efeito genético dado pelo somatório dos efeitos genéticos de cada loco e M_i o efeito ambiental, gerado segundo uma distribuição normal com média zero e variância compatível com a herdabilidade da característica simulada, dada pela seguinte expressão:

$$h^2 = \frac{\sigma_g^2}{\sigma_f^2}$$

Em que:

σ_g^2 refere-se a variância genética associada a característica simulada

σ_f^2 corresponde a variância fenotípica

O valor genético total expresso por um determinado indivíduo foi estimado a partir da expressão:

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

Em que:

$$a_i = \sum_{j=1}^{n'} p_j \alpha_j \quad d_i = 0$$

Sendo α_j o efeito do alelo favorável no loco j , considerado igual a 1, 0 ou -1 para as classes genotípicas AA, Aa e aa, respectivamente, p_j é a contribuição do loco j para a manifestação da característica considerada, sendo considerado contribuição uniforme dos locos, e n' é número de loco controlador da característica.

Devido à pressuposição de $gmd=0$, adotada no momento de simulação, d_i , que corresponde a relação intraalélica de dominância, foi considerado como sendo igual a 0.

3.2 Métodos para seleção de genitores

3.2.1 Métodos para escolha de genitores para formação de população de cruzamento

A seleção dos genitores para formar a população de cruzamento (cruzamento dialélico) foi baseada nas informações da população RILs.

A escolha seguiu abordagens clássicas e uma nova abordagem proposta neste estudo, a partir da GWS. Entre as abordagens clássicas, realizou-se a seleção univariada e multivariada com base em informações fenotípicas. Na metodologia de seleção de genitores empregando-se a GWS, os genitores foram selecionados a partir de seus valores genético genômicos estimados.

1. Escolha univariada com base em informação fenotípica da característica de interesse

Em uma primeira etapa, os genitores foram selecionados baseando-se exclusivamente no desempenho *per se* para a característica produtividade (característica de maior importância agrônômica). As 10 linhagens mais produtivas foram cruzadas entre si em esquema dialélico e tiveram seus híbridos avaliados segundo análise dialélica proposta por Griffing (1956).

Adicionalmente, analisou-se os dados fenotípicos *per se* dos híbridos F₁ para determinar o potencial deste critério de seleção em gerar populações híbridas fenotipicamente superiores.

2. Escolha multivariada com base em informação fenotípica

Uma análise multivariada da diversidade genética, com base nas informações de quatro características fenotípicas simuladas, foi realizada visando a escolha de genitores para formação da população de cruzamento. O objetivo foi explorar a diversidade genética existente entre os genitores.

A análise de dissimilaridade, que mede a distância par a par entre indivíduos, foi empregada através da distância Euclidiana média, cujo estimador é dado por:

$$d_{ii'} = \sqrt{\frac{1}{v} \sum (Y_{ij} - Y_{i'j})^2}$$

sendo v o número de características, Y_{ij} a observação no i -ésimo indivíduo ($i=1,2,\dots,p$), em referência a j -ésimo característica ($j=1,2,\dots,n$) estudada.

Em posse da matriz de dissimilaridade aplicou-se o método de Tocher para agrupamento das linhagens. Neste método, os indivíduos mais similares formam o grupo inicial. A inclusão de novos indivíduos é realizada adotando-se o critério de que a distância intragrupo deve ser menor que a distância média intergrupos. Desta forma, os indivíduos de um mesmo grupo são geneticamente mais similares entre si do que entre os indivíduos de outros grupos.

Visando a obtenção de informações adicionais de diversidade genética para aumentar a confiabilidade na seleção dos genitores, utilizou-se a técnica de componentes principais. Esta técnica realiza o agrupamento dos genótipos em dispersão gráfica. Na dispersão gráfica, aqueles indivíduos que estão situados próximos são geneticamente mais similares, sendo o oposto válido. Para realizar a análise de componentes principais

considera que x_{ij} é a média padronizada do j -ésimo caráter ($j=1,2,\dots,v$) avaliado no i -ésimo indivíduo ($i=1,2,\dots,g$) e R a matriz de correlação entre esses caracteres (ou matriz de correlação fenotípica entre os caracteres baseada em dados originais). Este conjunto de v variáveis ($x_{i1}, x_{i2}, \dots, x_{iv}$) é transformado em um novo conjunto ($Y_{i1}, Y_{i2}, \dots, Y_{iv}$), que são funções lineares dos $x_{i's}$ e independentes entre si. Entre todos os componentes principais, Y_{i1} apresenta a maior variância, Y_{i2} a segunda maior, e assim sucessivamente.

Assim como no critério univariado, os genitores selecionados a partir das análises de diversidade genética foram cruzados entre si, em todas as combinações possíveis, sem a realização de cruzamentos recíprocos, e o potencial das populações híbridas foi avaliada segundo o de análise dialélica proposta por Griffing (1956).

Posteriormente, foi realizada uma análise *per se* dos dados fenotípicos das populações híbridas visando estimar a capacidade desta metodologia de seleção de genitores em gerar híbridos fenotipicamente superiores.

3. Escolha univariada com base em valor genotípico predito dos híbridos

Em adição aos métodos de seleção de genitores anteriormente realizados, utilizou-se o valor genético genômico estimados das linhagens como critério univariado para seleção dos genitores a constituírem o bloco de cruzamento. Esta metodologia visa selecionar as melhores linhagens a partir de informações genômicas e não fenotípicas, como anteriormente realizado no critério fenotípico univariado, permitindo atenuar os efeitos ambientais e incrementar a variância genética selecionada.

Assim sendo, a partir dos dados genotípicos e fenotípicos para a característica produtividade, obtidos das 200 linhagens genitoras e dos 45 híbridos oriundos do cruzamento de genitores selecionados por critério fenotípico multivariado, foram estimados os efeitos de cada um dos locos marcadores que, somados, compõem o valor genético-genômico estimado de cada indivíduo. A inclusão de populações híbridas foi realizada visando elevar o tamanho da população de treinamento e adicionar informações de locos heterozigóticos, de forma a incrementar a acurácia preditiva da GWS para as populações híbridas. Para isso, foi utilizada o método estatístico RR-BLUP.

O método Ridge Regression-Best Linear Unbiased Prediction (RR-BLUP) é denominado regressão aleatória ou regressão cumeeira. Este método estima simultaneamente o efeito de todos os marcadores para a seleção genômica (WHITTAKER *et al.*, 2000; MEUWISSEN *et al.*, 2001). A regressão aleatória promove o *shrinkage* dos efeitos de todos os marcadores para zero, assumindo que os marcadores têm efeito aleatório com variância constante (BREIMAN, 1995; WHITTAKER *et al.*,

2000; MEUWISSEN *et al.*, 2001). De acordo com Breiman (1995), a regressão de cumeeira é mais apropriada para características com pouca ou nenhuma região de grande magnitude e muitos genes de pequeno efeito.

Os coeficientes de regressão ridge são definidos como aqueles que minimizam a soma de quadrados. O seguinte modelo linear misto geral foi usado para estimar os efeitos dos marcadores, conforme Resende *et al* (2012a).

$$y = Xb + Za + e$$

Em que:

y é o vetor de observações fenotípicas;

b é o vetor de efeitos fixos (média geral);

a é o vetor dos efeitos aleatórios dos marcadores;

e é o vetor dos erros aleatórios;

X e Z são as matrizes de incidência para b e a .

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

$$a \sim N(0, G = I\sigma_m^2)$$

$$E(y) = Xb$$

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

$$\text{Var}(y) = V = ZGZ' + R$$

$$G = I\sigma_e^2/n$$

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + 1 \frac{\sigma_e^2}{(\sigma_g^2/n)} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Em que σ_e^2 é a variância residual e σ_g^2 é a variância genética.

Desta forma, com os efeitos dos marcadores estimou-se os valores genético genômicos de cada indivíduo, nas populações em questão, por meio do seguinte estimador:

$$VGG = \hat{y}_i = \sum_i Z_{ij} \hat{\alpha}_i$$

As equações de predição apresentadas acima assumiram lócus com uniformidade na σ_g^2 . Desta forma, a variação genética explicada por cada loco é dado por σ_g^2/n , em que σ_g^2 é a variância genética total e n é o número de marcadores utilizados.

A partir dos efeitos de cada marcador, obtido da população de treinamento (linhagens potencialmente genitoras e amostra de híbridos), a equação de predição de valores genético genômicos foi testada para verificar a sua acurácia na população de validação (populações híbridas). O cômputo da acurácia foi dado pela correlação entre os

valores genético-genômico estimados com o valor fenotípico observado dividido pela raiz quadrada da herdabilidade da característica estudada.

Em posse dos valores estimados pela GWS para as linhagens genitoras, selecionaram-se 10 progenitores superiores para formarem o bloco de cruzamento, com base no seu valor genético-genômico estimado, de forma que os indivíduos com maior valor genético-genômico fossem selecionados. Posteriormente, realizou-se o cruzamento destes genitores e os híbridos F₁ resultantes, a partir de seus valores genético genômicos preditos, tiveram a capacidade geral e específica de combinação determinada segundo a metodologia de análise dialélica proposta por Griffing (1956). Neste caso, diferentemente dos dois dialelos realizados anteriormente, não foram realizados cruzamentos, ou seja, os valores dos híbridos foram estimados.

Em uma segunda etapa, para fins de comparação, foram considerados isoladamente os valores fenotípicos *per se* deste conjunto de híbridos F₁ para estimar o potencial da seleção univariada baseada no valor genético-genômico estimados em gerar híbridos fenotipicamente superiores.

3.2.2 Estabelecimento das melhores combinações híbridas

- Predição com base em efeitos da capacidade combinatória

A análise dialélica das populações híbridas, cujos genitores foram selecionados por meio de análises fenotípicas uni e multivariada e, foi realizada segundo o método IV, modelo I, da metodologia proposta por Griffing (1956). Para este método foi utilizado o modelo estatístico:

$$Y_{ij} = m + g_i + g_j + s_{ij} + \bar{\epsilon}_{ij}$$

em que:

Y_{ij} é o valor médio do híbrido ij ($i, j=1, 2, \dots, p$, $i < j$);

g_i, g_j são os efeitos da capacidade geral de combinação do i -ésimo e j -ésimo progenitor, respectivamente;

s_{ij} é o efeito da capacidade específica de combinação para os cruzamentos entre progenitores de ordem i e j ; e

$\bar{\epsilon}_{ij}$ é o erro experimental médio.

