

**PREDIÇÃO DO DESEMPENHO DE
HÍBRIDOS DE MILHO UTILIZANDO
MODELOS LINEARES MISTOS COM A
INFORMAÇÃO DE MARCADORES
MOLECULARES**

MARCIO BALESTRE

2009

MARCIO BALESTRE

**PREDIÇÃO DO DESEMPENHO DE HÍBRIDOS DE
MILHO UTILIZANDO MODELOS LINEARES
MISTOS COM A INFORMAÇÃO DE MARCADORES
MOLECULARES**

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

Orientador

Prof. Dr. Renzo Garcia Von Pinho

LAVRAS
MINAS GERAIS - BRASIL
2009

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

Balestre, Marcio.

Predição do desempenho de híbridos de milho utilizando
modelos lineares mistos com a informação de marcadores
moleculares / Marcio Balestre. – Lavras : UFLA, 2009.
64 p. : il.

Dissertação (Mestrado) – Universidade Federal de Lavras, 2009.

Orientador: Renzo Garcia Von Pinho.

Bibliografia.

1. BLUP. 2. Marcadores de microssatélites. 3. Parentesco. 4.
Acurácia genética I. Universidade Federal de Lavras. II. Título.

CDD – 633.153
– 581.3

MARCIO BALESTRE

**PREDIÇÃO DO DESEMPENHO DE HÍBRIDOS DE
MILHO UTILIZANDO MODELOS LINEARES
MISTOS COM A INFORMAÇÃO DE MARCADORES
MOLECULARES**

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

APROVADA 20 de fevereiro de 2009

Prof. Dr. Renato Ribeiro de Lima

UFLA

Prof. Dr. João Cândido de Souza

UFLA

Prof. Dr. Renzo Garcia Von Pinho

UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

A Deus,

por ser o Único digno de toda honra e Glória;

OFEREÇO

Aos meus pais, Laércio Balestre (*in memoriam*) e Maria Elena de Oliveira Balestre, pelos valiosos conselhos e incentivo;

Ao amigo Hamilton Peixe (*in memmoriám*),

DEDICO

AGRADECIMENTOS

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

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela concessão da bolsa de pesquisa.

Ao Prof. Renzo Garcia Von Pinho, pela orientação, amizade, incentivo e confiança depositada para a condução deste trabalho e, em especial, por me apoiar em um momento difícil durante a realização do curso.

Ao Prof. João Cândido de Souza, que me orientou durante a iniciação científica e participou da elaboração deste projeto.

Ao Prof. Julio Silvio de Souza Bueno Filho, pela ajuda e apoio durante a condução deste trabalho e ao Prof. Renato Ribeiro de Lima, pelas valiosas sugestões apresentadas.

À professora Anete Pereira de Souza e a Prianda Rios Laborta, do CBMEG/Unicamp, por disponibilizar os *primers* de SSR.

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

Aos funcionários do Departamento de Biologia, Elaine, Erondina, Heloisa, Léo, Lindolfo e, em especial, ao Lamartine pela amizade e grande auxílio no Laboratório de Genética Molecular.

A minha noiva, Fernanda, por fazer parte da minha vida e, a cada dia, me transformar em uma pessoa melhor e mais feliz.

As minhas maravilhosas irmãs, Rúbia e Vanessa, pelo incentivo e por me fazerem sentir o melhor irmão do mundo, ainda que não seja.

A todos os colegas do GEN e, em especial, aos amigos Rogerão, Léo, Helton, Jacaré, Passatempo, Reginaldo, Zé Luiz, Juarez, Rodela, Flávio (o Fino), Thaís, Joyce e Ester. E também aos colegas de “Brejão”, Fabão, Natalino, Eric, Adriano e Anselmo. A todos vocês, muito obrigado.

SUMÁRIO

	Página
RESUMO	i
ABSTRACT	iii
1 Introdução.....	1
2 Referências Bibliográficas.....	4
CAPÍTULO 1 Potencial do uso de marcadores moleculares na predição de valores genotípicos e no desempenho de híbridos de milho.....	6
1.1 Resumo	6
1.2 Introdução.....	7
1.3 Material e Métodos.....	10
1.4 Resultados.....	15
1.5 Discussões.....	18
2 Referências bibliográficas	26
3 Tabelas	30
4 Figuras	35
CAPÍTULO 2 Predição do desempenho de híbridos de milho utilizando modelos lineares mistos com informação de marcadores de microssatélites.....	37
1.1 Resumo	37
1.2 Introdução.....	38
1.3 Materiais e Métodos.....	41
1.4 Resultados.....	47
1.5 Discussões.....	50
2 Referências Bibliográficas.....	56
3 Tabelas	59
4 Figura	64

