

**PREDIÇÃO DO DESEMPENHO DE
HÍBRIDOS DE MILHO UTILIZANDO
INFORMAÇÕES DE SIMILARIDADE NO
ESTADO E PARENTESCO**

DENYS VITOR FERREIRA

2010

DENYS VITOR FERREIRA

**PREDIÇÃO DO DESEMPENHO DE HÍBRIDOS DE MILHO
UTILIZANDO INFORMAÇÕES DE SIMILARIDADE NO ESTADO E
PARENTESCO**

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

Orientador
Prof. Renzo Garcia Von Pinho

LAVRAS
MINAS GERAIS - BRASIL
2010

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

Ferreira, Denys Vitor.

Predição do desempenho de híbridos de milho utilizando informações de similaridade no estado e parentesco / Denys Vitor Ferreira. – Lavras: UFLA, 2010.

76 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2010.

Orientador: Renzo Garcia Von Pinho.

Bibliografia.

1. BLUP. 2. Melhor preditor linear não viesado. 3. Parentesco por marcadores moleculares. 4. Similaridade genética. 5. Modelos lineares mistos. I. Universidade Federal de Lavras. II. Título.

CDD – 633.1523

DENYS VITOR FERREIRA

**PREDIÇÃO DO DESEMPENHO DE HÍBRIDOS DE MILHO
UTILIZANDO INFORMAÇÕES DE SIMILARIDADE NO ESTADO E
PARENTESCO**

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

APROVADA em 26 de março de 2010

Prof. José Airton Rodrigues Nunes

UFLA

Prof. Renato Ribeiro de Lima

UFLA

Prof. Renzo Garcia Von Pinho
UFLA
(Orientador)

LAVRAS
MINAS GERAIS – BRASIL

A Deus,

Por estar sempre comigo iluminando os meus caminhos e dando-me serenidade nos momentos mais difíceis da minha vida.

OFEREÇO

Aos meus queridos pais Armando e Isa, que sempre me incentivaram a lutar pelos meus objetivos e pelo amor incondicional durante toda a minha vida; aos meus queridos irmãos Mauro Cesar, João e Aline, pelo apoio e incentivo durante toda a minha caminhada.

DEDICO

AGRADECIMENTOS

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, pela oportunidade de realização desse trabalho.

Aos meus queridos pais, Armando e Isa, pelo amor e apoio durante toda minha vida.

Aos meus queridos irmãos, Mauro Cesar, João e Aline, pelo incentivo nas horas difíceis.

À minha esposa, Adrienne, pelo incentivo, companheirismo e compreensão durante essa caminhada.

Ao professor Renzo Garcia Von Pinho, pela valiosa orientação, amizade e confiança depositada na condução desse trabalho.

Aos professores do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, César Brasil, Elaine Aparecida, João Bosco, João Cândido, José Airton e Magno Ramalho, pelos ensinamentos transmitidos ao longo de todo o curso.

Aos professores José Airton Rodrigues Nunes e Renato Ribeiro de Lima, pelas valiosas sugestões apresentadas.

Aos funcionários do Departamento de Biologia, Heloisa, Erondina, Elaine, Léo, Lindolfo e, em especial, ao Lamartine, pelo auxílio durante as análises laboratoriais.

Aos companheiros de república, Paulo Eduardo, Emiliano e Breguete, pela amizade e companheirismo. Valeu a pena.

Aos colegas do GEN, em especial, Márcio, Rogério, Jerônimo, Emiliano, Guilherme Abreu, Guilherme Ribeiro, Fernandão, Fernando Guedes, Karla, Constantino e Igor. A todos vocês, **MUITO OBRIGADO**.

SUMÁRIO

RESUMO GERAL	i
ABSTRACT GENERAL	iii
CAPÍTULO 1.....	4
1 Introdução geral	4
2 Referencial teórico.....	6
2.1 Marcadores moleculares e o melhoramento de plantas	6
2.2 Marcadores microssatélites	9
2.3 Coeficientes de similaridade e parentesco	11
2.4 Uso da metodologia de modelos lineares mistos com informação de marcadores moleculares para predição de híbridos de milho	15
3 Referências bibliográficas.....	22
CAPÍTULO 2: Predição do desempenho de híbridos de milho utilizando informações de similaridade no estado e parentesco	28
1 Resumo	28
2 Abstract.....	30
3 Introdução	31
4 Materiais e métodos	33
4.1 Descrição dos materiais	33
4.2 Delineamento experimental e correção das médias via modelos lineares mistos.....	34
4.2.1 Análise dialética via modelos lineares mistos	34
4.3 Predição de híbridos não Avaliados.....	36
4.3.1 Comparação das predições obtidas pelos coeficientes de similaridade no estado e parentesco	37
5 Resultados.....	39
6 Discussões.....	45
7 Referências bibliográficas.....	50
Anexos	54

RESUMO GERAL

FERREIRA, Denys Vitor. **Predição do desempenho de híbridos de milho utilizando informações de similaridade no estado e parentesco**. 2010. 76p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras, Lavras, MG.*

Em programas de melhoramento de milho a identificação das melhores combinações híbridas demanda muito esforço por parte dos melhoristas e recursos financeiros para realização dos cruzamentos e avaliação dos híbridos que possam apresentar desempenhos satisfatórios. Nesse contexto, algumas ferramentas têm surgido e podem auxiliar o trabalho dos melhoristas. Dentre estas, informações de similaridade no estado e parentesco entre genitores obtidos pelo uso de marcadores moleculares podem ser utilizadas no melhor preditor linear não viesado (BLUP) como método de predição do desempenho de híbridos não avaliados. Diante disso, este estudo foi realizado com o objetivo de avaliar a eficiência do melhor preditor linear não viesado (BLUP) e a influência do uso da similaridade no estado e do parentesco nas predições de híbridos não testados de milho. Foram utilizadas nove linhagens de milho que foram genotipadas com 48 marcadores microssatélites (SSR) associados a QTL's para produtividade de grãos. Estas linhagens foram inter cruzadas em esquema dialelo completo, com intuito de se obter os híbridos simples para estudos de correlações. Os resultados de similaridade no estado e parentesco foram utilizados para a construção das matrizes de relacionamento genético aditivo e de dominância que, posteriormente, foram empregadas no BLUP para predição dos valores genotípicos e de capacidade específica de combinação (CEC) dos híbridos não avaliados. Para isso, foram simulados cinco graus de desbalanceamento (3, 6, 9, 12 e 15 híbridos retirados). A similaridade no estado apresentou desempenho superior quanto aos valores de correlação e eficiência de seleção dos melhores híbridos em comparação ao parentesco. Isso demonstra que a similaridade no estado foi mais eficiente para o ranqueamento dos híbridos, demonstrando apresentar aplicação mais prática ao melhorista. Quanto ao parentesco, este apresentou menores valores para a soma do erro de predição (PRES), demonstrando que as predições obtidas apresentaram-se mais próximas do valor observado, ou seja, foram mais acuradas que aquelas obtidas pela similaridade no estado. A diferença no ranqueamento dos híbridos pode ser causada pelos baixos valores de parentesco detectados entre os genitores, o que permite inferir que, provavelmente estes sejam provenientes de diferentes

* Comitê Orientador: Renzo Garcia Von Pinho – UFLA (Orientador) e Márcio Balestre – UFLA.

populações. Nesse sentido, embora apresente um viés devido à superestimação do parentesco, a similaridade no estado pode ser utilizada quando forem detectados baixos valores de parentesco entre o conjunto de genitores cujo desempenho em combinações híbridas deseja-se prever.

ABSTRACT GENERAL

FERREIRA, Denys Vitor. **Prediction of maize hybrid performance using the information on the alikeness in state and coancestry among relatives**. 2010. 76p. Dissertation (Master's Degree in Genetics and Plant Breeding)-Universidade Federal de Lavras, Lavras, MG.*

The identification of better hybrid combinations demands lots of efforts from breeders and financial resources for the realization of the crosses and evaluation of the hybrid combinations that may have satisfactory performance on maize breeding programs. In this context, some tools have been arising and can help the breeders work. Among these, information on the state of similarity and relationship between relatives obtained by use of molecular markers can be used in the best linear unbiased predictor (BLUP) as a method of non evaluated hybrid prediction. The aim of this study was to evaluate the BLUP efficiency and the influence of the use of information on the state of similarity and relationship between relatives on non evaluated hybrid prediction. It was used nine inbred lines which were genotyped with 48 microsatellite markers (SSR) that were associated to QTLs for grain productivity. These inbred lines were crossed in a complete diallel scheme, and single hybrids were obtained for the correlation studies. Information on the alikeness in state and coancestry among relatives were used to construct the additive and dominance genetic matrices that were used for the prediction of the genotypic values on BLUP and the non evaluated hybrids specific combining ability (SCA). Five unbalanced degrees were simulated (3, 6, 9, 12 and 15 withdraw hybrids). The information on the alikeness in state compared to coancestry among relatives had a superior performance on the correlation of the better hybrids and selection efficiency, demonstrating that it was more efficient on hybrid ranking, besides it would have a more practical application for breeders. The relationship between relatives showed lower values for PRES, demonstrating that the obtained predictions are closer to the observed value, showing that they were more accurate than the one obtained from the information on the state of similarity. The hybrid ranking difference might be due to the lowest values of similarity detected among the inbred lines; which leads to the inference that they are from different genetic backgrounds. In this context, although it has a bias due to overestimation of the relationship, the similarity in the state can be used when it is detected lower values of relationship between the parents which hybrid performance one wants to predict.

* Guidance Committee: Renzo Garcia Von Pinho (Major professor) and Márcio Balestre (UFLA).

CAPÍTULO 1

1 INTRODUÇÃO GERAL

Em programas de melhoramento de milho ,a seleção e a avaliação dos melhores genótipos é extremamente importante para obtenção de híbridos cada vez mais produtivos e adaptados às mais diferentes condições de cultivo.

Algumas ferramentas podem auxiliar os melhoristas no intuito de direcionar os cruzamentos visando à obtenção das melhores combinações híbridas. Nesse sentido, muitas pesquisas têm concentrado esforços em estimar o desempenho dos híbridos através da determinação dos grupos heteróticos das linhagens, para direcionar os cruzamentos baseando-se nas distâncias genéticas (Reif et al., 2003b; Amorim, 2005; Guimarães, 2007). Pela teoria da genética quantitativa, os pares que são mais divergentes poderiam maximizar a heterose e revelar as melhores combinações híbridas (Melchinger, 1999).

Nesse contexto, o uso de marcadores moleculares na determinação de grupos heteróticos e na predição do desempenho de híbridos de milho tem sido proposto por vários autores (Reif et al., 2003a; Amorim, 2005). Contudo, até o momento, os resultados são ainda inconsistentes (Dias et al., 2004; Guimarães, 2007).

Dessa forma, uma alternativa para o uso de informações obtidas por marcadores moleculares seria a utilização dessas informações através da metodologia de modelos lineares mistos para a predição de cruzamentos não realizados (Bernardo, 1994; Charcosset et al., 1998). Para isso, torna-se necessário o conhecimento da genealogia ou do parentesco entre os indivíduos avaliados e indivíduos cujo desempenho deseja-se prever (Bernardo, 1994). Porém, nem sempre o melhorista dispõe de tal informação, o que poderia limitar a utilização dessa técnica.

Com intuito de contornar essa limitação, alguns autores sugerem a utilização da informação de marcadores moleculares (similaridade alélica no estado) como opção quando não se dispõe de dados de genealogias (Bernardo, 1994, 1995; André, 1999; Balestre et al., 2008). Segundo Nejati-Javeremi et al. (1997), o emprego da similaridade no estado possui a vantagem de utilizar toda a semelhança alélica existente entre os indivíduos avaliados.

No entanto, Bernardo (1993) propõe o uso do coeficiente de coancestria, que utiliza informação da similaridade genética, no intuito de corrigir possíveis vieses da similaridade no estado. Entretanto, essa metodologia somente é válida para dialelos parciais nos quais não ocorrem cruzamentos endogâmicos, ou seja, quando se possui informação prévia a respeito de grupos heteróticos.

Como alternativa a essa restrição, alguns autores (Queller & Goodnight, 1989; Ritland, 1996; Lynch & Ritland, 1999; Wang, 2002) têm sugerido procedimentos para estimação conjunta do parentesco por meio marcadores moleculares, que pode ser utilizado em dialelos completos e parciais. Esses estimadores, mesmo utilizando dados moleculares, consideram a probabilidade condicional de que dois indivíduos possuam alelos herdados de genitores em comum. Portanto, somente indivíduos que atendam a essa premissa participam da predição de híbridos não avaliados. Dessa forma, aqueles indivíduos que possuem parentesco igual a zero não contribuirão na predição, concorrendo para resultados mais coerentes.

Por disso, o objetivo desse estudo foi avaliar a eficiência do melhor preditor linear não viesado (BLUP) com a utilização de informações de similaridade no estado e parentesco entre genitores visando à obtenção de estimativas de predições de híbridos de milho não testados.

2 REFERENCIAL TEÓRICO

2.1 Marcadores moleculares e o melhoramento de plantas

O melhoramento genético de plantas tem proporcionado ganhos em produtividade e incrementos em adaptabilidade nas mais diversas culturas utilizadas pelo homem para sua sobrevivência. Fehr (1987) define o melhoramento como sendo a ciência e a arte de se fazer o melhoramento genético de plantas.

Esta ciência teve seu início com a domesticação de algumas espécies de plantas há mais de 10.000 anos e se intensificou com o aumento da demanda populacional por alimentos. Com o crescimento da população mundial, produzir alimentos em quantidade e qualidade necessárias exigirá muito empenho por parte de pesquisadores e da sociedade de modo geral.

Nesse sentido, o melhoramento necessita de ferramentas que possam auxiliá-lo na obtenção de cultivares cada vez mais produtivas e adaptadas às diferentes condições edafoclimáticas de cultivo.

Mais recentemente, no início da década de 80 do século XX, com o advento de novas técnicas de Biologia Molecular, o uso de marcadores moleculares tem se tornado extremamente útil dentro dos programas de melhoramento de plantas. Estes têm sido utilizados na identificação de genes de resistência a doenças e, ou, insetos-praga; na avaliação e caracterização de germoplasma; no melhoramento de parentais de híbridos; na introgressão gênica e seleção assistida por marcadores; no desenvolvimento de mapas genéticos; na determinação de grupos heteróticos e associação com regiões genômicas que afetam heterose; na reconstituição de *pedigrees*; nos testes de pureza genética; na seleção de resistência a patógenos exóticos ainda inexistentes em determinada região; na associação com caracteres quantitativos; nos estudos de interação genótipos x ambientes; nos processos

legais, entre outros (Rafalski & Tingey, 1993; Ferreira & Grattapaglia, 1996; Milach, 1998a,b).

A detecção de variabilidade genética diretamente no DNA permite encontrar diferenças que podem ser utilizadas em estudos detalhados de divergência entre indivíduos e substituem, com vantagens, medidas baseadas em genealogias.

Um marcador molecular de DNA é tipicamente uma pequena região do DNA mostrando polimorfismo (diferenças) entre indivíduos (Liu, 1998). Essas diferenças podem ser detectadas por hibridação de uma sequência do DNA marcada radioativamente e hibridizada com o DNA dos indivíduos em análise, usando o princípio da complementariedade da molécula do DNA. Um exemplo desse tipo de marcador é o RFLP.

Outros marcadores baseiam-se na amplificação de sequências específicas no DNA, através do emprego de reações para o aumento dessas sequências específicas. É o caso dos marcadores moleculares do tipo Microssatélites (SSR), o RAPD (“*Random Amplified Polymorphic DNA*”) e o AFLP (“*Amplified Fragment Length Polymorphism*”) (Liu, 1998).

O detalhamento de cada um dos tipos de marcadores acima citados pode ser encontrado em Ferreira & Grattapaglia (1996) e Milach (1998a).