Neste modelo é considerado que $s_{ij} = s_{ji}$.

Em seguida, estimou-se a capacidade geral e específica de combinação para cada genótipo e para cada cruzamento, respectivamente, para então determinar as combinações híbridas mais promissoras.

- Predição do valor genético de híbridos

As linhagens selecionadas para formarem o bloco de cruzamento com base unicamente em valor genético genômico estimado tiveram o seu potencial como genitores avaliado a partir do valor genético genômico predito dos híbridos, não sendo utilizados dados fenotípicos. Em outras palavras, a partir do efeito dos marcadores, oriundos da associação de valores genotípicos e fenotípicos, e da genotipagem dos híbridos não realizados (obtida pela associação das marcas moleculares das linhagens genitoras), foi estimado o desempenho de cada população híbrida.

Da mesma forma que para os critérios anteriores, a capacidade geral e específica de combinação dos híbridos preditos foi estimada pela mesma metodologia de análise dialélica proposta por Griffing (1956) supra explicada.

-Análise fenotípica das melhores combinações híbridas

Em adição ao estabelecimento das melhores populações híbridas a partir de análise dialélica, realizou-se a seleção dos 20% melhores híbridos para cada bloco de cruzamento, cuja média foi comparada entre as diferentes abordagens de seleção de genitores. Neste caso, almejou-se verificar o potencial das metodologias em gerar híbridos fenotipicamente superiores.

Desta forma, ao realizar a predição de um grande conjunto de populações híbridas a partir da GWS, é possível estimar quão superior ou inferior fenotipicamente os híbridos gerados por esta abordagem serão frente as populações híbridas oriundas de técnicas tradicionais de seleção de genitores.

3.3 Predição de híbridos não realizados

Em um programa de melhoramento de escala comercial, a realização de todas as combinações híbridas entre as 200 linhagens de soja é impraticável, tanto pela necessidade de um exuberante volume de hibridações quanto pela avaliação dos respectivos híbridos. Entretanto, a partir da genotipagem das linhagens genitoras é possível obter, pela associação par a par de cada marcador, a genotipagem dos híbridos, sem a necessidade de se realizar o híbrido, ou seja, sem custo com genotipagem, hibridação e condução dos materiais. Desta forma, utilizando os dados genotípicos dos $p(p-1)/2$ híbridos e os efeitos dos marcadores, estimados via GWS, torna-se possível a prever o valor genético genômico de todos os híbridos, os quais sequer foram realizados. Neste cenário, a grande vantagem e principal objetivo desta metodologia é a realização

exclusiva dos cruzamentos cujo híbridos são superiores. Esta superioridade está em função das marcas que o híbrido apresenta e o seu efeito sobre a característica sob seleção, que, somados, resultam no valor genético-genômico predito. Desta forma, uma nova abordagem de seleção de genitores baseada no valor genético genômico predito dos híbridos é proposta neste trabalho.

Para tanto, conforme descrito anteriormente, foram estimados os valores genético-genômicos dos híbridos oriundos do inter cruzamento entre todas as 200 linhagens genitoras. Diferentemente da etapa realizada anteriormente, em que foram estimados e analisados apenas um conjunto de 45 híbridos, o objetivo desta etapa foi de estimar os valores genético-genômicos de todos os híbridos possíveis. Além disto, a estimação dos valores genético-genômicos foi realizada considerando-se as quatro características fenotípicas simuladas (produtividade, altura de planta, acamamento e ramificações laterais).

Para cada característica fenotípica analisada, foram selecionados os híbridos com maiores valores genético-genômico preditos, de forma que estes representassem 20% do total de híbridos.

Em seguida, avaliou-se o desempenho fenotípico deste conjunto de híbridos superiores, almejando determinar o potencial da GWS em determinar os melhores híbridos a partir de seus valores genético genômicos. Neste caso, o objetivo deste estudo não foi aferir a acurácia preditiva da GWS para todos os híbridos da população de validação. De fato, almejou-se estimar a acurácia entre o valor genético genômico dos melhores híbridos com seus respectivos valores fenotípicos.

3.4 Aplicativos computacionais para análise dos dados

As simulações foram desenvolvidas pelo programa GENES (CRUZ, 2013) e as análises estatísticas através de uma interface do programa GENES com o software R, utilizando o pacote rr-BLUP (ENDELMAN *et al.*, 2011).

4. RESULTADOS E DISCUSSÕES

- Estratégia 1: Escolha de genitores e performance de híbridos a partir de análise univariada com base em informação fenotípica da característica de interesse.

As linhagens selecionadas para formarem o bloco de cruzamento foram aquelas com maior valor fenotípico para a característica produtividade. Assim, em ordem decrescente de produtividade, as linhagens selecionadas foram 57, 15, 37, 178, 103, 183,

131, 142, 49 e 33.

A análise de variância para os híbridos não foi realizada, visto que se obtinha somente informação de um único indivíduo por cruzamento, sem repetições. Entretanto, para fins de análise da capacidade combinatória com testes de hipóteses adotou-se como estimativa da variância residual o valor paramétrico utilizado na simulação.

Os valores de capacidade geral de combinação (CGC) das linhagens genitoras se encontram na Tabela 2. Os genitores 57 e 49 foram os que apresentaram os maiores valores de CGC, sendo 0.103 e 0.076, respectivamente. Isto indica que estes genitores tem uma elevada frequência de alelos favoráveis transmitidos à descendência contribuindo para o bom desempenho de seus híbridos. Isto é observado pela elevada média fenotípica dos híbridos que continham estas linhagens como genitoras. Genitores cujo híbridos apresentaram maiores médias também apresentaram elevada CGC, corroborando com Freire Filho (1988), Freire Filho & Vello (1989), Lopes *et al* (2002) e Vieira (2009).

A capacidade específica de combinação (CEC) indica se um cruzamento específico segue o modelo esperado pela CGC dos genitores, ou seja, estima os efeitos não aditivos dos genes (dominantes e epistáticos). Por meio da análise dos valores de CEC foi possível identificar híbridos com potencial para produtividade de grãos por planta (Tabela 2).

Tabela 2. Estimativas da capacidade geral combinação (CGC), capacidade específica de combinação (CEC) (acima da diagonal), dos valores fenotípicos de cada híbrido (abaixo da diagonal) e da média dos valores fenotípicos dos híbridos oriundos de cruzamentos envolvendo cada linhagem selecionada com base no desempenho *per se* para produtividade.

Genitores	57	15	37	178	103	183	131	142	49	33
57	-	-0,18	-0,02	0,36	0,17	-0,55	-0,18	-0,11	0,19	0,31
15	0,46	-	0,09	0,25	-0,25	-0,12	-0,10	0,14	-0,10	0,27
37	0,45	0,43	-	-0,36	-0,13	0,48	-0,17	0,19	-0,15	0,07
178	0,93	0,70	-0,08	-	0,22	0,07	-0,26	0,22	-0,35	-0,14
103	0,83	0,28	0,23	0,68	-	0,41	0,16	0,03	0,06	-0,67
183	0,13	0,43	0,86	0,56	0,98	-	0,40	-0,49	-0,2	-0,01
131	0,54	0,49	0,25	0,27	0,77	1,03	-	-0,03	0,18	0,02
142	0,60	0,71	0,60	0,73	0,64	0,12	0,63	-	0,13	-0,09
49	0,94	0,52	0,29	0,20	0,70	0,45	0,88	0,82	-	0,22
33	1,01	0,83	0,47	0,36	-0,08	0,60	0,67	0,56	0,89	-
CGC	0,10	-0,03	-0,20	-0,09	-0,01	0,01	0,06	0,04	0,08	0,03
Média dos híbridos	0,66	0,54	0,39	0,48	0,56	0,58	0,61	0,60	0,63	0,59

O cruzamento que proporcionou CEC positiva de maior magnitude foi 183 x 37. Neste cruzamento o híbrido apresentou desempenho muito superior ao esperado pela CGC, indicando complementariedade alélica entre os genitores. Entretanto, este cruzamento não é apropriado para a extração de linhagens superiores devido aos baixos valores de CGC dos genitores, indicando que estes portam poucos alelos favoráveis para a característica em questão. Desta forma, os híbridos destes cruzamentos apresentam relativamente poucos alelos favoráveis e estes se encontram em heterozigose nos loci em questão, sujeitos aos efeitos da segregação aleatória, dificultando a obtenção de uma linhagem com elevada frequência de alelos favoráveis.

Os cruzamentos mais promissores, com valores de alta magnitude e positivos para a CEC, foram 57 x 178, 57 x 33, 183 x 131 e 183 x 103. Estes cruzamentos revelaram boa complementariedade alélica entre os seus genitores, visto que eram combinações de um genitor portador de muitos alelos favoráveis e outro pouco promissor segundo análise dos valores de CGC, apontando o potencial destas populações segregantes em gerar combinações alélicas superiores. Valores contrastantes da CGC também proporcionaram híbridos com elevada CEC no trabalho de Freire Filho (1988) e Vieira (2009).

Outros cruzamentos apresentaram CEC positiva, porém de moderada magnitude (142 x 37, 142 x 178, 49 x 31 e 33 x 49). Estas populações também podem ser selecionadas, contudo, segundo a metodologia de análise dialélica, a variabilidade existente é menor, assim como a probabilidade de se extrair linhagens superiores.

Embora os cruzamentos anteriores apresentaram elevados valores de CEC, o híbrido mais promissor foi obtido do cruzamento entre os genitores 57 e 49, visto que ambos os genitores possuíam altas magnitudes de CGC, indicando que transmitem a descendência alelos favoráveis. Nesta situação, a baixa magnitude da CEC sugere moderada complementariedade alélica entre os genitores, ou seja, moderada dissimilaridade genética. Entretanto, o excelente desempenho fenotípico do híbrido F_1 , oriundo de um cruzamento relativamente aparentado, remete a uma moderada complementariedade de alelos favoráveis, gerando um híbrido com muitos loci fixados.