RESUMO

BALESTRE, Marcio. **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, MG. *

Nos últimos anos, várias metodologias têm sido empregadas no intuito de avaliar o desempenho de híbridos de milho. O objetivo deste trabalho foi verificar qual o potencial do melhor preditor linear não viciado (BLUP) e das distâncias genéticas na predição de híbridos de milho. Foram utilizadas noventa progênies $S_{0.2}$ provenientes de três populações oriundas de três híbridos simples comerciais. Utilizaram-se 30 progênies da população A (híbrido 30F45), 31 da população B (híbrido Dow 657) e 29 da população C (DKB333B). Todas as 90 progênies foram genotipadas com 25 marcadores de microssatélites (SSR), sendo 9 dessas marcas ligadas à QTL para produtividade de grãos. A partir das análises moleculares, foram obtidas as similaridades genéticas e o parentesco, empregando-se a distância modificada de Roger e o coeficiente de Lynch e Ritland, respectivamente. Com base nas similaridades genéticas, foram selecionadas 17 linhagens que foram inter cruzadas em um sistema de dialelo parcial, com o objetivo de recuperar os cruzamentos mais e menos divergentes. Assim, foram avaliados 60 híbridos em dois locais em látice triplo 8x8. Também foram utilizados os resultados de similaridade e parentesco para a construção da matriz genética aditiva e de dominância que, posteriormente, foram empregadas em associação com o BLUP, para a predição dos valores genotípicos e da CEC dos híbridos não avaliados. Para isso, foram simulados cinco graus de desbalanceamento (5, 10, 20, 30 e 40). Os valores de similaridades genéticas também foram utilizados na predição dos híbridos, ou seja, correlacionaram-se as DMR com a produtividade, CEC e heterose dos híbridos avaliados. Para teste de significância da correlação entre DMR com a produtividade utilizou-se um teste empírico via simulação Monte Carlo. Neste estudo, foi observado que 71% dos híbridos interpopulacionais apresentaram desempenho superior aos híbridos genitores (30F45, Dow657 e DKB333B). Além disso, observou-se que as distâncias genéticas apresentaram média capacidade preditiva da produtividade (0,546), da capacidade de combinação (0,567) e da heterose (0,661). A simulação Monte Carlo apresentou-se como uma alternativa viável para teste de correlação entre distâncias genética e produtividade, principalmente por verificar

* Comitê Orientador: Renzo Garcia Von Pinho-UFLA (Orientador).

se o tamanho da amostra utilizado para estudo de correlação é adequado. No que se refere às predições utilizando o BLUP, pode-se observar que os valores de correlações entre os valores genotípicos preditos e as médias observadas variaram de 0,55 a 0,70, dependendo do grau de desbalanceamento. Resultado semelhante foi observado para as predições da CEC, ou seja, variaram de 0,61 a 0,70. Além disso, constatou-se que o coeficiente de parentesco de Lynch e Ritland associado ao BLUP proporcionou predições mais acuradas, quando comparado às predições obtidas pela similaridade no estado, estimada pela distância modificada de Roger. O método (BLUP) demonstrou ser uma alternativa viável na predição de cruzamentos não realizados, porém, sua eficácia é maior quando coeficientes de similaridade adequados são utilizados para a construção da matriz de parentesco.

ABSTRACT

BALESTRE, Marcio. **Prediction of maize hybrids performance by linear mixed models with molecular markers information**. 2009. 64 p. Dissertation (Master Degree in Genetics and Plant Breeding)-Universidade Federal de Lavras, Lavras, MG *

In the last years, several methodologies have been used to predict the untested maize hybrids performance. Thus, the aim of this work was to ascertain the potential of the best linear unbiased prediction (BLUP) and the genetic distance in the prediction of maize hybrids performance. Ninety $S_{0:2}$ progenies from three populations of three single cross hybrids were utilized being thirty progenies from population A (P30F45 hybrid), thirty-one from population B (DOW657 hybrids) and twenty-nine from population C. The ninety progenies were genotyped with twenty-five SSR markers, with nine QTL's markers linked to grain yield. Starting from the molecular analysis, the genetic similarities and the relationship were obtained by Roger's modified distance and the Lynch and Ritland coefficient respectively. Based on the genetic similarities, 17 progenies were selected and crossed in a partial diallel design, to recover the most and least divergent crosses. Thus, sixty hybrids were evaluated in two environments using 8x8 lattice design with three replications. The similarity and relationship results were used to construct additive and dominance matrixes that were utilized in association with BLUP for prediction of the untested maize hybrids. Five unbalanced degrees were simulate (5, 10, 20, 30 and 40). The genetic similarities values were also used for hybrids predictions, in other words, they were correlated with yield, SCA and heterosis of the appraised hybrids. For significance test of correlation among DMR and yield, the Monte Carlo empirical test was applied. In this study it was observed that 71% of the interpopulation hybrids showed superior performance in relation to the parental hybrids (30F45, Dow657 e DKB333B). In addition, it was observed that the genetic distances presented a medium capacity to predict yield (0.546), specific combining ability (0.567) and heterosis (0.661). The Monte Carlo simulation demonstrated to be a viable alternative for correlation test between genetic distance and yield, mainly for verifying if the sample size is suitable. Considering the BLUP predictions, it can be observed that the correlations between the genotypic values predicted and the phenotypic means varied from 0.55 to 0.70 depending on the unbalanced degree. Similar results were observed