O princípio da análise de marcadores consiste na utilização de uma matriz de valores binários, em que as colunas dessa matriz representam os diferentes genótipos e as linhas representam os diferentes locos a serem comparados. Cada linha corresponde à identificação de uma marca numa região específica do DNA. A ausência ou presença dessa marca é identificada pelo número 0 ou 1, respectivamente.

Por exemplo, ${}_m M_g = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 1 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 1 & \dots & 0 \end{bmatrix}$

em que m é o número de marcas e g é o número de genótipos.

Para a comparação de dois genitores ocorrem as seguintes situações:

- i) presença de bandas em ambos os indivíduos (1, 1);
- ii) presença e ausência de bandas nos indivíduos e vice-versa (1, 0 ou 0, 1);
- iii) ausência de bandas em ambos os indivíduos (0, 0).

Baseado nessa matriz de valores binários e com o uso de alguns coeficientes é possível estabelecer a similaridade genética entre os indivíduos e conseqüentemente, determinar grupos heteróticos e direcionar os cruzamentos.

Da mesma forma, a determinação do parentesco entre genitores com o uso de marcadores moleculares também tem apresentado grande utilidade em milho, tornando-se uma alternativa quando não se dispõe de informação de *pedigree* dos genitores (Bernardo, 1994). Essa informação pode ser utilizada no melhor preditor linear não viesado (BLUP) como método de predição do desempenho de híbridos não avaliados (Bernardo, 1993, 1994, 1995, 1996a,b; Balestre, 2009). Esses autores demonstram que dados de parentesco por meio de marcadores moleculares tornaram-se uma ferramenta extremamente útil aos programas de melhoramento, uma vez que permitem redução nos custos com cruzamentos e avaliação de híbridos comerciais que possam vir a não apresentar desempenhos satisfatórios.

2.2 Marcadores microssatélites

Microssatélites ou SSR, do inglês *Simple Sequence Repeats*, são sequências de DNA com dois a seis pares de bases de comprimento repetidas em tandem. Os primeiros relatos da sua existência foram descritos no genoma humano, no qual estão presentes em um grande número de cópias (Litt & Luty, 1989).

No genoma vegetal, os microssatélites são largamente distribuídos com uma frequência de 1 a cada 50 mil pares de bases, sendo o dinucleotídeo AT o mais comum (Morganti & Olivieri, 1993).

O número de repetições em tandem de um microssatélite geralmente é variável, enquanto a sequência de bases adjacentes ao microssatélite pode ser única no genoma e conservada entre diferentes indivíduos da mesma espécie. Sendo assim, pode-se desenhar *primers* específicos para as sequências adjacentes a um dado microssatélite, de forma que, por meio de uma reação de PCR (*“Polymerase Chain Reaction”*), seja possível amplificar este loco em diferentes genótipos. Como o número de unidades repetidas em tandem em um microssatélite pode ser variável entre diferentes genótipos, os produtos de amplificação dos diferentes indivíduos exibirão polimorfismo no tamanho do fragmento amplificado, o que é a base da genotipagem com microssatélites.

Não existe uma explicação definitiva sobre a variação encontrada no número de unidades repetidas. Acredita-se que ela seja causada por falhas na recombinação do DNA ou no processo de replicação. Esta variação leva à existência de vários alelos em uma população, podendo o nível de heterozigidade ser extremamente alto. Aliado a isso, os SSR possuem herança mendeliana e comportamento codominante, ou seja, permitem a identificação de genótipos homozigóticos e heterozigóticos na população. Além disso, já existe um grande conjunto de microssatélites disponíveis para a utilização em milho, muitos deles identificados como sendo associados à QTL's para produtividade de grãos (Sibov et al., 2003).

Existem também conjuntos de microssatélites disponíveis para a utilização em estudos genéticos que podem ser acessados por meio do *Maize Genome Database Page* (www.maizedp.org/SSR).

O desenvolvimento de marcadores microssatélites requer alto investimento inicial, sendo realizado, na maioria das vezes, por instituições públicas como, por exemplo, a Universidade de Columbia/Missouri, EUA (*primers* UMC) ou instituições privadas, como o *Brookhaven National Laboratory* (*primers* BNLG). No entanto, a partir da obtenção de um conjunto de SSR, a sua aplicabilidade se torna extremamente viável.

Estudos realizados com linhagens de milho obtidas das populações BR-105 e BR-106, correlacionando produtividade e divergência genética estimada através de marcadores moleculares do tipo RFLP, RAPD, AFLP e SSR (Lanza et al., 1997; Benchimol et al., 2000) indicaram que os melhores marcadores para alocar linhagens em grupos heteróticos foram os SSR e AFLP, sendo ambos complementares.

Bernardo et al. (2000), em estudos visando a obter estimativas sobre o parentesco entre linhagens de milho, concluíram que os marcadores microssatélites são superiores aos marcadores RFLP para a obtenção dessas estimativas de parentesco.

Alguns resultados utilizando marcadores microssatélites não associados a QTL's, ou seja, marcadores que não estão associados à característica alvo da seleção, apresentam resultados não consistentes entre os valores fenotípicos e suas distâncias genéticas. No entanto, o uso desses marcadores associados a QTL's tem apresentado resultados promissores em estudos de divergência genética (Amorim, 2005; Balestre et al., 2008a).

2.3 Coeficientes de similaridade e parentesco

Os coeficientes de similaridade são medidas que quantificam o quanto dois indivíduos são parecidos. Esses coeficientes podem ser divididos em duas categorias: medidas de similaridade e de dissimilaridade. Para a primeira, quanto maior o valor observado, mais parecidos são os indivíduos; para a segunda, quanto maior o valor observado, menos parecidos são os indivíduos (Meyer, 2002). Para cada tipo de variável (quantitativa, qualitativa, ordinais, qualitativas nominais e mistas) são definidos diferentes coeficientes de similaridade (Sneath & Sokal, 1973; Johnson & Wichern, 1988; Bussab et al., 1990).

De modo geral, os coeficientes de similaridade baseiam-se em comparações entre a ocorrência de bandas comuns (indicado por aquelas bandas em comum em uma matriz de dados) e bandas diferentes (indicado por um e zero ou zero e um). Alguns coeficientes também consideram a ocorrência de zeros em comum. Esses coeficientes apresentam valores que comumente variam entre zero e um (Skroch et al., 1992).

No melhoramento de plantas, esses coeficientes têm sido utilizados para a identificação de populações que possibilitam maior efeito heterótico quando cruzadas ou que possibilitam maior probabilidade de seleção de genótipos superiores nas gerações segregantes. Além disso, têm sido úteis na avaliação de bancos de germoplasma, no estabelecimento de relações entre a diversidade genética e geográfica e também para evitar a vulnerabilidade genética das culturas.

Para esses estudos que utilizam coeficientes de similaridade, normalmente empregam-se diferentes coeficientes de marcadores dominantes, como o RAPD e RFLP, bem como aqueles multialélicos, como os microsatélites. Alguns destes coeficientes são específicos para as variáveis dicotômicas, e seu uso é sugerido para estudos de divergência com base em marcadores moleculares (Duarte et al., 1999).

Segundo Reif et al. (2005), a escolha de um coeficiente para estudos de divergência depende das propriedades do marcador envolvido, da genealogia do germoplasma, dos objetivos do estudo e das condições necessárias para a análise multivariada. De acordo com esses autores, muitos estudos ignoram essas questões, especialmente aquelas relacionadas com as propriedades do coeficiente, que estão intimamente ligados ao objetivo do estudo e às condições para realização de análise multivariada.

Amorim (2005) trabalhou com marcadores microssatélites com o objetivo de estimar o desempenho de linhagens de milho em cruzamentos, baseando-se em distâncias genéticas. Esse autor utilizou linhagens parcialmente endogâmicas ($S_{0:1}$) e as distâncias genéticas foram estimadas pela Distância Modificada de Rogers. Houve alta correlação entre distância genética e desempenho dos híbridos resultantes dos cruzamentos.

Em estudo semelhante, trabalhando com marcadores AFLP e microssatélites, Guimarães (2007) utilizou o coeficiente de similaridade de Jaccard para estimar a divergência genética entre linhagens de milho. Essa autora encontrou baixa correlação entre distância genética e desempenho de híbridos simples resultantes dos cruzamentos. Esse fato demonstra que para os mesmos objetivos e com utilização dos mesmos marcadores, não há uma explicação simples para a escolha dos coeficientes de similaridade genética.

Por isso, é importante estudar os coeficientes de similaridade no estado rotineiramente utilizados, visto que, os resultados de agrupamento de genótipos pode ser afetado pela escolha do coeficiente de similaridade (Jackson et al., 1989; Duarte et al., 1999). Desse modo, as mudanças no ranqueamento dos genótipos pode alterar os valores de correlação em estudo.

Balestre et al. (2008a) compararam sete diferentes coeficientes de similaridade e um coeficiente de dissimilaridade genética, utilizando progênies parcialmente endogâmicas ($S_{0:1}$) derivadas de híbridos simples de milho.

Observaram que todos os coeficientes foram eficazes na distinção dos três diferentes grupos heteróticos. No entanto, dentro de cada grupo, houve grande variação nos resultados de agrupamento dos genótipos de um coeficiente para outro. Essa observação se deu através do índice de consenso proposto por alguns autores (Duarte et al., 1999; Meyer et al., 2004). Esse índice compara os dendrogramas de uma forma mais refinada, uma vez que utiliza a proporção de subgrupos em comum entre os dendrogramas (Duarte et al., 1999; Meyer et al., 2004). A partir daí, os autores concluíram que a escolha dos coeficientes de similaridade de Jaccard, Rogers e Tanimoto e Russel e Rao para estudos de divergência genética e predição de híbridos não avaliados merecem atenção especial, uma vez que apresentaram elevados valores de estresse (Kruskal, 1964). Emygdio et al. (2003) também observaram que os coeficientes de Rogers e Tanimoto e Russel e Rao apresentaram maiores estimativas dos valores de estresse no agrupamento de genótipos de feijão comum (61 e 78%, respectivamente). Resultados semelhantes também foram encontrados por Duarte et al. (1999), demonstrando que esses coeficientes não deveriam ser utilizados. A Distância Modificada de Rogers (DMR) apresentou um agrupamento completamente diferente de todos os outros, demonstrando que, para valores de dissimilaridade baixa, há uma mudança no agrupamento dos genótipos (Duarte et al., 1999; Meyer et al., 2004).

Há evidências de que a utilização de dados de similaridade genética apresenta vantagem na acurácia seletiva quando não se dispõe, ou não se utiliza qualquer informação de parentesco (Balestre et al., 2008a).

No entanto, conforme descrito anteriormente, os trabalhos realizados com os coeficientes de similaridade e dissimilaridade utilizados (Jaccard, Simple Matching, DMR, Sorensen-Dice, entre outros) fornecem estimativas viesadas do parentesco, por considerarem apenas a similaridade no estado (evolutiva), quando na verdade, a metodologia dos modelos mistos foi desenvolvida para

utilização do parentesco, baseando-se na probabilidade de genótipos compartilharem alelos herdados de genitores em comum (Henderson, 1984).

A similaridade no estado pode superestimar o parentesco entre indivíduos, podendo detectar uma similaridade entre genitores quando, na realidade, estes não apresentam relação genética ou parentesco. Logo, não é uma medida “confiável” da relação genética entre genitores e, portanto, seu uso deve ser analisado de acordo com a finalidade do estudo.

Por outro lado, segundo Falconer & Mackay (1996) o parentesco é uma medida clássica da relação genética entre linhagens. É a probabilidade de que, em determinado *locus*, duas linhagens tenham alelos que são idênticos por descendência, ou seja, que sejam cópias de um mesmo alelo ancestral. Dessa forma, o coeficiente de parentesco (r_{xy}) entre dois indivíduos (X e Y) é definido como a probabilidade de que um alelo, tomado ao acaso do indivíduo X seja idêntico por descendência ao alelo do indivíduo Y (Falconer & Mackay, 1996; Lynch & Walsh, 1998).

Diante disso, o uso de informações de parentesco entre os genitores deve ser preterido em estudos que visam à predição do desempenho de híbridos não avaliados ou aqueles cujas combinações híbridas foram perdidas.

Balestre (2009) detectou que a utilização da similaridade no estado possui algumas desvantagens para predição dos valores genotípicos, principalmente por recuperar informação de genótipos não-aparentados, causando perda de acurácia de predição dos valores genéticos, quando comparado a estimadores que eliminam esse viés. Nesse estudo, para a predição dos valores genéticos e desempenho de híbridos não avaliados, o autor utilizou a Distância Modificada de Rogers (DMR) e o coeficiente de parentesco proposto por Lynch & Ritland (1999) (LR). Nesse trabalho, foi possível verificar que, em todas as análises comparativas, o uso da DMR apresentou resultados inferiores ao coeficiente de LR. Além dos menores valores de correlações e maior desvio

padrão, a DMR apresentou maior erro padrão de predição, demonstrando que as predições obtidas por meio desse coeficiente foram menos estáveis em relação ao coeficiente de LR. De modo semelhante, comparando-se os valores genotípicos observados e os valores de capacidade específica de combinação (CEC), constatou-se que o uso da DMR acarretou em valores de predição mais distantes do valor observado. Essa menor eficiência também foi ressaltada considerando a eficiência de seleção das melhores combinações híbridas.

Portanto, com intuito de contornar esses efeitos da similaridade no estado, alguns coeficientes têm sido propostos para se obter as estimativas do parentesco por meio da utilização de informação de marcadores moleculares (Queller & Goodnight, 1989; Ritland, 1996; Lynch & Ritland, 1999; Wang, 2002).

2.4 Uso da metodologia de modelos lineares mistos com informação de marcadores moleculares para predição de híbridos de milho

Vários trabalhos têm sido realizados utilizando informações de marcadores moleculares com o objetivo de prever o desempenho de híbridos por meio de genotipagem dos seus genitores. De maneira geral, os resultados encontrados na literatura são bastante inconsistentes (Amorim, 2005; Guimarães, 2007; Paterniani et al., 2008).

No entanto, o emprego de marcadores moleculares tem-se mostrado muito importante na avaliação da diversidade genética entre linhagens e na determinação de grupos heteróticos entre elas. Contudo, a correlação entre divergência genética e o comportamento do híbrido não é consistente, o que mostra que análises de divergência genética nem sempre são úteis para prever o desempenho dos híbridos obtidos (Barbosa et al., 2003).

O uso de metodologias adequadas para a predição do desempenho de híbridos de milho por meio da determinação de parentesco entre linhagens com

o uso da informação de marcadores, torna-se uma ferramenta extremamente útil dentro de programas de melhoramento visando à produção de híbridos comerciais de milho. Segundo Bernardo (1994, 1995) a utilização do melhor preditor linear não viesado (BLUP) no melhoramento genético vegetal consiste em prever o desempenho genético de híbridos que não estejam presentes ou que tenham sido perdidos. Esse autor considera relevante o uso de técnicas preditivas como uma alternativa viável para diminuir os custos e também o tempo de avaliação de híbridos pela diminuição substancial do número de híbridos a serem avaliados.

A metodologia dos modelos mistos foi inicialmente proposta por Henderson (1949), e consiste na predição de valores genéticos tratados como efeitos aleatórios corrigidos para os demais efeitos fixos contidos no modelo. Por meio dessa metodologia pode-se obter o BLUP dos valores genéticos de cada indivíduo, além do melhor estimador linear não viesado (BLUE) dos efeitos fixos (Henderson, 1974, 1975).

As propriedades do BLUP, descritas por Henderson (1984) evidenciam sua superioridade teórica como preditor de valores aleatórios. Essa superioridade tem sido evidenciada principalmente através do uso de simulações (Bueno Filho, 1997; André, 1999).