A análise fenotípica das populações híbridas exibiu resultados promissores. A seleção das 9 melhores populações híbridas (20% superiores) com base isoladamente no desempenho fenotípico *per se* apontou que cinco destes híbridos foram classificados como superiores também pela análise dialélica (híbrido e ao menos um genitor com elevada e positiva $\hat{\sigma}_{ij}$ e \hat{g}_i , respectivamente). Este resultado permite associar valores fenotípicos com informações genotípicas, visto que a capacidade de combinação representa a frequência de alelos favoráveis. Assim, as cinco populações híbridas

apresentam bom desempenho fenotípico devido à presença de alelos favoráveis. Estes híbridos apresentaram média elevada, de 0,93, sendo muito superior à média global dos híbridos deste bloco de cruzamento, igual a 0,56.

- Estratégia 2: Escolha de genitores e híbridos a partir de sistemas dialélicos estabelecidos por análise multivariada com base em informação fenotípica

A matriz de dissimilaridade, utilizando-se a distância Euclidiana média, apontou a maior dissimilaridade entre os genótipos 142 e 155, com 0,709. A maior similaridade foi observada entre os genótipos 118 e 148, com valor de 0,006. Em média, a dissimilaridade entre as 200 linhagens de soja foi de 0,25.

Tabela 3. Agrupamento dos genótipos de acordo com o método de otimização de Tocher utilizando a distância Euclidiana média.

I	118, 148, 156, 7, 2, 48, 158, 50, 168, 164, 99, 128, 73, 110, 146, 124, 144, 20, 6, 66, 157, 188, 123, 100, 195, 135, 18, 161, 171, 191, 151, 181, 159, 126, 177, 60, 165, 114, 92, 25, 67, 166, 160, 162, 87, 186, 180, 176, 174, 192, 120, 187, 47, 94, 200, 68, 182, 75, 149, 163, 143, 76, 38, 91, 52, 84, 8, 134, 45, 93, 86, 197, 62
II	33, 132, 49, 78, 9, 139, 44, 103, 61, 22, 71, 109, 42, 137, 108, 175, 28, 27, 104, 32, 111, 79, 121, 150, 55, 185, 102, 4, 39, 152
III	46, 173, 140, 133, 98, 138, 53, 129, 131, 41, 97, 12, 193, 56, 77, 85
IV	19, 30, 127, 90, 167, 112, 81, 196, 64, 59, 35, 147, 145, 101, 82, 119, 107, 170, 23
V	43, 184, 40, 88, 11, 117, 3, 54, 13, 172, 179, 26
VI	36, 65, 24, 95, 10, 199, 14
VII	125, 194, 51, 142, 80, 198
VIII	29, 136, 17, 169, 122, 154, 83, 69
IX	5, 190, 1, 16, 74
X	58, 116, 105
XI	70, 141, 155
XII	72, 89, 113
XIII	37, 178
XIV	63, 183
XV	34, 96
XVI	106, 189
XVII	130, 153
XVIII	31
XIX	115
XX	21
XXI	15
XXII	57

O método de agrupamento de otimização de Tocher separou as 200 linhagens em 22 grupos, em que cinco eram constituídos de um único genótipo (Tabela 3). Observou-se que a grande maioria dos genótipos ficou agrupada no primeiro e segundo grupos, os quais representam 36,5% e 15% de todos os genótipos, respectivamente. Similarmente, o agrupamento de Tocher de cultivares de soja realizado por Villela *et al* (2014) demonstrou que o primeiro grupo continha 74,3% das cultivares avaliadas, num total de 74, e que três grupos continham apenas uma cultivar, num total de sete grupos formados.

Por outro lado, a análise de componentes principais permitiu inferir sobre quatro complexos de variáveis. Nesta análise, os autovalores são utilizados para determinar quantos componentes devem ser explorados. Estes podem ser considerados como um índice que afeta a intensidade dos componentes principais. Se o valor do componente principal é menor do que a unidade então a capacidade explanatória do conjunto de dados originais é maior do que a do componente. Geralmente, se o autovalor é maior do que um ele pode ser utilizado como um critério de inclusão. Neste estudo, somente o primeiro componente principal apresentou valor superior a um, absorvendo grande parte da variação total presente no conjunto analisado (Tabela 4).

Tabela 4. Autovalores, porcentagem de variação e variação acumulada por componente principal (CP) obtida da análise de componentes principais entre 200 linhagens de soja para quatro características fenotípicas.

	CP1	CP2	CP3	CP4
Autovalor	2,087	0,789	0,619	0,504
Porcentagem de Variação	52,19	19,73	15,48	12,59
Variação Acumulada	52,19	71,92	87,4	100

A análise de componentes principais realizada por Hashash (2016) mostrou que somente os dois primeiros componentes principais apresentavam autovalores superiores a unidade, sendo os cinco restantes inferiores a unidade. Isto indica que a utilização dos últimos autovetores seria desvantajoso frente ao uso do conjunto de dados originais, visto que os dois primeiros componentes principais explicaram 82,55% da variação total existente entre 10 genótipos de soja.

Em estudo apresentado por Vianna *et al* (2013), quatro autovalores se apresentaram superiores a unidade, contendo 71,6% da variação total das 11 características de 95 progênies de soja avaliadas. Já Khan *et al* (2014) reportaram que três autovalores foram superiores a um, os quais contribuíam com 71,11% da variação original

de 115 genótipos de soja.

Ainda de acordo com a Tabela 4, os dois primeiros componentes explicaram somente 71,92 % da variação total, sendo inferior ao valor recomendado por Cruz *et al* (2012), de 80%. Desta forma, utilizou-se a informação dos três primeiros componentes principais, que juntos explicaram 87,4% da variação total, para representar graficamente, em plano tridimensional, a diversidade genética entre os genótipos (Figura 1).

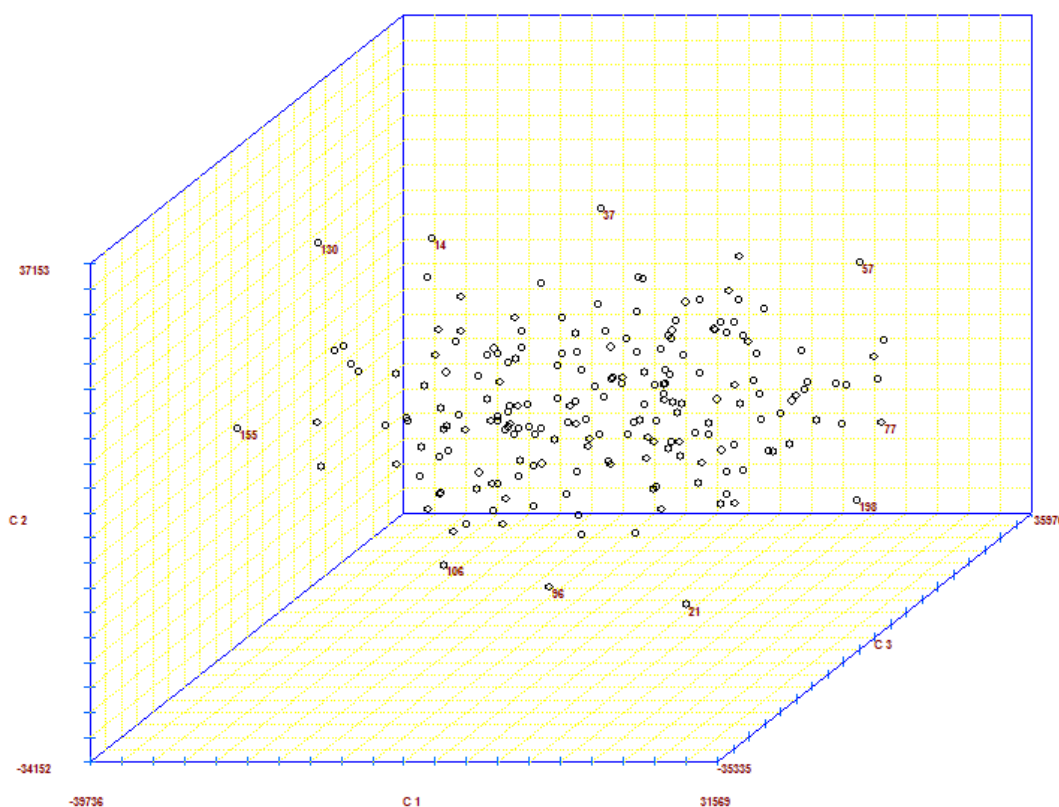


Figura 1. Dispersão gráfica das 200 linhagens de acordo com os três primeiros componentes principais. Os indivíduos identificados são os selecionados para compor o bloco de cruzamento.

Por outro lado, em estudo realizado por Villela *et al* (2014) os autores mostraram que quatro componentes principais absorveram quase 81% da variação total de 74 cultivares de soja, em que o primeiro componente expressava somente 36,65% da variação observada. Narjesi *et al* (2007) também necessitaram de mais de três componentes principais (cinco) para explicar 80,2% da variação observada em 30 genótipos de soja.

Com base na análise visual da dispersão gráfica dos genitores segundo o método de componentes principais e no agrupamento de Tocher, buscou-se selecionar genótipos que se apresentavam distantes graficamente (indivíduos identificados na Figura 1) e que estivessem alocados em grupos diferentes, respectivamente. Desta forma, seria explorado

o máximo da diversidade genética entre as linhagens. Seguindo este critério, foram selecionadas as seguintes linhagens para formarem o bloco de cruzamento: 14, 21, 37, 57, 77, 96, 106, 130, 155 e 198, respectivamente alocados nos grupos VI, XX, XIII, XXII, III, XV, XVI, XVII, XI e VII.

Os valores de CGC das linhagens genitoras selecionadas são apresentadas na Tabela 5. Nota-se que as linhagens 37, 57, 77, 96 e 198 apresentaram valores positivos para a CGC, demonstrando a capacidade destas em transmitir para a descendência, aqui avaliada nos híbridos, seus alelos favoráveis.