* Guidance Committee: Renzo Garcia Von Pinho (Major Professor).

for the SCA predictions, which varied from 0.61 to 0.70. Besides, it was verified that the Lynch and Ritland coefficient associated with BLUP provided more accurate predictions when compared to the similarity in state predictions estimated by the Roger's modified distance. Thus, it is possible to infer that BLUP methodology demonstrated to be a viable alternative to predict the performance of untested hybrids, however, this efficiency is larger when appropriate similarity coefficients are used to construct the relationship matrix.

1 INTRODUÇÃO

A seleção e a avaliação de linhagens de milho visando à obtenção de híbridos é um processo rotineiro e um dos principais objetivos dos programas de melhoramento dessa cultura. Porém, devido ao grande efeito de dominância presente na produtividade de grãos, o desempenho de um híbrido de milho não deve ser predito com base no desempenho *per se* da linhagem. Sendo assim, é necessário avaliar essas linhagens em cruzamentos (Bernardo, 1994).

Baseado nesse fato, os melhoristas têm utilizado cruzamentos dialélicos para a seleção de linhagens promissoras. No entanto, nas fases iniciais de seleção nas quais a quantidade de genótipos é expressivamente maior, a utilização de dialelos completos em linhagens parcialmente endogâmicas pode tornar-se inexecutável, devido à grande quantidade de genótipos que podem ser obtidos com a utilização de poucos genitores.

Nesse sentido, várias metodologias têm sido propostas para prever o desempenho de cruzamentos visando diminuir do número de genótipos a serem avaliados. Conforme relatado por Dias et al. (2004), nos últimos anos, vários trabalhos têm sido realizados com a finalidade de prever o desempenho de híbridos com base nas distâncias genéticas por meio do uso de marcadores moleculares. Nesse caso, seria possível realizar uma pré-seleção com base apenas na divergência entre os genótipos, de modo a combinar em cruzamentos dialélicos apenas os “genótipos promissores”.

Vários estudos têm sido realizados com o objetivo de prever a heterose, a capacidade de combinação e a produtividade com base na distância genética dos genitores (Melchinger, 1999; Reif et al., 2003; Balestre et al., 2008a). No entanto, os resultados, até o momento, demonstram pouca eficiência e repetibilidade (Dias et al., 2004).

Uma alternativa seria a utilização da metodologia dos modelos lineares mistos (BLUP) para a predição de cruzamentos não avaliados (Bernardo, 1994; 1995; Charcosset et al., 1998). A utilização do melhor preditor linear não viciado (BLUP) tem sido vista como uma metodologia promissora para a predição de híbridos não avaliados (Piepho et al., 2007). Para isso, torna-se necessário o conhecimento da genealogia ou parentesco entre os indivíduos avaliados e indivíduos cujo desempenho deseja-se conhecer (Bernardo, 1994). Porém, nem sempre o melhorista dispõe de tal informação, o que poderia limitar o emprego da técnica.

Para contornar essa limitação, alguns autores têm sugerido a utilização da informação de marcadores moleculares (similaridade alélica no estado) como opção quando não se dispõem de dados de pedigree (Bernardo, 1994; 1995; André, 1999; Balestre et al., 2008b). Segundo Nejavati-Javameri et al. (1997), o emprego da similaridade no estado possui a “vantagem” de se utilizar toda a semelhança alélica existente entre os indivíduos avaliados.

No entanto, a similaridade no estado tende a superestimar o parentesco. Com base neste fato, Bernardo (1994) propõe o uso de um coeficiente de coancestria que utiliza a informação da similaridade genética, no intuito de corrigir possíveis vieses da similaridade genética no estado. Entretanto, essa metodologia somente é válida para dialelos parciais nos quais não ocorrem cruzamentos endogâmicos, ou seja, pode ser utilizada apenas quando já se possui prévia informação a respeito de grupos heteróticos, o que limita sua utilização. Como alternativa a essa restrição, Lynch & Ritland (1999) propõem um procedimento não viesado para estimação conjunta de parentesco por meio de marcadores moleculares, que pode ser utilizado em dialelos completos e parciais. Esse coeficiente tem sido o mais recomendado, pois apresenta menor variância de amostragem quando comparado aos demais coeficientes sugeridos na literatura (Lynch & Ritland, 1999; Bouvet et al., 2008).

Com isso os objetivos deste trabalho foram:

- a) verificar qual o potencial preditivo das distâncias genéticas estimadas via marcadores de microssatélites no desempenho de híbridos provenientes de linhagens $S_{0:2}$;
- b) averiguar o potencial do uso de marcadores moleculares na predição de valores genotípicos;
- c) analisar a aplicabilidade do método Monte Carlo no estudo de correlação entre as distâncias genéticas e a produtividade;
- d) avaliar a eficiência do melhor preditor linear não-viesado (BLUP) e a influência do uso da similaridade no estado e parentesco de Lynch & Ritland (1999), nas predições de híbridos não testados.

