

FRANCYSE EDITE DE OLIVEIRA CHAGAS

**ANÁLISES GENÔMICAS E BIOMÉTRICAS PARA ESCOLHA DE
GENITORES E PREDIÇÃO DE HÍBRIDOS NÃO REALIZADOS**

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

VIÇOSA
MINAS GERAIS - BRASIL
2018

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

T

C433a
2018

Chagas, Francyse Edite de Oliveira, 1993-
Análises genômica e biométrica para escolha de genitores e
predição de híbridos não realizados / Francyse Edite de Oliveira
Chagas. – Viçosa, MG, 2018.
xiv, 49 f. : il. (algumas color.) ; 29 cm.

Orientador: Cosme Damião Cruz.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 45-49.

1. Genômica. 2. Biometria. 3. Hibridação vegetal -
Previsão. I. Universidade Federal de Viçosa. Departamento de
Biologia Geral. Programa de Pós-Graduação em Genética e
Melhoramento. II. Título.

CDD 22. ed. 576.5

FRANCYSE EDITE DE OLIVEIRA CHAGAS

**ANÁLISES GENÔMICA E BIOMÉTRICA PARA ESCOLHA DE GENITORES
E PREDIÇÃO DE HÍBRIDOS NÃO REALIZADOS**

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

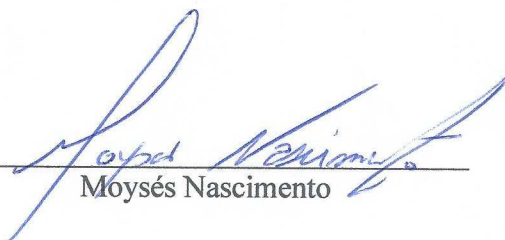
APROVADA: 16 de fevereiro de 2018.



Felipe Lopes da Silva



Renato Domiciano Silva Rosado



Moysés Nascimento



Cosme Damião Cruz
(Orientador)

AGRADECIMENTOS

Início agradecendo primeiramente a Deus, por todo amparo e por ser o Senhor da minha vida.

Aos meus pais e minha irmã por acreditarem em mim, pelo apoio e pela força.

Ao meu orientador Professor Dr. Cosme Damião Cruz por toda a ajuda e dedicação destinada durante o tempo do mestrado. Obrigada pela paciência, incentivo e pelo exemplo de profissionalismo dado pelo senhor, estas competências serão levadas comigo ao longo de minha caminhada acadêmica da qual o senhor sempre fará parte.

Aos membros da banca, Dr. Renato Domiciano Silva Rosado, Prof. Dr. Felipe Lopes da Silva e Prof. Dr. Moyses Nascimento por se disporem tão prontamente a participarem deste momento tão importante em minha vida.

A toda equipe da secretaria do Programa de Pós-graduação por todo o auxílio prestado durante o mestrado.

Aos meus colegas de laboratório por toda a ajuda nas análises dos dados e críticas ao trabalho que estava sendo realizado.

A Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas pela oportunidade concedida.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – pela bolsa concedida.

A todos que contribuíram direta e indiretamente para construção deste trabalho: muito obrigada!

DEDICATÓRIA

A Deus, meus pais e minha irmã Francielle, minha família, dedico.

BIOGRAFIA

Francyse Edite de Oliveira Chagas, filha de Adalberto Custódio das Chagas e Maria Elesbão Neves de Oliveira Chagas, nasceu em 09 de dezembro de 1993, na cidade de Viçosa, Minas Gerais.

Em janeiro de 2016, graduou-se em Licenciatura em Ciências Biológicas pela Universidade Federal de Viçosa, Minas Gerais, Campus Viçosa.

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

SUMÁRIO

LISTA DE FIGURAS	vi
LISTA DE TABELAS	viii
RESUMO	x
ABSTRACT	xii
1. Introdução	1
2. Referencial teórico	5
Escolha de genitores a partir de informações fenotípicas	8
Escolha de genitores agregando informações sobre a diversidade genética	9
Escolha de genitores agregando informações sobre a progênie em cruzamentos dialélicos	10
Abordagem de escolha de genitores e híbridos envolvendo informações moleculares	10
3. Material e Métodos	15
3.1 Materiais	15
Obtenção de informações fenotípicas e genotípicas de genitores	15
Obtenção de informações fenotípicas e genotípicas de híbridos	17
Valores reais contínuos e discretos dos híbridos e genitores	17
3.2 Estratégias de escolhas de genitores para fins de hibridação	18
3.2.1 Abordagem tradicional (apenas informação fenotípica)	18
Análise do desempenho per se acompanhado de análise de diversidade genética	19
Desempenho per se acompanhado de análise dialélica	22
Desempenho per se acompanhado de análise de diversidade genética e de análise dialélica	22
3.2.2 Abordagem com informação molecular	23
3.3 Comparação dos resultados de escolha de genitores	24
3.4 Capacidade de predição do modelo frente as outras características	24

3.5 Predição dos híbridos não realizados	25
4. Resultados e Discussão.....	26
4.1 Estratégias de escolhas de genitores para fins de hibridação	26
4.1.1 Abordagem tradicional (apenas informação fenotípica)	26
Escolha de genitores com base em informação de ensaios de competição acompanhado de análise de diversidade genética	26
Ensaio de competição acompanhado de análise dialélica	30
Ensaio de competição acompanhado de análise de diversidade genética e de análise dialélica	32
4.1.2. Abordagem com informação molecular	37
4.2 Capacidade de predição do modelo aplicado às outras características.	40
4.3 Predição dos híbridos não realizados	41
5. Conclusões	44
6. Referências Bibliográficas	45

LISTA DE FIGURAS

- Figura 1:** Dispersão gráfica dos 20 melhores genitores segundo o ensaio de competição. Os genótipos mostrados estão codificados para números entre 1 a 20 sendo correspondentes, respectivamente, a 171, 167, 65, 187, 150, 175, 196, 172, 138, 133, 37, 149, 137, 96, 8, 152, 83, 22, 67 e 121.....29
- Figura 2:** Distribuição nas cinco classes de desempenho das linhagens escolhidas pela metodologia ensaio de competição somado a análise de divergência genética e dos híbridos formados pelo cruzamento entre essas linhagens.....30
- Figura 3:** Distribuição nas cinco classes de desempenho das linhagens escolhidas pela metodologia ensaio de competição somado a análise dialélica e dos híbridos formados pelo cruzamento entre essas linhagens.....32
- Figura 4:** Dispersão gráfica dos 40 melhores genitores segundo o ensaio de competição. Os genótipos mostrados estão codificados para números entre 1 a 40, sendo correspondentes, respectivamente, a 2, 8, 11, 21, 22, 34, 37, 44, 55, 60, 61, 65, 67, 70, 83, 89, 92, 121, 131, 133, 137, 138, 139, 149, 150, 152, 153, 156, 167, 171, 172, 173, 175, 177, 179, 185, 187, 192 e 196.....34
- Figura 5:** Distribuição nas cinco classes de desempenho das linhagens escolhidas pela metodologia ensaio de competição somado a análise de diversidade genética e análise dialélica e dos híbridos formados pelo cruzamento entre essas linhagens.....37
- Figura 6:** Distribuição nas cinco classes de desempenho das linhagens escolhidas por meio do valor genotípico predito fornecido por meio da seleção

genômica ampla e dos híbridos formados pelo cruzamento entre essas linhagens.....39

Figura 7 Dispersão dos valores genéticos reais e os valores genômicos preditos mensurados nos genitores, com estimativa de correlação de 0,917.....40

Figura 8. Distribuição nas cinco classes de desempenho dos híbridos apontados como melhores quando usado o valor genotípico predito fornecido pelo segundo modelo de predição criado.....42

Figura 9: Dispersão dos valores genéticos reais e os valores genômicos preditos mensurados nos híbridos, com estimativa de correlação de 0,7591.....43

LISTA DE TABELAS

- Tabela 1** - Valores estabelecidos, para fins de simulação, tomando como referência algumas características mensuradas na espécie modelo *Zea mays*.....16
- Tabela 2** - Correlação fenotípica (acima da diagonal) e genotípica (abaixo da diagonal) entre as características. Em negrito as correlações com a característica rendimento.....27
- Tabela 3** - Agrupamento de genitores, pré-selecionados com base no rendimento, de acordo com o método de otimização de “Tocher invertido” utilizando a matriz de similaridade expressa pelo complemento da distância Euclidina média padronizada.....28
- Tabela 4** - Porcentagem de variação simples e variação acumulada por componente principal (CP) obtida da análise de componentes principais entre 20 linhagens para cinco características fenotípicas.....28
- Tabela 5** - Capacidade geral de combinação (CGC) das 20 melhores linhagens pré-selecionadas segundo o ensaio de competição.....31
- Tabela 6** - Agrupamento dos genótipos de acordo com o método de otimização de Tocher invertido utilizando a matriz de similaridade.....33
- Tabela 7** - Porcentagem de variação simples e variação acumulada por componente principal (CP) obtida da análise de componentes principais entre 40 linhagens para cinco características fenotípicas.....34

Tabela 8 - Capacidade geral de combinação das 20 linhagens mais divergentes segundo a análise dialélica.....	36
---	-----------

RESUMO

CHAGAS, Francyse Edite de Oliveira, M.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Análises genômica e biométrica para escolha de genitores e predição de híbridos não realizados.** Orientador: Cosme Damião Cruz.

O lançamento de uma nova cultivar ou híbrido é uma tarefa complexa e que envolve várias etapas, sendo a primeira delas a escolha dos genitores para montar a população de onde o novo cultivar será retirado. Este trabalho tem como objetivo realizar um estudo comparativo entre várias estratégias de escolha de genitores, para fins de hibridação, baseadas em análises fenotípicas ou em análises que agregam informações moleculares. Também avaliou-se a eficácia da predição do desempenho de híbridos não realizados por meio de abordagem de predição genômica. Para a escolha dos genitores, com base em informação fenotípica, foram consideradas as análises de ensaios de competição para a característica de interesse agregando informações multivariadas do estudo da diversidade genética ou, ainda, informações do valor genético aditivo predito por análise dialélica em cruzamento de genitores pré-selecionados. Para a escolha de genitores, com base em informações moleculares adicionais, foi empregado modelo de predição genômica a partir de dados fenotípicos e genotípicos de todos os genitores. No estudo adicional, para a predição dos híbridos não realizados, também foi usado a abordagem de seleção genômica em que o conjunto de dados para treinamento foi constituído com as informações genotípicas e fenotípicas de todas as linhagens analisadas, acrescida da informação de 190 híbridos gerados para fins do estudo da análise dialélica anterior. O modelo de predição foi utilizado numa população de validação constituída por informações genotípicas de todos os híbridos não realizados. Os dados foram obtidos através de simulação utilizando o programa computacional GENES. Das linhagens escolhidas pelos métodos convencionais (com base em informação fenotípica) de 40 a 70% coincidiu na classe ótima e quando cruzadas formaram híbridos em que de 37,8 a 53,3% também estavam nessa classe. Pela análise considerando as informações genotípicas, todas as linhagens escolhidas foram

classificadas na melhor classe e dos híbridos formados 66,7% estavam também nesse grupo. Sobre a predição dos híbridos não realizados, usando um modelo genômico, foi previsto o valor genômico desses e escolhidos os 45 melhores para uso no programa de melhoramento. Quando comparados os valores genômicos, 82,2% dos híbridos coincidiram na classe de ótimos. Por meio desse trabalho foi visto que a seleção genômica consiste em um método útil e eficaz para predição dos valores genéticos genômicos de linhagens e de híbridos, devendo ser explorada em maior magnitude.

ABSTRACT

CHAGAS, Francyse Edite de Oliveira, M.Sc., Universidade Federal de Viçosa, February, 2018. **Genomic and biometric analyzes for parental selection and prediction of hybrid of unrealized hybrids.** Advisor: Cosme Damião Cruz.

The development of a new cultivar or hybrid is a complex task and take several steps, the first being the parental selection for the breeding population where the cultivar will be made. The objective of this work has the objective of evaluate a comparative study among several strategies of parental selecting, for hybridization purposes, based only in phenotypic data or including molecular information. The efficacy of the prediction of the performance of unrealized hybrids through a genomic prediction approach was also evaluated. For parental selection, based in phenotypic information, the analyses were carried out using the information of competition tests for the phenotypic trait adding multivariate information from the study of genetic diversity or, also, information of the additive genetic value predicted by diallel analysis with the parents that were selected. For the parental selection adding molecular information, a genomic prediction model was used based on phenotypic and genotypic data of all the parents. In the additional study, for the prediction of unrealized hybrids, the genomic selection approach was also used, in which the training data set consisted of the genotypic and phenotypic information of all the inbred lines analyzed, plus the information of 190 hybrids generated for study of the previous diallel cross analysis. The prediction model was used in a validation population consisting of genotypic information of all unrealized hybrids. The data were generated from simulation using the software GENES. From the inbred lines chosen by conventional methods (based only on phenotypic information) from 40 to 70% coincided in the optimal class and from 37.8 to 53.3% the unrealized hybrids from them were also in this class. By analyzing the genotypic information, all the selected inbred lines were classified in the best class and 66.7% of the hybrids formed were also in this group. In addition, from the genomic selection prediction of unrealized hybrids, the best 45 hybrids (based in genomic breeding value estimated) were chosen in order to use in the inbreeding program. When comparing the genomic values, 82.2% of

the hybrids are also in the optimal class. Through this work it was seen that genomic selection is a useful and effective method to predict genetic values of inbred lines and hybrids, and should be explored in greater magnitude

1. Introdução

Os programas de melhoramento genéticos com o objetivo de lançamento de uma nova cultivar realizam uma tarefa complexa que envolve várias etapas. Primeiramente é feita a escolha dos genitores para montar a população de onde o novo cultivar será retirado. São feitos os cruzamentos e analisados o ganho de seleção e a relação entre os caracteres do material analisado. A última etapa está relacionada a recomendação do cultivar para o ambiente que será usado, sendo feitos estudos da interação genótipos x ambientes e de estabilidade e adaptabilidade (CRUZ , REGAZZI & CARNEIRO, 2004).