White & Hodge (1989) apresentam várias propriedades úteis do BLUP quando utilizado no melhoramento genético vegetal, dentre as quais se destacam:

a) a estimação e predição não viesada através de um procedimento único (BLUE e BLUP);

b) a possibilidade de se levar em conta os efeitos de seleção e endogamia ao longo das gerações desde que o grau de parentesco entre os indivíduos avaliados seja conhecido;

c) a correlação entre os valores genéticos verdadeiros e os preditos é máxima entre as classes de preditores lineares não viesados;

d) a possibilidade de se prever o valor genético de indivíduos observados ou não;

e) as variâncias e os erros do BLUP são menores em relação aos outros métodos.

A predição dos valores genéticos requer o conhecimento prévio dos componentes de variância e covariâncias envolvidos. Existem vários métodos de estimação de componentes de variância que podem ser utilizados (Searle et al., 1992; Lynch & Walsh, 1998; Resende, 2002). Dentre estes podemos destacar o método da máxima verossimilhança restrita (REML) proposto por Patterson & Thompson (1971).

Dentre as principais características do BLUP está a utilização de informações de parentesco genético aditivo entre os pares de indivíduos. Os indivíduos mais aparentados têm uma proporção maior de alelos em comum, o que está relacionado ao grau de parentesco (por exemplo, o parentesco genético aditivo entre pai e filho é de 0,50, entre meios irmãos 0,25, e assim por diante). Esse parentesco é levado em consideração na análise (André, 1999), o que pode ser considerada uma vantagem desse método. O termo parentesco genético aditivo refere-se à proporção de alelos idênticos por descendência entre dois indivíduos, sendo equivalente ao dobro do valor do coeficiente de Malecot (Vleck, 1993b).

O BLUP possibilita, além da predição mais acurada do mérito genético, a inclusão do coeficiente de endogamia para todos os indivíduos avaliados, se toda a informação de parentesco for conhecida (Henderson, 1974, 1984). Quando se dispõe de pouca ou nenhuma informação sobre o indivíduo, informação de parentes pode contribuir para a predição do seu valor genético. A magnitude dessa contribuição é determinada pelo grau de parentesco entre os

indivíduos envolvidos na análise, sendo que, quanto maior o grau de relacionamento genético, maior será a contribuição (Wood et al., 1991; Vleck, 1993a; Panter & Allen, 1995).

Segundo Bernardo (1994, 1995), a partir de covariâncias genéticas entre as linhagens a serem cruzadas, pode-se prever o desempenho genotípico de alguns dos seus cruzamentos, sem que eles tenham sido obtidos e avaliados. Essa técnica tem grande potencial para auxiliar o melhoramento genético, devido ao elevado número de híbridos possíveis de serem obtidos com poucas linhagens. Apesar da correlação entre o desempenho genotípico do cruzamento predito e o seu valor fenotípico ser baixo em muitos casos, essa abordagem apresenta vantagens (Bernardo, 1994). Dessa forma, a utilização do BLUP tem mostrado eficiência considerável para tal finalidade (Bernardo, 1994, 1995, 1996a,b).

Bernardo (1994) comparou os resultados obtidos para o BLUP por meio de informações de parentesco de dois grupos de linhagens de milho. Para tal, empregou genealogias e dados de marcadores moleculares, com objetivo de avaliar a eficiência do BLUP em prever cruzamentos não realizados. Não houve alterações significativas nos resultados obtidos por dados de genealogia e dados de marcadores moleculares. Desta forma, a utilização da informação de marcadores moleculares seria uma alternativa viável, desde que não se disponha de informações confiáveis sobre a genealogia das linhagens avaliadas, uma situação muito comum na prática.

A habilidade do BLUP para predição de cruzamentos não realizados, bem como o número de preditores (cruzamentos a serem realizados) necessários para obter predições suficientemente confiáveis, depende do grau de parentesco dos pais utilizados como preditores (Bernardo, 1995). Nesse contexto, o emprego de marcadores moleculares surge como alternativa para obtenção desse grau de parentesco com maior confiabilidade.

Alguns trabalhos demonstraram a eficiência do BLUP em prever valores genéticos. Panter & Allen (1995) estudaram a eficiência dos métodos dos quadrados mínimos ordinários e do BLUP na identificação de linhagens superiores em cruzamentos de variedades de soja. Os autores simularam situações de balanceamento e desbalanceamento de dados e observaram que em todas as situações estudadas, o BLUP apresentou menor erro padrão, maiores valores de correlação entre os valores preditos e o desempenho dos cruzamentos avaliados em campo, bem como maior porcentagem de identificação de cruzamentos superiores dentre os realizados.

Bernardo (1994), usando a metodologia de modelos mistos, com estimativas de parentesco obtidas através de marcadores moleculares do tipo RFLP, para prever o desempenho de alguns cruzamentos não realizados em um grupo de 54 híbridos de milho provenientes de dois grupos de linhagens avaliadas em um dialelo parcial, obteve estimativas de correlação entre valor genotípico predito e o desempenho do cruzamento em campo para produção de grãos, variando de 0,65 a 0,80. A eficiência na predição de cruzamentos não realizados aumentou à medida que se empregou um maior número de cruzamentos como preditores. Isso demonstra a viabilidade da utilização dessa metodologia como forma de identificação de cruzamentos promissores antes de sua realização.

Bernardo (1996a,b), utilizando um conjunto de 4099 híbridos de 16 combinações entre nove grupos heteróticos de milho avaliados em até 15 locais dos Estados Unidos, Canadá, Itália e França, encontrou estimativas de correlação entre o valor predito e o desempenho fenotípico para a produção de grãos de híbridos de milho variando entre 0,49 a 0,76. Segundo o autor, estes valores podem ser considerados moderadamente elevados e sugerem que o melhor preditor linear não viesado seja empregado de forma eficiente para identificação de cruzamentos não realizados com desempenho superior.

André (1999), utilizando simulações e dados reais de um dialelo entre linhagens de feijão genotipadas com marcador RAPD (“*Random Amplified Polymorphic DNA*”), concluiu que o melhor preditor linear não viesado é mais eficiente que o método dos quadrados mínimos ordinários para predição das capacidades de combinação e do efeito de cruzamento, principalmente quando existem informações sobre a similaridade genética entre os materiais a serem cruzados.

Iemma (2003) comparou os resultados de análises dialélicas obtidas considerando o modelo fixo e modelo misto e avaliou a eficiência do BLUP em prever cruzamentos não realizados entre linhagens de milho, utilizando informações de marcadores RFLP para a estimação da matriz de parentesco. Verificou-se moderada correlação entre os valores dos BLUP’s para produção de grãos e os valores preditos a partir dos híbridos observados, indicando que essa metodologia pode ser utilizada quando há interesse em prever o desempenho de híbridos não avaliados.

Nesse contexto, Nejati-Javeremi et al. (1997) sugerem o emprego da similaridade no estado, pois segundo esses autores, esta possui a vantagem de se utilizar toda a semelhança alélica existente entre os indivíduos avaliados. No entanto, pelo fato de levar em consideração a co-ocorrência de bandas (0 e 1), e não a probabilidade de genitores possuírem alelos herdados de um genitor em comum, a similaridade no estado pode superestimar o parentesco.

Portanto, Bernardo (1993) propõe o uso de um coeficiente de coancestria que utiliza informação de similaridade genética, no intuito de corrigir possíveis vieses da similaridade no estado. Entretanto, essa metodologia somente é válida para dialelos parciais nos quais não ocorrem cruzamentos endogâmicos, podendo ser utilizada somente quando se tem informação prévia a respeito de grupos heteróticos. Como alternativa a essa restrição, alguns autores (Queller & Goodnight, 1989; Ritland, 1996; Lynch & Ritland, 1999; Wang,

2002) sugerem procedimentos não viesados para estimação conjunta de parentesco por meio de marcadores moleculares que podem ser utilizados em dialelos completos e parciais. Segundo Balestre (2009), o coeficiente de parentesco proposto por Lynch & Ritland (1999) pode ser utilizado em dialelos completos, pois as medidas de parentesco obtidas por esse estimador levam em consideração todos os indivíduos em avaliação. Além disso, mesmo utilizando dados moleculares, esse estimador de parentesco considera a probabilidade condicional de que dois indivíduos possuam alelos herdados de genitores em comum. Desse modo, somente indivíduos que atendam a essa premissa, participam da predição de híbridos não avaliados. Logo, aqueles indivíduos que possuem parentesco igual a zero não contribuirão na predição, concorrendo para resultados mais coerentes.

3 REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, E.P. **Produtividade de híbridos de milho derivados de populações S₀, e associação com distância genética baseada em microssatélites**. 2005. 97p. Tese (Doutorado em Genética e Melhoramento de Plantas)-Universidade Federal de Lavras, Lavras.

ANDRÉ, C.M.G. **Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores moleculares na análise dialética**. 1999. 101p. Dissertação (Mestrado em Genética e Melhoramento de Plantas)-Universidade Federal de Lavras, Lavras.

BALESTRE, M. **Predição do desempenho de híbridos de milho utilizando modelos lineares mistos com a informação de marcadores moleculares**. 2009. 64p. Dissertação (Mestrado em Genética e Melhoramento de Plantas)-Universidade Federal de Lavras, Lavras.

BALESTRE, M.; PINHO, R.G. von; SOUZA, J.C.; LIMA, J.L. Comparison of similarity and dissimilarity genetic coefficients based on microsatellites markers. **Genetics and Molecular Research**, Ribeirão Preto, v.7, n.4, p.695-705, jul./ago. 2008.

BARBOSA, A.M.M.; GERALDI, I.O.; BENCHIMOL, L.L.; GARCIA, A.A.F.; SOUZA, C.L.; SOUZA, A.P. Relationship of intra-and interpopulation tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. **Euphytica**, Netherlands, v.130, n.1, p.87-99, Mar. 2003.

BENCHIMOL, L.L.; SOUZA JUNIOR, C.L.; GARCIA, A.A.F.; KONO, P.M.S.; MANGOLIN, C.A.; BARBOSA, A.M.M.; COELHO, A.S.G.; SOUZA, A.P. Genetic diversity in tropical maize inbred lines: heterotic group assignment and hybrid performance determined by RFLP markers. **Plant Breeding**, Berlin, v.119, n.6, p.491-496, Dec. 2000.

BERNARDO, R. Estimation of coefficient of coancestry using molecular markers in maize. **Theoretical and Applied Genetics**, Berlin, v.85, n.8, p.1055-1062, Feb. 1993.

BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science**, Madison, v.34, n.1, p.20-25, Jan./Feb. 1994.

- BERNARDO, R. Genetic models for predicting maize performance in unbalanced yield trial data. **Crop Science**, Madison, v.35, n.1, p.141-147, Jan./Feb. 1995.
- BERNARDO, R. Best linear unbiased prediction of maize single cross performance. **Crop Science**, Madison, v.36, n.1, p.50-56, Jan./Feb. 1996a.
- BERNARDO, R. Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. **Crop Science**, Madison, v.36, n.4, p.862-866, Aug./Sept. 1996b.
- BERNARDO, R.; ROMERO-SEVERSON, J.; ZIEGLE, J.; HAUSER, J.; JOE, L.; HOOKSTRA, G.; DOERGE, R.W. Parental contribution and coefficient coancestry among maize inbreds: pedigree, RFLP, and SSR data. **Theoretical and Applied Genetics**, Berlin, v.100, n.3/4, p.552-556, Feb. 2000.
- BUENO FILHO, J.S. de S. **Modelos mistos na predição de valores genéticos aditivos em testes de progênies florestais**. 1997. 118p. Tese (Doutorado em Estatística e Experimentação Agronômica)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.
- BUSSAB, W. de; MIAZAKI, E.S.; ANDRADE, D.F. **Introdução à análise de agrupamentos**. São Paulo: Associação Brasileira de Estatística, 1990. 105p.
- CHARCOSSET, A.; BONNISSEAU, B.; TOUCHEBEUF, O.; BURSTIN, J.; DUBREUIL, P.; BARRIERE, Y.; GALLAIS, A.; DENIS, J.B. Prediction of maize hybrid silage performance using marker data: comparison of several models for specific combining ability. **Crop Science**, Madison, v.38, n.1, p.38-44, Feb. 1998.
- DIAS, L.A. dos S.; PICOLI, E.A. de T.; ROCHA, R.B.; ALFENAS, A.C. A priori choice of hybrid parents in plants. **Genetics and Molecular Research**, Ribeirão Preto, v.3, n.3, p.356-368, May/June 2004.
- DUARTE, J.M.; SANTOS, J.B. dos; MELO, L.C. Comparison of similarity coefficients based on RAPD markers in the common bean. **Genetic and Molecular Biology**, Ribeirão Preto, v.22, n.3, p.427-432, Sept. 1999.
- EMYGDIO, B.M.; ANTUNES, I.F.; CHOER, E.; NEDEL, J.L. Eficiência de coeficientes de similaridade em genótipos de feijão mediante marcadores RAPD. **Pesquisa Agropecuária Brasileira**, Brasília, v.38, n.2, p.243-250, fev. 2003.

FALCONER, D.S.; MACKAY, T.F. **Introduction to quantitative genetics**. 4.ed. Essex: Longman, 1996. 464p.

FEHR, W.R. **Principles of cultivar development**. New York: Mcmillian, 1987. 761p.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3.ed. Brasília: EMBRAPA-CENARGEN, 1996. 220p. (Documentos, 20).

GUIMARÃES, P. de S. **Desempenho de híbridos simples de milho (*Zea mays* L.) e correlação entre heterose e divergência genética entre linhagens parentais**. 2007. 111p. Dissertação (Mestrado em Melhoramento Vegetal)-Instituto Agrônômico de Campinas, Campinas.

HENDERSON, C.R. Estimates of changes in herd environment. **Journal of Dairy Science**, Champaign, v.32, p.706, 1949. Abstract.

HENDERSON, C.R. General flexibility of linear models for sire evaluation. **Journal of Dairy Sciences**, Champaign, v.57, n.8, p.963-972, Aug. 1974.

HENDERSON, C.R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Washington, v.31, n.2, p.423-447, Apr./June 1975.

HENDERSON, C.R. **Applications of linear models in animal breeding**. Ontario: University of Guelph, 1984. 462p.

IEMMA, M. **Uso do melhor preditor linear não viesado (BLUP) em análises dialélicas e predição de híbridos**. 2003. 81p. Dissertação (Mestrado em Estatística e Experimentação Agronômica)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

JACKSON, A.A.; SOMERS, K.M.; HARVEY, H.H. Similarity coefficients: measures for co-occurrence and association or simply measures of occurrence. **American Naturalist**, Chicago, v.133, n.3, p.436-453, Mar. 1989.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall, 1988. 607p.

KRUSKAL, J.B. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. **Psychometrika**, Williamsburg, v.29, n.1, p.1-27, Mar. 1964.

LANZA, L.L.; SOUZA JUNIOR, C.L.; OTTOBONI, L.M.M.; VIEIRA, M.L.C.; SOUZA, A.P. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. **Theoretical and Applied Genetics**, Berlin, v.94, n.8, p.1023-1030, June 1997.

LITT, M.; LUTY, J.A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle action gene. **American Journal of Human Genetics**, Chicago, v.44, n.1, p.398-401, Mar. 1989.

LIU, B.H. **Statistical genomics: linkage, mapping and QTL analysis**. Boca Raton: CRC, 1998. 611p.

LYNCH, M.; RITLAND, K. Estimation of pairwise relatedness with molecular markers. **Genetics**, Baltimore, v.152, n.4, p.1753-1766, Aug. 1999.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 948p.

MELCHINGER, A.E. Genetic diversity and heterosis. In: COORS, J.G.; PANDEY, S. **The genetics and exploitation of heterosis in crops**. Madison: ASA, 1999. p.99-118.

MEYER, A.S. **Comparação dos coeficientes de similaridade usados em análise de agrupamento com dados de marcadores moleculares dominantes**. 2002. 106p. Dissertação (Mestrado em Agronomia)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

MEYER, A.S.; GARCIA, A.A.F.; SOUZA, A.P.; SOUZA JUNIOR, C.L. Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays L.*). **Genetic and Molecular Biology**, Ribeirão Preto, v.27, n.1, p.83-91, Jan./Feb. 2004.