2 REFERÊNCIAS BIBLIOGRÁFICAS

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, MG.

BALESTRE, M.; MACHADO, J.C.; LIMA, J.L.; SOUZA, J.C.; FILHO, L.N. Genetics distances estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids **Genetics and Molecular Research**, Ribeirão Preto, v.7, n. 1, p. 65-73, 2008a.

BALESTRE, M.; VON PINHO, R.G.; SOUZA, J.C.; MACHADO J.C. Potential of maize single-cross hybrids for extraction of inbred lines using the mean components and mixed models with microsatellite marker information **Genetics and Molecular Research**, Ribeirão Preto v.7, n. 4, p.1106-1118, 2008b.

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. 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

BOUVET, J.M.; KELLY, B.; SANOU, H.; ALLAL, F. Comparison of marker and pedigree based methods for estimating heritability in an agroforestry population of *Vitellaria paradoxa* C.F. Gaertn. (shea tree) **Genet Resour Crop Evol**, 2008, (In Press).

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, 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 Pretov. 3, n.3, p. 356-368, 2004.

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. p. 99-118, 1999.

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

PIEPHO, H.P.; MOHRING, J.; MELCHINGER, A.E. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, 2007. (In press).

REIF, J. C.; MELCHINGER, A. E.; XIA, X. C.; WARBURTON, M. L.; HOISINGTON, D. A.; VASAL, S. K.; SRINIVASAN, G.; BOHN, M.; FRISCH, M Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. **Crop Science**, Madison, v. 43, n. 4, p. 1275-1282. 2003.

CAPÍTULO 1

O capítulo número 1 será transcrito em formato de artigo e encaminhado para submissão do Periódico Científico **Genetics and Molecular Research** com o seguinte título:

Potencial do uso de marcadores moleculares na predição de valores genéticos e no desempenho de híbridos de milho

Marcio Balestre¹, Renzo Garcia Von Pinho², João Cândido de Souza¹, Rogério Lunezzo de Oliveira¹.

¹Departamento de Biologia, Universidade Federal de Lavras, Lavras, MG, Brasil. (Fax +35 3829 1368/1341; e-mail: marciobalestre@hotmail.com)

²Departamento de Agricultura, Universidade Federal de Lavras. CP 3037. Lavras. MG. Brazil.

1.1 Resumo

Os objetivos deste trabalho foram verificar qual o potencial preditivo das distâncias genéticas estimadas via marcadores de microssatélites no desempenho de híbridos provenientes de linhagens $S_{0.2}$ e avaliar o potencial do uso de marcadores moleculares na predição de valores genotípicos e a aplicabilidade do método Monte Carlo no estudo de correlação entre distâncias genéticas e produtividade. Foram utilizadas 90 progênies $S_{0.2}$ provenientes de três populações oriundas de três híbridos simples comerciais. Utilizaram-se 30 progênies da população A (híbrido 30F45), 31 da população B (híbrido Dow 657) e 29 da população C (DKB333B). Todas as 90 progênies foram

genotipadas com 25 marcadores de microssatélites (SSR), sendo 9 dessas marcas ligadas à QTL para produtividade de grãos. Os dados de similaridade genética foram empregados na construção das matrizes genética aditiva e de dominância que, posteriormente, foram utilizadas para obter o BLUP dos valores de CEC e CGC. As similaridades genéticas também foram correlacionadas com a produtividade, CEC e heterose dos híbridos avaliados. Observou-se que as distâncias genéticas apresentaram moderada capacidade preditiva da produtividade (0,546), da capacidade de combinação (0,567) e da heterose (0,661). A simulação por Monte Carlo apresentou-se como uma alternativa viável para o teste de correlação entre distâncias genética e produtividade. A estimativa de parentesco por marcadores moleculares demonstra ser uma metodologia promissora na predição de valores genéticos utilizando modelos lineares mistos (BLUP), principalmente quando não se dispõe de nenhuma informação de pedigree.

Palavras-chave. BLUP, parentesco, dialelo, acurácia genética.

1.2 Introdução.

A avaliação e a seleção de linhagens visando à obtenção de híbridos de milho constituem um processo rotineiro e um dos principais objetivos em programas de melhoramento dessa cultura. No entanto, a fase de avaliação pode ser considerada a mais onerosa, pois demanda maior quantidade de recursos, mesmo com poucas linhagens a serem testadas. Isso ocorre porque a avaliação do desempenho *per se* das linhagens nem sempre é garantia de obtenção de bons híbridos, sendo necessário avaliar essas linhagens em combinações híbridas (Hallauer et al., 1988; Troyer, 1994).