Como foi relatado, tudo começa com a escolha correta dos pais para serem trabalhados durante o programa de melhoramento. Erros em outros procedimentos são graves, mas não tanto como nesse passo inicial por causa da irreversibilidade do erro (BOREM & MIRANDA, 2013).

Muito subsídio é encontrado quando se deseja conhecer as fases intermediária e final de um programa de melhoramento, mas quanto a escolha correta dos genitores a literatura científica ainda deixa muito a desejar, como abordado por Borem e Miranda (2013). Estes autores relatam que um dos maiores dilemas dos melhoristas é como escolher corretamente os genitores para um programa de desenvolvimento de variedades, pois, embora de importância indiscutível, muito pouco foi elucidado sobre as bases científicas da seleção do germoplasma. Mencionam que a literatura científica está repleta de estudos sobre métodos de melhoramento e seleção, mas sobre a escolha de genitores não.

Com a escolha correta dos genitores aumenta-se a probabilidade de obtenção de características de interesse na população segregante por maximizar a utilização de genes desejáveis (BOREM & MIRANDA, 2013).

Os cruzamentos realizados para lançar novo cultivar podem envolver genitores mais próximos ou mais distantes geneticamente que formam, nessa ordem, populações segregantes com baixa e alta variância genética. As populações segregantes com baixa variância genética são usadas quando se deseja a maximização dos ganhos genéticos, enquanto que as com alta

variância são usadas quando se deseja ampliar a possibilidade de manifestação de indivíduos de alta performance na população (BOREM & MIRANDA, 2013)

Para escolher os pais é preciso conhecer o potencial, sendo feitos ensaios de competição, em que a média dos genitores estabelece o desempenho da progênie. São analisadas várias características, sendo uma das mais importantes a produtividade (BOREM & MIRANDA, 2013).

Uma outra maneira de escolher os pais é por meio de técnicas preditivas da diversidade genética. Toma-se por base diferenças morfológicas, fisiológicas ou moleculares quantificando-as em alguma medida de dissimilaridade que expresse o grau de diversidade genética entre os genitores (CRUZ, FERREIRA & PESSONI, 2011).

Unindo o conhecimento da diversidade genética com o ensaio de competição consegue-se escolher como pais genótipos bons e distantes geneticamente que seriam os mais interessantes.

Pode-se ainda, e isso é muito usado quando se tem número muito grande de genitores, implementar uma técnica de análise dialélica. Segundo Cruz et al (2004), o termo dialelo tem sido utilizado para expressar um conjunto de $p(p-1)/2$ híbridos, resultante do acasalamento entre p progenitores (linhagens, variedades, clones, etc.), podendo incluir, além dos respectivos pais, os híbridos recíprocos (quando há efeito materno) e, ou, outras gerações relacionadas, como F₂'s, retrocruzamentos, etc.

Com o dialelo é possível obter estimativas de parâmetros úteis para seleção de progenitores para hibridação e entender efeitos genéticos envolvidos na determinação dos caracteres. Existem várias metodologias de análise dialélica como as propostas por Griffing, em que se estima a capacidade geral e a capacidade específica de combinação; Gardner e Eberhart, em que se avalia os efeitos de variedade e heterose varietal; e Hayman, em que são obtidas informações sobre o caráter em estudo, os valores genéticos dos progenitores utilizados e o limite de seleção (CRUZ REGAZZI & CARNEIRO, 2004).

As metodologias expostas até o momento são usadas nos programas de melhoramento convencionais, todas baseadas nos valores fenotípicos dos indivíduos. Uma maneira de melhorar o processo de escolha dos pais seria se

basear no valor genotípico dos mesmos por meio da herdabilidade, um parâmetro que mede a proporção da variabilidade total que tem origem genética. Ela poderia ser usada com esse objetivo, por estimar a proporção da variância fenotípica que tem origem genética (BOREM & MIRANDA, 2013).

Sendo a herdabilidade alta, haverá alta correlação entre o valor genético e fenotípico, mas caso contrário, haverá baixa correlação. Deseja-se em um programa de melhoramento que o primeiro caso ocorra. É possível, por meio da genômica, obter um critério mais acurado ao se construir um fenótipo enriquecido (F^*) que se aproximaria mais do genótipo. Realmente, as duas tecnologias que tiveram maior impacto no melhoramento de plantas, como o do milho, nos últimos tempos, foram o uso de duplo haploides e o emprego de informações moleculares como critério de seleção (CHAKRADHAR et al, 2017).

O primeiro método proposto para o uso de marcadores no melhoramento ficou conhecido como seleção assistida por marcadores (MAS). Essa técnica tinha várias limitações, sendo novas metodologias propostas como a genética de associação (GWAS) e a seleção genômica ampla ou seleção genômica (GWS) (CRUZ, SALGADO & BHERING, 2013).

Segundo Cruz, Salgado e Bhering (2013) a seleção genômica ampla pode ser definida, de modo simplificado como um método de predição de fenótipo ou valor genético, visando à seleção de indivíduos superiores para uma ou mais características, com base exclusivamente em informações genotípicas. Escolhe-se os melhores indivíduos que servirão como parentais para formar a próxima geração com base em informações de marcadores moleculares.

Para realizar a seleção genômica ampla é necessária, primeiramente, a determinação dos efeitos genéticos dos marcadores que serão utilizados na seleção sobre as características de interesse. Os marcadores atualmente usados são do tipo SNP_S (Single Nucleotide Polimorphism) que consistem em nucleotídeos substituídos em pontos específicos do ácido desoxirribonucleico (DNA). A importância biológica da substituição está relacionada ao dogma central da biologia (CRICK & WATSON, 1956).

A informação contida na molécula de DNA se expressa por meio de proteínas e a maquinaria para formação delas está no citoplasma. É necessária

uma molécula que carregue a informação do DNA para esse espaço. Essa molécula é o ácido ribonucleico (RNA) (WATSON et al, 2015).

Os ribossomos, principal peça da maquinaria que converte a informação do DNA em proteínas, relaciona cada trio de nucleotídeos presentes no RNA mensageiro (códon) a um aminoácido. As trocas de nucleotídeos no DNA podem levar a formação de proteínas diferentes que influenciarão na expressão do fenótipo do indivíduo analisado (WATSON et al, 2015).

Os efeitos dos marcadores sobre as características de interesse são estimados simultaneamente e modelos para prever o valor genético genômico dos indivíduos em gerações futuras da população de melhoramento são elaborados. O número de marcadores utilizados é considerável, cobrindo amplamente o genoma, possibilitando que quase toda a variação genética do caráter quantitativo seja capturada (CRUZ, SALGADO & BHERING, 2013).

Posteriormente, são utilizadas duas populações, uma chamada de população de validação e outra chamada de população de treinamento. A primeira consiste na parcela do conjunto de dados utilizado para construção de um modelo que por meio do genótipo e valor fenotípico prevê o valor genômico. A população de treinamento consiste na parcela usada para aplicar o modelo.

O presente trabalho visa oferecer subsídio aos melhoristas quanto a escolha dos genitores, mostrando a GWS como uma ferramenta válida para a escolha de genitores e também para determinar a capacidade genética de híbridos que não foram gerados usando-se como modelo a planta de milho, portando tendo como referência a espécie *Zea mays*. Essa cultura foi escolhida por ser ela a que mais se beneficia com novas informações sobre a produção de híbridos simples, por ser a que mais trabalha com esse tipo de material de cultivo.

2. Referencial teórico

A maioria dos caracteres de importância econômica e agrícola é de natureza quantitativa. Esses caracteres são controlados por muitos genes e bastante influenciados pelo ambiente (CRUZ, 2012).

As técnicas usuais de melhoramento das espécies consistem na análise fenotípica dos dados. No estudo da herança de caracteres quantitativos a média de um conjunto de indivíduos será uma medida mais confiável, pois os efeitos do ambiente tende a se cancelar. Outra medida que torna os dados mais confiáveis é a variância do conjunto de dados (CRUZ, 2012).

Na escolha de um determinado cultivar para ser lançado em uma região são feitos experimentos em que se varia os genótipos testados e mantêm-se constante as condições ambientais como fertilidade do solo e disponibilidade hídrica (FILHO et al, 2005).

Uma outra forma de se realizar estudos genéticos é pelo uso de dados simulados. Os benefícios do uso de simulações já haviam sido abordados por Bhering (2008) e Correa (2001). Segundo estes autores, uma boa alternativa para se obter um conjunto de dados que permitam a comparação de eficácia de métodos e modelos é aquela fundamentada na utilização das técnicas de simulação que permite a obtenção de um grande volume de informações, sob um sistema que se conhece valores paramétricos em que hipóteses podem ser convenientemente testadas. Estes dados podem ser obtidos em um curto período de tempo sem os custos de implantação e condução de experimentos.

O modo de reprodução da espécie influencia diretamente no método de melhoramento que será realizado. As plantas de reprodução sexuada, ou seja, em que há formação de gametas para formação de novos indivíduos, podem ser divididas em três grupos: autógamas, alógamas e autógamas com frequente alogamia. No primeiro ocorre a autofecundação com frequência superior a 99%, no segundo caso ocorre a fecundação cruzada e no último caso há uma taxa intermediária de fecundação cruzada em relação as espécies autógamas e alógamas (BOREM & MIRANDA, 2013).

O milho, planta usada como modelo para simulação dos dados, enquadra-se no grupo das alógamas. Ele é o cereal preferido na África do Sul e Oriental,

América Central e México. Há grande diversidade de cores dessa espécie (branco, amarelo, vermelho e preto), sendo que a cor preferida para consumo varia de região para região (RANUM et al, 2014). Sua produção está em constante crescimento chegando a atingir 875,226,630 de toneladas no mundo em 2012, sendo o Brasil responsável por 8% dessa produção (FAO, 2012). Além do melhoramento genético, o uso de pacotes tecnológicos durante o plantio como alteração da densidade de plantas por hectare e do espaçamento entre as plantas também tem influencia sobre o ganho por produção (SILVA et al, 2014). O produto é procurado tanto para alimentação animal como humana, sendo vendido desde a forma de espiga até produtos mais simples como farinha de milho, fubá e farelo (OLIVEIRA & BEZERRA, 2013).

Ele é uma planta C4, tendo altos níveis de rendimento em regiões temperadas. Apesar das regiões tropicais ocuparem 73% da área em que é feito o cultivo do milho, a sua produção é menor (2 toneladas por hectare) quando comparada com as regiões temperadas (9 toneladas por hectare) (PRAZANNA, 2014). Seu centro de origem é o México. De lá ele se espalhou pelo mundo devido a tribos indígenas que o levaram para outras regiões da América Latina, do Caribe e depois dos Estados Unidos e do Canadá. Os exploradores europeus levaram o milho para a Europa e, mais tarde, os comerciantes levaram o milho para a Ásia e a África (RANUM, 2014).

Quando se trabalha com plantas alógamas o objetivo do programa de melhoramento é obtenção de variedades e, principalmente, de híbridos visando a utilização da heterose, que consiste na superioridade da geração F1 em relação a média dos seus pais. Teorias que tentam explicar a ocorrência da heterose se fundamentam na heterozigosidade e na dominância, que afirma que a presença de alelos diferentes confere um efeito conjunto favorável para o indivíduo ou que um alelo dominante (ou sobredominante) confere maior superioridade genotípica (CRUZ, 2012).

Há três tipos de híbridos: simples, triplo e duplo, sendo essa geralmente a ordem decrescente de uniformidade e produtividade. No trabalho de Emygdio et al (2007), com milho, foram encontrados alguns híbridos duplos com potencial produtivo superior a híbridos simples e triplos, indicando que a

generalização da superioridade do desempenho de híbridos de acordo com o cruzamento que o originou não é apropriada.

Os híbridos simples são obtidos do cruzamento entre duas linhagens. Os híbridos triplos são o resultado do cruzamento entre um híbrido simples e uma linhagem. Os híbridos duplos envolvem o cruzamento de dois híbridos simples (BOREM & MIRANDA, 2013).

Para formação do cultivar híbrido a ser lançado, o programa de melhoramento precisa de boas linhagens. Linhagens são organismos que possuem quase todos os locos em homozigose. São obtidas por ciclos sucessivos de autofecundação de populações, variedades sintéticas, híbridos comerciais, etc. Nas espécies autógamas o cultivar a ser lançado é uma linhagem, mas também é necessária a formação de híbridos para obtenção de genótipos de alto desempenho ao final do programa (TESSELE, 2017).

Os custos de produção dos híbridos simples são muito grandes, refletindo em seu preço de venda. Pequenos agricultores os substituem por híbridos triplos e duplos devido a esses custos. A diminuição da média da população das alógamas devido as autofecundações é chamada de depressão por endogamia (BOREM & MIRANDA, 2013).

Independentemente se as plantas são autógamas ou alógamas ou autógamas com frequente alogamia, os programas de melhoramento consistem em várias etapas, sendo a primeira e mais importante a escolha de genitores, como abordado por Borem e Miranda (2013). Estes autores relatam que os programas de melhoramento envolvem recursos humanos, infraestrutura para testes e avaliações, métodos adequados de melhoramento e germoplasma. A escolha inadequada de qualquer um desses componentes limita o progresso do programa, mas a escolha inapropriada do germoplasma talvez seja o fator mais crítico e limitante.

Os objetivos do programa variam, sendo assim, a escolha dos cruzamentos também deve variar. Usualmente, os melhoristas realizam cruzamentos com diferentes objetivos. Os critérios de escolha de genitores, bem como a metodologia de condução das populações segregantes, variam com os diversos objetivos dos cruzamentos (BOREM & MIRANDA, 2013).