MILACH, S.C.K. **Marcadores moleculares em plantas**. Porto Alegre: UFRGS, 1998a. 141p.

MILACH, S.C.K. Marcadores de DNA. **Biotecnologia, Ciência e Desenvolvimento**, Brasília, v.1, n.5, p.14-17, 1998b.

MORGANTE, M.; OLIVIERI, A.M. PCR-amplification microsatellites as markers in plant genetics. **The Plant Journal**, Oxford, v.3, n.1, p.175-182, Jan. 1993.

NEJATI-JAVEREMI, A.; SMITH, C.; GIBSON, J.P. Effect of total allelic relationship on accuracy of evaluation and response to selection. **Journal of Animal Science**, Champaign, v.75, n.7, p.1735-1738, July 1997.

PANTER, D.M.; ALLEN, F.L. Using best linear unbiased predictions to enhance breeding for yield in soybean: I., choosing parents. **Crop Science**, Madison, v.35, n.2, p.397-405, Mar./Apr. 1995.

PATERNIANI, M.E.A.G.Z.; GUIMARÃES, P. de S.; LUDERS, 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. **Bragantia**, Campinas, v.67, n.3, p.639-648, maio/jun. 2008.

PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, London, v.58, n.3, p.545-554, Dec. 1971.

QUELLER, D.C.; GOODNIGHT, K.F. Estimating relatedness using genetic markers. **Evolution**, Houston, v.43, n.2, p.258-275, 1989.

RAFALSKI, J.A.; TINGEY, S.V. Genetic diagnostics in plant breeding: RAPDs, microsatellites, and machines. **Trends in Genetics**, London, v.9, n.8, p.275-280, Aug. 1993.

REIF, J.C.; MELCHINGER, A.E.; FRISCH, M. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. **Crop Science**, Madison, v.45, n.1, p.1-7, Jan./Feb. 2005.

REIF, J.C.; MELCHINGER, A.E.; XIA, X.C.; WARBURTON, M.L. Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. **Crop Science**, Madison, v.43, n.4, p.1275-1282, July 2003a.

REIF, J.C.; MELCHINGER, A.E.; XIA, X.C.; WARBURTON, M.L.; HOISINGTON, D.A.; VASAI, S.K.; BECK, D.; BOHN, M.; FRISCH, M. Use of SSRs for establishing heterotic groups in subtropical maize. **Theoretical and Applied Genetics**, Berlin, v.107, n.5, p.947-957, Sept. 2003b.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: EMBRAPA Informação Tecnológica, 2002. 975p.

RITLAND, K. Estimators for pairwise relatedness and individual inbreeding coefficients. **Genetical Research**, London, v.67, n.2, p.175-185, Apr. 1996.

SEARLE, S.; CASELA, G.; McCULLOCH, C.E. **Variance components**. New York: J.Wiley, 1992. 501p.

SIBOV, S.T.; SOUZA JUNIOR, C.L.; GARCIA, A.A.F.; SILVA, A.R.; GARCIA, A.F.; MANGOLIN, C.A.; BENCHIMOL, L.L.; SOUZA, A.P. Molecular mapping in tropical maize (*Zea mays L.*) using microsatellite markers: 2., quantitative trait loci (QTL) for grain yield, plant height, ear height and grain moisture. **Hereditas**, Copenhagen, v.139, n.2, p.107-115, 2003.

SKROCH, P.; TIVANG, J.; NIENHUIS, J. **Analysis of genetic relationships using RAPD marker data in:** applications of RAPD technology to plant breeding symposia series. Madison: CCSA, 1992. 30p.

SNEATH, S.; SOKAL, R.R. **Numeric taxonomy:** the principles and practice of numerical classification. Sacramento: W.H. Freeman, 1973. 573p.

VLECK, L.D. van. **Selection index in introduction to mixed model methods for genetic improvement of animals:** the green book. Boca Raton: CRC, 1993a. 481p.

VLECK, L.D. van. Variance of prediction error with mixed models equations when relationships are ignored. **Theoretical and Applied Genetics**, Berlin, v.85, n.5, p.545-549, Jan. 1993b.

WANG, J. An estimator for pairwise relatedness using molecular markers. **Genetics**, Berlin, v.160, n.1, p.1203-1215, Mar. 2002.

WHITE, T.L.; HODGE, G.R. **Predicting breeding values with applications in forest tree improvement**. Dordrecht: Kluwer Academic, 1989. 376p.

WOOD, C.M.; CHRISTIAN, L.L.; ROTHSFIELD, M.F. Use of animal model in situations of limited numbers and high degrees of relationships. **Journal of Animal Science**, Champaign, v.69, n.6, p.1420-1427, June 1991.

CAPÍTULO 2

Predição do desempenho de híbridos de milho utilizando informações de similaridade no estado e parentesco

1 RESUMO

Este estudo foi realizado com o objetivo de avaliar a eficiência do melhor preditor linear não viesado (BLUP) e a influência do uso da similaridade no estado e o parentesco na predição de híbridos de milho não testados. Foram utilizadas nove linhagens provenientes do programa de melhoramento de milho da empresa Geneseeds Recursos Genéticos Ltda. Estas linhagens foram genotipadas com 48 marcadores microssatélites (SSR) associados a QTL's para produtividade de grãos. Essas linhagens foram intercruzadas em esquema dialelo completo, com intuito de se obter os híbridos simples para estudos de correlações. Os resultados de similaridade no estado e parentesco foram utilizados para a construção das matrizes de relacionamento genético aditivo e de dominância que, posteriormente, foram empregadas no BLUP para predição dos valores genotípicos e de capacidade específica de combinação (CEC) dos híbridos não avaliados. Para a obtenção das estimativas dos valores preditos para a produtividade de grãos e CEC foram simulados cinco graus de desbalanceamento (3, 6, 9, 12 e 15 híbridos retirados). Os valores de correlações entre os valores genotípicos preditos e as médias observadas, dependendo do grau de desbalanceamento, variaram de 0,48 a 0,99 para a similaridade no estado e de 0,40 a 0,99 utilizando informações de parentesco. Os resultados obtidos para a CEC variaram de 0,26 a 0,98 para a similaridade no estado e de 0,001 a 0,990 utilizando as informações de parentesco. A similaridade no estado apresentou desempenho superior quanto aos valores de correlação e eficiência de seleção dos melhores híbridos em comparação ao parentesco. Isso demonstra que a similaridade no estado foi mais eficiente para o ranqueamento dos híbridos, demonstrando apresentar aplicação mais prática ao melhorista. O parentesco apresentou menores valores para a soma do erro de predição (PRES), demonstrando que as predições obtidas apresentaram-se mais próximas do valor observado, ou seja, foram mais acuradas que aquelas obtidas pela similaridade no estado. A diferença no ranqueamento dos híbridos pode ser causada pelos baixos valores de parentesco detectados entre os genitores, o que permite inferir que, provavelmente, estes sejam provenientes de diferentes populações. Nesse sentido, embora apresente um viés devido à superestimação do parentesco, a similaridade no estado pode ser utilizada quando forem detectados baixos

valores de parentesco entre o conjunto de genitores cujo desempenho em combinações híbridas deseja-se prever.

Palavras-chave: BLUP; parentesco por marcadores moleculares; similaridade no estado.

2 ABSTRACT

This study was conducted with the purpose of evaluating the efficiency of both the best linear unbiased predictor (BLUP) and the influence of the use of similarity in the status and relationship in the prediction of untested corn hybrids. Nine lines coming from the corn breeding program of Geneseeds Recursos Genéticos Ltda enterprise were utilized. These lines were genotyped with 48 microsatellite markers (SSR) associated with QTL's for kernel yield. Those lines were intercrossed in a complete diallel scheme with the goal of obtaining the single hybrids for correlation studies. The similarity results in the status and relationship were utilized for the construction of the matrices of additive and dominance genetic relationship which, afterward, were employed in BLUP for prediction of the genotypic values and specific combining capacity (SCC) of the non-evaluated hybrids. For the obtaining of the estimates of the predicted values for the kernel yield and SCC were simulated five unbalancing degrees (3, 6, 9, 12 and 15 removed hybrids). The correlation values among the predicted genotypic values and the observed means, depending upon the unbalancing degree, ranged from 0.48 to 0.99 for the similarity in the status and from 0.40 to 0.99 utilizing relationship information. The results found for SCC ranged from 0.26 to 0.98 for the similarity in the status and from 0.001 to 0.990 by utilizing the relationship information. The similarity in the status presented a higher performance as regards the values of correlation and efficiency of selection of the best hybrids as compared with relationship. That demonstrates that the similarity in the status was more efficient for the ranking of the hybrids, demonstrating to present more practical application to the breeder. Relationship presented lower values for the prediction error sum (PRES), demonstrating that the prediction obtained proved closer to the observed value, that is, they were more accurate than those obtained by the similarity in the status. The difference in the ranking of the hybrids may be caused by the low values of relationship detected between the parents, which allows to infer that, likely, these are coming from different populations. In this sense, although, it presents a bias due to the overestimation of the relationship, the similarity in the status can be utilized when low values of relationship were detected between the set of parents the performance of which in hybrid combinations is wanted to predict.

Key words: BLUP; relationship by molecular markers; similarity in the status.

3 INTRODUÇÃO

Em programas de melhoramento de milho a seleção e a avaliação dos melhores genótipos é extremamente importante para obtenção de híbridos cada vez mais produtivos e adaptados às mais diferentes condições de cultivo.

Algumas ferramentas podem auxiliar os melhoristas no intuito de direcionar os cruzamentos visando à obtenção das melhores combinações híbridas. Por isso, muitas pesquisas têm concentrado esforços em estimar o desempenho dos híbridos através da determinação dos grupos heteróticos das linhagens para direcionar os cruzamentos baseando-se nas distâncias genéticas (Reif et al., 2003b; Amorim, 2005; Guimarães, 2007). Pela teoria da genética quantitativa, os pares que são mais divergentes poderiam maximizar a heterose e revelar as melhores combinações híbridas (Melchinger, 1999).

Nesse contexto, o uso de marcadores moleculares na determinação de grupos heteróticos e na predição do desempenho de híbridos de milho tem sido proposto por vários autores (Reif et al., 2003a; Amorim, 2005). Contudo, até o momento, os resultados são ainda inconsistentes (Dias et al., 2004; Guimarães, 2007).

Dessa forma, uma alternativa para o uso de informações obtidas por marcadores moleculares seria a utilização dessas informações através da metodologia de modelos lineares mistos para a predição de cruzamentos não realizados (Bernardo, 1994; Charcosset et al., 1998). Para isso, torna-se necessário o conhecimento da genealogia ou do parentesco entre os indivíduos avaliados e indivíduos cujo desempenho deseja-se prever (Bernardo, 1994). Porém, nem sempre o melhorista dispõe de tal informação, o que poderia limitar a utilização dessa técnica.

Com intuito de contornar essa limitação, alguns autores sugerem a utilização da informação de marcadores moleculares (similaridade alélica no

estado) como opção quando não se dispõe de dados de genealogias (Bernardo, 1994, 1995; André, 1999; Balestre et al., 2008). Segundo Nejati-Javeremi et al. (1997), o emprego da similaridade no estado possui a vantagem de utilizar toda a semelhança alélica existente entre os indivíduos avaliados.

No entanto, Bernardo (1993) propõe o uso de um coeficiente de coancestria, que utiliza informação da similaridade genética, no intuito de corrigir possíveis vieses da similaridade no estado. Entretanto, essa metodologia somente é válida para dialelos parciais nos quais não ocorrem cruzamentos endogâmicos, ou seja, quando se possui informação prévia a respeito de grupos heteróticos.

Como alternativa à essa restrição, alguns autores (Queller & Goodnight, 1989; Ritland, 1996; Lynch & Ritland, 1999; Wang, 2002) têm sugerido procedimentos para estimação conjunta do parentesco por meio de marcadores moleculares, que podem ser utilizados em dialelos completos e parciais. Esses estimadores, mesmo utilizando dados moleculares, consideram a probabilidade condicional de que dois indivíduos possuam alelos herdados de genitores em comum, de modo que, somente indivíduos que atendam à essa premissa participam da predição de híbridos não avaliados. Dessa forma, aqueles indivíduos que possuem parentesco igual a zero não contribuirão na predição, concorrendo para resultados mais coerentes.

Portanto, o objetivo desse estudo foi avaliar a eficiência do melhor preditor linear não viesado (BLUP) com a utilização de informações de similaridade no estado e parentesco entre genitores visando à obtenção de estimativas de predição de híbridos de milho não testados.

4 MATERIAIS E MÉTODOS

4.1 Descrição dos materiais

Foram utilizadas nove linhagens (Tabela 1A) provenientes do programa de melhoramento de milho da empresa Geneseeds Recursos Genéticos Ltda.

Para a realização das análises com marcadores microssatélites (SSR), uma amostra de aproximadamente vinte sementes de cada linhagem foi semeada em canteiro em casa de vegetação. Após a emergência das plântulas, quando estas apresentavam o segundo par de folhas, procedeu-se a coleta de tecido foliar para extração do DNA.

A extração do DNA, o preparo das reações e as estimativas da similaridade genética seguiram os mesmos procedimentos descritos por Balestre et al. (2008).

Posteriormente, todas as progênies foram genotipadas utilizando 48 marcadores de microssatélites (Tabela 2A), sendo todas essas marcas ligadas a QTL's para componentes de produtividade de grãos.

Todo o trabalho para a obtenção das informações de marcadores microssatélites foi desenvolvido no Laboratório de Biologia Molecular do Departamento de Biologia (LBM/DBI) da Universidade Federal de Lavras. Com as informações dos marcadores microssatélites, foi construída a matriz de valores binários ("0" e "1") para comparação dos diferentes coeficientes de similaridade (Tabela 3A). Os valores binários foram convertidos em frequências alélicas com a utilização do Programa NTSYS-PC 2.1 (Rohlf, 2000).

Os coeficientes de parentesco apresentados na Tabela 4A foram determinados com aporte do Programa Mark (Ritland, 2006).

4.2 Delineamento experimental e correção das médias via modelos lineares mistos

As nove linhagens foram inter cruzadas em esquema de dialelo completo para obtenção dos respectivos híbridos simples, excluindo-se os recíprocos. Devido à pequena quantidade de sementes, dos 36 híbridos simples possíveis no dialelo, foram obtidos 29 híbridos.

Esses híbridos, juntamente com os híbridos utilizados como testemunhas (P30F35, DKB 390 e DOW 2B587), foram avaliados em dois locais no município de Lavras/MG (Campus da Universidade Federal de Lavras e Fazenda Vitorinha) na safra 2008/2009. O delineamento experimental utilizado foi alfa-látice 8 x 4, com três repetições. Cada parcela foi constituída de 2 linhas de 4,0 metros de comprimento, com espaçamento entre linhas de 0,80 metros. O plantio foi realizado em meados de novembro de 2008 e a colheita em meados de abril de 2009.

Utilizou-se 450 Kg.ha⁻¹ da formulação 08-28-16 (N, P₂O₅ e K₂O) na semeadura e 300 Kg.ha⁻¹ da formulação 30-00-20 (N, P₂O₅ e K₂O) em cobertura. Os demais tratamentos culturais foram aqueles comumente recomendados para a cultura do milho na região.

As médias ajustadas dos híbridos para cada local foram corrigidas levando-se em consideração a recuperação da informação interblocos pela análise de modelos lineares mistos.

4.2.1 Análise dialélica via modelos lineares mistos

A partir das médias fenotípicas ajustadas dos híbridos em cada local realizou-se a análise dialélica conjunta utilizando o método IV, proposto por Griffing (1956). As análises foram realizadas utilizando o Software SAS[®] System, Módulo IML (Statistical Analysis Software Institute-SAS Institute, 2000). As estimativas dos efeitos fixos e BLUP's para capacidade geral de

combinação (CGC) e a capacidade específica de combinação (CEC) foram obtidas de maneira semelhante à apresentada por Bernardo (1995), proposta para dialelos parciais. Entretanto, nesse trabalho utilizou-se o modelo de dialelo completo, considerando todos os genótipos dentro de um mesmo grupo.