Tabela 5. Estimativas da capacidade geral combinação (CGC), capacidade específica de combinação (CEC) (acima da diagonal), dos valores fenotípicos de cada híbrido (abaixo da diagonal) e da média dos valores fenotípicos dos híbridos oriundos de cruzamentos envolvendo cada linhagem selecionada com base em análise de diversidade genética.

Genitores	14	21	37	57	77	96	106	130	155	198
14	-	0,31	0,23	0,15	-0,04	0,21	-0,55	-0,08	-0,11	-0,13
21	0,64	-	-0,32	-0,44	0,31	0,20	0,49	-0,04	-0,27	-0,23
37	0,62	0,19	-	-0,28	-0,21	-0,05	0,15	-0,09	0,16	0,41
57	0,69	0,24	0,45	-	0,14	-0,18	0,10	-0,02	0,52	0,01
77	0,67	1,16	0,70	1,20	-	0,053	-0,1	-0,06	-0,21	0,11
96	0,74	0,86	0,66	0,69	1,10	-	-0,10	0,24	-0,28	-0,08
106	-0,51	0,67	0,38	0,49	0,46	0,27	-	-0,26	0,15	0,13
130	0,06	0,19	0,20	0,43	0,56	0,66	-0,31	-	0,27	0,04
155	0,19	0,17	0,65	1,17	0,62	0,35	0,29	0,48	-	-0,24
198	0,55	0,59	1,28	1,03	1,31	0,93	0,66	0,62	0,55	-
CGC	-0,19	-0,05	0,01	0,16	0,34	0,14	-0,34	-0,28	-0,08	0,30
Média dos híbridos	0,40	0,52	0,57	0,71	0,87	0,69	0,27	0,31	0,50	0,83

Foram observados cruzamentos promissores de acordo com a CEC (Tabela 5). Os híbridos oriundos do cruzamento 21 x 106, 14 x 21 e 130 x 155 apresentaram elevada CEC, entretanto envolveram genitores com baixa CGC, ou seja, com poucos alelos favoráveis para o aumento do valor fenotípico da característica. Desta forma, estes cruzamentos são pouco promissores para serem mantidos em um programa de melhoramento.

As populações 21 x 77, 96 x 14, 96 x 21, 96 x 130, 57 x 155 e 198 x 37 seriam selecionadas por apresentarem elevada capacidade específica de combinação e ao menos um genitor com elevada capacidade geral de combinação. Desta forma, estar-se-ia explorando a complementariedade de alelos favoráveis em ambos os genitores, favorecendo populações híbridas com elevada \hat{s}_{ij} e elevando a probabilidade de

ocorrência de linhagens superiores.

Novamente, empregou-se a seleção de 20% dos híbridos com maiores valores fenotípicos. As 9 populações híbridas com maiores valores fenotípicos observados foram 198 x 77, 198 x 37, 77 x 57, 155 x 57, 77 x 21, 96 x 77, 198 x 57, 198 x 96 e 98 x 21. A performance média destes híbridos foi de 1,11, sendo, novamente, muito superior à média dos demais híbridos gerados neste bloco de cruzamento. Além disso, o desempenho médio dos melhores híbridos foi superior aos melhores híbridos cujo genitores foram selecionados por critério fenotípico univariado.

Em contrapartida, um menor número de populações híbridas (quatro) foi identificado como superiores segundo critério fenotípico *per se* e pela metodologia de análise dialélica (198 x 37, 55 x 157, 21 x 77 e 21 x 96). A corroboração das melhores populações híbridas pelos dois critérios realizados assegura ao melhorista a seleção de indivíduos portadores de alelos favoráveis e que, fenotipicamente, expressam seu potencial.

Uma pertinente observação a ser ressaltada é que as linhagens 37 e 57 foram selecionadas para comporem o bloco de cruzamento nas duas abordagens de seleção de genitores empregadas. Entretanto, foi possível notar uma substancial superioridade da linhagem 57 em gerar híbridos promissores em relação a linhagem 37, atestando a capacidade deste genitor em transmitir alelos favoráveis a descendência.

- Estratégia 3: Escolha de genitores e híbridos baseada no valor genético genômico predito de linhagens

O uso da seleção genômica ampla para estimar os efeitos dos 5400 marcadores, com base nos dados fenotípico de produtividade, permitiu estimar os valores genético genômico das linhagens, que foram utilizados para seleção daquelas com elevado potencial produtivo. As dez linhagens com maior valor genético genômico selecionadas foram: 15, 33, 35, 37, 44, 49, 51, 56, 57 e 62.

Conhecendo as linhagens promissoras, realizou-se a predição do desempenho dos híbridos (valor genético genômico predito), em esquema dialélico, sem recíproco. Estes valores foram utilizados para determinar a capacidade geral e específica de combinação das mesmas e, conseqüentemente, reportar os cruzamentos mais promissores.

Quatro linhagens apresentaram valores positivos de CGC, sendo as linhagens 51 e 57 as de valores de maiores magnitudes (Tabela 6). Assim, os híbridos oriundos de cruzamentos envolvendo esses genitores apresentaram elevados valores genético genômicos médios. Vieira (2009) observou que os valores da CGC possuíam uma relação

direta na média das populações híbridas, ou seja, maiores médias estavam associadas a populações originadas de genitores com maiores CGC.

Analisando os dados da capacidade específica de combinação, verificou-se que quatro cruzamentos, sendo eles 15 x 49, 15 x 56, 49 x 56 e 49 x 62, apresentaram valores positivos com altas magnitudes (Tabela 6). Estes cruzamentos apresentaram-se promissores, pois ao menos um genitor envolvido no cruzamento apresentou valores positivos para a capacidade geral de combinação.

O cruzamento 15 x 49 favoreceu a obtenção de um híbrido com desempenho superior ao esperado para ambos genitores, indicando a complementariedade alélica. Além disso, envolveu genitores com CGC positiva, revelando que os mesmos possuem alelos distintos favoráveis ao aumento da produtividade. Portanto, este híbrido pode ser utilizado para melhor explorar a variabilidade genética existente conduzindo à obtenção de futuras linhagens superiores para a característica em questão.

Tabela 6. Estimativas da capacidade geral combinação (CGC), capacidade específica de combinação (CEC) (acima da diagonal), dos valores fenotípicos de cada híbrido (abaixo da diagonal) e da média dos valores fenotípicos dos híbridos oriundos de cruzamentos envolvendo cada linhagem selecionada com base no valor genético genômico predito.

	15	33	35	37	44	49	51	56	57	62
15	-	-0,001	-0,001	-0,001	-0,001	0,005	-0,001	0,007	-0,001	-0,002
33	0,83	-	0,002	0,002	0,001	-0,002	0,002	-0,001	0,002	0
35	0,94	0,16	-	0,001	0,002	-0,002	0,001	-0,001	0,002	0
37	0,43	0,47	0,36	-	0,001	-0,002	0,001	-0,001	0,001	0
44	1,07	0,92	0,51	1,47	-	-0,002	0,002	-0,001	0,002	0
49	0,52	0,89	0,64	0,29	0,58	-	-0,002	0,005	-0,002	0,006
51	0,79	0,86	0,09	0,70	0,62	0,89	-	-0,001	0,001	0
56	0,65	1,11	0,58	1,03	1,10	0,39	0,91	-	-0,001	-0,002
57	0,46	1,01	0,82	0,45	1,08	0,94	0,48	0,82	-	0
62	1,44	0,42	0,86	0,55	0,73	0,44	0,46	0,97	0,91	-
CGC	0,006	-0,066	-0,049	-0,084	-0,021	0,001	0,17	-0,037	0,089	-0,009
Média dos híbridos	1,29	1,22	1,24	1,21	1,26	1,28	1,43	1,25	1,36	1,27

O cruzamento entre os genitores 51 e 57, os quais portam a maior frequência de alelos favoráveis, verificada pela alta magnitude da CGC, não resultou em elevada CEC na população híbrida. Isto pode ser consequência do compartilhamento do mesmo conjunto de alelos favoráveis para a característica produtividade.

As informações fenotípicas dos híbridos cujos valores genético-genômicos foram utilizados na análise dialélica ressaltam o potencial da GWS na seleção de genitores

(Tabela 6).

A análise dos valores da capacidade específica de combinação das populações híbridas apontou um reduzido número de cruzamentos promissores, revelando um baixo potencial desta abordagem de seleção de genitores. No entanto, a análise dos valores fenotípicos das populações híbridas apontou a ocorrência de um elevado número de híbridos superiores. Desta forma, nota-se que o uso de valores genético-genômicos para análise dialélica não necessariamente reflete o potencial das populações híbridas.

A seleção dos 20% melhores híbridos pelo valor fenotípico para a característica produtividade não corroborou com nenhum híbrido superior segundo a análise dialélica. Entretanto, o valor fenotípico médio das populações híbridas selecionadas foi igual a 1,14, sendo superior à média dos melhores híbridos oriundo do cruzamento de genitores selecionados por critérios fenotípicos uni e multivariados. Além disso, a média global dos híbridos também se apresentou superior, sendo igual a 0,68.

Embora a seleção de genitores com base no valor genômico estimado das linhagens seja similar ao procedimento empregado na seleção fenotípica univariada, ou seja, considerando apenas uma variável para seleção, esta primeira apresenta relativa superioridade por considerar informações genômicas, visto que modelo estatístico da GWS atenua os efeitos do ambiente e eleva a herdabilidade da característica a nível populacional (RESENDE *et al*, 2013). Em contrapartida, ao considerar somente os valores fenotípicos das linhagens o efeito ambiental pode induzir a seleção de um indivíduo portador de poucos alelos favoráveis, cuja superioridade fenotípica é decorrente do melhor aproveitamento da influência ambiental.

Predição de híbridos não realizados

A averiguação do nível de sucesso na predição do desempenho dos híbridos para a produtividade, característica com baixa herdabilidade e controlada por vários loci, se encontra na Figura 2. Observa-se que 30% dos híbridos preditos apresentaram ótimo desempenho fenotípico, enquanto outros 24% foram ligeiramente inferiores, apresentando desempenho classificado como bom.