Um dos objetivos do programa é o desenvolvimento de cultivares, de germoplasma ou estudos de aspectos genéticos de uma característica em dada espécie (BOREM & MIRANDA, 2013).

Para o desenvolvimento de germoplasma são necessários genitores que contenham características que se deseja introduzir. Muitas vezes esses genitores não possuem as características comerciais adequadas, sendo necessários vários ciclos de melhoramento (BOREM & MIRANDA, 2013).

No estudo de aspectos genéticos, os genitores precisam ser bem divergentes em relação a característica analisada, possibilitando-se o estudo do número de genes, herdabilidade, etc (BOREM & MIRANDA, 2013).

Para escolher com maior segurança as combinações que irão resultar em uma população em que se poderá ser retirado o novo cultivar é necessário analisar as características dos genitores. Deve-se sempre ter em mente que os genitores escolhidos devem ser agronomicamente superiores e adaptados à região de cultivo (BOREM & MIRANDA, 2013).

Escolha de genitores a partir de informações fenotípicas

A escolha de genitores poderá ser feita por observações do bom desempenho dos materiais em condições naturais de campo por meio de ensaios de competição. Essa análise consiste na observação das médias, obtidas a partir de informações de parcelas experimentais, de vários indivíduos, recomendando-se as mais promissoras.

Outra forma de escolha seria a análise do histórico dos genótipos como pais, vendo se frequentemente originam uma prole com alto desempenho.

A análise de características relacionadas com a característica principal do programa também é uma forma de escolha útil. No trabalho de Entringer et al (2014) foi verificado, pela análise de trilha, que dois caracteres, o volume do grão e o volume de espiga apresentaram efeito direto sobre a produção, sendo essas características eficientes no aumento do peso médio de espiga de milho superdoce.

Escolher genitores com base unicamente em seu desempenho pode ser questionável, pois este critério não contempla informação sobre a complementariedade gênica dos genitores além de ter confiabilidade

questionável quando se trabalha com característica de baixa herdabilidade. A herdabilidade é a porção da variância fenotípica que se deve a causas genéticas.

Escolha de genitores agregando informações sobre a diversidade genética

O conhecimento das distâncias genéticas entre os genitores podem ser calculadas desde a análise de genealogias ou de técnicas multivariadas que se utilizam de características morfológicas como a análise de diversidade genética (FILHO, 2004).

A análise de diversidade consiste na formação de grupos onde os genótipos similares são organizados. Pretende-se com isso que as maiores diferenças estejam concentradas entre os grupos (CRUZ, FERREIRA & PESSONI, 2011). A distância entre os genótipos pode ser medida pela Distância Euclidiana Média, em que não são necessárias repetições ou pela Distância Generalizada de Mahalanobis, que leva em consideração as variâncias e covariâncias residuais que existem entre as características em análise (CRUZ, FERREIRA & PESSONI, 2011).

Como o número de estimativas estudadas é muito grande, é praticamente impossível formar grupos pela simples análise das distâncias. Por conta disso são usados métodos de agrupamento ou de projeções de distâncias em gráficos. (CRUZ, FERREIRA & PESSONI, 2011)

Outra possibilidade de avaliar a diversidade genética entre os genitores é por meio da análise dos componentes principais (ACP) ou de variáveis canônicas. Estes procedimentos têm por objetivo analisar os dados a partir de combinações lineares das variáveis originais. Quanto menos componentes forem criados, melhor será, pois eles representarão com mais clareza o comportamento dos indivíduos (CRUZ, FERREIRA & PESSONI, 2011).

Apesar de toda importância da diversidade genética, Moose e Mumm (2008) afirmaram que a necessidade de diversidade genética deve ser equilibrada pelo desempenho de elite, porque escolher os melhores pais é a chave para maximizar a probabilidade de uma melhoria bem-sucedida.

Escolha de genitores agregando informações sobre a progênie em cruzamentos dialélicos

Interessa ao melhorista, geralmente, bons genitores que ao serem cruzados proporcionem combinações híbridas que apresentam elevada capacidade específica de combinação, com ao menos um dos genitores apresentando elevada estimativa de capacidade geral de combinação (CANÇADO, 2002).

Os cruzamentos dialélicos podem ser definidos como todas as combinações duas a duas envolvendo n genitores. Há vários métodos de análise dialélica como a de Gardner e Eberhart, em que se avalia os efeitos de variedade e heterose varietal; de Hayman, em que são obtidas informações sobre o caráter em estudo, os valores genéticos dos progenitores utilizados e o limite de seleção e o modelo de Griffing que estima a capacidade geral de combinação (CGC) e a capacidade específica de combinação (CEC) (CRUZ REGAZZI & CARNEIRO, 2004). Pela CGC são obtidas informações sobre a concentração de genes aditivos, enquanto que com os efeitos da CEC são medidos os efeitos gênicos não aditivos (CRUZ, REGAZZI & CARNEIRO, 2004). Em alguns casos deve-se abrir mão de genitores com alta CGC para se ter uma alta CEC.

Griffing propôs quatro modelos de dialelo em que a presença dos cruzamentos recíprocos, genitores e híbridos varia. Enquanto no tipo I todos esses elementos estão presentes, no tipo II exclui-se os cruzamentos recíprocos, no tipo III os genitores e no tipo IV os genitores e os cruzamentos recíprocos (FILHO, 2004).

Abordagem de escolha de genitores e híbridos envolvendo informações moleculares

Até o século XX somente essas metodologias baseadas na informação fenotípica eram usadas. No fim desse século, estratégias de seleção em que eram usadas marcadores passaram a ser usadas. Marin (2005) definiu os marcadores moleculares como segmentos de DNA que estão fisicamente ligados a locos que determinam características de interesse. Podem ser evidenciados por métodos que combinam o uso de enzimas de restrição à hibridização entre sequências complementares de DNA, como no caso do

“Restriction Fragment Length Polymorphisms” (RFLP) ou pela técnica de “Polymerase Chain Reaction” (PCR). O grande potencial do uso de marcadores moleculares no melhoramento reside no fato de eles serem praticamente ilimitados em número, de fácil detecção e se comportarem como “caracteres” de herança simples e previsível, não sendo afetados pelo meio.

Vale lembrar que os marcadores não são afetados pelo meio, mas os indivíduos sim, tendo diferentes genes sendo expressos de acordo com o local em que estão situados. Um exemplo é o milho cultivado nas regiões tropicais que está mais exposto a fatores de estresse abiótico, como pragas e excesso de chuvas, causando inundações, tendo mais alelos favoráveis relacionados a sobrevivência nessas condições. Essa diferença entre o milho tropical e temperado precisa ser levada em consideração quando analisado os dados genômicos (CHAKRADHAR et al, 2017).

Existem vários tipos de marcadores, como os Simple Sequence Repeats (SSR), o Polymerase Chain Reaction -Random Amplified Polymorphism DNA (PCR – RAPD), os Single Nucleotide Polimorphism (SNPS) e os RFLP. Os primeiros a serem usados nos programas de melhoramento foram os RFLP, na década de 1980 (GUIMARAES, 2009). Atualmente é utilizado em maior quantidade os SNPs.

Há milhares de SNPs espalhados pelo genoma dos organismos. A importância desses vai muito além do melhoramento. Eles foram usados na identificação de mecanismos de resposta imune em plantas, algo que nunca antes havia sido identificado nos vegetais (ZORZATTO et al, 2015).

Para a identificação e mapeamento de um marcador é necessário uma população segregante para a característica de interesse que pode ser uma população F2, uma população derivada de retrocruzamento ou mesmo uma população de linhagens endogâmicas recombinantes (RILs) (MARIN, 2005).

As RILs podem ser definidas como um conjunto de linhas homozigóticas derivadas de um ancestral comum, por um processo endogâmico. O desenvolvimento de RILs envolve muitas gerações de recombinação, sendo os genótipos homozigotos normalmente fixados por autofecundação (SCHUSTER & CRUZ, 2013).

As informações moleculares são obtidas pela genotipagem dos organismos. No caso das plantas podem ser retirados pedaços de tecidos das

folhas e realizados os procedimentos adequados para extração do DNA a ser analisado (EMBRAPA, 2010).

Esse DNA é levado as empresas de genotipagem como a Illumina. Estando o genoma da espécie sequenciado são produzidos adaptadores universais de DNA ou RNA que se complementam com o DNA enviado. Esses adaptadores são chamados de SNPchip. As bases nitrogenadas dos nucleotídeos da amostra são corados diferentemente por meio de técnicas modernas de fluorescência. Essas cores são detectadas por comprimentos de ondas diferentes (RAFFAN & SEMPLE, 2011).

Os fragmentos de sequência estão alinhados com a sequência de referência. Onde há discordância com a sequência de referência, sua frequência é usada para determinar se a variante é heterozigótica ou homozigótica. Bancos de dados de referência são pesquisados para determinar se as variantes são novas ou previamente reconhecidas como SNPs (RAFFAN & SEMPLE, 2011).

Tendo-se identificados os SNPs, é possível seu uso nos programas de melhoramento. A primeira técnica de uso dos marcadores no melhoramento foi chamada de seleção assistida por marcadores (MAS). Ela se baseia na detecção de locos de características quantitativas (QTLs) para identificar genes envolvidos na expressão de características complexas.

Essa técnica não foi muito útil, pois nem todos os QTLs segregados são identificados em um estudo e estão superestimados (LADO, 2017). Estudos envolvendo a MAS incluem a alteração da proporção de aminoácidos presentes no endosperma do milho. Há baixa quantidade dos aminoácidos lisina e triptofano e se tem como objetivo criar um milho com o dobro da quantidade desses aminoácidos presentes no milho comum. Utilizando-se marcadores microssatélites, por meio da MAS, foi possível acelerar o ritmo dos programas de melhoramento que tem esse objetivo (BABU et al, 2014).

Devido aos problemas inerentes à MAS, outros métodos usando marcadores foram criados, como a seleção genômica ou seleção genômica ampla (GWS). Para a GWS é importante apenas a presença ou ausência de SNP na amostra analisada.

Ela é uma técnica melhor por fazer a predição do desempenho genotípico analisando simultaneamente todos os efeitos do marcador no

genoma. O objetivo da GWS é obter um modelo que prediz o valor genético do indivíduo, mas que não necessariamente determina genes específicos envolvidos no controle do caráter. Ela trabalha com duas populações: de treinamento e de validação (CRUZ, SALGADO & BHERING, 2013).

A população de treinamento é usada para a genotipagem dos marcadores moleculares e fenotipagem das características de interesse, obtendo-se equações de predição dos valores genético-genômicos.

Na predição dos valores genéticos genômicos dos indivíduos por meio da análise dos marcadores ocorre o problema de multicolinearidade e dimensionalidade. A multicolinearidade deve-se ao número de marcadores usados ser bem maior do que o número de indivíduos. A dimensionalidade é devido a falta de independência entre os marcadores, por muitos destes estarem no mesmo grupo de ligação (desequilíbrio de ligação). Devido a esses problemas não é possível a realização de uma análise de regressão multivariada para predição dos valores genéticos dos indivíduos.

A predição é então realizada usando-se os efeitos estimados com base na população de treinamento e submetidos à análise de correlação com os valores genéticos obtidos via análise dos dados fenotípicos, conseguidos por metodologias-padrão (BLUP) (CRUZ, SALGADO & BHERING, 2013).

A população de validação é usada para testar as equações de predição e avaliar a acurácia do modelo previamente estabelecido. Sendo a acurácia a proximidade entre os resultados obtidos pelos modelos preditivos e os valores reais. Ela também é composta por indivíduos analisados quanto aos marcadores moleculares e a fenotipagem (CRUZ, SALGADO & BHERING, 2013).

Para ser possível analisar a acurácia das equações de predição é necessário que a população de validação e de treinamento sejam independentes, pois assim os erros associados aos valores preditos e observados são também independentes, e toda a correlação entre esses valores é de natureza genética e indica a capacidade preditiva da GWS (CRUZ, SALGADO & BHERING, 2013).

Sendo o modelo acurado é possível utilizá-lo para predição dos valores genético-genômicos dos indivíduos em gerações futuras da população de melhoramento, sem que haja a necessidade de realizar-se a fenotipagem deles

para as características de interesse. A acurácia dos indivíduos nas gerações futuras é equivalente à acurácia estimada na população de validação (CRUZ, SALGADO & BHERING, 2013).

As equações de predição dos valores genéticos genômicos obtidas com os dados fenotípicos e genotípicos associam a cada intervalo (delimitado pelos marcadores moleculares), ou a cada marcador individual, o efeito na característica de interesse (CRUZ, SALGADO & BHERING, 2013).

Vários fatores atuam sobre a capacidade preditiva do modelo criado pela seleção genômica ampla como o desequilíbrio de ligação populacional, a herdabilidade, número de QTLs e número de locos que controlam a característica. Quanto menor o tamanho populacional, maior será o desequilíbrio de ligação notado. O número de QTLs influencia negativamente a acurácia do modelo quando é usada baixa densidade de marcadores, o que faz com que nem todos os QTLs estejam em desequilíbrio de ligação com pelo menos um marcador (CRUZ, SALGADO & BHERING, 2013).

É possível adaptar os modelos preditivos da GWS em estudos de predição de valores genômicos de genitores e principalmente prever o comportamento de híbridos de um conjunto grande de genitores. A determinação da capacidade genética de híbridos que não foram realizados se deve a obtenção gratuita do genótipo de todos os indivíduos híbridos devido à combinação dois a dois das marcas moleculares presentes nos pais, obtida inicialmente para construção do modelo de predição para construção dos melhores híbridos. Com o sucesso dessa metodologia, a recomendação das linhagens será melhorada, podendo-se recomendar as superiores, mesmo que estas estejam dentro de um grupo muito grande. Acredita-se que, além disso será possível recomendar diretamente os melhores híbridos, fazendo-se somente os cruzamentos que irão resultá-los, sem gastos desnecessários. Tais fatos deverão ser investigados e proporcionarão valiosas informações para o melhoramento de plantas cujo objetivo é a utilização de híbridos.