O modelo linear considerado foi o seguinte:

$$y = X\beta + Z_1g + Z_2s + e$$

em que y é o vetor das médias ajustadas das combinações híbridas; β é o vetor dos efeitos fixos (média geral e locais); g é o vetor de efeitos da capacidade geral de combinação das linhagens; s é o vetor da capacidade específica de combinação dos híbridos; e é o vetor dos erros e X , Z_1 e Z_2 são as matrizes de incidência dos efeitos β , g e s , respectivamente.

A solução conjunta para os efeitos fixos e aleatórios foi obtida pelo seguinte sistema de equações, conforme Henderson (1984):

$$\begin{bmatrix} \hat{\beta} \\ \hat{g} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1Z_1 + A_1^{-1}\gamma_1 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + A_2^{-1}\gamma_2 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix}$$

com $\gamma_1 = \frac{\sigma_e^2}{\sigma_{CGC}^2}$ e $\gamma_2 = \frac{\sigma_e^2}{\sigma_{CEC}^2}$, em que: A_1 é a matriz de similaridade genética

aditiva entre as linhagens; A_2 é a matriz de dominância entre as linhagens.

Os componentes de variância foram estimados utilizando o método de máxima verossimilhança restrita com uso do algoritmo EM (Expectation - Maximization).

As matrizes de similaridade genética aditiva foram construídas de acordo com os coeficientes de similaridade no estado (Tabela 3A) e parentesco (Tabela 4A).

As matrizes de dominância (A_2) foram construídas como sugerido por Henderson (1984) e Vleck (1993).

4.3 Predição de híbridos não Avaliados

As predições dos valores genotípicos e da CEC dos cruzamentos não avaliados foram realizadas com base em perdas simuladas ou validação cruzada. Assim, de um total de 29 híbridos avaliados, simularam-se cinco situações considerando desbalanceamentos aleatórios de 3, 6, 9, 12 e 15 híbridos. Para cada situação, o processo foi repetido 2000 vezes. O modelo utilizado foi semelhante ao apresentado por Bernardo (1994).

Considerando que \hat{y}_p é o vetor das médias dos híbridos avaliados corrigidos para os efeitos fixos (β), tem-se:

$$\hat{y}_p = (Z'Z)^{-1}Z'(y - X\hat{\beta})$$

em que: Z é a matriz de incidência dos efeitos dos híbridos avaliados.

As produtividades dos híbridos não avaliados foram preditas por meio da seguinte expressão:

$$\hat{y}_m = CV^{-1}\hat{y}_p$$

em que \hat{y}_m é o vetor do desempenho dos híbridos com perda simulada; C é a matriz de covariância genética entre os cruzamentos com perda simulada e cruzamentos avaliados; V é a matriz de variâncias e covariâncias fenotípicas entre os cruzamentos avaliados.

De maneira semelhante, os valores da CEC dos cruzamentos com perda simulada foram preditos por meio da seguinte expressão:

$$d_m = CV^{-1}s_p$$

em que d_{nt} é o vetor da CEC dos cruzamentos com perda simulada; C é a matriz de covariância genética entre os cruzamentos com perda simulada e cruzamentos avaliados; V é a matriz de variâncias e covariâncias fenotípicas entre os cruzamentos avaliados; s_p é o vetor da CEC dos híbridos avaliados.

4.3.1 Comparação das predições obtidas pelos coeficientes de similaridade no estado e parentesco

Para a predição dos híbridos com perda simulada, utilizaram-se informações de similaridade no estado e parentesco. Para a comparação das predições, a partir da similaridade genética determinada pelos diferentes coeficientes, utilizou-se a correlação de Pearson entre os valores observados de CEC e as médias fenotípicas com os valores preditos dos híbridos não testados.

Utilizou-se também a soma do erro de predição (PRES) (Allen, 1974), através da seguinte expressão:

$$PRES_p = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

em que y_i é o valor observado (efeito da CEC ou de cruzamento) do i -ésimo híbrido não testado; n é o número de parâmetros a serem preditos e \hat{y}_i é a i -ésima predição de y_i . Essa expressão mede a soma do erro de predição em relação ao seu valor observado.

A eficiência do BLUP na predição dos valores da CEC e dos valores genotípicos foi avaliada por meio da magnitude dos valores de correlações.

Também foi aplicada a eficiência de seleção de Hamblim & Zimmermann (1986), comparando os nove maiores valores preditos de CEC e valores genotípicos, com os nove maiores valores de CEC e produtividade observados. Foram consideradas as médias de todos os valores genotípicos dos

híbridos não testados, em todas as combinações e condições de desbalanceamento apresentadas.

5 RESULTADOS

O conjunto de genótipos avaliados apresentou valores médios de similaridade no estado de 0,251, 0,153, 0,646 e 0,636, para os coeficientes de Sorensen-Dice (SD), Jaccard (J), Distância Modificada de Rogers (DMR) e Simple Matching (SM), respectivamente (Tabelas 3A, 4A, 5A e 6A). Quanto ao valor médio de parentesco, observou-se valores de 0,005, 0,006, 0,186 e 0,001, correspondentes ao parentesco de Lynch-Ritland (LR), Queller-Goodnight (QG), Ritland (R) e Wang (W), respectivamente (Tabelas 7A, 8A, 9A e 10A). É importante ressaltar que os valores médios de parentesco, quando comparados à similaridade no estado, podem ser considerados superestimados, visto que os valores negativos obtidos por esses coeficientes foram considerados iguais a zero.

Na Tabela 9A encontram-se os valores médios de correlação entre as médias preditas e observadas obtidos em 2000 simulações, em diferentes níveis de desbalanceamento (3, 6, 9, 12 e 15 híbridos retirados), para os coeficientes de similaridade no estado SD, J, DMR e SM. Observa-se que as correlações foram de moderada magnitude, quando consideramos os valores da média de distribuição, variando de 0,439 a 0,494, 0,449 a 0,505, 0,442 a 0,503 e 0,434 a 0,472. Quando se considera a mediana da distribuição, os valores de correlação também demonstraram moderada magnitude, variando de 0,468 a 0,694, 0,468 a 0,728, 0,466 a 0,722 e 0,463 a 0,667, dependendo do nível de desbalanceamento, para os referidos coeficientes. Além disso, observou-se que o desvio padrão das correlações diminui à medida que o número de híbridos preditores também diminui, ou seja, o desvio padrão das correlações foi menor quando o nível de desbalanceamento foi maior.

Entretanto, vale ressaltar que quando ocorre alta assimetria da distribuição de probabilidade (Figuras 1 e 2), a média e a mediana podem não

refletir uma boa medida de posição visto que os valores mais verossímeis não estão próximos a esses estimadores, ou seja, a média e a mediana não refletem os valores de máxima probabilidade ou verossimilhança. Segundo Dalenius (1965), a moda se aplica em algumas situações, dentre elas, como medida de posição em distribuições assimétricas, de modo que expressa o valor mais provável de uma distribuição de probabilidade, principalmente no caso de previsão de valores futuros. Da mesma forma, Bickel (2002, 2003) afirma que a moda é uma medida de posição mais robusta que a média e a mediana em casos de distribuições de probabilidade assimétricas.

Assim, obtiveram-se as densidades das amostras e, dessa forma, determinou-se a moda da distribuição dos valores de correlação. Levando-se em consideração a moda da distribuição, os valores de correlações mais prováveis variaram de 0,480 a 0,990, 0,500 a 0,990, 0,520 a 0,990 e 0,500 a 0,990 dependendo dos níveis de desbalanceamento, para os coeficientes de SD, J, DMR e SM, respectivamente.

Isso permite inferir que, com desbalanceamento de 10%, ou seja, três híbridos retirados, é possível prever o desempenho desses híbridos com alta acurácia. Com níveis de desbalanceamento de 20% ou acima desse valor, ou seja, 6, 9, 12 e 15 híbridos retirados, os valores de correlações demonstram moderada magnitude.

Na Tabela 10A, encontram-se os valores de correlações entre as médias preditas e observadas obtidos em 2000 simulações, considerando os mesmos níveis de desbalanceamento (3, 6, 9, 12 e 15 híbridos retirados), referentes à produtividade de grãos, utilizando os coeficientes de parentesco (LR, QG, R e W). Nesse caso, assim como na similaridade no estado, a utilização dos valores de correlação referentes à moda foram preteridos, pelo mesmo fato descrito anteriormente. Desse modo, obtiveram-se correlações entre os valores preditos e observados de moderada a elevada magnitude, variando de 0,410 a 0,990 (LR),

0,400 a 0,980 (QG), 0,450 a 0,990 (R) e 0,410 a 0,980 (W), de acordo com os diferentes níveis de desbalanceamento. O desvio padrão foi maior quando o nível de desbalanceamento foi menor.

É importante ressaltar que quando se analisa a produtividade de grãos, é possível inferir que os valores de correlação entre as médias preditas e as médias observadas demonstraram ligeira superioridade da similaridade no estado em relação à similaridade por descendência. Essa situação fica mais pronunciada quando se analisam as correlações com níveis de desbalanceamento de 6, 9, 12 e 15 híbridos retirados.

Na Tabela 11A, encontram-se os valores de correlação entre a CEC dos híbridos não testados e os seus valores observados. De modo semelhante ao que ocorreu para produtividade de grãos, observa-se que à medida que o desbalanceamento aumenta, os valores de correlações tendem a diminuir para todos os coeficientes analisados. Da mesma forma, observa-se que a similaridade no estado apresentou valores de correlações com magnitude superior aos valores de correlações correspondentes ao parentesco de LR, QG, R e W (Tabela 12A). A magnitude desse fato foi maior quando foram analisados os níveis de desbalanceamento a partir de 6 híbridos retirados.

Uma provável explicação para essas observações pode estar associada aos baixos valores de parentesco encontrados entre as linhagens genitoras, o que dificulta a predição quando os níveis de desbalanceamento são maiores (Tabelas 7A e 8A). Por exemplo, considerando o coeficiente de parentesco de LR, este determinou parentesco entre L31 e L37 (0,147), L43 e L57 (0,002), L12 e L84 (0,033). Situação semelhante foi observada para os coeficientes de QG e, principalmente, para o parentesco de W. Para esse último estimador, detectou-se parentesco somente entre L31 e L37, cujo cruzamento não foi possível de se obter no campo.

Entretanto, é possível inferir que, mesmo com baixos valores de parentesco detectados, os coeficientes de parentesco permitiram prever com moderada acurácia o desempenho dos híbridos não testados (Tabela 10A). Pode-se inferir que esses coeficientes demonstraram boa coerência entre o parentesco dos genótipos, visto que, dentre os híbridos que apresentaram algum parentesco entre as linhagens, estes não apresentaram bom desempenho quanto à produtividade de grãos em relação aos híbridos cujas linhagens não apresentaram qualquer parentesco. Como exemplo, temos os híbridos L12XL84, L12XL43, L43XL57 e L12XL30 cujos desempenhos foram abaixo da média geral (Tabela 13A).

Ainda considerando o parentesco, para os valores correspondentes à CEC (Tabela 12A), o desempenho foi inferior se comparado à similaridade no estado, principalmente quando o nível de desbalanceamento aumentou. Mesmo assim, o maior valor de CEC observado (L32XL43) encontra-se entre as nove maiores CEC's preditas para os coeficientes de QG e R (Tabela 20A). Da mesma forma, o segundo maior valor de CEC observado (L63XL84) encontra-se presente entre as nove maiores CEC's preditas para o coeficiente de LR. Ainda considerando os valores preditos para a CEC, vale ressaltar que o coeficiente de parentesco proposto por Wang (2002) não conseguiu prever as CEC's (dados não apresentados). Uma explicação para esse fato pode estar relacionada aos baixos valores de parentesco detectados por esse estimador (Tabela 8A).

Na Tabela 13A encontram-se os valores médios referentes às produtividades de grãos dos vinte e nove híbridos simples obtidos do cruzamento entre as nove linhagens. Conforme citado no tópico Material e Métodos, dos trinta e seis híbridos possíveis de se obter no dialelo, vinte e nove foram obtidos e avaliados. Nesse caso, observa-se que a média fenotípica foi de 8,780 ton.ha⁻¹; sendo que o cruzamento com melhor desempenho foi L32XL43 com produtividade de 11,547 ton.ha⁻¹.

Os nove maiores valores médios de predição dos híbridos não testados obtidos para cada coeficiente, considerando todos os níveis de desbalanceamento, estão apresentados nas Tabelas 14A, 15A, 16A e 17A. Pode-se observar que os valores fenotípicos correspondentes a esses nove maiores valores preditos, em sua grande maioria, foram de alta magnitude, com vantagem em relação à média geral de $0,878 \text{ ton.ha}^{-1}$ para os coeficientes de similaridade no estado e em média $0,628 \text{ ton.ha}^{-1}$ para os coeficientes de parentesco. A eficiência de seleção demonstrada pelo coeficiente de Hamblim & Zimmermann (1986), levando-se em consideração os nove maiores valores fenotípicos observados, foi de 65% para os coeficientes de similaridade no estado, de 53% para o coeficiente de parentesco R e de 42% para os demais coeficientes de parentesco (LR, QG e W).

Outra constatação importante é que o híbrido com maior média ajustada (L32XL43) está presente entre as nove maiores médias preditas para todos os coeficientes de parentesco estudados (Tabelas 16A e 17A); fato esse também identificado quando se analisa os coeficientes de similaridade no estado (Tabelas 14A e 15A).

Isso demonstra a eficiência do melhor preditor linear não viesado (BLUP) em prever o desempenho de híbridos de milho não testados, através da utilização de informação de similaridade no estado e parentesco. Mesmo com detecção de pouco parentesco entre as linhagens, é possível prever o desempenho de híbridos não testados com relativa eficiência de predição.

Quanto aos valores referentes à soma do erro de predição (PRES) para produtividade de grãos, pode-se inferir que os coeficientes de parentesco apresentaram uma ligeira superioridade (Tabelas 16A e 17A) perante os coeficientes de similaridade no estado (Tabelas 14A e 15A). Este fato fica mais evidente quando analisamos os valores referentes à soma do erro de predição das

CEC's (Tabelas 18A, 19A e 20A), demonstrando que o parentesco proporcionou estimativas de predições mais acuradas dos valores observados (Tabela 21A).

6 DISCUSSÕES

A utilização do melhor preditor linear não viesado (BLUP) na predição de híbridos não avaliados tem sido recomendada por muitos autores (Bernardo, 1994, 1995, 1996a,b; Charcosset et al., 1998). Nesse contexto, o uso de informação de marcadores moleculares surge como alternativa, quando não se dispõe de genealogias sobre os genitores, contribuindo para que se obtenha estimativas de predições mais acuradas dos valores genotípicos de híbridos não testados.

Nesse estudo, os resultados observados utilizando informações de similaridade no estado e parentesco, permitiram obter estimativas de correlação de moderada magnitude para produtividade de grãos e capacidade específica de combinação (CEC) de híbridos não testados de milho. Isso permite inferir que o melhor preditor linear não viesado (BLUP) apresentou moderada capacidade preditiva do desempenho de híbridos não testados, corroborando com os resultados obtidos por Bernardo (1996a,b) que encontrou correlações em estudos de predição de híbridos não testados de milho, para o caráter rendimento de grãos, variando de 0,43 a 0,76.

Através da análise dos resultados obtidos, é possível inferir que a similaridade no estado apresentou desempenho superior ao parentesco, no que diz respeito aos valores de predição encontrados por meio da utilização de simulações. Isso fica evidente quando analisamos os valores de correlação entre os valores genotípicos preditos e os valores observados para os caracteres produtividade de grãos e capacidade específica de combinação (CEC).