Estes resultados indicam que a GWS é capaz de corretamente predizer mais da metade dos melhores híbridos (categorias ótimo e bom). Todavia, do conjunto de melhores híbridos preditos, foram contabilizados 16% e 10% que apresentaram desempenho fenotípico ruim e péssimo, respectivamente. Assim sendo, ao se utilizar da predição da seleção genômica ampla para posicionamento dos cruzamentos a serem realizados, o melhorista aposta na probabilidade de que metade dos híbridos apresentarão

desempenho superior, porém podendo originar um quarto de híbridos pouco promissores.

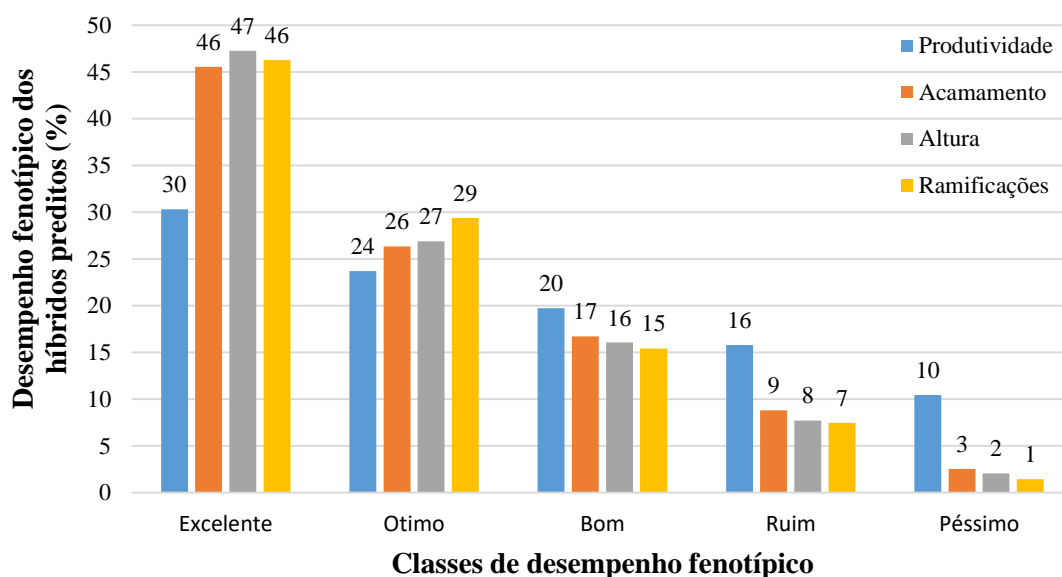


Figura 2. Desempenho fenotípico para as características produção de grãos, acamamento, altura de plantas e ramificações laterais dos melhores híbridos preditos via GWS. Valores arredondados.

Estes resultados evidenciam a superioridade da GWS frente as análises dialélicas uni e multivariadas. Os modelos convencionais proporcionaram 26% e 20% das populações híbridas com desempenho ótimo, respectivamente, conforme média fenotípica das populações híbridas dos blocos de cruzamento. Em contrapartida, a predição de todos cruzamentos pela GWS rendeu 30% de populações híbridas superiores.

Essa superioridade, de quase 15% e 34%, respectivamente, poderia ainda ser maior caso fossem consideradas informações genômicas dos genitores para seleção dos híbridos a serem realizados. Isto ocorre, pois, o desempenho das populações híbridas esteve associada ao valor genético genômico estimado dos genitores. Uma amostra destes valores apontou que os híbridos agrupados na classe ótima apresentaram maior valor genético genômico para os parentais, de 0,571, enquanto que as demais classes apresentaram valor média amostrado igual a 0,532, 0,531, 0,522 e 0,448, respectivamente para as classes bom, regular, ruim e péssima. A pertinência desses valores é consequência do elevado r^2 (correlação entre valor genético-genômico estimado e valor fenotípico) dos valores genético-genômico dos genitores (0,676), visto que a acurácia dos híbridos foi substancialmente menor (0,075) e o seu emprego isoladamente reduz a probabilidade de sucesso. Desta forma, ao ponderar o uso de informações dos genitores para seleção dos híbridos a serem realizados, estar-se-á aumentando o volume e a pertinência de informações consideradas, incrementando a confiabilidade e o ganho com seleção.

Um dos fatores que contribuíram para o maior percentual de híbridos superiores apresentados pela GWS foi a realização exclusiva dos híbridos considerados promissores. Nas análises dialélicas, geralmente se selecionam um conjunto de genitores que são intercruzados, independente de alguns cruzamentos apresentarem potencial reduzido. Por potencial reduzido, cita-se o cruzamento de indivíduos aparentados, baseando-se no desempenho *per se*, e, mesmo na análise multivariada, com informações de diversidade genética, o cruzamento de indivíduos relativamente similares (como os indivíduos 21 e 198).

Por outro lado, a seleção de genitores na GWS é baseada na predição do desempenho do híbrido. Desta forma, os indivíduos selecionados para cruzamento são aqueles cujo híbrido, em cruzamento específico, apresenta elevado valor genético genômico predito. Neste caso, uma linhagem pode ser selecionada para cruzamento apenas uma única vez, caso o seu potencial como genitor seja favorável em somente um cruzamento.

Um segundo fator que justifica a performance superior da GWS reside no uso de informações moleculares associadas à valores fenotípicos. Estes dados permitem explorar a complementariedade alélica entre genitores, pela predição do desempenho híbrido a partir de informações loco a loco. A complementariedade gênica é o que melhoristas buscam ao realizar análises de diversidade genética. Entretanto, a diversidade genética é estimada com base em dados fenotípicos de várias características, não refletindo necessariamente a diversidade para a característica de maior interesse, ou a partir de dados genotípicos isolados, que não apresentam associação com o fenótipo.

Segundo Ramalho & Lambert (2004), sendo disponíveis marcadores moleculares específicos para a característica sob seleção, a medida de divergência genética poderá ser útil se associada com o desempenho médio dos pais. Assim, poder-se-á selecionar genitores com elevada média e complementares para a característica de interesse. Portanto, a GWS na seleção de genitores não deve ser entendida apenas como uma análise de diversidade genética, mas uma ferramenta que permite explorar a complementariedade alélica para o caráter sob seleção.

Muito embora os resultados deste estudo são promissores e superiores às técnicas convencionais, ainda é possível aperfeiçoar a acurácia da GWS e conseqüentemente aprimorar a predição do desempenho dos híbridos, visto que a acurácia da predição de híbridos via GWS foi de 7,5% (Figura 2). Segundo Resende *et al* (2013), a acurácia da GWS depende da proporção (r_{mq}^2) da variação genética explicada pelos marcadores e da acurácia da predição dos efeitos dos marcadores em desequilíbrio de ligação com os

QTLs (r_{mm}). O parâmetro r_{mq}^2 depende da densidade de marcadores e da extensão e padrão do desequilíbrio de ligação existente na população. Em resumo, depende do número e espaçamento dos marcadores e do tamanho do genoma. Esses mesmos autores reportaram que, em eucalipto, espécie com genoma de 1300 cM e com tamanho efetivo igual a 20, 1300 marcadores equidistantes proporcionaram uma acurácia de seleção de 0,56 e, utilizando-se 2600 marcadores equidistantes, o valor esperado seria de 0,71.

Todavia, VanRanden *et al* (2011) reportaram que uma maior densidade de marcadores não necessariamente aumenta a acurácia da GWS. Xavier *et al* (2016) avaliaram as propriedades preditivas da GWS na cultura da soja e testaram subconjuntos dos dados genotípicos, utilizando o painel completo, meio painel e um quarto de painel de marcadores, correspondendo a 4077, 2039 e 1020 SPNs, respectivamente. Os autores relataram que 1020 marcadores foram suficientes para prover uma predição consistente, enquanto que o aumento da densidade de marcadores promoveu apenas ganhos marginais na capacidade de predição. Possivelmente, este resultado está associado às propriedades genômicas da soja, como o grande desequilíbrio de ligação existente (HYTEN *et al.*, 2007) e a distribuição desproporcional dos marcadores SNPs (LI *et al.*, 2014).

A importância do emprego de marcadores SNPs densamente dispersos no genoma aumenta ao passo que a estrutura populacional é desconhecida, o número de gerações de seleção acresce e o desequilíbrio de ligação entre marcadores e QTL decai (BASTIAANSEN *et al.*, 2012; DAETWYLER *et al.*, 2013)

Valente *et al* (2016) destacaram que o emprego de maiores densidades de SNPs é necessário em populações com baixo desequilíbrio de ligação e com maior variância genética. Além disso, destacaram que a seleção de um caráter com elevada herdabilidade requer uma alta densidade de marcadores no genoma para que a acurácia da GWS fosse superior à seleção fenotípica.

Já o parâmetro r_{mm} depende da quantidade e precisão dos dados disponíveis para estimar os efeitos dos marcadores, além da eficiência da estratégia e dos métodos estatísticos usados na predição, ou seja, depende do número de indivíduos na população de treinamento e do tamanho efetivo populacional.

Xavier *et al* (2016) avaliaram o impacto do tamanho populacional na acurácia da GWS para algumas características de importância agrônômica na soja. Foi observado que o aumento da população de treinamento de 250 para 4000 indivíduos aumentou a acurácia em 94,8% (de 0,384 para 0,747). Entretanto, os aumentos em acurácia decaíram rapidamente a partir de 2000 indivíduos. Os autores afirmaram que conjuntos de 1000-2000 indivíduos podem representar as populações de treinamento com melhor custo-

benefício, visto que os ganhos se tornam marginais a partir de 2000 indivíduos.

Segundo Grattapaglia & Resende (2011), o tamanho efetivo populacional apresenta maior influência somente se a densidade de marcadores empregada for insuficiente para amostrar o genoma.