3. Material e Métodos

3.1 Materiais

Os dados utilizados neste trabalho foram oriundos de simulação realizada no Laboratório de Bioinformática da Universidade Federal de Viçosa, localizado no Instituto de Biotecnologia aplicada a Agropecuária (BIOAGRO) utilizando o aplicativo computacional Genes (CRUZ, 2013).

Obtenção de informações fenotípicas e genóticas de genitores

Foi simulado um experimento de avaliação de 200 linhagens, em relação a cinco características e feita a genotipagem de 1000 marcadores. A espécie modelo usada foi *Zea mays*, conhecida popularmente como milho. Ela é uma espécie diploide ($2n = 2x = 20$) estando seu material genético concentrado em dez grupos de ligação que possuem tamanhos variados.

Foi considerado um grupo de marcas distribuídas nos dez grupos de ligação com distância entre locos variável e que o marcador utilizado era do tipo codominante. De todos os gametas possíveis de serem realizados (2^{1000}) foram tomados 5000 e esses foram combinados aleatoriamente formando cada indivíduo de uma população que, por sucessivas gerações de autofecundação, derivou 200 linhagens endogâmicas recombinantes (RILs). Foi suposto que esses gametas originaram-se de F1s obtidas de dois genitores homocigotos contrastantes.

O valor genético total expresso por um determinado indivíduo foi estimado a partir da expressão (CRUZ, SALGADO & BHERING, 2013):

$$G_i = \mu + a_i + d_i \quad (1)$$

Em que:

d_i são os efeitos de dominância no genótipo i , ou seja, o efeito dado pela presença de dois alelos diferentes no mesmo loco em cromossomos homólogos;

a_i são os efeitos aditivos presentes no genótipo i , dado por:

$$a_i = \sum_{j=1}^g p_j \alpha_j \quad (2)$$

Em que:

α_j é o efeito do alelo favorável no loco j, considerado igual a 1, 0 ou -1 para as classes genotípicas AA, Aa e aa, respectivamente;

e

p_j é a contribuição do loco j para a manifestação da característica estabelecida no trabalho a partir de pesos gerados por valores de uma distribuição binomial $(p+q)^n$ sendo $p=q = 0,5$ e n igual a $g-1$ (g = número de genes que controlam o caráter). Assim, no processo de simulação foi admitido que há genes que tem maior efeito sobre a característica do que outros.

As características que foram simuladas estão detalhadas na Tabela 1. As informações estabelecidas para fins de simulação foram obtidas tomando por base informações disponíveis na literatura para um conjunto de características consideradas relevantes para a cultura do milho. Trata-se de apenas valores referenciais utilizados para fins de comparação do desempenho das metodologias utilizadas sem prejuízo na extrapolação dos resultados obtidos. Alguns dados foram estabelecidos levando em conta os outros parâmetros da característica analisada.

Tabela 1 - Valores estabelecidos, para fins de simulação, tomando como referência algumas características mensuradas na espécie modelo *Zea mays*

Característica	Número de genes	Média	Coefficiente de variação	Grau médio de dominância	Herdabilidade
X ₁	12	4,38	19,06	0,47	0,86
X ₂	23	7,03	7,42	0,56	0,7
X ₃	31	27,67	9,99	1	0,5
X ₄	14	238,94	11,05	0,89	0,4
Y	60	149,16	12,56	0,3	0,33

Fonte: SOLFERINI 2010, BRITO 2010, FALUBA et al 2010, SOUZA et al 2016.

O grau médio de dominância (gmd) tem influência sobre os indivíduos heterozigotos, situando-os em uma posição diferente do ponto médio entre os

homozigotos dominante e recessivo. Podendo, nos casos em que o gmd é superior a 1, segundo a teoria da sobredominância, causar a heterose, ou seja, a superioridade do F1 em relação a média de seus pais (CRUZ, 2012).

Os fenótipos dos indivíduos (i) foram gerados segundo o modelo (CRUZ, SALGADO & BHERING, 2013):

$$F_i = G_i + M_i \quad (3)$$

Em que:

G_i é o efeito genético dado pelo somatório dos efeitos genéticos de cada loco e pelo desvio de dominância, conforme exposto na equação (1)

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:

$$h^2 = \frac{\sigma_g^2}{\sigma_f^2} \quad (4)$$

Em que:

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

σ_f^2 corresponde a variância fenotípica.

Foram atribuídos entre quatro e oito genes em comum entre as características e há muitos genes que estão no mesmo grupo de ligação, ou seja, há genes que determinam tanto a característica x_1 como a característica x_2 e/ou que estão localizados no mesmo cromossomo, indicando a existência de correlações genéticas proporcionadas tanto por ligação gênica quanto por pleiotropia.

Obtenção de informações fenotípicas e genotípicas de híbridos

A partir do conjunto de informações genotípicas e fenotípicas de cada linhagem foram gerados um novo conjunto de dados representados pelas combinações híbridas, num total de 19900 híbridos (C_{200}^2).

Valores reais contínuos e discretos dos híbridos e genitores

Para fins de comparação foram utilizados os dados referentes aos valores genotípicos simulados dos híbridos e dos genitores. Outra alternativa utilizada foi a transformação destes dados com distribuição contínua em valores discretos. Assim, os valores genotípicos dos híbridos foram

organizados equitativamente em cinco grupos em relação a sua qualidade, tomando por base a característica principal (Y), em ótimos, bons, médios, ruins e péssimos. Para organização das informações das linhagens nesses grupos foi usada a informação da capacidade geral de combinação das mesmas, obtidas por meio do dialelo de Griffing a partir da análise dos valores genotípicos.

O modelo de dialelo de Griffing usado foi do tipo II, em que se exclui os F1's recíprocos (CRUZ, SALGADO & BHERING, 2013).

O modelo estatístico usado nesse dialelo está representado a seguir:

$$Y_{ij} = m + g_i + g_j + s_{ij} + \bar{e}_{ij} \quad (5)$$

Sendo:

Y_{ij} o valor médio da combinação híbrida ($i \neq j$) ou do progenitor ($i = j$);

m a média geral;

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

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

\bar{e}_{ij} o erro experimental médio.

A capacidade geral de combinação para cada um dos 200 genitores foi dada por:

$$\hat{g}_i = \frac{1}{(p+2)} [Y_{ii} + Y_{i.} - \frac{2}{p} Y_{..}] \quad (6)$$

Onde:

\hat{g}_i a capacidade geral de combinação do genitor i ;

p o número de progenitores analisados;

Y_{ii} total do cruzamento do genitor i com o genitor i ;

$Y_{i.}$ total da característica analisada para o genótipo i ;

e

$Y_{..}$ total para a característica analisada em relação a todos os genótipos.

3.2 Estratégias de escolhas de genitores para fins de hibridação

3.2.1 Abordagem tradicional (apenas informação fenotípica)

O método tradicional de escolha dos pais leva em consideração apenas o valor fenotípico dos mesmos. Foram abordadas três metodologias

tradicionais diferentes para fins de escolha de genitores destinado a hibridação para derivação de populações segregante: baseada apenas na informação da característica de interesse obtida em ensaio de competição; baseada na informação fenotípica da variável de interesse agregando-se informações sobre a diversidade genética multivariada; e baseada em análise dialélica.

Análise do desempenho per se acompanhado de análise de diversidade genética

Apesar do interesse ser a escolha de 10 linhagens superiores tendo-se como referência a característica Y, adotou-se como critério fazer uma pré-seleção. Assim, primeiramente foram consideradas as informações obtidas no desempenho *per se* a partir das quais foram pré-selecionados 20 genitores levando-se em consideração seu desempenho em relação a característica Y. Posteriormente foi realizado a análise de diversidade genética apenas com este grupo pré-selecionado. A distância entre os genótipos foi medida pela distância Euclidiana média. Essa distância pode ser medida pela seguinte expressão:

$$d_{ii'} = \sqrt{\frac{\sum_j (Y_{ij} - Y_{i'j})^2}{v}} \quad (7)$$

Em que:

$d_{ii'}$: distância euclidiana média entre os genitores i e i';

v : número de características estudadas;

e

Y_{ij} : valor médio do i-ésimo genitor em relação a j-ésima variável.

Foi suposto durante a simulação que havia genes em comum e no mesmo grupo de ligação entre as características e, portanto, uma certa correlação. Apesar disso, foi usada a distância euclidiana média porque em estudos de melhoramento praticamente não existem características que não tenham nenhuma relação (CRUZ, FERREIRA & PESSONI, 2011).

Com a matriz de distâncias em mãos foi realizado o método de agrupamento de Tocher. Segundo Cruz, Ferreira & Pessoni (2011) esse método consiste na identificação do par de indivíduos mais similares para formar o grupo inicial. Posteriormente é avaliada a possibilidade de inclusão de novos indivíduos, tendo-se como critério que a distância média intragrupo deve ser menor que a distância média intergrupo. Como critério para inclusão dos genótipos em um grupo é analisado se o valor médio das distâncias dentro do grupo ($d_{(grupo)k}/n$) está abaixo do valor máximo da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo (θ). Assim:

Se $d_{(grupo)k}/n \leq \theta$, inclui-se o indivíduo k no grupo;

Se $d_{(grupo)k}/n > \theta$, o indivíduo k não é incluído no grupo.

Como o objetivo da análise era escolher os pais mais distantes geneticamente, a matriz de dissimilaridade foi transformada em similaridade estando as menores distâncias associadas aos genótipos mais diferentes, tendo-se um Tocher invertido. Deve-se portanto fazer a escolha dos genótipos mais distantes analisando-se o interior dos grupos.

Para análise complementar foi usada a análise de componentes principais que consiste em uma técnica da estatística multivariada que transforma um conjunto de variáveis originais em outro conjunto de variáveis de mesma dimensão denominadas de componentes principais com o objetivo de redistribuir a variação observada nos eixos originais de forma a se obter um conjunto de eixos ortogonais não correlacionados. Essa análise é muito útil nas análises genéticas por agrupar os indivíduos de acordo com sua variação, ou seja, segundo seu comportamento dentro da população, representado pela variação do conjunto de características que define o indivíduo, portanto, a técnica agrupa os indivíduos de uma população segundo a variação de suas características (VARELLA, 2008).

Para realizar a análise de componentes principais é conveniente padronizar as variáveis, pois normalmente elas estão em unidades de medida distintas. Uma das formas é estabelecer que a variância entre as características seja 1:

$$z_{ij} = \frac{x_{ij}}{s(x_j)}, i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, p \quad (8)$$

Em que:

$s_{(x_j)}$ é o desvio padrão da característica j;

z_{ij} é o valor padronizado da característica j no indivíduo i

e

x_{ij} é o valor da característica.

Em seguida faz-se necessário ter a matriz de correlação entre as características R. Os componentes principais são determinados resolvendo-se a equação característica da matriz R, isto é:

$$\det[R - \lambda I] = 0 \text{ ou } |R - \lambda I| = 0 \quad (9)$$

Se a matriz R for de posto completo igual a p, isto é, não apresentar nenhuma coluna que seja combinação linear de outra, a equação $R - \lambda I = 0$ terá p raízes chamadas de autovalores ou raízes características da matriz R.

Sejam $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$ as raízes da equação característica da matriz R, então:

$$\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_p$$

Para cada autovalor λ_i existe um autovetor a_i :

$$\tilde{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \cdot \\ \cdot \\ a_{ip} \end{bmatrix} \quad (10)$$

Os autovetores \tilde{a}_i são normalizados, isto é, a soma dos quadrados dos coeficientes é igual a 1, e ainda são ortogonais entre si. Devido a isso apresentam as seguintes propriedades:

$$\sum_{j=1}^p a_{ij}^2 = 1 \quad (\tilde{a}_i \cdot \tilde{a}_i = 1) \quad (11)$$

$$E \sum_{j=1}^p a_{ij} \cdot a_{kj} = 0 \quad (\tilde{a}_i \cdot \tilde{a}_k = 0 \text{ para } i \neq k)$$

Sendo \tilde{a}_i o autovetor correspondente ao autovalor λ_i , então o i-ésimo componente principal é dado por:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (12)$$

O primeiro componente é o que apresenta maior variância e assim por diante (VARELLA, 2008).

Dos 20 genitores disponíveis, considerados de bom desempenho para Y, os dez mais distantes foram escolhidos. Com esses dois procedimentos pretendia-se obter pais bons e distantes, seu uso é recomendado pelos melhoristas, como mencionado por Cruz, Ferreira & Pessoni (2011) que relatam que o sucesso de um programa de melhoramento reside na existência de variabilidade na população de trabalho. Melhoristas têm recomendado, para a formação de uma população base (ou população de melhoramento), o intercruzamento entre cultivares de desempenho superiores e divergentes entre si.

Desempenho per se acompanhado de análise dialélica

Foram pré-selecionados 20 genitores levando-se em consideração seu desempenho em relação a característica principal Y. Posteriormente estes genitores e suas combinações híbridas foram avaliados em sistema dialélico. Os dados obtidos foram submetidos à análise dialélica e à análise da capacidade combinatória geral com objetivo de identificar os 10 genitores com maior valor genético aditivo.