Os cruzamentos dialélicos podem ser utilizados para avaliação da capacidade geral (CGC) e específica de combinação (CEC) dessas linhagens. Porém, com poucos genótipos é possível obter grande quantidade de cruzamentos, o que torna a técnica pouco utilizada nas fases iniciais dos programas de melhoramento.

Uma das alternativas a essa grande quantidade de genótipos nas fases iniciais de avaliação seria a realização de cruzamentos “top cross”. Contudo, esse método é menos preciso na estimativa da CGC em relação aos dialelos parciais (Araújo 2000). Além disso, não é possível obter as estimativas da CEC entre as linhagens que auxiliam o melhorista na seleção de genótipos superiores.

Outra alternativa seria a realização de uma pré-seleção, ou *screening*, com base nas distâncias genéticas em linhagens parcialmente endogâmicas com objetivo de cruzar apenas os genótipos promissores em cruzamentos dialélicos. A vantagem de se realizar dialelos em linhagens parcialmente endogâmicas está no seu potencial na obtenção de híbridos que podem ser diretamente utilizados, muito antes que essas linhagens estejam com alto grau de endogamia como linhagens S_6 ou S_7 (Carlone & Russel, 1988; Amorim et al., 2006). Além disso, existe alta correlação entre o desempenho dos indivíduos em gerações precoces e gerações avançadas, o que viabiliza ainda mais a avaliação de cruzamentos dialélicos em linhagens com moderada endogamia (Bernardo, 1991).

Segundo Falconer & Mackay (1996), pela teoria da genética quantitativa, existe relação linear entre a distância genética, na presença de locos com dominância, e a heterose. Alguns estudos têm sido realizados com o objetivo de prever a heterose, a capacidade de combinação e a produtividade com base na distância genética dos genitores, conforme destacado por Dias et al. (2004). Além disso, diversos tipos de germoplasma têm sido utilizados, desde linhagens endogâmicas e parcialmente endogâmicas até híbridos simples (Amorim et al., 2006; Guimarães et al., 2007; Balestre et al., 2008a). No entanto,

os resultados até o momento demonstram pouca eficiência e repetibilidade (Dias et al., 2004).

Vários fatores podem contribuir para essa baixa eficiência na predição por meio das distâncias genéticas por marcadores moleculares. Dentre esses fatores, Bernardo (1992) destaca a presença de um pequeno efeito de dominância nos locos que controlam o caráter, baixa herdabilidade, marcadores não ligados à QTL ou mais que 70% dessas marcas não ligadas aos locos quantitativos, além de grande variação nas frequências alélicas dos parentais. Além disso, a escolha do coeficiente de distância genética adequado aos dados e aos objetivos do estudo pode contribuir para a magnitude da correlação entre a distância genética e a heterose (Reif et al., 2005; Balestre et al., 2008b).

No estudo de correlação entre distância genética e produtividade de grãos, normalmente, se utilizam testes de hipóteses, como teste de t de Student, que não são robustos aos desvios de normalidade dos dados ou assimetria da amostra. Essa assimetria dos dados pode ocorrer com as estimativas de distâncias genéticas, acarretando em violação das pressuposições do teste e aumentando a taxa do erro tipo I e erro tipo II, ou seja, de rejeitar uma hipótese verdadeira, ou de aceitar uma hipótese falsa, respectivamente. Com base nisso, testes alternativos ao teste t têm sido propostos para o estudo de correlação por estatística não paramétrica. Como exemplo, pode-se citar um teste frequentemente utilizado, denominado como o teste Z de Mantel (Mantel, 1967). No entanto, vários pesquisadores, em diversas áreas, têm sugerido o teste empírico de correlação entre variáveis via simulação Monte Carlo (May et al., 2001; Fan et al., 2002).

Este trabalho foi realizado com os objetivos de verificar qual o potencial preditivo das distâncias genéticas estimadas via marcadores de microssatélites no desempenho de híbridos provenientes de linhagens $S_{0.2}$ e avaliar o potencial do uso de marcadores moleculares na predição de valores genotípicos e a

aplicabilidade do método Monte Carlo no estudo de correlação entre distância genéticas e produtividade.

1.3 Material e métodos

Descrição dos materiais

Foram utilizadas 90 progênies $S_{0,2}$, provenientes de três populações oriundas de três híbridos simples comerciais. Utilizaram-se 30 progênies da população A (híbrido P30F45), 31 da população B (híbrido Dow 657) e 29 da população C (híbrido DKB333B).

Todas as 90 progênies foram genotipadas com 25 marcadores de microssatélites (SSR), sendo 9 dessas marcas ligadas à QTL para produtividade de grãos e as demais, selecionadas de acordo com resultado de análise “bootstrap” para essas mesmas populações (Amorim et al., 2006). A extração de DNA, o preparo das reações e as estimativas das distâncias genéticas foram descritos no trabalho desenvolvido por Balestre et al. (2008b). Com base nos resultados desses autores, utilizaram-se as distâncias modificadas de Roger (DMR), para a seleção de 17 linhagens, com o objetivo de recuperar as cinco combinações mais e menos divergentes. A distância modificada de Roger é a expressão mais recomendada para esse tipo de análise (Reif et al., 2005), além de possuir valores aceitáveis de distorção e estresse para essas mesmas populações (Balestre et al., 2008b).