Por disso, observa-se que o coeficiente proposto por Jaccard apresentou maior estabilidade dos valores de correlação, considerando os diferentes níveis de desbalanceamento. Esse coeficiente não considera a ausência comum de bandas (Tabela 3A) na construção da matriz de dados binários, colaborando para

valores de similaridade no estado menos viesados em comparação com coeficientes que consideram essa informação, como por exemplo, o coeficiente de Simple Matching (Tabela 3A). Segundo alguns autores (Meyer, 2002; Iemma, 2003), em estudos com marcadores dominantes e codominantes, não é correto o emprego de coeficientes que consideram a ausência comum de bandas. Esses autores argumentam que essa ausência comum não significa necessariamente que as regiões do DNA são idênticas, o que poderia levar a estimativas mais viesadas de similaridade entre genitores.

Partindo dessa condição, analisando-se os coeficientes que não apresentam ocorrência comum de bandas, observamos que o coeficiente de Sorensen-Dice pode superestimar a similaridade no estado, porque dá um peso maior à ocorrência comum de bandas entre os genitores (Tabela 3A) em detrimento do coeficiente de Jaccard. Portanto, em estudos de filogenia o uso do coeficiente de Sorensen-Dice tem apresentado bons resultados. Em situações onde o intuito é a predição de híbridos não testados, o uso desse coeficiente, por proporcionar peso maior à similaridade no estado, pode levar à obtenção de estimativas de predição com menor precisão.

Os baixos valores de parentesco detectados entre os genitores talvez tenham sido uma das principais razões do desempenho inferior da similaridade por descendência em relação à similaridade no estado. Isso porque as matrizes de parentesco correspondem às matrizes A^1 's usadas como matrizes dos coeficientes de variâncias e covariâncias associadas aos efeitos da CGC. Em tal condição, pouca informação é recuperada entre os níveis dos efeitos aditivos do modelo e, por esse motivo, parte da informação recuperada pela metodologia talvez seja conseqüência da participação do coeficiente de dominância associada aos efeitos da CEC.

Esse baixo parentesco pode ser devido à ocorrência de possíveis vieses na obtenção dessas estimativas. Ritland (1996) enfatiza que o uso dos

coeficientes pode levar a estimativas negativas de parentesco, refletindo em grande erro estatístico com pequeno tamanho da amostra da população em estudo, introduzindo viés nas estimativas e distorcendo as associações entre variáveis. Isso pode ocorrer pelo fato desses coeficientes terem sido desenvolvidos para estudos de populações naturais, onde o tamanho da amostra pode interferir nos resultados das estimativas de parentesco.

Outra provável explicação para os baixos valores detectados de parentesco pode ser o fato das linhagens genitoras dos híbridos presentes nesse estudo serem, em sua grande maioria, oriundas de diferentes *backgrounds*, ou seja, advindas de diferentes populações. Quando analisamos a genealogia das linhagens (Tabela 1A), esse fato fica bem evidente. É possível observar que as linhagens L31 e L37 são oriundas da mesma população, demonstrando a coerência do parentesco detectado por meio dos estimadores de parentesco Queller & Goodnight (1989) e Lynch & Ritland (1999). Por isso, como os coeficientes de parentesco consideram a probabilidade condicional de que dois indivíduos possuam alelos herdados de genitores em comum, somente indivíduos que atendam a essa premissa participam da predição de híbridos não testados.

Ainda com relação à ausência de parentesco, Bernardo (1996b) verificou que a predição BLUP é robusta quando são desconhecidos até 25% de parentesco, ou quando existem vieses nas estimativas. No caso desse trabalho, a porcentagem de cruzamentos avaliados que apresentam coeficiente de parentesco igual a zero é bem maior que os valores descritos por esse autor.

Da mesma forma, vale ressaltar que nesse trabalho a participação da CGC e CEC para a predição dos valores genotípicos dos híbridos não testados foi significativa, considerando todos os coeficientes de similaridade no estado e similaridade por descendência (Tabelas 5A a 8A). Como esses valores de CGC e CEC influenciam na predição dos híbridos não testados por meio do uso do

melhor preditor linear não viesado (BLUP), mesmo com baixos valores de parentesco detectados, foi possível prever os valores genotípicos dos híbridos não avaliados com moderada magnitude.

É importante salientar que, embora a similaridade no estado tenha apresentado valores superiores de correlação e de eficiência de seleção, esta apresentou valores inferiores para a PRES, ou seja, as previsões utilizando a similaridade por descendência ficaram mais próximas do valor observado, levando a previsões mais estáveis ou mais precisas dos valores genotípicos dos híbridos não testados. Já a similaridade no estado proporcionou um melhor ranqueamento dos híbridos, que pode ser confirmado através dos valores de eficiência de seleção. Pode-se, então, inferir que a similaridade no estado tem aplicação mais prática ao melhorista, visto que, o que se busca na grande maioria dos casos é a identificação e ranqueamento dos melhores híbridos.

Essa diferença no ranqueamento pode ser causada pelos baixos valores de parentesco identificados pela similaridade por descendência devido aos diferentes *backgrounds* dos genitores envolvidos nesse estudo. No entanto, isso demonstra o quanto o método que utiliza informações da similaridade por descendência é robusto. Desse modo, deve ser utilizado em situações onde se obtém maior quantidade de medidas de parentesco entre os genitores, o que concorreria para valores mais precisos de híbridos não testados.

Portanto, conclui-se nesse trabalho que a similaridade no estado, embora apresente um viés devido à superestimação do parentesco, pode ser utilizada em situações onde a similaridade por descendência apresente baixos valores de parentesco entre o conjunto de genitores avaliados. Além disso, torna-se uma ferramenta de grande utilidade dentro dos programas de melhoramento de milho, de modo que, por meio da utilização dessas informações, podemos prever o desempenho de híbridos antes destes serem efetivamente avaliados. Isso permite ao melhorista reduzir o número de híbridos a serem avaliados e propicia a

avaliação desses híbridos em mais ambientes e condições de cultivo, levando à seleção de híbridos com maior adaptabilidade e potencial produtivo.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- ALLEN, D.M. The relationship between variable selection and data augmentation and a method for prediction. **Technometrics**, Washington, v.16, n.1, p.125-127, Feb. 1974.
- AMORIM, E.P. **Produtividade de híbridos de milho derivados de populações S₀, e associação com distância genética baseada em microssatélites**. 2005. 97p. Tese (Doutorado em Genética e Melhoramento de Plantas)-Universidade Federal de Lavras, Lavras.
- ANDRÉ, C.M.G. **Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores moleculares na análise dialélica**. 1999. 101p. Dissertação (Mestrado em Genética e Melhoramento de Plantas)-Universidade Federal de Lavras, Lavras.
- BALESTRE, M.; PINHO, R.G. von; SOUZA, J.C.; LIMA, J.L. Comparison of similarity and dissimilarity genetic coefficients based on microsatellites markers. **Genetics and Molecular Research**, Ribeirão Preto, v.7, n.4, p.695-705, jul./ago. 2008.
- BERNARDO, R. Estimation of coefficient of coancestry using molecular markers in maize. **Theoretical and Applied Genetics**, Berlin, v.85, n.8, p.1055-1062, Feb. 1993.
- BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science**, Madison, v.34, n.1, p.20-25, Jan./Feb. 1994.
- BERNARDO, R. Best linear unbiased prediction of maize single cross performance. **Crop Science**, Madison, v.36, n.1, p.50-56, Jan./Feb. 1996a.
- BERNARDO, R. Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. **Crop Science**, Madison, v.36, n.4, p.862-866, Aug./Sept. 1996b.
- BERNARDO, R. Genetic models for predicting maize performance in unbalanced yield trial data. **Crop Science**, Madison, v.35, n.1, p.141-147, Jan./Feb. 1995.

BICKEL, D.R. Robust estimators of the mode and skewness of continuous data. **Computational Statistical & Data Analysis**, New York, v.39, n.2, p.153-163, Apr. 2002.

BICKEL, D.R. Robust and efficient estimation of the mode of continuous data: the mode as a viable measure of central tendency. **Journal of Statistical Computation and Simulation**, New York, v.73, n.12, p.899-912, Dec. 2003.

CHARCOSSET, A.; BONNISSEAU, B.; TOUCHEBEUF, O.; BURSTIN, J.; DUBREUIL, P.; BARRIERE, Y.; GALLAIS, A.; DENIS, J.B. Prediction of maize hybrid silage performance using marker data: comparison of several models for specific combining ability. **Crop Science**, Madison, v.38, n.1, p.38-44, Jan./Feb. 1998.

DALENIUS, T. The mode: a neglected statistical parameter. **Journal Of The Royal Statistical Society Series A (General)**, London, v.128, n.1, p.110-117, 1965.

DIAS, L.A. dos S.; PICOLI, E.A. de T.; ROCHA, R.B.; ALFENAS, A.C. A priori choice of hybrid parents in plants. **Genetics and Molecular Research**, Ribeirão Preto, v.3, n.3, p.356-368, maio/jun. 2004.

GRIFFING, B.A. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian Journal of Biological Sciences**, Melbourne, v.9, n.4, p.463-493, 1956.

GUIMARÃES, P. de S. **Desempenho de híbridos simples de milho (*Zea mays* L.) e correlação entre heterose e divergência genética entre linhagens parentais**. 2007. 111p. Dissertação (Mestrado em Melhoramento Vegetal)-Instituto Agrônomo de Campinas, Campinas.

HAMBLIN, J.; ZIMMERMANN, M.J. Breedings common bean for yield in mixtures. **Plant Breeding Reviews**, Berlin, v.4, n.4, p.245-272, 1986.

HENDERSON, C.R. **Applications of linear models in animal breeding**. Ontario: University of Guelph, 1984. 462p.

IEMMA, M. **Uso do melhor preditor linear não viesado (BLUP) em análises dialélicas e predição de híbridos**. 2003. 81p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica)-Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba.

LYNCH, M.; RITLAND, K. Estimation of pairwise relatedness with molecular markers. **Genetics**, Baltimore, v.152, n.4, p.1753-1766, Aug. 1999.

MELCHINGER, A.E. Genetic diversity and heterosis. In: COORS, J.G.; PANDEY, S. **The genetics and exploitation of heterosis in crops**. Madison: ASA, 1999. p.99-118.

MEYER, A.S. **Comparação dos coeficientes de similaridade usados em análise de agrupamento com dados de marcadores moleculares dominantes**. 2002. 106p. Dissertação (Mestrado em Estatística e Experimentação Agronômica)-Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

NEJATI-JAVEREMI, A.; SMITH, C.; GIBSON, J.P. Effect of total allelic relationship on accuracy of evaluation and response to selection. **Journal of Animal Science**, Champaign, v.75, n.7, p.1735-1738, July 1997.

QUELLER, D.C.; GOODNIGHT, K.F. Estimating relatedness using genetic markers. **Evolution**, Houston, v.43, n.2, p.258-275, 1989.

REIF, J.C.; MELCHINGER, A.E.; XIA, X.C.; WARBURTON, M.L. Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. **Crop Science**, Madison, v.43, n.4, p.1275-1282, July 2003a.

REIF, J.C.; MELCHINGER, A.E.; XIA, X.C.; WARBURTON, M.L.; HOISINGTON, D.A.; VASAI, S.K.; BECK, D.; BOHN, M.; FRISCH, M. Use of SSRs for establishing heterotic groups in subtropical maize. **Theoretical and Applied Genetics**, Berlin, v.107, n.5, p.947-957, Sept. 2003b.

RITLAND, K. Estimators for pairwise relatedness and individual inbreeding coefficients. **Genetical Research**, London, v.67, n.2, p.175-185, Apr. 1996.

RITLAND, K. **Mark genetic marker analysis program**. Version 3.1. Columbia: University of British Columbia, 2006. Software.

ROHLF, F.J. **NTSYSpc numerical taxonomy and multivariate analysis system**. Version 2.1. Setauket: Exeter Software, 2000. Software.

STATISTICAL ANALYSIS SOFTWARE INSTITUTE. **SAS language and procedures: usage**. Version 8.1. Cary, 2000. 1 CD-ROM.

VLECK, L.D. van. **Selection index in introduction to mixed model methods for genetic improvement of animals: the green book.** Boca Raton: CRC, 1993. 481p.

WANG, J. An estimator for pairwise relatedness using molecular markers. **Genetics**, Austin, v.160, n.1, p.1203-1215, Mar. 2002.

ANEXOS

ANEXO A	Página
TABELA 1A Caracterização das nove linhagens de milho disponíveis nesse estudo. Genesseds Recursos Genéticos Ltda. Lavras, MG, 2010.	58
TABELA 2A Descrição dos <i>primers</i> utilizados para análises de marcadores microssatélites.	58
TABELA 3A Coeficientes de similaridade no estado utilizados para análises de dados genéticos obtidos com marcadores microssatélites.	59
TABELA 4A Coeficientes de parentesco utilizados para análises de dados genéticos obtidos com marcadores microssatélites.	59
TABELA 5A Estimativas de similaridade no estado entre nove linhagens de milho obtidas por meio do coeficiente de similaridade de Sorensen-Dice (acima da diagonal) e de Jaccard (abaixo da diagonal).	60
TABELA 6A Estimativas de similaridade no estado entre nove linhagens de milho obtidas por meio da Distância Modificada de Rogers (acima da diagonal) e coeficiente de similaridade de Simple Matching (abaixo da diagonal).	60
TABELA 7A Estimativas de parentesco entre nove linhagens de milho obtidas por meio do coeficiente de parentesco proposto por Lynch & Ritland (1999) (acima da diagonal) e Queller & Goodnight (1989) (abaixo da diagonal).	61
TABELA 8A Estimativas de parentesco entre nove linhagens de milho obtidas por meio do coeficiente de parentesco proposto por Ritland (2006) (acima da diagonal) e Wang (2002) (abaixo da diagonal).	62
TABELA 9A Parâmetros relacionados aos valores de correlação obtidos em 2000 simulações, em diferentes níveis de desbalanceamento, referentes à produtividade de híbridos simples não avaliados e seus valores observados.	63
TABELA 10A Parâmetros relacionados aos valores de correlação obtidos em 2000 simulações, em diferentes níveis de desbalanceamento, referentes à produtividade de híbridos simples não avaliados e seus valores observados.	64
TABELA 11A Parâmetros relacionados aos valores de correlação obtidos em 2000 simulações em diferentes níveis de desbalanceamento referentes à capacidade específica de combinação (CEC) de	

	híbridos simples não avaliados e seus valores observados.	65
TABELA 12A	Parâmetros relacionados aos valores de correlação obtidos em 2000 simulações em diferentes níveis de desbalanceamento referentes à capacidade específica de combinação (CEC) de híbridos simples não avaliados e seus valores observados.	66
TABELA 13A	Valores referentes às produtividades médias observadas para cada cruzamento, obtidos através de avaliações em dois locais (Campus da Universidade Federal de Lavras e Fazenda Vitorinha, Lavras, Minas Gerais, Safra 2008/2009).	67
TABELA 14A	Valores referentes aos nove maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 2000 simulações, em relação às nove maiores produtividades obtidas na avaliação de vinte e nove híbridos simples de milho.	68
TABELA 15A	Valores referentes aos nove maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 2000 simulações, em relação às nove maiores produtividades obtidas na avaliação de vinte e nove híbridos simples de milho.	69
TABELA 16A	Valores referentes aos nove maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 2000 simulações, em relação às nove maiores produtividades obtidas na avaliação de vinte e nove híbridos simples de milho.	70
TABELA 17A	Valores referentes aos nove maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 2000 simulações, em relação às nove maiores produtividades obtidas na avaliação de vinte e nove híbridos simples de milho.	71
TABELA 18A	Valores referentes aos nove maiores valores preditos de capacidade específica de combinação (CEC ¹) e seus respectivos valores observados (CEC ²), considerando 2000 simulações, em relação aos nove maiores valores de capacidade específica obtidos na avaliação de vinte e nove híbridos simples de milho.	72
TABELA 19A	Valores referentes aos nove maiores valores preditos de capacidade específica de combinação (CEC ¹) e seus respectivos valores observados (CEC ²), considerando 2000 simulações, em relação aos nove maiores valores de capacidade específica obtidos na avaliação de vinte e nove híbridos simples de milho.	73
TABELA 20A	Valores referentes aos nove maiores valores preditos de capacidade específica de combinação (CEC ¹) e seus	

	respectivos valores observados (CEC ²), considerando 2000 simulações, em relação aos nove maiores valores de capacidade específica obtidos na avaliação de vinte e nove híbridos simples de milho.....	74
TABELA 21A	Valores referentes à soma do erro de predição (PRES) de acordo com os diferentes coeficientes de similaridade no estado e parentesco, em relação aos valores genotípicos preditos e seus respectivos valores observados para produtividade de grãos (PROD) e capacidade específica de combinação (CEC).....	75

- FIGURA 1A Distribuição freqüência de correlações entre valores preditos e observados, considerando desbalanceamento de 3 híbridos. Coeficiente de similaridade de Sorensen-Dice. (Média: 0,49; Mediana: 0,69; Moda: 0,99 e Desvio Padrão: 0,58). 76
- FIGURA 2A Distribuição freqüência de correlações entre valores preditos e observados, considerando desbalanceamento de 3 híbridos. Coeficiente de parentesco de Lynch & Ritland (1999). (Média: 0,46; Mediana: 0,69; Moda: 0,99 e Desvio Padrão: 0,58). 76

TABELA 1A Caracterização das nove linhagens de milho disponíveis nesse estudo. Geneseds Recursos Genéticos Ltda. Lavras, MG, 2010.