Desempenho per se acompanhado de análise de diversidade genética e de análise dialélica

Neste caso realizou-se uma primeira pré-seleção de forma que os 40 melhores genitores em relação a característica Y foram identificados e também analisados quanto a diversidade genética usando-se o método de Tocher invertido e a análise de componentes principais. Desses genitores, realizou-se nova pré-seleção de forma que os 20 de bom desempenho e divergentes foram escolhidos. Posteriormente foi realizada a análise dialélica a partir das informações dos 20 pais que foram cruzados entre si formando as combinações híbridas. Assim, finalmente obteve-se dez genitores com bom desempenho, divergentes e de maior valor genético aditivo tendo em vista a sua maior capacidade geral de combinação quanto ao caráter Y.

3.2.2 Abordagem com informação molecular

Neste caso o critério de escolha dos genitores foi o valor genético-genômico predito obtido a partir da abordagem utilizada na seleção genômica ampla. A Seleção Genômica Ampla consiste em um modelo que prediz o valor genético do indivíduo, sem determinar genes específicos envolvidos no caráter.

Para modelar o efeito dos marcadores foi usado o Ridge Regression-Best Linear Unbiased Prediction (RR-BLUP), denominado de regressão aleatória. O RRBLUP assume que os efeitos dos marcadores são aleatórios e apresentam distribuição normal com variância constante. Foi usado, segundo Cruz, Salgado e Bhering (2013) o seguinte modelo para modelar os efeitos dos marcadores:

$$y = Xg + e \quad (13)$$

em que:

y é o fenótipo;

X é uma matrix de incidência de dimensões N (número de indivíduos) x n (número de marcadores);

g é o vetor de coeficientes de regressão, com distribuição normal (0, σ^2_g);

e

e é o erro aleatório com distribuição normal (0, σ^2_e).

Os efeitos dos marcadores podem ser estimados pela equação:

$$g = (X'X + I\lambda)^{-1} X'y \quad (14)$$

em

$\lambda = k = \sigma^2_e / \sigma^2_g$ é constante para todos os marcadores. A variância dos marcadores é constante, e em geral k deve ser considerada como função da variância genética aditiva.

Tendo-se o efeito predito de cada marcador é possível obter o valor genético genômico (VGG) predito dos indivíduos:

$$\widehat{VGG} = \sum_i^n X_i \hat{g}_i \quad (14)$$

em que:

n é o número de marcadores dispostos no genoma,

X_i é a linha da matriz de incidência que aloca o genótipo do i-ésimo marcador para cada indivíduo

e

\hat{g}_i é o efeito estimado do i-ésimo marcador.

No presente trabalho, as populações de treinamento e validação foram todas as linhagens, genotipadas para os marcadores e fenotipadas para as características de interesse. Os 10 melhores genitores em relação ao valor genômico predito foram escolhidos.

Como as populações de validação e de treinamento são as mesmas será usado para medir a acurácia do modelo a simples correlação de Pearson entre o valor genético real dos indivíduos e o valor genético genômico:

$$\rho_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)\sigma_X\sigma_Y} \quad (15)$$

Em que:

ρ denota o coeficiente de Pearson;

n número de termos;

x_i e y_i são os valores das variáveis X e Y para o indivíduo i;

\bar{x} e \bar{y} são as médias dos variáveis X e Y;

σ_X e σ_Y são os desvios padrão das variáveis X e Y.

A significância dessa correlação foi analisada pelo teste t a 1% de probabilidade.

3.3 Comparação dos resultados de escolha de genitores

Os resultados das técnicas apresentadas anteriormente relativas a escolha de genitores foram comparados com os resultados genotípicos guardados inicialmente, analisando-se em qual das cinco classes (ótima, boa, média, ruim, péssima) as linhagens escolhidas estavam.

3.4 Capacidade de predição do modelo frente as outras características

Foi analisado se a capacidade de predição do modelo variava alterando-se a herdabilidade e o número de genes usando-se as outras características simuladas. Para isso, avaliou-se a correlação entre os valores genéticos reais e preditos pelo modelo, também analisando-se a significância dessa correlação pelo teste t a 1% de probabilidade.

3.5 Predição dos híbridos não realizados

Foi também utilizada a abordagem da GWS porém, neste caso, a população de treinamento foi constituída pelas 200 RILs acrescida da informação de 190 híbridos formados da combinação dois a dois de 20 indivíduos do primeiro dialelo, em que os pais foram submetidos apenas ao ensaio de competição. A inclusão dos híbridos foi realizada visando adicionar informações de locus heterozigóticos a população de treinamento, de forma a incrementar a acurácia preditiva da GWS para as populações híbridas. O modelo de predição foi utilizado numa população de validação constituída por todos híbridos. Ele é um modelo simples que considera a presença de efeitos de dominância (g_{md}) igual a zero para todas as características.

Os valores genômicos obtidos para os híbridos via GWS foram comparados com valores genômicos reais, guardados inicialmente e novamente pela correlação de Pearson entre os valores genéticos reais e preditos, foi analisada a acurácia do modelo de predição. A significância da correlação foi avaliada pelo teste t a 1% de probabilidade.

4. Resultados e Discussão

4.1 Estratégias de escolhas de genitores para fins de hibridação

4.1.1 Abordagem tradicional (apenas informação fenotípica)

Escolha de genitores com base em informação de ensaios de competição acompanhado de análise de diversidade genética

Nos programas de melhoramento a análise fenotípica de dados costuma ser rotineiramente realizada com o objetivo de escolha dos melhores genótipos, mas na maioria das vezes, principalmente em características muito influenciadas pelo ambiente e controlada por um grande número de genes (características quantitativas), nem sempre os melhores fenótipos correspondem aos melhores genótipos.

Uma das formas de se fazer a escolha dos pais por métodos fenotípicos é a análise do desempenho *per se* em que são analisadas as características dos genitores e escolhidos os melhores. As características analisadas são mensuradas em relação a suas médias e normalmente a escolha leva o critério univariado, ou seja, é dado foco em uma característica.

Apesar da escolha univariada, normalmente os genitores têm desempenho parecido em relação as outras características, por haver correlação entre elas, seja dado por ligação gênica ou pleiotropia. Sendo a ligação gênica a presença de genes em um mesmo cromossomo e a pleiotropia, um gene determinando duas ou mais características diferentes. A correlação devido à ligação gênica muitas vezes é perdida no processo de recombinação para formação de gametas, mas como a população de trabalho simulada é RILS os eventos de recombinação estão fixados e esse problema não existiu. Por isso que esse tipo de população costuma ser muito usada em estudos genéticos. Brondani (2003) afirmou esse uso dizendo que, apesar de ser bastante trabalhoso obter uma população RIL, elas são úteis para a análise genética, por terem fixado os eventos de recombinação, originando um mapa de ligação que pode servir como referência para análises de QTLs para

várias características, locais e anos, aumentando a confiabilidade dos resultados.

Esta relação entre as características foi vista no trabalho, como mostrado na Tabela 2.

Tabela 2 - Correlação fenotípica (acima da diagonal) e genotípica (abaixo da diagonal) entre as características. Em negrito as correlações com a característica Y.

Características	X ₁	X ₂	X ₃	X ₄	Y
X ₁	-	0,0829 ^{ns}	0,1772*	-0,0324 ^{ns}	0,2047**
X ₂	0,0594 ^{ns}	-	0,1188 ^{ns}	0,2449**	0,0825^{ns}
X ₃	0,2400**	0,2517**	-	0,1518*	0,2861**
X ₄	0,0063 ^{ns}	0,387**	0,4608**	-	0,2457**
Y	0,4691**	0,2137**	0,5498**	0,4758**	-

** , * significativo a 1 e 5% de probabilidade pelo teste t

A diversidade genética geralmente é quantificada por uma medida multivariada, por levar em consideração mais de uma característica ao calcular a distância entre os genitores. Com os 20 genitores pré-selecionados por ter melhor desempenho em relação ao ensaio de competição em Y foi empregado o método de “Tocher invertido” uma vez que a análise foi feita a partir da matriz de similaridade e não da matriz de dissimilaridade como normalmente é realizado. Assim, com esta estratégia os grupos formados representavam agrupamento de genitores divergentes tornando fácil o reconhecimento dos pais mais distantes geneticamente. Foram formados nove grupos por esse método, estando na Tabela 3, os resultados da distribuição dos genótipos dentro desses grupos.

Pode-se notar que há bastantes grupos formados por apenas um ou dois genitores. Isso ocorre, segundo Vasconcelos *et al.* (2007), porque o método de Tocher utiliza de um critério global de agrupamento se baseando na maior entre as menores distâncias encontradas na matriz de dissimilaridade (nesse caso de similaridade por envolver uma modificação na análise) durante todo o processo.

Tabela 3 - Agrupamento de genitores, pré-selecionados com base no rendimento, de acordo com o método de otimização de Tocher invertido utilizando a matriz de similaridade expressa pelo complemento da distância Euclidiana média padronizada.

I(*)	171 152 67 167 149 150 8 175 96 137
II	187 121
III	65 22
IV	196
V	83
VI	172
VII	133
VIII	138
IX	37

(*) Cada grupo agrega genitores com alta dissimilaridade

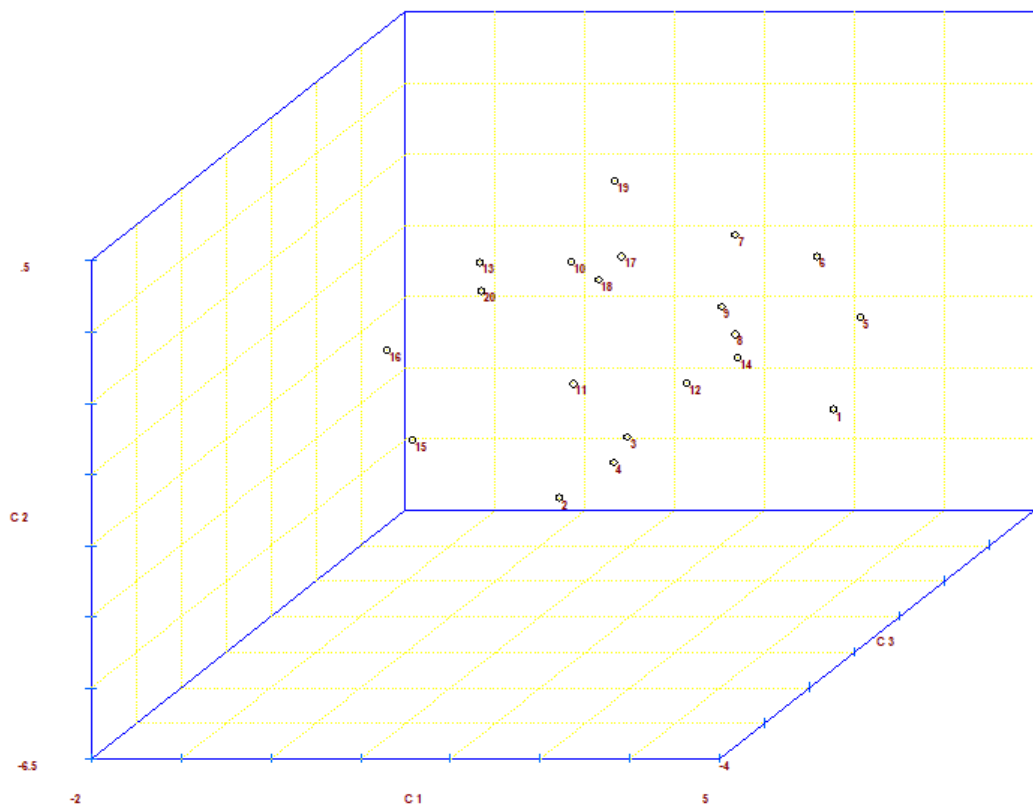
Após o método de Tocher invertido foi realizada a análise dos componentes principais. O número de componentes é sempre menor ou igual ao número de variáveis originais. As variâncias simples e acumulada dos cinco componentes principais estão descritas na Tabela 4.

Tabela 4 - Porcentagem de variação simples e variação acumulada por componente principal (CP) obtida da análise de componentes principais entre 20 linhagens para cinco características fenotípicas (x_1 , x_2 , x_3 , x_4 e x_5).

	CP1	CP2	CP3	CP4	CP5
Varição Simples	39,59	24,02	18,67	13,12	4,6
Varição Acumulada	39,59	63,61	82,28	95,4	100

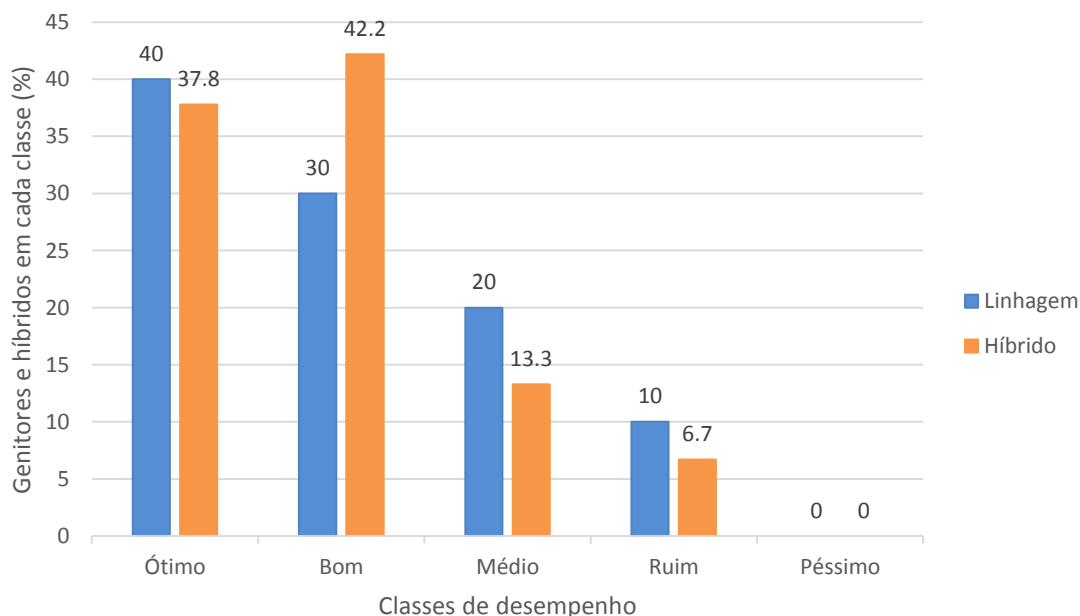
O componente principal 1 retém a máxima variabilidade dos dados. Foram escolhidos, para a análise de componentes principais, os componentes 1, 2 e 3, sendo, portanto, tridimensional o gráfico formado pela projeção dos escores obtidos pelos componentes (Figura 1). A escolha desses componentes deve-se ao envolvimento deles com um mínimo de 70% da variação disponível que, Regazzi (2000) determinou como ideal.