A partir das 17 linhagens selecionadas, realizou-se um dialelo parcial (11 x 6) tendo, dentro de cada grupo, as linhagens sido alocadas de forma a recuperar os cruzamentos entre as linhagens mais e menos divergentes.

Delineamento experimental e correção das médias via modelo misto

Devido à pequena quantidade de sementes, dos 66 cruzamentos possíveis no dialelo, apenas 60 foram avaliados, juntamente com quatro testemunhas, em dois locais no município de Lavras (câmpus da Universidade Federal de Lavras e Fazenda Vitorinha), na safra 2007/2008. Das quatro testemunhas utilizadas, uma foi o híbrido DKB199 e as três restantes, os híbridos simples que originaram as populações A (P30F45), população B (Dow 657) e população C (DKB 333B). O delineamento experimental utilizado foi um látice 8x8, com três repetições. Cada parcela foi constituída de 2 linhas de 3m com aplicação de 450 kg.ha⁻¹ de NPK (8-28-16) na semeadura e 100 kg.ha⁻¹ de nitrogênio de cobertura.

O modelo linear misto, na forma matricial, utilizado para análise dos dados, levando em consideração a recuperação da informação interblocos, seguiu a seguinte expressão:

$$y = X\beta + Z_1a + Z_2g + \delta$$

em que y é o vetor de dados observados, β é o vetor dos efeitos fixos (médias de repetições), a é o vetor aleatório dos valores genotípicos, g é o vetor aleatório dos efeitos de blocos dentro de repetições, δ é o erro aleatório e X , Z_1 e Z_2 são as matrizes de incidência relacionadas a β , a e g respectivamente.

As médias dos tratamentos ajustadas para cada local foram estimadas pelo procedimento BLUP, conforme Resende (2002), por meio da seguinte expressão:

$$y = \mu + \hat{a}/h^2$$

em que μ é a média geral do experimento, \hat{a} é o valor genotípico predito + valor referente à recuperação interblocos e h^2 é a herdabilidade do caráter produtividade de grãos.

Segundo Resende (2002), esse estimador é equivalente às médias corrigidas na recuperação interblocos via análise de variância.

Análise dialélica via modelos misto

A partir dos dados corrigidos para cada local realizou-se a análise dialélica conjunta utilizando-se o método IV proposto por Griffing (1956). As análises foram realizadas utilizando-se o software SAS[®] PROC IML System (SAS Institute, 2000). As estimativas dos efeitos fixos e os BLUPs das capacidades de combinação (CEC e CGC) foram obtidos de maneira semelhante à apresentada por Bernardo (1995), proposta para dialelos parciais. No entanto, neste estudo, optou-se por utilizar o modelo de dialelo completo, considerando todos os genótipos dentro de um mesmo grupo.

O modelo linear considerado foi dado por:

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

em que y é o vetor das observações; β é o vetor dos efeitos fixos (média geral, e locais); g é o vetor de efeitos da capacidade geral de combinação; s é o vetor da capacidade específica de combinação; 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 sistema de equações conforme Henderson (1984):

$$\begin{bmatrix} \beta^0 \\ \hat{g} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1Z_1 + A_1^{-1}\gamma_1 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + A_2^{-1}\gamma_2 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix}$$

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

aditiva entre as linhagens. A matriz de parentesco aditiva foi obtida, conforme Lynch & Ritland (1999), pela seguinte expressão:

$$\hat{r}_{xy(k)} = \frac{p_a(S_{bc} + S_{bd}) + p_b(S_{ac} + S_{sd}) - 4p_a p_b}{(1 + S_{ab})(p_a + p_b) + 2p_a p_b}$$

em que:

$\hat{r}_{xy(k)}$ é o estimador do parentesco 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 dos alelos a e b ao longo das linhagens para um dado loco k ;

Considerando que as estimativas de parentescos são obtidas ao longo de muitos locos, é sugerida a utilização de pesos às estimativas, além da alternância dos indivíduos tomados como referência, ou seja, ora utiliza o indivíduo X como referência e ora o indivíduo Y (Lynch & Ritland, 1999). A expressão final, levando em consideração todos os locos, os pesos atribuídos a cada loco e o parentesco estimado para cada loco, é dada por:

$$\hat{r}_{xy} = \frac{\sum_1^L w_{r,x(k)} \hat{r}_{x,y(k)} + \sum_1^L w_{r,y(k)} \hat{r}_{y,x(k)}}{W_{r,x} + W_{r,y}}$$

em que $w_{r,x(k)}$ e $w_{r,y(k)}$, são os pesos para o k-ésimo loco das estimativas de $\hat{r}_{xy(k)}$ e $W_{r,x}$ e $W_{r,y}$ os somatório dos pesos atribuídos a todos os locos envolvidos.