Linhagem	População Original	Background	Maturidade	Textura grãos	Cor grãos
L30	POP GNS 10	Cateto/(Tuxpeno/Caribbean)	T	F+	A1
L31	POP GNS 10	Cateto/(Tuxpeno/Caribbean)	M	F+	A1
L32	POP GNS 30	Eto/Cateto	T	F+	A2
L37	POP GNS 10	Cateto/(Tuxpeno/Caribbean)	M	F-	A3
L63	POP GNS 20	Tuxpeno	P	F+	A1
		Nicarilho/Autopionn			
L12	POP GNS 50	Tuxpeno	T	D-	A3
L43	POP GNS 70	Tuxpeno	T	MD	A2
L57	POP USA 80	B73/Tuxpeno	P	D-	A3
L84	POP GNS 60	Tuxpeno/Autoembrapa	P	MD	A2

Maturidade: P (precoce), M (médio) e T (tardio).

Textura de grãos: D- (denatdo), MD (meio dente), F+ (muito duro) e F- (duro).

Cor de grãos: A1 (vermelho), A2 (laranja) e A3 (amarelo).

TABELA 2A Descrição dos *primers* utilizados para análises de marcadores microssatélites.

<i>Primers</i>	<i>Primers</i>	<i>Primers</i>	<i>Primers</i>	<i>Primers</i>	<i>Primers</i>
UMC 1025	UMC 1550	UMC 1762	BNLG 1017	BNLG 1360	BNLG 1805
UMC 1040	UMC 1576	UMC 1792	BNLG 1055	BNLG 1521	BNLG 1810
UMC 1042	UMC 1614	UMC 1812	BNLG 1057	BNLG 1526	BNLG 1863
UMC 1066	UMC 1630	UMC 1827	BNLG 1082	BNLG 1583	BNLG 1890
UMC 1112	UMC 1653	BNLG 0197	BNLG 1108	BNLG 1600	BNLG 2057
UMC 1426	UMC 1659	BNLG 0238	BNLG 1194	BNLG 1662	BNLG 2235
UMC 1528	UMC 1737	BNLG 0252	BNLG 1208	BNLG 1700	BNLG 2248
UMC 1532	UMC 1738	BNLG 0657	BNLG 1272	BNLG 1721	BNLG 2305

TABELA 3A Coeficientes de similaridade no estado utilizados para análises de dados genéticos obtidos com marcadores microsatélites.

Coeficientes	Expressão
Sorense-Dice (SD)	$^{(1)} \frac{2a}{2a + b + c}$
Jaccard (J)	$\frac{a}{a + b + c}$
Simple Matching (SM)	$\frac{a + d}{a + b + c + d}$
Distância Modificada de Rogers (DMR)	$^{(2)} dij = \sqrt{\frac{1}{2n} \sum_k (x_{ki} - x_{kj})^2}$

⁽¹⁾ a = 1 e 1; b = 1 e 0; c = 0 e 1; d = 0 e 0;

⁽²⁾ n = número de locos; x_{ki} e x_{kj} = frequência do k-ésimo alelo das linhagens i e j;

TABELA 4A Coeficientes de parentesco utilizados para análises de dados genéticos obtidos com marcadores microsatélites.

Coeficientes	Expressão
Lynch-Ritland (LR)	$\hat{r}_{xy(k)} = \frac{p_a(S_{bc} + S_{bd}) + p_b(S_{ac} + S_{ad}) - 4p_a p_b}{(1 + S_{ab})(p_a + p_b) + 2p_a p_b}$
Queller-Goodnight (QG)	$\hat{r}_{xy} = \frac{0,5(S_{ac} + S_{ad} + S_{bc} + S_{bd}) - p_a - p_b}{1 + S_{ab} - p_a - p_b}$
Ritland (R)	$\hat{r}_{xy} = \frac{2(S'_{xy} - J_0)}{1 - J_0}$
Wang (W)	$\hat{r}_{xy} = \frac{(1 + S_{ab})(p_a + p_b) - 4p_a p_b}{2p_a p_b}$

$\hat{r}_{xy(k)}$ é o estimador do parentesco genético aditivo entre o indivíduo X (tomado como referência) com alelos a e b e o indivíduo Y com alelos c e d no loco k;

S_{ab}: igual a 1 se a e b forem idênticos e igual a 0, caso contrário; S_{ac}: igual a 1 se a e c forem idênticos e igual a 0, caso contrário; S_{ad}: igual a 1 se a e d forem idênticos e igual a 0, caso contrário; S_{bc}: igual a 1 se b e c forem idênticos e igual a 0, caso contrário; S_{bd}: igual a 1 se b e d forem idênticos e igual a 0, caso contrário; p_a e p_b são as frequências alélicas dos alelos a e b ao longo das linhagens para um dado loco k;

S'_{xy} = 1,0 (x=ii; y=ii) ou (x=ij; y=ij); S'_{xy} = 0,75 (x=ii; y=ij); S'_{xy} = 0,5 (x=ij; y=ik) e S'_{xy} = 0,0 (x=ij; y=kl);

J₀ é a homozigotidade esperada, obtida por $J_0 = \sum_{i=1}^n p_i^2$, sendo n o número de locos.

TABELA 5A Estimativas de similaridade no estado entre nove linhagens de milho obtidas por meio do coeficiente de similaridade de Sorensen-Dice (acima da diagonal) e de Jaccard (abaixo da diagonal).

	L30	L31	L32	L37	L63	L12	L43	L57	L84	
L30	1,000	0,250	0,267	0,143	0,500	0,375	0,250	0,333	0,286	
L31	0,143	1,000	0,211	0,667	0,143	0,100	0,300	0,250	0,111	
L32	0,154	0,118	1,000	0,118	0,462	0,105	0,105	0,267	0,118	
L37	0,077	0,500	0,063	1,000	0,167	0,222	0,333	0,286	0,125	
L63	0,333	0,077	0,300	0,091	1,000	0,000	0,125	0,000	0,000	
L12	0,231	0,053	0,056	0,125	0,000	1,000	0,500	0,300	0,444	
L43	0,143	0,176	0,056	0,200	0,067	0,333	1,000	0,300	0,444	
L57	0,200	0,143	0,154	0,167	0,000	0,176	0,176	1,000	0,429	
L84	0,167	0,059	0,063	0,067	0,000	0,286	0,286	0,273	1,000	
Sorensen-Dice					Jaccard					
σ_{cgc}^2	0,390								σ_{cgc}^2	0,361
σ_{cec}^2	0,669								σ_{cec}^2	0,617
σ_e^2	0,743								σ_e^2	0,742
h^2	0,637								h^2	0,618

TABELA 6A Estimativas de similaridade no estado entre nove linhagens de milho obtidas por meio da Distância Modificada de Rogers (acima da diagonal) e coeficiente de similaridade de Simple Matching (abaixo da diagonal).

	L30	L31	L32	L37	L63	L12	L43	L57	L84	
L30	1,000	0,646	0,661	0,646	0,750	0,677	0,646	0,711	0,677	
L31	0,647	1,000	0,605	0,750	0,646	0,567	0,618	0,646	0,592	
L32	0,645	0,605	1,000	0,605	0,730	0,579	0,579	0,661	0,605	
L37	0,586	0,833	0,545	1,000	0,677	0,618	0,646	0,677	0,618	
L63	0,769	0,600	0,741	0,600	1,000	0,592	0,618	0,646	0,618	
L12	0,706	0,561	0,553	0,611	0,529	1,000	0,646	0,618	0,677	
L43	0,647	0,659	0,553	0,667	0,588	0,750	1,000	0,618	0,677	
L57	0,692	0,636	0,633	0,643	0,538	0,650	0,650	1,000	0,711	
L84	0,667	0,568	0,559	0,563	0,533	0,730	0,730	0,724	1,000	
Distância Modificada Rogers					Simple Matching					
σ_{cgc}^2	0,751								σ_{cgc}^2	0,748
σ_{cec}^2	0,816								σ_{cec}^2	0,855
σ_e^2	0,751								σ_e^2	0,753
h^2	0,720								h^2	0,725

TABELA 7A Estimativas de parentesco entre nove linhagens de milho obtidas por meio do coeficiente de parentesco proposto por Lynch & Ritland (1999) (acima da diagonal) e Queller & Goodnight (1989) (abaixo da diagonal).

	L30	L31	L32	L37	L63	L12	L43	L57	L84
L30	1,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
L31	0,000	1,000	0,000	0,147	0,000	0,000	0,000	0,000	0,000
L32	0,000	0,000	1,000	0,000	0,000	0,000	0,000	0,000	0,000
L37	0,000	0,180	0,000	1,000	0,000	0,000	0,000	0,000	0,000
L63	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	0,000
L12	0,005	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,033
L43	0,000	0,000	0,000	0,000	0,000	0,023	1,000	0,002	0,000
L57	0,000	0,000	0,000	0,000	0,000	0,000	0,015	1,000	0,000
L84	0,000	0,000	0,000	0,000	0,000	0,024	0,000	0,000	1,000
Lynch & Ritland (1999)							Queller & Goodnight (1989)		
σ_{cgc}^2	0,325					σ_{cgc}^2	0,327		
σ_{cec}^2	0,583					σ_{cec}^2	0,583		
σ_e^2	0,754					σ_e^2	0,754		
h^2	0,598					h^2	0,599		

TABELA 8A Estimativas de parentesco entre nove linhagens de milho obtidas por meio do coeficiente de parentesco proposto por Ritland (2006) (acima da diagonal) e Wang (2002) (abaixo da diagonal).

	L30	L31	L32	L37	L63	L12	L43	L57	L84
L30	1,000	0,171	0,190	0,153	0,192	0,226	0,165	0,172	0,039
L31	0,000	1,000	0,176	0,582	0,244	0,066	0,125	0,059	0,128
L32	0,000	0,000	1,000	0,176	0,231	0,201	0,218	0,191	0,093
L37	0,000	0,038	0,000	1,000	0,164	0,156	0,207	0,187	0,145
L63	0,000	0,000	0,000	0,000	1,000	0,108	0,097	0,067	0,121
L12	0,000	0,000	0,000	0,000	0,000	1,000	0,312	0,279	0,394
L43	0,000	0,000	0,000	0,000	0,000	0,000	1,000	0,328	0,186
L57	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000	0,168
L84	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
	Ritland								
								Wang	
σ_{cgc}^2	0,408							σ_{cgc}^2	0,319
σ_{cec}^2	0,638							σ_{cec}^2	0,584
σ_e^2	0,746							σ_e^2	0,756
h^2	0,633							h^2	0,597

TABELA 9A Parâmetros relacionados aos valores de correlação obtidos em 2000 simulações, em diferentes níveis de desbalanceamento, referentes à produtividade de híbridos simples não avaliados e seus valores observados.

SD					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,457	0,494	0,479	0,462	0,439
Mediana	0,694	0,556	0,512	0,492	0,468
Desvio Padrão	0,582	0,296	0,228	0,204	0,191
Moda	0,990	0,700	0,620	0,560	0,480

J					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,495	0,505	0,492	0,478	0,449
Mediana	0,728	0,569	0,525	0,501	0,468
Desvio Padrão	0,562	0,298	0,222	0,197	0,185
Moda	0,990	0,650	0,580	0,570	0,500

DMR					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,493	0,503	0,486	0,471	0,442
Mediana	0,722	0,551	0,514	0,492	0,466
Desvio Padrão	0,558	0,286	0,209	0,187	0,175
Moda	0,990	0,650	0,570	0,550	0,520

SM					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,434	0,472	0,471	0,453	0,435
Mediana	0,667	0,533	0,516	0,479	0,463
Desvio Padrão	0,594	0,303	0,229	0,202	0,188
Moda	0,990	0,620	0,560	0,500	0,540

SD: Sorensen-Dice; J: Jaccard; DMR: Distância Modificada de Rogers e SM: Simple Matching.

TABELA 10A Parâmetros relacionados aos valores de correlação obtidos em 2000 simulações, em diferentes níveis de desbalanceamento, referentes à produtividade de híbridos simples não avaliados e seus valores observados.

LR					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,455	0,483	0,454	0,414	0,358
Mediana	0,692	0,522	0,476	0,429	0,371
Desvio Padrão	0,585	0,287	0,208	0,182	0,172
Moda	0,990	0,540	0,510	0,450	0,410
QG					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,461	0,479	0,456	0,415	0,366
Mediana	0,708	0,529	0,482	0,434	0,381
Desvio Padrão	0,583	0,295	0,214	0,182	0,171
Moda	0,980	0,590	0,520	0,460	0,400
R					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,445	0,448	0,434	0,409	0,373
Mediana	0,674	0,499	0,464	0,433	0,391
Desvio Padrão	0,583	0,315	0,234	0,208	0,191
Moda	0,990	0,550	0,460	0,490	0,450
W					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,445	0,475	0,465	0,427	0,375
Mediana	0,682	0,525	0,481	0,444	0,388
Desvio Padrão	0,585	0,298	0,205	0,181	0,167
Moda	0,980	0,570	0,510	0,430	0,410

LR: Lynch-Ritland; QG: Queller-Goodnight; R: Ritland e W: Wang.

TABELA 11A Parâmetros relacionados aos valores de correlação obtidos em 2000 simulações em diferentes níveis de desbalanceamento referentes à capacidade específica de combinação (CEC) de híbridos simples não avaliados e seus valores observados.

SD					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,291	0,373	0,388	0,374	0,343
Mediana	0,563	0,441	0,431	0,412	0,374
Desvio Padrão	0,679	0,384	0,273	0,232	0,194
Moda	0,980	0,650	0,540	0,490	0,440
J					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,370	0,433	0,447	0,401	0,369
Mediana	0,628	0,515	0,478	0,432	0,384
Desvio Padrão	0,643	0,368	0,264	0,223	0,193
Moda	0,980	0,660	0,640	0,490	0,520
DMR					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,112	0,155	0,183	0,178	0,164
Mediana	0,231	0,208	0,227	0,223	0,191
Desvio Padrão	0,704	0,413	0,315	0,263	0,222
Moda	0,980	0,260	0,380	0,290	0,270
SM					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,145	0,171	0,196	0,217	0,206
Mediana	0,343	0,235	0,253	0,271	0,239
Desvio Padrão	0,711	0,445	0,328	0,261	0,214
Moda	0,970	0,450	0,350	0,370	0,330

TABELA 12A Parâmetros relacionados aos valores de correlação obtidos em 2000 simulações em diferentes níveis de desbalanceamento referentes à capacidade específica de combinação (CEC) de híbridos simples não avaliados e seus valores observados.