Jannink *et al* (2010) relataram que um bom conjunto de treinamento deve ser representativo do germoplasma sob avaliação para poder capturar a estrutura populacional e ter um tamanho suficiente para uma acurada estimação dos efeitos dos alelos. Todavia, uma população de treinamento suficientemente grande deve ser empregada quando objetivo final é a seleção de materiais não fenotipados (HEFFNER *et al.*, 2009)

Comumente, para se obter satisfatória acurácia na estimação dos valores genético genômicos em populações de autopolinização são necessárias menores populações de treinamento, em consequência da menor diversidade genética. Todavia, em populações com maior diversidade genética, especialmente aquelas de polinização cruzada, o tamanho da população de treinamento requerido é maior para que se obtenha resultado satisfatório (MUJIBI *et al*, 2011).

Além do exposto, a acurácia da GWS ainda depende de mais dois fatores que não estão sob controle do melhorista. O primeiro é a herdabilidade do caráter e o segundo é o número de loci controladores do caráter e a distribuição dos seus efeitos. Esta diferença é claramente notada ao comparar as magnitudes das acurácias dos valores genético genômicos preditos para produtividade e altura de plantas (Figura 2). A característica altura de plantas foi simulada considerando-se 73 loci controlando a expressão do caráter e herdabilidade alta, de 0,7, resultando em maior proporção de bons híbridos preditos via GWS (74%) em comparação aos obtidos para a característica produtividade de grãos.

Analisando a acurácia da predição da seleção genômica ampla, Valente *et al* (2016) reportaram que, quando os efeitos dos marcadores forem estimados e utilizados na mesma população, o aumento da acurácia da predição foi proporcional ao aumento da herdabilidade, embora nem sempre associado a grandes incrementos.

Em características complexas, como a produtividade, cuja variação genotípica explica uma reduzida fração da variação fenotípica, ou seja, possui baixa herdabilidade, a acurácia da GWS é, impreterivelmente, reduzida. Isto ocorre, pois, a GWS utiliza informação genotípica para predição do desempenho. Desta forma, quando a própria variação genotípica não explica grande parte da variação, a GWS fica à mercê da influência ambiental, a qual é imprevisível. Goddard & Hayes (2007) afirmaram que com

a redução da herdabilidade, uma menor proporção da variação aditiva é explicada pelos marcadores genéticos e, portanto, menores acurácias são observados.

Todavia, em estudo apresentado por Valente *et al* (2016), os autores reportaram que a eficiência da seleção genômica, comparada com a seleção fenotípica, foi inversamente proporcional a herdabilidade. Assim, quanto menor a herdabilidade, maior a superioridade da GWS frente a seleção fenotípica, visto que a herdabilidade proporciona menor correlação com o fenótipo observado em relação ao valor genético-genômico estimado.

Isto é decorrente do emprego de um elevado número de indivíduos, que aumenta o conjunto amostral do qual os marcadores terão seus efeitos estimados. Com mais repetições experimentais por locus, os efeitos dos marcadores serão estimados minimizando a influência ambiental, que, devido a elevada repetitividade, proporciona um aumento da precisão mesmo para características com baixa herdabilidade (RESENDE *et al.*, 2013). Assim, mesmo com a herdabilidade de cada marcador efetivo seja muito baixa, com várias repetições essa herdabilidade é aumentada (RESENDE *et al.*, 2013).

A obtenção de valores genético genômicos robustos, especialmente para características complexas e de baixa herdabilidade, dependem de uma acurada coletada dos dados fenotípicos e bom controle ambiental pelo emprego de repetições, testemunhas, informação de parcelas vizinhas, divisão da área experimental em blocos e um design de campo bem planejado (HEFFNER *et al.*, 2009; ENDELMAN *et al.*, 2014). Assim, reduzindo os efeitos ambientais, maior parte da variação fenotípica recai sobre a variação genotípica.

Desta forma, mesmo enfrentando a barreira da herdabilidade, resultados indicam que até as características com baixa herdabilidade podem prover acurácia razoável (XAVIER *et al.*, 2016). Muir (2007) apontou que características com baixa herdabilidade tem um grande potencial a serem exploradas pela GWS para incrementar a acurácia de seleção e, por conseguinte, os ganhos genéticos.

Analisando a capacidade preditiva da GWS para o acamamento, característica controlada por 32 locus e herdabilidade de 50%, o desempenho fenotípico dos melhores híbridos preditos foi muito similar à altura de plantas (Figura 2). Isto demonstra que o modelo de predição da GWS tem sua acurácia aumentada quando o número de genes é pequeno e reduzido quando a herdabilidade é menor. Neste sentido, o valor da acurácia proporcionou a predição dos valores genético genômicos indicando a seleção de 72 % de bons híbridos para serem explorados no programa de melhoramento.

A acurácia na predição do desempenho fenotípico da característica ramificações

laterais por plantas também apresentou resultados promissores (Figura 2), porém abaixo do esperado. Por se tratar de uma característica com elevada herdabilidade (0,8) e com reduzido número de loci controladores (21), esperava-se um incremento na predição dos melhores híbridos em comparação com a característica altura de planta. Neste caso, um reduzido número de genes explicaria grande parte da variação fenotípica expressada e a predição dos valores genético genômicos via GWS seria mais eficiente.

Uma das grandes vantagens da utilização da predição do desempenho via GWS para características controladas por poucos loci e com elevada herdabilidade é o pequeno percentual de híbridos com performance ruim e péssima. Em média, para altura de planta, acamamento e ramificações laterais, a ocorrência média destes piores híbridos foi menor que 10%. Contudo, independentemente da característica avaliada, a predição de bons híbridos variou de 54 a 75%, evidenciando a eficiência do uso da GWS nesta predição.

5. CONSIDERAÇÕES FINAIS

O uso da seleção genômica ampla na seleção de genitores para serem utilizados em cruzamentos dialélicos apresentou híbridos com maior potencial para serem explorados em programas de melhoramento quando comparados aos obtidos via sistemas dialélicos estabelecidos por análise univariada com base em informação fenotípica da característica de interesse e por análise multivariada com base em informação fenotípica.

A inclusão dos valores genético-genômicos dos genitores na determinação dos cruzamentos a serem realizados aumenta a probabilidade de gerar híbridos fenotipicamente superiores, especialmente para características com baixa herdabilidade, em que a acurácia de seleção é reduzida.

O uso da GWS como ferramenta na seleção de genitores permite explorar a complementariedade alélica do caráter sob seleção, utilizando proveitosamente informações moleculares e fenotípicas associadas.

A acurácia na predição do desempenho dos híbridos reduziu conforme a herdabilidade das características sob seleção era reduzida.

O número de locos controladores das características não influenciou notavelmente na acurácia de predição de híbridos via GWS.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ADITYA, J. P.; BHARTIYA, P.; BHARTIYA, A. Genetic variability, heritability and character association for yield and component characters in soybean (*G. max* (L.) Merrill). **Journal of Central European Agriculture**, v. 12, n. 1, p. 27-34, 2011.
- ALMEIDA LOPES, Â.C.; VELLO, N.A.; PANDINI, F.; DE MOURA, M. Variabilidade e correlações entre caracteres em cruzamentos de soja. **Scientia Agrícola**, v. 59, n. 2, p. 341-348, 2002.
- ALMEIDA, R.D.; PELUZIO, J.M.; AFFÉRI, F.S. Divergência genética entre cultivares de soja, sob condições de várzea irrigada, no sul do Estado Tocantins. **Revista Ciência Agronômica**, v. 42, n. 1, p. 108-115, 2011.
- ASSUNÇÃO FILHO, J. R. **Potencial de cruzamentos dialélicos de soja para precocidade, produtividade de grãos e tolerância à ferrugem**. 2016. 92 p. Tese Doutorado - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba.
- BARASKAR, V. V.; KACHHADIA, V. H.; VACHHANI, J. H.; BARAD, H. R.; PATEL, M. B.; DARWANKAR, M. S. Genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. **Electronic Journal of Plant Breeding**, v. 5, n. 4, p. 802-806, 2014.
- BARROSO, P.A.V.; HOFFMANN, L.V. **Métodos de predição do comportamento de populações de melhoramento**. Campina Grande: Embrapa Algodão, 38p. (Embrapa Algodão. Documentos, 108). 2003.
- BASTIAANSEN, J. W.; COSTER, A.; CALUS, M. P.; VAN ARENDONK, J. A.; BOVENHUIS, H. Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. **Genet. Sel. Evol**, v.44, n.3, 2012.
- BEKELE, A.; ALEMAHU, G. Desirable Traits Influencing Grain Yield in Soybean (*Glycine max* (L.) Merrill). **Innovative Systems Design and Engineering**, v.2, n.3, 2011.
- BERNINI, C. S., PATERNIANI, M. E. A. G. Z., DUARTE, A. P., GALLO, P. B., DE SOUZA GUIMARÃES, P., ROVARIS, S. R. S. Depressão endogâmica e heterose de híbridos de populações F₂ de milho no estado de São Paulo. **Bragantia**, v. 72, n. 3, p. 217-223, 2013.
- BERTAN, I.; CARVALHO, F.I.F.; OLIVEIRA, A.C. Parental selection strategies in plant breeding programs. **Journal of Crop Science and Biotechnology**, v.10, p.211-222, 2007.
- BORÉM, A.; MIRANDA, G. V. **Melhoramento de Plantas**. 523p. UFV, 2013.
- BREIMAN, L. Better subset regression using the nonnegative garrote. **Technometrics**, v. 37, n. 4, p. 373-384, 1995.
- BRIM, C.A.; COCKERHAM, C.C. Inheritance of quantitative characters in soybeans. **Crop Science**, Madson, v.1, n.3, p. 187-190, 1961.

BRUULSEMA, T.; HEFFER, P.; WELCH, R.; CAKMAK, I.; MORAN, K. Fertilizing crops to improve human health: a scientific review. **International Plant Nutrition Institute/International Fertilization Industry Association**, Norcross, GA. 8 p. 2012.

CAIXETA, E. T.; OLIVEIRA, A. C. B.; BRITO, G. G.; SAKIYAMA, N. S. **Tipos de marcadores moleculares**. In: BORÉM, A.; CAIXETA, E. T. (Ed.). Marcadores moleculares. 2. ed. Viçosa: UFV, p. 11-93. 2009.