A matriz de parentesco de dominância (A_2) foi construída como sugerido por Henderson (1984) e Van Vleck (1993).

Também foram utilizadas as distâncias de Roger para a construção das matrizes A_1 e A_2 , pela expressão $s_{ij} = 1 - d_{ij}$, em que d_{ij} é o valor da distância modificada de Roger. Com isso, foram comparadas as duas estimativas de similaridade para verificar qual a influência do uso da similaridade em estado (Distâncias de Roger) e similaridade por descendência (LR), na acurácia das predições dos valores de CGC CEC. As estimativas de acurácia da CGC e CEC foram obtidas por:

$$r_{CGC} = \sqrt{(1 - d_i \gamma_1)} \quad r_{CEC} = \sqrt{(1 - d_i \gamma_2)}$$

em que d_i é o i-ésimo elemento da diagonal de $Z_1 Z_1 + A_1^{-1} \gamma_1$ e $Z_2' Z_2 + A_2^{-1} \gamma_2$ resultante da inversa da matriz conjunta de efeitos fixos e aleatórios.

Estudo de correlação entre a distância modificada de Roger com a produtividade de grãos, capacidade específica de combinação e heterose

Para o estudo de correlação de Pearson, utilizaram-se as variáveis produtividade de espigas despalhadas (PED), CEC e heterose com a distância modificada de Roger (DMR). As análises foram realizadas por meio do SAS[®], utilizando-se o PROC CORR. Utilizou-se o teste de Mantel para estudo da significância das correlações (Manly, 1991). Os valores de heterose foram estimados com base no desempenho dos híbridos simples que originaram as populações A, B e C, conforme sugerido por Amorim et al. (2006).

Além do teste de Mantel, a significância da correlação entre DMR e PED foi testada via simulação Monte Carlo. Para isso utilizaram-se os

algoritmos recomendados por Fan et al. (2002). Esses autores sugerem a construção de dados moldados a partir das características de uma amostra qualquer, com o objetivo de verificar o quanto o coeficiente de correlação pode variar quando se retiram inúmeras amostras de tamanhos diferentes desse conjunto de dados.

Basicamente, o algoritmo consiste em: a) obter os primeiros, segundos, terceiros e quartos momentos da amostra, ou seja, a média, variância, assimetria e curtose, b) verificar se ocorre algum grau de dependência entre as amostras ao qual se deseja realizar inferências (correlação) e c) gerar dados moldados conforme as características da amostra (Tabela 1).

Para a construção de dados aleatórios com desejada assimetria e curtose, utilizou-se o algoritmo proposto por Fleishman (1978) e, para dados com desejado valor de dependência (correlação), o algoritmo proposto por Kaiser & Dickman (1962).

Foram simulados quatro tamanhos de amostras diferentes (10, 30, 60 e 100). Cada amostra foi repetida 10.000 vezes, totalizando 40.000 amostras com 2 milhões de dados. A probabilidade de $r > 0$ foi testada pela expressão sugerida por May et al. (2001) e Fan et al. (2002). Além disso, foi construído um intervalo de confiança para o valor de correlação entre a variável produção de espigas despalhadas e a DMR para um tamanho de amostra 60, tamanho este igual ao obtido neste experimento.

1.4 Resultados

Os BLUPs da produtividade de espigas despalhadas de cada híbrido e das quatro testemunhas se encontram na Tabela 2. Pode-se verificar que três cruzamentos superaram a testemunha de maior produtividade (DKB 199). No caso do híbrido interpopulacional 1BX28C, observa-se que este cruzamento foi

superior em $1,66 \text{ t.ha}^{-1}$ em produtividade de espigas despalhadas (PED) em relação ao híbrido DKB199. Além disso, constatou-se que seis cruzamentos apresentaram melhor desempenho quando comparados ao híbrido DKB333B, que foi um dos utilizados na formação de uma das três populações base (população C) e que apresentou o melhor desempenho em relação aos demais híbridos (DOW657 e P30F45).

Verifica-se que ocorreram híbridos intrapopulacionais tão produtivos quanto aos híbridos interpopulacionais (Tabela 2). Contudo, as maiores produtividades foram verificadas nos híbridos interpopulacionais, tendo esses híbridos apresentado desempenho médio de $9,84 \text{ t.ha}^{-1}$ em relação à produtividade média dos intrapopulacionais que foi de $8,24 \text{ t.ha}^{-1}$.