LR					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,365	0,269	0,185	0,141	0,105
Mediana	0,585	0,266	0,157	0,098	0,075
Desvio Padrão	0,624	0,349	0,255	0,209	0,182
Moda	0,990	0,270	0,065	0,024	0,002
QG					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,399	0,307	0,243	0,194	0,151
Mediana	0,655	0,316	0,237	0,173	0,125
Desvio Padrão	0,612	0,342	0,267	0,211	0,188
Moda	0,980	0,160	0,096	0,037	0,009
R					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,401	0,384	0,331	0,276	0,215
Mediana	0,645	0,421	0,338	0,287	0,221
Desvio Padrão	0,618	0,346	0,256	0,218	0,195
Moda	0,980	0,430	0,320	0,280	0,200
W					
Número de híbridos retirados					
	3	6	9	12	15
Média	0,348	0,216	0,156	0,107	0,084
Mediana	0,528	0,203	0,126	0,073	0,046
Desvio Padrão	0,606	0,359	0,279	0,224	0,193
Moda	0,980	0,110	0,048	0,012	0,001

TABELA 13A Valores referentes às produtividades médias observadas para cada cruzamento, obtidos através de avaliações em dois locais (Campus da Universidade Federal de Lavras e Fazenda Vitorinha. Lavras, Minas Gerais. Safra 2008/2009).

Ordem	Cruzamento	*M.A.	Ordem	Cruzamento	*M.A.
1	L32XL43	11,547	16	L30XL43	8,861
2	L63XL84	10,348	17	L43XL84	8,857
3	L32XL57	10,249	18	L30XL84	8,824
4	L32XL84	10,017	19	L12XL37	8,718
5	L12XL32	9,798	20	L12XL30	8,475
6	L37XL43	9,359	21	L31XL57	8,355
7	L37XL84	9,283	22	L12XL63	8,311
8	L30XL32	9,276	23	L31XL84	8,283
9	L43XL63	9,134	24	L31XL32	8,259
10	L32XL37	9,114	25	L43XL57	8,106
11	L30XL57	9,113	26	L43XL84	7,618
12	L30XL37	9,041	27	L30XL31	7,487
13	L31XL43	9,039	28	L12XL43	5,902
14	L37XL57	9,004	29	L12XL84	5,273
15	L57XL63	8,984	Média		8,780

* M.A. : Médias ajustadas obtidas em avaliações em dois locais. Lavras, MG. 2009.

TABELA 14A Valores referentes aos nove maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 2000 simulações, em relação às nove maiores produtividades obtidas na avaliação de vinte nove híbridos simples de milho.

SD				
Ordem	Cruzamento	VG	Média Ajustada	PRES
1	L32XL37	14,436	9,114	5,738
2	L31XL32	14,322	8,259	6,451
3	L57XL63	13,731	8,984	4,427
4	L30XL32	13,346	9,276	4,772
5	L32XL57	13,211	10,249	3,327
6	L63XL84	12,084	10,348	2,207
7	L43XL63	11,794	9,134	3,161
8	L32XL84	11,375	10,017	2,418
9	L32XL43	10,235	11,548	2,488
Média		12,726	9,658	3,379
ES%	65,00			
J				
Ordem	Cruzamento	VG	Média Ajustada	PRES
1	L32XL37	14,956	9,114	6,542
2	L30XL32	13,348	9,276	4,922
3	L31XL32	13,347	8,259	5,961
4	L32XL57	13,140	10,249	3,456
5	L57XL63	12,779	8,984	3,887
6	L32XL84	10,939	10,017	2,353
7	L43XL63	10,933	9,134	2,632
8	L63XL84	10,918	10,348	1,680
9	L32XL43	10,310	11,547	2,406
Média		12,296	9,658	3,323
ES%	65,00			

TABELA 15A Valores referentes aos nove maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 2000 simulações, em relação às nove maiores produtividades obtidas na avaliação de vinte e nove híbridos simples de milho.

DMR				
Ordem	Cruzamento	VG	Média Ajustada	PRES
1	L32XL37	16,226	9,114	6,515
2	L30XL32	14,353	9,276	4,984
3	L31XL32	14,144	8,259	6,191
4	L32XL57	13,909	10,249	3,620
5	L57XL63	13,428	8,984	4,515
6	L32XL84	11,673	10,017	2,348
7	L43XL63	11,630	9,134	2,977
8	L63XL84	11,563	10,348	2,429
9	L32XL43	11,552	11,547	2,153
Média		13,164	9,658	3,239
ES %	65,00			
SM				
Ordem	Cruzamento	VG	Média Ajustada	PRES
1	L32XL37	14,148	9,114	5,506
2	L31XL32	14,052	8,259	6,446
3	L30XL32	12,824	9,276	4,322
4	L57XL63	12,706	8,984	3,794
5	L32XL57	12,522	10,249	2,759
6	L63XL84	11,320	10,348	2,081
7	L43XL63	10,617	9,134	2,587
8	L32XL84	10,438	10,017	2,023
9	L32XL43	9,772	11,547	2,656
Média		12,044	9,658	3,302
ES %	65,00			

TABELA 16A Valores referentes aos nove maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 2000 simulações, em relação às nove maiores produtividades obtidas na avaliação de vinte e nove híbridos simples de milho.

LR				
Ordem	Cruzamento	VG	Média Ajustada	PRES
1	L32XL37	14,570	9,114	6,106
2	L30XL32	13,094	9,276	4,355
3	L32XL57	12,850	10,249	3,427
4	L31XL32	12,528	8,259	4,869
5	L57XL63	12,096	8,984	3,604
6	L43XL57	11,284	8,106	3,658
7	L37XL57	10,999	9,004	2,233
8	L32XL84	10,600	10,017	2,351
9	L32XL43	10,525	11,547	2,602
Média		12,060	9,395	3,102
ES %	42,00			
QG				
Ordem	Cruzamento	VG	Média Ajustada	PRES
1	L32XL37	15,306	9,114	6,263
2	L30XL32	13,062	9,276	4,486
3	L32XL57	12,854	10,249	3,527
4	L31XL32	12,548	8,259	5,124
5	L57XL63	12,096	8,984	3,433
6	L43XL57	11,232	8,106	3,768
7	L37XL57	10,686	9,004	2,318
8	L32XL84	10,503	10,017	2,326
9	L32XL43	10,446	11,547	2,621
Média		12,081	9,395	3,138
ES %	42,00			

TABELA 17A Valores referentes aos nove maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 2000 simulações, em relação às nove maiores produtividades obtidas na avaliação de vinte e nove híbridos simples de milho.

R				
Ordem	Cruzamento	VG	Média Ajustada	PRES
1	L32XL37	14,702	9,114	6,187
2	L31XL32	13,198	8,259	5,778
3	L30XL32	13,039	9,276	4,635
4	L32XL57	12,998	10,249	3,341
5	L57XL63	11,919	8,984	3,084
6	L43XL57	10,790	8,106	3,925
7	L43XL63	10,596	9,134	2,608
8	L32XL43	10,443	11,547	2,775
9	L63XL84	10,007	10,348	1,400
Média		11,966	9,446	3,244
ES %	53,00			
W				
Ordem	Cruzamento	VG	Média Ajustada	PRES
1	L32XL37	15,636	9,114	6,770
2	L30XL32	13,138	9,276	4,636
3	L32XL57	12,830	10,249	3,303
4	L31XL32	12,499	8,259	5,029
5	L57XL63	12,174	8,984	3,614
6	L43XL57	11,343	8,106	3,835
7	L37XL57	10,994	9,004	2,519
8	L32XL84	10,631	10,017	2,235
9	L32XL43	10,264	11,547	2,600
Média		12,168	9,395	3,156
ES %	42,00			

TABELA 18A Valores referentes aos nove maiores valores preditos de capacidade específica de combinação (CEC¹) e seus respectivos valores observados (CEC²), considerando 2000 simulações, em relação aos nove maiores valores de capacidade específica obtidos na avaliação de vinte e nove híbridos simples de milho.

SD				
Ordem	Cruzamento	CEC¹	CEC²	PRES
1	L43XL63	0,665	0,071	0,594
2	L57XL63	0,633	-0,135	0,715
3	L32XL84	0,617	0,549	0,444
4	L12XL63	0,609	-0,044	0,554
5	L32XL57	0,556	0,453	0,389
6	L12XL32	0,444	0,699	0,533
7	L30XL32	0,389	-0,061	0,426
8	L63XL84	0,384	0,649	0,451
9	L32XL43	0,305	1,361	1,174
Média		0,511	0,394	0,618
ES %	53,00			
J				
Ordem	Cruzamento	CEC¹	CEC²	PRES
1	L57XL63	0,446	-0,098	0,563
2	L32XL84	0,346	0,485	0,357
3	L43XL63	0,345	0,044	0,343
4	L32XL57	0,320	0,002	0,279
5	L12XL63	0,279	0,399	0,260
6	L12XL32	0,260	0,677	0,532
7	L30XL32	0,200	-0,132	0,323
8	L32XL43	0,189	1,262	1,127
9	L63XL84	0,168	0,659	0,541
Média		0,284	0,366	0,542
ES %	53,00			

TABELA 19A Valores referentes aos nove maiores valores preditos de capacidade específica de combinação (CEC¹) e seus respectivos valores observados (CEC²), considerando 2000 simulações, em relação aos nove maiores valores de capacidade específica obtidos na avaliação de vinte e nove híbridos simples de milho.

DMR				
Ordem	Cruzamento	CEC¹	CEC²	PRES
1	L32XL57	0,678	0,450	0,715
2	L30XL32	0,555	-0,061	0,835
3	L32XL37	0,497	-0,253	1,016
4	L57XL63	0,477	-0,108	0,764
5	L32XL84	0,373	0,499	0,432
6	L43XL63	0,328	0,038	0,462
7	L63XL84	0,317	0,689	0,511
8	L31XL32	0,295	-0,543	0,961
9	L32XL43	0,281	1,299	1,090
Média		0,422	0,223	0,681
ES %	42,00			
SM				
Ordem	Cruzamento	CEC¹	CEC²	PRES
1	L32XL57	0,583	0,458	0,599
2	L57XL63	0,448	-0,177	0,786
3	L30XL32	0,447	-0,060	0,691
4	L32XL37	0,378	-0,252	0,848
5	L63XL84	0,323	0,650	0,485
6	L32XL84	0,309	0,513	0,440
7	L43XL63	0,273	0,011	0,413
8	L31XL32	0,271	-0,628	1,064
9	L32XL43	0,150	1,355	1,242
Média		0,353	0,208	0,757
ES %	42,00			

TABELA 20A Valores referentes aos nove maiores valores preditos de capacidade específica de combinação (CEC¹) e seus respectivos valores observados (CEC²), considerando 2000 simulações, em relação aos nove maiores valores de capacidade específica obtidos na avaliação de vinte e nove híbridos simples de milho.

LR				
Ordem	Cruzamento	CEC¹	CEC²	PRES
1	L57XL63	0,345	-0,068	0,516
2	L63XL84	0,031	0,664	0,632
3	L31XL84	0,027	0,075	0,066
4	L37XL43	0,025	0,203	0,192
5	L12XL31	0,022	0,027	0,032
6	L31XL43	0,018	0,361	0,348
7	L32XL84	0,012	0,409	0,400
8	L37XL84	0,010	0,325	0,325
9	L12XL32	0,001	0,631	0,625
Média		0,055	0,292	0,533
ES %	53,00			
QG				
Ordem	Cruzamento	CEC¹	CEC²	PRES
1	L57XL63	0,359	-0,070	0,525
2	L37XL43	0,034	0,211	0,196
3	L31XL84	0,033	0,073	0,064
4	L63XL84	0,032	0,663	0,633
5	L12XL31	0,030	0,028	0,034
6	L31XL43	0,022	0,364	0,348
7	L12XL32	0,020	0,633	0,617
8	L32XL57	0,013	0,326	0,317
9	L32XL43	0,012	1,169	1,161
Média		0,062	0,377	0,532
ES %	53,00			
R				
Ordem	Cruzamento	CEC¹	CEC²	PRES
1	L57XL63	0,436	-0,059	0,569
2	L32XL57	0,394	0,404	0,318
3	L43XL63	0,247	0,038	0,235
4	L12XL32	0,176	0,656	0,565
5	L63XL84	0,148	0,707	0,605
6	L30XL32	0,146	-0,146	0,307
7	L32XL43	0,139	1,213	1,116
8	L32XL84	0,138	0,503	0,443
9	L31XL84	0,112	0,092	0,146
Média		0,215	0,379	0,555
ES %	53,00			

TABELA 21A Valores referentes à soma do erro de predição (PRES) de acordo com os diferentes coeficientes de similaridade no estado e parentesco, em relação aos valores genotípicos preditos e seus respectivos valores observados para produtividade de grãos (PROD) e capacidade específica de combinação (CEC).

SIMILARIDADE NO ESTADO				
PRES				
	SD	J	DMR	SM
PROD	3,379	3,323	3,239	3,302
CEC	0,618	0,542	0,681	0,757
PARENTESCO				
PRES				
	LR	QG	R	W
PROD	3,102	3,138	3,244	3,156
CEC	0,533	0,532	0,555	-

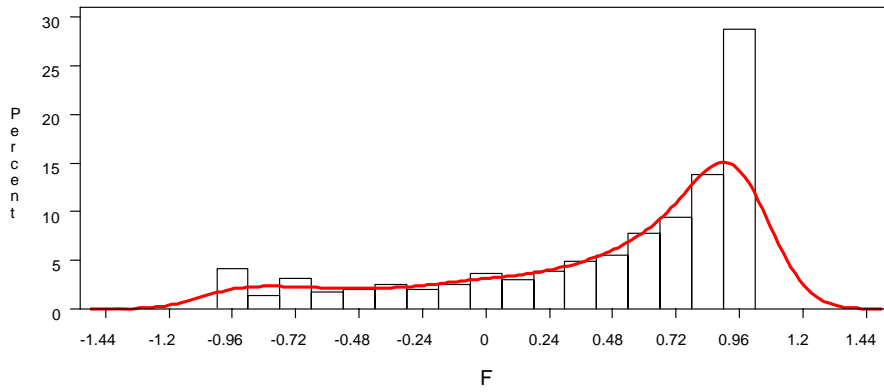


FIGURA 1A Distribuição freqüência de correlações entre valores preditos e observados, considerando desbalanceamento de 3 híbridos. Coeficiente de similaridade de Sorensen-Dice. (Média: 0,49; Mediana: 0,69; Moda: 0,99 e Desvio Padrão: 0,58).

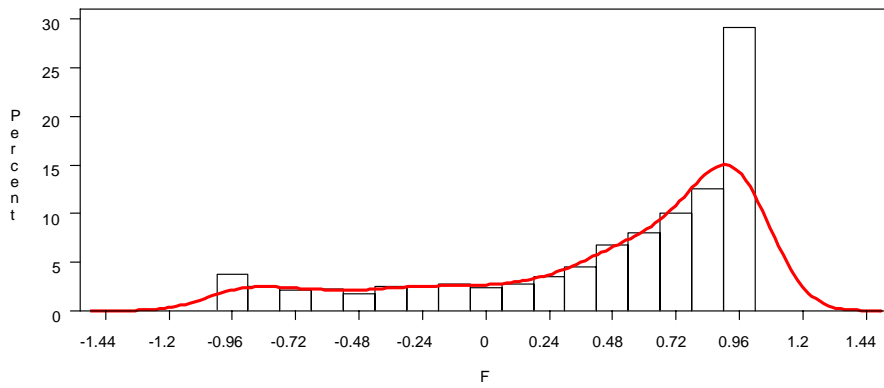


FIGURA 2A Distribuição freqüência de correlações entre valores preditos e observados, considerando desbalanceamento de 3 híbridos. Coeficiente de parentesco de Lynch & Ritland (1999). (Média: 0,46; Mediana: 0,69; Moda: 0,99 e Desvio Padrão: 0,58).