CANTELLI, D. A. V.; HAMAWAKI, O. T.; ROCHA, M. R.; NOGUEIRA, A. P. O.; HAMAWAKI, R. L.; SOUSA, L. B.; HAMAWAKI, C. D. L. Analysis of the genetic divergence of soybean lines through hierarchical and Tocher optimization methods. **Genetics and molecular research: GMR**, v.15, n.4, 2016.

CARPENTIERI-PÍPOLO, V.; PÍPOLO, A. E.; SILVA, F. A. M. D.; PETEK, M. R. Soybean parent selection based on genetic diversity. **Brazilian Archives of Biology and Technology**, v. 43, n. 3, p. 295-300, 2000.

CARVALHO, L. P.; LANZA, M. A.; FALLIERI, J.; DOS SANTOS, J. W. Análise da diversidade genética entre acessos de banco ativo de germoplasma de algodão. **Pesq. agropec. bras.**, Brasília, v. 38, n. 10, p. 1149-1155, 2003.

CHO, Y.; SCOTT, R.A. Combining ability of seed vigor and seed yield in soybean. **Euphytica**, v. 112, n. 2, p. 145-150, 2000.

COSTA, M. M.; UNÊDA-TREVISOLI, S. H.; PINHEIRO, J. B.; DE SOUZA KIIHL, R. A.; CALVO, É. S.; DI MAURO, A. O. Marcadores RAPD para detecção de resistência à ferrugem-asiática-da-soja. **Pesquisa agropecuária brasileira**, v. 43, n. 12, p. 1733-1739, 2008.

CROSBIE, T. M.; EATHINGTON, S. R.; JOHNSON, G. R.; EDWARDS, M.; REITER, R.; STARK, S.; DOBERT, R. Plant breeding: past, present, and future. In: **Plant breeding: the Arnel R**.

CRUZ, C. D.; VENCOVSKY, R. Comparação de alguns métodos de análise dialélica. **Revista Brasileira de Genética**, v. 12, n. 2, p. 425-438, 1989.

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

CRUZ, C.D.; FERREIRA, F.M.; PESSONI, L.A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Suprema, 2011. 620 p.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético** - ISBN 9788572694339. 4. ed. Viçosa: Editora UFV, 2012. v. 2. 514p.

DAETWYLER, H. D.; CALUS, M. P.; PONG-WONG, R.; DE LOS CAMPOS, G.; HICKEY, J. M. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, v.193, n.2, p.347-365, 2013.

DEKKERS, J.C.M. Commercial application of marker-and gene-assisted selection in livestock: strategies and lessons. **Journal of animal science**, v. 82, n. 13_suppl, p. E313-

E328, 2004.

DESTRO, D. **Capacidade de combinação de genótipos de soja (*Glycine max* (L.) Merrill) apropriados para o consumo humano**. 158 p. Tese Doutorado - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba. 1991.

DIAS, L.A.S. Análises Multidimensionais. In: ALFENAS, A.C. (Ed.). **Eletroforese de isoenzimas e proteínas afins**. Viçosa-MG: Ed. UFV, 1998. P. 405-475.

DILNESAW, Z.; ABADI, S.; GETAHUN, A. Genetic variability and heritability of soybean (*Glycine max* (L.) Merrill) genotypes in Pawe district, Metekel zone, Benishangule Gumuz regional state, northwestern Ethiopia. **Wudpecker Journal of Agricultural Research**, v. 2, n.9, p.240 – 245, 2013.

DUDLEY, J. W. Theory for transfer of alleles. **Crop science**, v.22, n.3, p. 631-637, 1982.
ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **The Plant Genome**, v.4, n.3, p.250-255, 2011.

FALCONER, D.S. **Introduction to quantitative genetics**. 2. ed. London: Longman, 340 p., 1981.

FAO - **Trade reforms and food security: conceptualizing the linkages**. Food and Agriculture Organization of the United Nations, Rome. 2003.

FERNANDO, R.L.; GROSSMAN, M. Marker assisted selection using best linear unbiased prediction. **Genet. Sel. Evol.** v.21, p. 467-477, 1989.

FREIRE FILHO, F.R. **Análise genética de um dialelo entre genótipos precoces de soja (*Glycine max* (L.) Merrill)**. Piracicaba. 224p. Tese Doutorado - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba. 1988.

FREIRE FILHO, F.R.; VELLO, N.A. Combining ability, heterosis, heterobeltiosis in early soybean genotypes. In: **IV World Soybean Research Conference**. Proceedings Buenos Aires: WSBC, v.2, p.953-957, 1989.

FREITA-JÚNIOR, A.J.; PEREIRA, M. G.; CRUZ, C.D.; SCAPIM, C. A.Capacidade combinatória de milho pipoca por meio de dialelo circulante. **Pesq. Agropecu. Bras.**, Brasília, v. 41, n. 11, p. 1599-1607, 2006.

FUGANTI, R.; BENEVENTI, M. A.; SILVA, J. F. V.; ARIAS, C. A. A.; MARIN, S. R.; BINNECK, E.; NEPOMUCENO, A. L. Identificação de marcadores moleculares de microssatélites para seleção de genótipos de soja resistentes a *Meloidogyne javanica*. **Nematologia Brasileira**, v. 28, n. 2, p. 125-130, 2004.

FUMES, L.A.A. **Resposta à seleção em duas populações de soja com diferentes proporções de cada genitor**. 48 p. Tese de Doutorado. Escola Superior de Agricultura “Luiz de Queiroz. 2013.

GARDNER, C. O.; EBERHART, S. A. Analysis and interpretation of the variety cross diallel and related populations. **Biometrics**, v.22, n.3, p. 439-452, 1966.

GHODRATI, G. Study of genetic variation and broad sense heritability for some

- qualitative and quantitative traits in soybean (*Glycine max* L.) genotypes. **Current Opinion in Agriculture**, v. 2, n.1, 2013.
- GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics**, v.124, p.323-330, 2007.
- GRANT, D.; NELSON, R.T.; CANNON, S.B.; SHOEMAKER, R.C. **SoyBase, the USDA-ARS soybean genetics and genomics database**. 2010.
- GRATTAPAGLIA, D. Marcadores moleculares em espécies florestais: Eucalyptus como modelo. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-ILGLIS, M.C. **Recursos genéticos e melhoramento: plantas**. Rondonópolis: Fundação MT, p. 967-1010. 2001.
- GRATTAPAGLIA, D.; RESENDE, M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v.7, p.241-255, 2011.
- GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian journal of biological sciences**, v. 9, n. 4, p. 463-493, 1956.
- GUIMARÃES, E. P. Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish. **Food & Agriculture Org.**, 2007.
- HAKIM, L.; SUYAMTO, S.; PATUROHMAN, E. Genetic variability, heritability and expected genetic advances of quantitative characters in F₂ progenies of soybean crosses. **Indonesian Journal of Agricultural Science**, v.15, n.1, p.11-16, 2014.
- HALLAUER, A. R. Methods used in developing maize inbreds. **Maydica**, v.35, n.1, p.1-16, 1990.
- HARTMAN, G.L.; WEST, E.D.; HERMAN, T.K. Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. **Food Security**, v.3, n.1, p.5-17, 2011.
- HASHASH, E.F.E. Genetic Diversity of Soybean Yield Based on Cluster and Principal Component Analyses. **Journal of Advances in Biology & Biotechnology**, v.10, n.3, 2016.
- HAYMAN, B. I. The theory and analysis of diallel crosses. **Genetics**, v.39, n.6, p.789, 1954.
- HEFFNER, E.L.; SORRELLS, M.E.; JANNINK, J. Genomic selection for crop improvement. **Crop Science**, v.49, n.1, p.1-12, 2009.
- HYTEN, D. L.; SONG, Q.; ZHU, Y.; CHOI, I. Y.; NELSON, R. L. Impacts of genetic bottlenecks on soybean genome diversity. **Proc. Natl. Acad. Sci. USA**, v.103, n.45, p.16666–16671. 2006.
- ISLEIB, T. G. Recovery of superior homozygous progeny from biparental crosses and backcrosses. **Crop science**, v.39, n.2, p.558-563, 1999.
- JANNINK, J. L.; LORENZ, A. J.; IWATA, H. Genomic selection in plant breeding: from

theory to practice. *Brief. Funct. Genomics*, v.9, n.2, p.166–177, 2010.

JENKINS, M. T. The effect of inbreeding and of selection within inbred lines of maize upon the hybrids made after successive generations of selfing. *Iowa State J. Sci*, v.3, p.429-450, 1935.

KAMIKOGA, M.K. **Herança do acamamento e associações com outros caracteres em soja (*Glycine max* (L.) Merrill)**. 130p. Dissertação (Mestrado em genética e melhoramento de plantas) – Escola Superior de Agronomia “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1989.

KEIM, P.; SHOEMAKER, R. C.; PALMER, R. G. Restriction fragment length polymorphism diversity in soybean. *Theoretical and applied genetics*, v.77, n.6, p.786-792, 1989.

KHAN, M. S. A.; KARIM, M. A.; HAQUE, M. M.; KARIM, A. J. M. S.; MIAN, M. A. K. Variations in agronomic traits of Soybean genotypes. *SAARC Journal of Agriculture*, v.12, n.2, p.90-100, 2015.

KWON, S. H.; TORRIE, J. H. Heritability and interrelationship among traits of two soybean populations. *Crop sci*, v.4, n.2, p.196-198, 1964.

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

LI, Y. H.; LIU, Y. L.; REIF, J. C.; LIU, Z. X.; LIU, B. Biparental resequencing coupled with SNP genotyping of a segregating population offers insights into the landscape of recombination and fixed genomic regions in elite soybean. *G3: Genes|Genomes|Genetics*, v.4, n.4, p.553–560, 2014.

MAPHOSA, M.; TALWANA, H.; GIBSON, P.; TUKAMUHABWA, P. Combining ability for resistance to soybean rust in F 2 and F 3 soybean populations. *Field Crops Research*, v.130, p.1-7, 2012.