Figura 1 - Dispersão gráfica dos 20 melhores genitores segundo o ensaio de competição. Os genótipos mostrados estão codificados para números entre 1 a 20 sendo correspondentes, respectivamente, a 171, 167, 65, 187, 150, 175, 196, 172, 138, 133, 37, 149, 137, 96, 8, 152, 83, 22, 67 e 121.



Por meio dessas análises foram selecionadas dez linhagens do grupo I formado pelo Tocher invertido (152, 137, 8, 167, 96, 171, 150, 149, 175 e 67) e, ao se formar os 45 híbridos (combinação dos dez pais dois a dois), tem-se, respectivamente, que 40% e 37,8% estariam no grupo dos classificados como ótimos (Figura 2).

Figura 2 - Distribuição, nas cinco classes de desempenho, das linhagens escolhidas pela metodologia ensaio de competição somado a análise de divergência genética e dos híbridos formados pelo cruzamento entre essas linhagens.



Ensaio de competição acompanhado de análise dialélica

Com o estudo da diversidade genética é possível a obtenção de pais distantes que deverão se complementar para produzir os cruzamentos, mas não de pais com um alto valor genético aditivo, que seria o melhor para um programa de melhoramento que se preocupa com a obtenção do híbrido e com a manutenção dos genitores tendo em vista a obtenção comercial do híbrido em campos de produção. Uma das formas de reconhecer o potencial genético aditivo dos genitores é por meio da análise dialélica de Griffing, observando, por meio deste procedimento, a capacidade geral de combinação dos pais, que é uma medida expressa por uma função do efeito aditivo dos genes.

No experimento em questão, antes de ser realizada a análise dialélica foi realizado uma pré-seleção por meio do ensaio de competição, escolhendo-se as 20 melhores linhagens fenotipicamente para serem cruzadas. Os resultados da capacidade geral de combinação de cada linhagem, obtidas por meio dos híbridos formados do cruzamento dois a dois entre as 20 linhagens

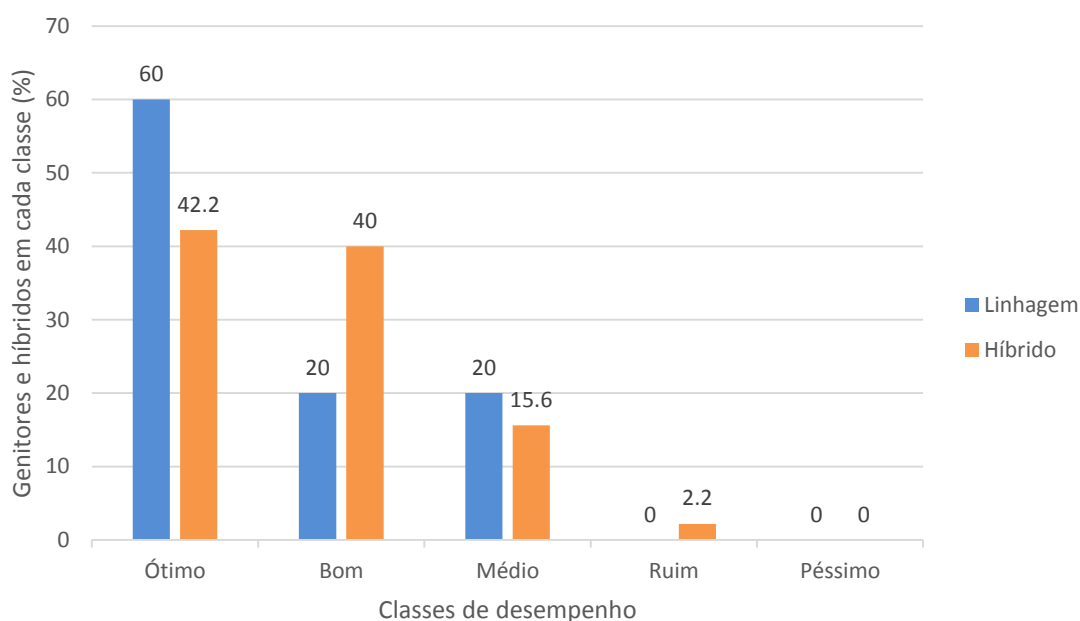
anteriores e pertinentes somente a esta situação, são mostrados na Tabela 5. Foram escolhidas as 10 linhagens de maior CGC.

Tabela 5 - Capacidade geral de combinação (CGC) das 20 melhores linhagens pré-selecionadas segundo o ensaio de competição.

Linhagens	Efeito da CGC
8	-0,00557
22	-0,06383
37	0,010185
65	-0,12805
67	-0,10216
83	-0,18714
96	0,101184
121	-0,08809
133	0,134214
137	0,021983
138	0,365462
149	-0,14545
150	0,021529
152	-0,12627
167	-0,13085
171	-0,06996
172	0,088535
175	0,070472
187	-0,06522
196	0,299011

Dentre as linhagens escolhidas (8, 37, 96, 133, 137, 138, 150, 172, 175 e 196) e dos híbridos formados, 60% e 42,2%, respectivamente, estavam no grupo dos ótimos (Figura 3), mais do que nas duas estratégias descritas anteriormente, mostrando que a informação sobre a CGC realmente conduz a uma melhor escolha de pais e híbridos.

Figura 3 - Distribuição nas cinco classes de desempenho das linhagens escolhidas pela metodologia ensaio de competição somado a análise dialélica e dos híbridos formados pelo cruzamento entre essas linhagens.



Com a análise dialélica foram obtidos pais com uma alta concentração de alelos favoráveis gerando uma população de trabalho com melhor desempenho, mas esse desempenho poderia ser melhorado ao ser incorporar uma análise de diversidade genética, pois os pais obtidos terão alelos favoráveis diferentes, havendo complementariedade nos cruzamentos.

Ensaio de competição acompanhado de análise de diversidade genética e de análise dialélica

Com o objetivo de se ter pais bons, com efeitos aditivos e complementares foi feito o ensaio de competição somado a diversidade genética e a análise dialélica. Os 40 melhores pais fenotipicamente em relação à característica Y foram submetidos a análise de diversidade genética por meio do método de Tocher invertido. Foram formados vinte e quatro grupos por esse método, estando na Tabela 6 a distribuição dos genitores dentro desses grupos.

Tabela 6 - Agrupamento dos genótipos de acordo com o método de otimização de Tocher invertido utilizando a matriz de similaridade.

I(*)	171 177 60 21 175 152 149 167 156 153 61 150 55
II	65 70
III	96 137
IV	8 67
V	92 172
VI	187
VII	185
VIII	133
IX	37
X	11
XI	44
XII	196
XIII	192
XIV	34
XV	131
XVI	2
XVII	139
XVIII	173
XIX	121
XX	138
XXI	83
XXII	89
XXIII	179
XXIV	22

(*) Cada grupo agrega genitores com alta dissimilaridade

Novamente, poucos grupos concentraram a maior parte dos genótipos e muitos grupos foram formados por apenas um genótipo devido ao critério global de agrupamento do método de Tocher (VASCONCELOS *et al.*, 2007). Após o método de Tocher foi realizada a análise dos componentes principais que utiliza uma transformação ortogonal para converter um conjunto de

variáveis possivelmente correlacionadas em um conjunto de valores de variáveis linearmente não correlacionadas. A variância simples e acumulada dos cinco componentes principais está na Tabela 7.

Tabela 7 - Porcentagem de variação simples e variação acumulada por componente principal (CP) obtida da análise de componentes principais entre 40 linhagens para cinco características fenotípicas.

	CP1	CP2	CP3	CP4	CP5
Varição Simples	34,44	22,8	22,31	13,57	6,88
Varição Acumulada	34,44	57,24	79,55	93,12	100

Os componentes 1, 2 e 3 juntos envolveram mais de 70% da variação disponível, sendo esses escolhidos para a análise e devido a isso, novamente o gráfico formado é tridimensional (Figura 4).

Com os 20 genitores escolhidos como mais divergentes, considerando-se os grupos do Tocher invertido e a análise gráfica, foi realizada a análise dialética, estando a capacidade geral de combinação desses na Tabela 8.

Figura 4 - Dispersão gráfica dos 40 melhores genitores segundo o ensaio de competição. Os genótipos mostrados estão codificados para números entre 1 a 40, sendo correspondentes, respectivamente, a 2, 8, 11, 21, 22, 34, 37, 44, 55, 60, 61, 65, 67, 70, 83, 89, 92, 121, 131, 133, 137, 138, 139, 149, 150, 152, 153, 156, 167, 171, 172, 173, 175, 177, 179, 185, 187, 192 e 196.

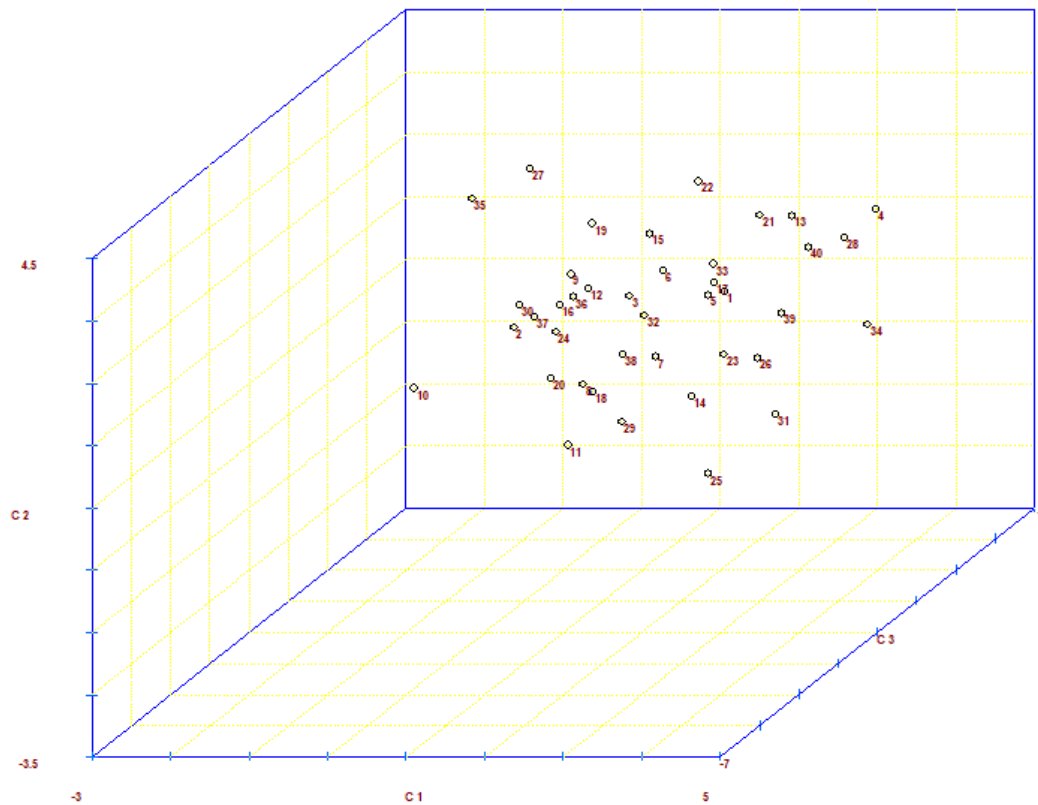


Tabela 8 - Capacidade geral de combinação das 20 linhagens mais divergentes segundo a análise dialélica.

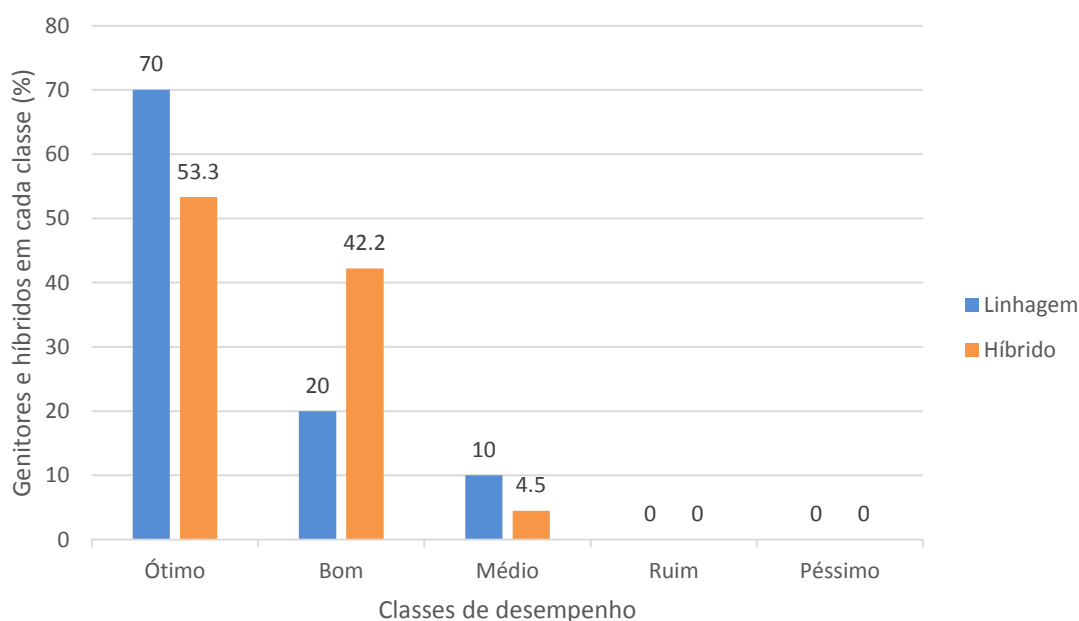
Linhagens	Efeito da CGC
8	-0,06361
21	0,046135
55	-0,35756
60	-0,16322
61	-0,12744
65	0,011768
67	0,04504
70	-0,12974
96	0,065484
137	0,269265
149	-0,1244
150	-0,14762
152	0,018256
153	0,283423
156	-0,02198
167	-0,04458
171	0,238543
172	0,150209
175	0,074564
177	-0,02253

Foram selecionadas as dez linhagens de maior capacidade geral de combinação (21, 65, 67, 96, 137, 152, 153, 171, 172 e 175) e formados 45 híbridos estando, respectivamente, 70% e 53,3% no grupo dos ótimos (Figura 5) mais do que nas outras três metodologias apresentadas. Além disso, nenhum dos híbridos se enquadrou na classificação de ruim, como havia acontecido anteriormente. Esses resultados mostram a potencialidade das técnicas biométricas apontando que a análise em que se utiliza de três metodologias (seleção visual, diversidade genética e análise dialélica) é melhor do que a dupla. Essa conclusão é sustentada por Dudley (1982) que afirmou que 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, sendo esses resultados possíveis utilizando-se análises de diversidade genética e dialélica.

Ferreira (1995) em seus trabalhos com milho constatou realmente que fazer preliminarmente uma análise de diversidade genética antes do dialelo seria útil para determinar variedades que teriam uma boa capacidade de combinação. Assim, seriam selecionados os pais mais divergentes, sendo feitos somente os cruzamentos necessários que teriam maior chance de sucesso.

Figura 5 - Distribuição nas cinco classes de desempenho das linhagens escolhidas pela metodologia ensaio de competição somado a análise de diversidade genética e análise dialélica e dos híbridos formados pelo cruzamento entre essas linhagens.



4.1.2. Abordagem com informação molecular

O critério de escolhas dos genitores para formação da população de trabalho pode ser melhorado ao se fazer uso de informações moleculares para prever o valor genético genômico dos pais. Assim, o melhorista poderá gerar

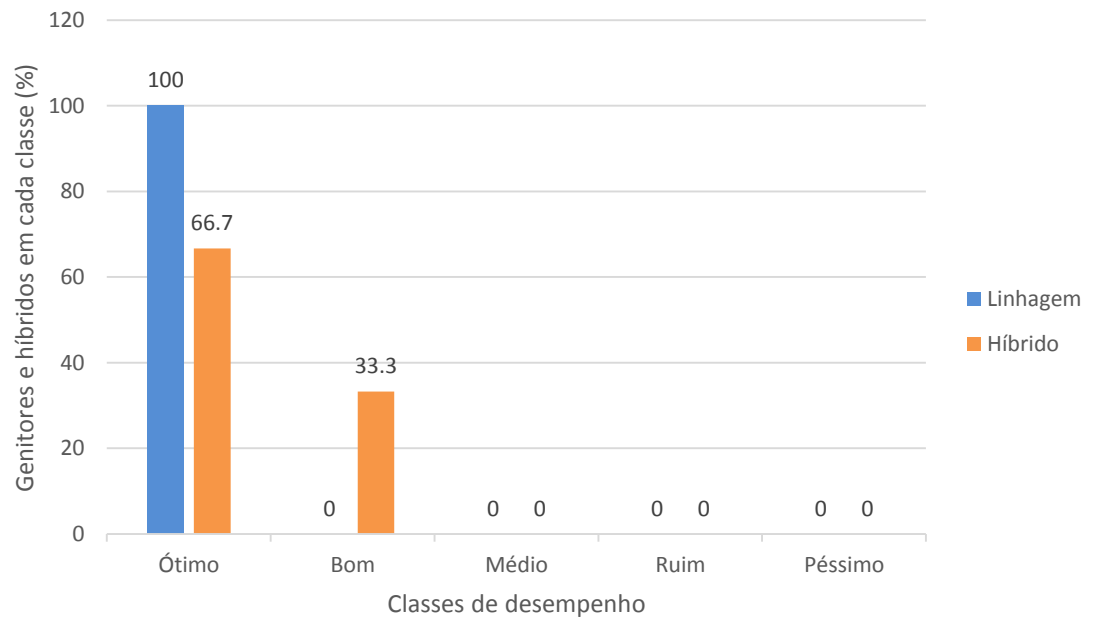
um novo valor fenotípico (valor genético genômico) predito a partir de análises segundo modelos obtidos pela abordagem da seleção genômica ampla.

Para constatar a capacidade dos marcadores em capturar a variação genética do caráter quantitativo para construção do modelo de predição foi analisado o comportamento de segregação dos marcadores, dado que a população analisada representava um conjunto de RILs, encontrando-se que 1,6% apresentavam segregação distorcida em relação à esperada (1:1). Essa frequência bem baixa indica que os marcadores eram de boa qualidade para a genotipagem. Quando os dados são reais, os marcadores que apresentam segregação distorcida são descartados, como no trabalho de Sibov et al (2003) de mapeamento molecular de milho tropical usando marcadores microssatélites.

Na predição, a população de treinamento e de validação foram todas as linhagens, genotipadas para os marcadores e fenotipadas para as características de interesse. As melhores linhagens obtidas por meio do modelo de predição criado se enquadravam totalmente no grupo das ótimas, indicando a capacidade desse método de obter bons genitores.

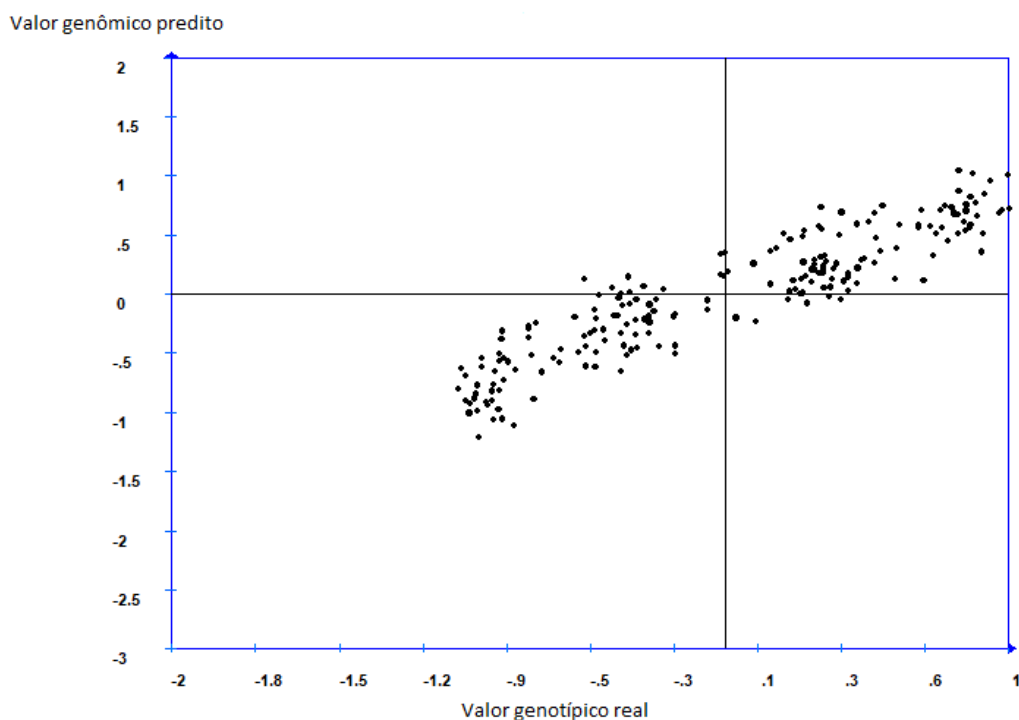
Sendo as linhagens escolhidas cruzadas, tem-se 66,7% das populações híbridas na classe dos ótimos (Figura 6), mais do que em todas as metodologias anteriores, evidenciando que o uso da modelo de predição genômico é útil para na formação de híbridos.

Figura 6 - Distribuição nas cinco classes de desempenho das linhagens escolhidas por meio do valor genotípico predito fornecido por meio da seleção genômica ampla e dos híbridos formados pelo cruzamento entre essas linhagens.



A correlação de Pearson entre o valor genético genômico e o valor genético real foi significativa pelo teste t a 1% de probabilidade, com valor 0,917 (Figura 7). Como anteriormente dito, essa correlação indica a acurácia do modelo de predição.

Figura 7 - Dispersão dos valores genéticos reais e os valores genômicos preditos mensurados nos genitores, com estimativa de correlação de 0,917.



Apesar do critério de seleção univariado, como o ensaio de competição e a análise dialélica, a seleção genômica apresentou resultados muito superiores, pois ao se utilizar dos marcadores moleculares para a escolha dos pais, foi formado o fenótipo enriquecido (F^*), que se aproximou do genótipo, como era desejado. Ao se utilizar apenas de informações fenotípicas pode-se estar selecionando um indivíduo que simplesmente aproveita de forma melhor o ambiente e não necessariamente que possua alelos favoráveis, que seria o ideal (TESSELE, 2017).

4.2 Capacidade de predição do modelo aplicado às outras características

Com o objetivo de avaliar se a acurácia dos modelos de predição se alteram em função da herdabilidade e do número de genes foi analisada a correlação entre os valores genéticos preditos e reais para as outras características simuladas. Os valores genômicos preditos foram obtidos por

meio da criação de modelos de predição para essas outras quatro características.

Para x1, x2 e x3, os valores foram, respectivamente, 0,9711, 0,9274 e 0,8893, podendo ser considerados elevados. Já para a característica x4, a correlação foi de 0,401307. Comparando-se as características x1 e x4, que possuem quase o mesmo número de genes (12 e 14, respectivamente), pode-se considerar que a diferença de predição das duas deve-se a herdabilidade diferenciada, que foi bem menor para x1. Resultado parecido foi encontrado por Tessele (2017), que percebeu para a característica produtividade, que tinha baixa herdabilidade, a acurácia da GWS tinha sido reduzida.

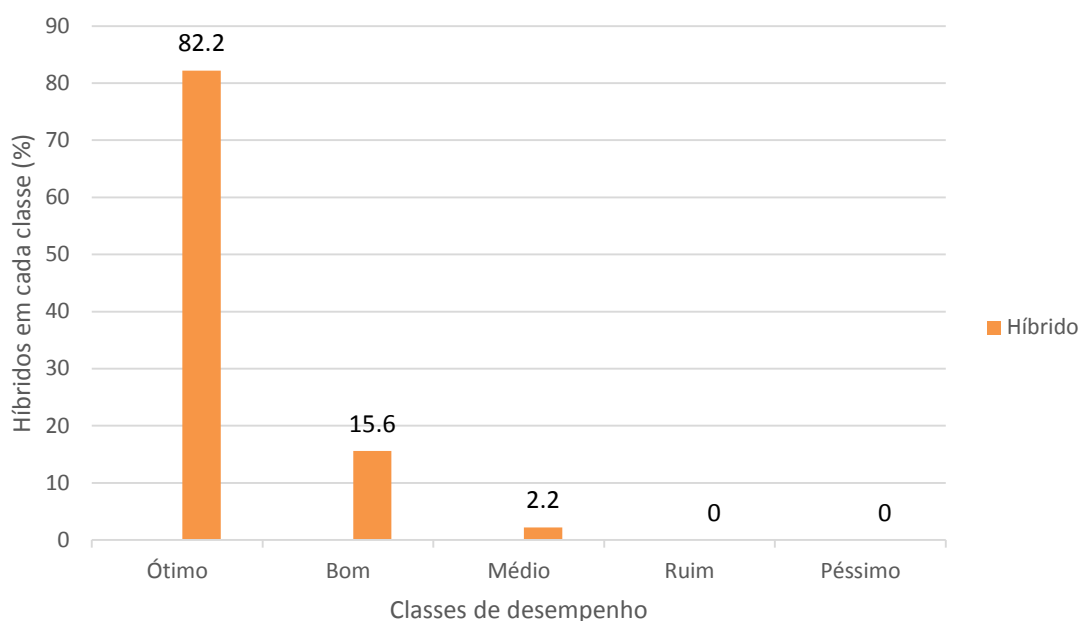
4.3 Predição dos híbridos não realizados

Com a genotipagem das linhagens genitoras em mãos, realizada para construção do modelo anterior, foi possível, pela associação par a par de cada marcador, obter a genotipagem de todos os híbridos (19900) que seriam formados pelo cruzamento das 200 linhagens.

A população de híbridos genotipada foi usada como população de treinamento de um modelo de predição e como população de validação o conjunto de informações relativos às 200 RILS mais as 190 combinações híbridas do dialelo em que os pais foram submetidos apenas ao ensaio de competição.

Com o modelo foi obtido o valor genético genômico predito de todos os híbridos e dos 45 apontados por ele como melhores em relação a característica rendimento (14643, 18005, 19596, 19518, 14518, 14585, 14622, 14619, 17984, 13680, 17981, 19497, 13581, 19545, 13622, 13659, 13656, 14620, 19400, 17685, 14615, 14580, 17982, 19524, 19495, 19490, 13791, 14540, 17977, 19379, 19376, 14618, 13692, 17627, 17664, 17661, 14441, 17816, 19725, 13657, 17980, 13733, 19469, 13770 e 19466), 82,2% estavam na classe dos ótimos (Figura 8).