MARTÍNEZ, B.A.; SANINT, R.; CABRERA, F.A.V.. Análisis de la heterosis y de la habilidad combinatoria entre diferentes cultivares de tomate, *Lycopersicon esculentum* Mill. a partir de un cruzamiento dialélico. *Acta Agronômica*, v.39, n.1-2, p.24-33, 1989.

MAUGHAN, P. J.; MAROOF, M. S.; BUSS, G. R.; HUESTIS, G. M. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. *Theoretical and Applied Genetics*, v.93, n.3, p.392-401, 1996.

MELO, PCT. **Heterose e capacidade combinatória em um cruzamento dialélico parcial entre seis cultivares de tomate (*Lycopersicon esculentum* Mill.)**. 127p. Tese de Doutorado. Escola Superior de Agricultura “Luiz de Queiroz. 1987.

MEUWISSEN, T.H.E.; HAYES, B.J. ; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, v.157, n.4, p.1819–1829, 2001. MILLIGAN, G.W. A Monte Carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika*, v.46, n.2, p.187-199, 1981.

- MILLIGAN, G.W.; COOPER, M.C. An examination of procedures for determining the number of clusters in a data set. **Psychometrika**, v.50, n.2, p.159-179, 1985.
- MIRANDA, J. E. C.; COSTA, C. P.; CRUZ, C. D. Análise dialéctica em pimentão. I. Capacidade combinatória. **Revista Brasileira de genética**, Ribeirão Preto, v.11, n.2, p.431-440, 1996.
- MUIR, W. M. Comparison of genomic and traditional BLUPestimated breeding value accuracy and selection response under alternative trait and genomic parameters. **J. Anim. Breed. Genet**, v.124, n.6, 2007.
- MUJIBI, F.D.N.; NKUMAH, J.D.; DURUNNA, O.N.; STOTHARD, P.; MAH, J.; WANG, Z.; BASARAB, J.; PLASTOW, G.; CREWS JR., D.H.; MOORE, S.S. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. **Journal of Animal Science**, v.89, p.3353-3361, 2011.
- MULATO, B.M. **Diversidade genética em germoplasma de soja identificada por marcadores SSR, EST-SSR e caracteres agromorfológicos**. 87 p. Tese de Doutorado. Escola Superior de Agricultura "Luiz de Queiroz". 2009.
- NARJESI, V; KHANEGHAH, H. Z.; ZALI, E.E. Assesment of genetical relationship in few important agronomic characters with grain yield in soybean by multivariate statistical analysis. **Agri. Nat. Res. Sci.**, v.41, n.11, p.227-234, 2007.
- OLIVEIRA, V.R.; CASALI, V.W.D.; CRUZ, C.D.; PEREIRA, P.R.G.; BRACCINI, A.L. Avaliação da diversidade genética em pimentão através de análise multivariada. **Horticultura Brasileira, Brasília**, v. 17, n. 1, p. 19-24, março 1999.
- PIMENTEL, A. J. B.; RIBEIRO, G.; SOUZA, M. D.; MOURA, L. M.; ASSIS, J. D.; MACHADO, J. C. Comparação de métodos de seleção de genitores e populações segregantes aplicados ao melhoramento de trigo. **Bragantia**, v. 72, n. 2, p. 113-121, 2013.
- RAMALHO, M.A. P.; LAMBERT, E. S. Biometria e o melhoramento de plantas na era da genômica. **Revista Brasileira de Milho e Sorgo**, v. 3, n. 2, p. 228-249, 2004.
- RANGEL, P. H. N.; CRUZ, C. D.; VENCOSKY, R.; FERREIRA, R. D. P. Selection of local lowland rice cultivars based on multivariate genetic divergence. **Revista Brasileira de Genética**, v.14, n.2, p.437-453, 1991.
- RESENDE, M. D. V.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, n. 56, p. 63, 2008.
- RESENDE, M. F. R.; MUNOZ, P.; ACOSTA, J. J.; PETER, G. F.; DAVIS, J. M.; GRATAPAGLIA, D.; KIRST, M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. **New Phytologist**, v. 193, n. 3, p. 617-624, 2012a.
- RESENDE, M. D.; RESENDE, M. F.; SANSALONI, C. P.; PETROLI, C. D.; MISSIAGGIA, A. A.; AGUIAR, A. M.; PAPPAS, G. J. Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, 194(1), 116-128. 2012b.

RESENDE, M.D.V.; SILVA, F.F.; RESENDE JÚNIOR, M.F.R. **Seleção Genômica Ampla (GWS)**. In: Biotecnologia Aplicada ao Melhoramento de Plantas. BORÉM, A.; FRITSCH NETO, R. Visconde do Rio Branco, MG: Suprema, 2013. Cap. 5, p.151-188. SCHMIDT, M. R.; WEI, W. Loss of Agro-Biodiversity, Uncertainty, and Perceived Control: A Comparative Risk Perception Study in Austria and China. **Risk analysis**, v. 26, n. 2, p. 455-470, 2006.

SÁNCHEZ, C.F.B. **Seleção genômica ampla em populações derivadas de acasalamento ao acaso ou de autofecundação**. 76 p. Tese de doutorado, Universidade Federal de Viçosa. 2013.

SILVA, J. F. V.; FERRAZ, L. C. B. C.; ARIAS, C. A.; ABDELNOOR, R. V. Identificação de marcadores moleculares de microssatélite associados à resistência de genótipos de soja a *Meloidogyne javanica*. **Nematologia brasileira**, v. 25, n. 1, p. 79-83, 2001.

SINGH, G. **The soybean: botany, production and uses**. CABI Publishing, Wallingford. 2010.

SINGH, P.; KUMAR, R.; SABAPATHY, S.N.; BAWA, A.S. Functional and edible uses of soy protein products. **Comprehensive reviews in food science and food safety**, v. 7, n. 1, p. 14-28, 2008.

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

TEODORO, P. E.; RIGON, J. P. G.; TORRES, F. E.; RIBEIRO, L. P.; CORREA, C. C. G.; SILVA, F. A.; SOUZA, E. C. (a) Comparison of clustering methods for study of genetic dissimilarity in soybean genotypes. **African Journal of Agricultural Research**, v. 10, n. 11, p. 1331-1337, 2015.

TORRES, F. E.; DAVID, G. V.; TEODORO, P. E.; RIBEIRO, L. P.; CORREA, C. G.; JÚNIOR, R. A. L. (b) Desempenho agronômico e dissimilaridade genética entre genótipos de soja. **Revista de Ciências Agrárias**, v. 38, n. 1, p. 111-117, 2015.

USDA. United States Department of Agriculture. **World Agriculture Supply and Demand Estimates**. 2017. Disponível em: 31/01/2017 <<https://www.usda.gov/oce/commodity/wasde/latest.pdf>> . Acesso em: 25 Mar. 2017.

VAL, B. H. P.; JÚNIOR, J. A. F.; BIZARI, E. H.; DI MAURO, A. O.; UNÊDA-TREVISOLI, S. H. Diversidade genética de genótipos de soja por meio de caracteres agromorfológicos. **Ciência & Tecnologia**, v. 6, n. 1, 2014.

VALENTE, B. D.; MOROTA, G.; PEÑAGARICANO, F.; GIANOLA, D.; WEIGEL, K.; et al. The causal meaning of genomic predictors and how it affects construction and comparison of genome-enabled selection models. **Genetics**, v.200, n.2, p.483–494, 2015.

VANRADEN, P. M.; O'CONNELL, J. R.; WIGGANS, G. R.; WEIGEL, K. A. Genomic evaluations with many more genotypes. **Genet. Sel. Evol.**, v.43, n.1, 10, 2011.

VIANNA, V. F.; DESIDERIOSUP, J. A.; DE SANTIAGOSUP, S.; JUNIORSUP, J. A. F.; FERRAUDOSUP, A. S. The multivariate approach and influence of characters in

selecting superior soybean genotypes. **African Journal of Agricultural Research**, v. 8, n. 30, p. 4162-4169, 2013.

VIEIRA, P.F.M.J. **Dialelo entre genitores de soja tolerantes à ferrugem asiática**. 86 p. Tese de Doutorado. Escola Superior de Agricultura “Luiz de Queiroz”. 2009.

VILLELA, O. T.; UNÊDA-TREVISOLI, S. H.; DA SILVA, F. M.; JUNIOR, L. S. B. A.; DI MAURO, A. O. Genetic divergence of roundup ready (RR) soybean cultivars estimated by phenotypic characteristics and molecular markers. **African Journal of Biotechnology**, v. 13, n. 26, 2014.

VILLELA, O.T. **Diversidade fenotípica e molecular de cultivares brasileiras de soja portadoras de gene RR**. 67p. Dissertação de mestrado. Faculdade de Ciências Agrárias e Veterinárias – UNESP, Campus de Jaboticabal, Jaboticabal. 2013.

WHITTAKER, J.C.; THOMPSON, R.; DENHAM, M.C. Marker-assisted selection using ridge regression. **Genetical research**, v. 75, n. 02, p. 249-252, 2000.

WIGGINS, B. T. **Heritability and genetic gain of seed protein, oil, and yield among ril of soybean**. Master of science degree. University of Tennessee, 83 p., 2012.

XAVIER, A.; MUIR, W.M.; RAINEY, K.M. Impact of imputation methods on the amount of genetic variation captured by a single-nucleotide polymorphism panel in soybeans. **BMC bioinformatics**, v. 17, n. 1, p. 55, 2016.

XU, Y.; CROUCH, J.H. Marker-assisted selection in plant breeding: from publications to practice. **Crop Science**, v. 48, n. 2, p. 391-407, 2008.

ZHAO, Y.; GOWDA, M.; LIU, W.; WÜRSCHUM, T.; MAURER, H. P.; LONGIN, F. H.; REIF, J. C. Accuracy of genomic selection in European maize elite breeding populations. **Theoretical and Applied Genetics**, 124(4), 769-776. 2012.

ZHAO, Y.; ZENG, J.; FERNANDO, R.; REIF, J. C. Genomic prediction of hybrid wheat performance. **Crop Science**, 53(3), 802-810. 2013.