Figura 8 - Distribuição nas cinco classes de desempenho dos híbridos apontados como melhores quando usado o valor genotípico predito fornecido pelo segundo modelo de predição criado.



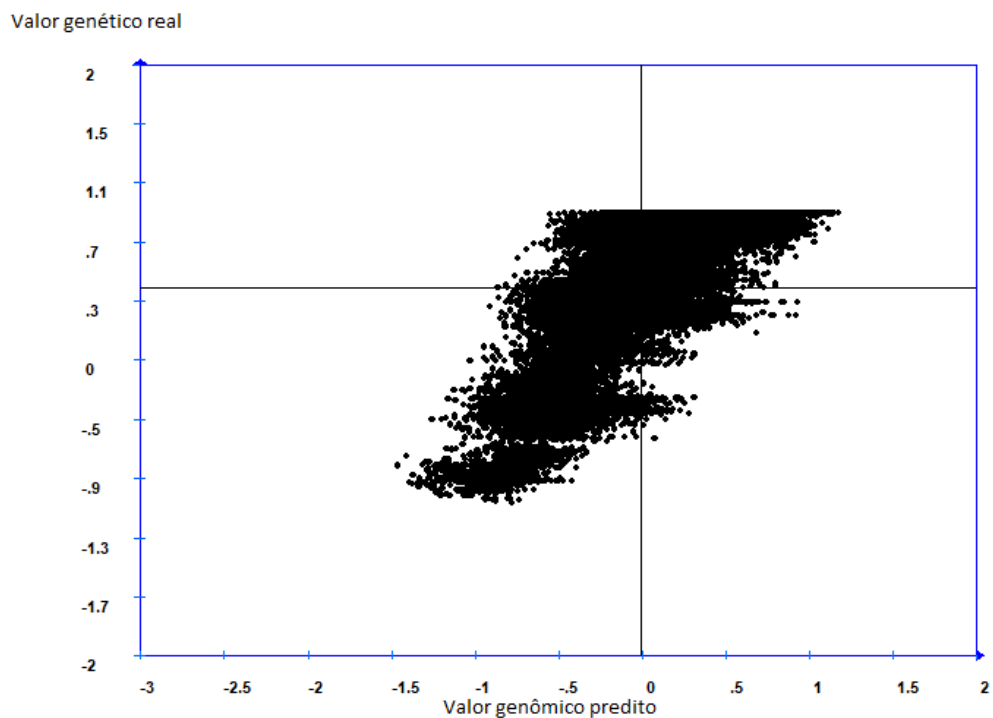
Essa informação é muito útil, pois mostra que o modelo de predição construído foi adequado para prever o desempenho de híbridos que não foram realizados e pode ser aplicado em programas de melhoramento, poupando a realização de cruzamentos que não resultariam em um híbrido vigoroso.

Os híbridos formados foram oriundos do cruzamento de 17 diferentes linhagens, algumas sendo usadas nos cruzamentos mais do que outras. Esse possível resultado havia sido abordado por Tessele (2017). Ao se fazer a escolha dos pais por meio dos melhores híbridos é possibilitado que uma linhagem apareça poucas ou somente uma vez, desde que resulte em híbrido com alto valor genético genômico predito.

A correlação de Pearson entre o valor genético genômico e o valor genético real também foi significativa a 1% de probabilidade pelo teste t, com valor 0,7591 (Figura 9). A menor magnitude dessa correlação em relação a encontrada entre os pais pode ser creditada ao grau médio de dominância, que para todas as características foi diferente de zero, mas que não foi levada em consideração pelo modelo de predição RR-BLUP utilizado. Esse grau médio não influencia as linhagens, uma vez que são indivíduos homozigotos, mas

influencia os híbridos. Os melhores híbridos obtidos pelos valores genéticos reais ocorrem quando os cruzamentos entre os pais maximizam a concentração de alelos aditivos, a divergência e os efeitos de dominância. Em análises futuras seria interessante considerar esse fator no modelo de predição.

Figura 9: Dispersão dos valores genéticos reais e os valores genômicos preditos mensurados nos híbridos, com estimativa de correlação de 0,7591. Os pontos superiores do gráfico formam uma linha reta porque há vários genótipos com o mesmo valor genético predito devido a inúmeros cruzamentos que chegam a um mesmo genótipo heterozigoto.



Foi visto que a seleção genômica ampla consiste em um método útil e eficaz para seleção de linhagens e também para predição do desempenho de híbridos que não foram realizados.

Apesar de todos os seus benefícios, deve-se salientar que a seleção genômica ampla não visa substituir a análise fenotípica dos dados, obtidas via campo, mas complementar o processo, para realização de uma escolha melhor.

5. Conclusões

Neste estudo foi considerada escolha de genitores, para fins de hibridação, por meio das estratégias da análise fenotípica sobre a característica de interesse, análise da diversidade genética e análise dialéctica. Conclui-se que:

- Ao se utilizar informações de análise de diversidade genética incorporada ao ensaio de competição, o número de linhagens ótimas e de híbridos ótimos foi relativamente baixo;

- Ao incorporar a análise dialéctica ao ensaio de competição observou-se mais linhagens no grupo das melhores do que na análise anterior, mostrando a superioridade desse método;

- Utilizando-se informações fenotípicas e abordagens biométricas combinadas os resultados foram ainda mais satisfatórios;

Foi avaliada a eficácia da incorporação da informação molecular para fins de escolha de genitores tendo em vista a produção de híbridos. Concluiu-se que o modelo de predição foi eficaz para predizer o valor genético genômico das linhagens em análise;

Por fim, foi avaliada a eficácia do uso de modelos genômicos na predição de híbridos não realizados. Conclui-se que a abordagem também foi útil para prever o desempenho de híbridos que não foram realizados.

6. Referências Bibliográficas

Babu R, Prasanna BM (2014). **Molecular breeding for quality protein maize (QPM)**. In: Genomics of plant genetic resources. Tuberosa R, Graner A, Frison E (eds.) Springer Dordrecht Heidelberg, New York, London, pp. 489–505.

BHERING, L. L. **Mapeamento genético em famílias simuladas de irmãos completos**. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, 166p, 2008.

BORÉM, A.; MIRANDA, G. V. **Melhoramento de Plantas**. 2013.

BRITO, A. H. **Controle genético e químico de doenças foliares e grãos ardidos em milho**. 2010.

CANÇADO, G. M. **Avaliação de nove linhagens de milho em cruzamentos dialélicos quanto à tolerância ao alumínio**. Pesquisa agropecuária Brasileira, Brasília. 2002.

CHAKRADHAR, T.. HINDU, V.. REDDY, P. S.. **Genomic-based-breeding tools for tropical maize improvement**. 2017.

CRICK, F.H. WATSON, J.D. Structure of Small Viruses. **Nature**.1956. Vol.177 pp.473-5 ref.19.

CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Capítulo 2: Diversidade Genética Baseada em Informações Fenotípicas, p. 29 a 188. 2011.

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. . **Princípios de Genética Quantitativa**. 2ª edição, capítulo 1. 2012.

CRUZ, C. D. ; REGAZZI, A. J.; CARNEIRO, P. C. S.. **Modelos biométricos aplicados ao melhoramento genético**. Volume 1, 4ª edição, capítulo 7. 2004.

CRUZ, C. D.; SALGADO, C. C.; BHERING, L. L.. **Genômica Aplicada**. Capítulo X: Seleção Genômica Ampla. Pag. 210. 2013. 13.

EMYGDIO, B. M.. IGNACZAK, J. C.. FILHO, A. C.. **Potencial de rendimento de grãos de híbridos comerciais simples, triplos e duplos de milho**. Revista Brasileira de Milho e Sorgo. 2007.

ENTRINGER, G. C.. SANTOS, P. H. A. D.. VETTORAZZI, J. C. F. CUNHA, K. S.. PEREIRA, M. G.. **Correlação e análise de trilha para componentes de produção de milho superdoce**. Rev. Ceres, Viçosa, v. 61, n.3, p. 356-361, mai/jun, 2014.

FAOSTAT. **Food and Agriculture Organization of the United Nations**. 2012. Disponível em <<http://faostat.fao.org/site/345/default.aspx>>. Acesso em 18 de março de 2018.

FALUBA, J. de S.; MIRANDA, G. V.; DeLIMA, R. O.; SOUZA, L. V. de; Debem, E. A.; OLIVEIRA, A. M. C.. **Potencial genético da população de milho UFV 7 para o melhoramento em Minas Gerais**. 2010.

FERREIRA, D. F.; OLIVEIRA, A. C. de; SANTOS, M. X. dos; RAMALHO, M. A. P. . **Métodos de avaliação da divergência genética em milho e suas relações com os cruzamentos dialélicos**. 1995.

FILHO, J. L. da S. . **Estratégias de predição de cruzamentos não realizados em análise dialélica, utilizando informações de marcadores moleculares**. 2004.

FILHO, J. P. R. do A. . **R. Bras. Ci. Solo**, 29: 467-473, 2005.

GUIMARAES, C. T. **Marcadores moleculares e suas aplicações no melhoramento genético**. 2009.

LANA, U. G. de P.; GOMES, P. C.; TINOCO, C. F. da S.; OLIVEIRA, B. C. F. S.; GUIMARAES, C. T.; MAGALHAES, J. V. de; OLIVEIRA, B. C. F. S. **Procedimento da Embrapa Milho e Sorgo para extração de DNA de tecido vegetal em larga escala**. Sete Lagoas: Embrapa Milho e Sorgo, 2010. 19 p. (Embrapa Milho e Sorgo. Documentos, 104).

MACHADO, C. de F. **Procedimentos para escolha de genitores de feijão**. 1999. Universidade Federal de Lavras, Lavras, MG.

MOOSE, S. P.. MUMM, R. H.. **Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement**. 2008.

OLIVEIRA, M. N. de; BEZERRA, R. . **Cultura do Milho**. XIII JORNADA DE ENSINO, PESQUISA E EXTENSÃO – JEPEX 2013 – UFRPE: Recife, 09 a 13 de dezembro.

PATERNIANI, M. E. A. G. Z.; GUIMARÃES, P. de S.; LÜDERS, R. R.; GALLO, P. B.; SOUZA, A. P. de; LABORDA, P. R; OLIVEIRA, K. M. . **Capacidade combinatória, divergência genética entre linhagens de milho e correlação com heterose**. 2008.

PRASANNA, B.. BABU, R. NAIR, S.. SEMAGN, K.. CHIKAM, V. CAIRNS, J. (2014). **Molecular marker-assisted breeding for tropical maize improvement**. In: Ramakrishna et al (eds) Genetics, genomics and breeding of maize. CRC, New York, pp 89–119.

PRASANNA, B.M.. PIXLEY, K.. WARBURTON, M.L.. XIE, C.X.. 2010. **Molecular marker-assisted breeding options for maize improvement in Asia**. Mol Breed 26:339–356 Prasanna B, Babu R, Nair.

RAFFAN, E. . SEMPLE, R. K. **Next generation sequencing—implications for clinical practice.** 2011.

RANUM, P.. PEÑA – ROSAS, J. P.o. GARCIA – CASAL, M. N.. **Global maize production, utilization, and consumption.** 2014.

REGAZZI, A.J. **Análise multivariada, notas de aula INF 766.** Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.

SCHUSTER, I.; CRUZ, C. D.. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados.** 2013.

SIBOV, S. T.. JR, C. L. S.. GARCIA, A. A. F.. GARCIA, A. F.. SILVA, A. R.. MANGOLIN, C. A.. BENCHIMOL, L. L.. SOUZA, A. P. . **Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers.**
1. Map construction and localization of loci showing distorted segregation. 2003.

SILVA, A. F. . SCHONINGER, E. L.. CAIONE, G.. KUFFEL, C.. CARVALHO, M. A. C.. **Produtividade de híbridos de milho em função do espaçamento e da população de plantas em sistema de plantio convencional.** 2014.

SOLFERINI, O. B. . **Milho alto óleo – parâmetros genéticos, correlações entre modelos fixo e misto e seleção.** Universidade Estadual Paulista “Júlio de Mesquita Filho” Faculdade de Ciências Agrárias e Veterinárias. Campus de Jaboticabal. 2010.

SOUZA, A. R. R. **Potencial de Ganho Genético em Raça Local de Milho-Branco.** Universidade Estadual do Norte Fluminense Darcy Ribeiro. Campos dos Goytacazes. Rio de Janeiro – Brasil. Dezembro, 2016.

TESSELE, A. . **Seleção genômica ampla para escolha de genitores de soja e predição do desempenho de populações híbridas**. 2017.

VASCONCELOS, E. S.; CRUZ, C. D.; BHERING, L. L.; RESENDE JÚNIOR, M. F. R. **Método alternativo para análise de agrupamento**. *Pesq. agropec. bras.*, v.42, n.10, p.1421-1428, 2007.

WATSON, J.. D. BAKER, T. A.. BELL, S. P.. GANN, A.. LEVINE, M.. LOSICK, R.. **Biologia Molecular do gene**. Capítulo 2: Os ácidos nucleicos transportam as informações genéticas. 7ª edição. 2015.

ZORZATTO, C.. MACHADO, J. P. B.. LOPES, K. V.G. . NASCIMENTO, K. J. T.. PEREIRA, W. A.. BRUSTOLINI, O.J.B.. REIS, P. A. B.. CALIL, I. P.. DEGUCHI, M.. MARTINS, G. S.. GOUVEIA, B. C.. LORIATO, V. A. P.. SILVA, Marcos A. C. SILVA, Fabyano F. SANTOS, Anésia A. CHORY, J.. FONTES, E. P. B.. NIK1-mediated translation suppression functions as a plant antiviral immunity mechanism. **Nature**, volume 520, páginas 679–682. 2015. Disponível em: <<https://www.nature.com/articles/nature14171>>. Acesso em 30 de março de 2018.