No que se refere à heterose, apenas 30% dos híbridos interpopulacionais apresentaram heterose negativa, ao passo que todos os intrapopulacionais obtiveram desempenho abaixo do híbrido simples original, refletindo em estimativas negativas de heterose. Esse resultado demonstra que a obtenção de híbridos interpopulacionais a partir de progênies $S_{0,2}$ é extremamente viável, pois 70% dos híbridos interpopulacionais superaram o desempenho médio dos híbridos simples utilizados na formação da população base (Tabela 2). Vale destacar que a heterose aplicada nesse estudo não reflete o conceito clássico de heterose, pois foram utilizados como referência os híbridos simples comerciais que deram origem às populações e não às linhagens $S_{0,2}$.

As estimativas de CEC e CGC encontram-se na Tabela 3. Pode-se verificar que as estimativas da CGC foram extremamente baixas, ou seja, variaram de $-0,009$ a $0,005$, refletindo um baixo valor de variância da CGC (Tabela 2). Por outro lado, as estimativas da CEC variaram de $3,339$ a $-3,399$, convergindo para um alto valor da variância da capacidade específica de combinação (Tabela 3).

Na Figura 1 é possível verificar que o modelo considerando os genótipos em um único grupo (dialelo completo desbalanceado) foi mais acurado que o modelo de dialelo parcial com e sem informação de parentesco, ou seja, modelo que considera os genitores separados em grupos distintos.

Quanto à acurácia utilizando os coeficientes de LR e DMR, verifica-se que a utilização de similaridade no estado pode acarretar em aumento da variância do erro de predição e, conseqüentemente, perda da acurácia dos valores genéticos, quando comparada a estimadores que eliminam esse viés (Figura 1). Além disso, é possível constatar que, no modelo sem a informação de parentesco, ocorreu perda de acurácia nos BLUPs da CGC e CEC em relação ao coeficiente de LR.

Entretanto, no que se refere aos valores de acurácia utilizando a DMR, houve vantagem apenas nos BLUPs da CGC em relação aos dois modelos de dialelos parciais. Contudo, para os valores de acurácia referentes à CEC, a DMR proporcionou valores inferiores em comparação aos modelos de dialelo parcial com e sem a informação de parentesco (Figura 1).

Pela Tabela 4 pode-se verificar que os doze híbridos mais divergentes, em média, foram bastante produtivos e apresentaram, em média, estimativas positivas de CEC e heterose, ao passo que os doze cruzamentos menos divergentes apresentaram baixo desempenho em produtividade e valores negativos de CEC e heterose.

No entanto, o híbrido mais divergente não foi o mais produtivo. Pelo contrário, apresentou desempenho abaixo da média geral. O mesmo se aplica à combinação híbrida menos divergente, ou seja, esse híbrido não foi o de menor desempenho no conjunto de híbridos avaliados (Tabela 4).

Porém, em média, os híbridos mais divergentes foram os mais produtivos, concorrendo para valores de correlação de média magnitude entre a DMR com a produtividade, CEC e a heterose (Tabela 5), valores esses, todos

diferentes de zero pelo teste de Mantel. Entretanto, quando considerados separadamente os híbridos intra e interpopulacionais, não houve correlação significativa entre a DMR e as variáveis produtividade, CEC e heterose.

O coeficiente de correlação (r) entre a DMR e produtividade (0,546) variou de -0,64 a 0,97, dependendo do tamanho da amostra (Figura 2). Para amostra de tamanho 10, a probabilidade de r ser maior que zero foi de 95,7%. Para os outros tamanhos de amostras, a probabilidade de r ser menor ou igual a zero $p(r \leq 0 | r)$ foi praticamente zero, demonstrando que, independentemente do tamanho da amostra, houve associação positiva entre a DMR e a produtividade de espigas.

Utilizando como referência a amostra de tamanho 60, pode-se verificar que o valor de correlação entre DMR e produtividade, em 98% das amostras, variou de 0,35 a 0,70, com média de 0,538, utilizando 600.000 dados amostrados aleatoriamente, dentro das populações simuladas e moldadas conforme os valores fenotípicos e de distância genética. Esse resultado corrobora os valores de correlação obtidos para cada ambiente (0.402 e 0.603).

Em média, todas as amostras apresentaram valores de correlação bem próximos aos valores amostrais (Figura 2), resultado este já esperado pelo resultado da população teste (Tabela 1).

1.5 Discussões

A utilização de híbridos provenientes de linhagens parcialmente endogâmicas obtidas a partir de híbridos simples comerciais vem sendo recomendada por vários pesquisadores (Souza Jr., 1992; Carvalho et al., 2003; Amorim et al., 2006). Em todos esses trabalhos foram encontrados genótipos com desempenho superior ao das testemunhas (híbridos progenitores). Neste

estudo, os resultados não foram diferentes, ou seja, grande parte dos híbridos interpopulacionais superou a média dos híbridos genitores (71%).

Esse resultado foi superior ao obtido por Amorim et al. (2006) com híbridos S_0 para essas mesmas populações, que encontrou um valor de 11,65%. Além disso, seis cruzamentos foram superiores ao DKB333B, ao passo que Amorim et al. (2006), utilizando 163 híbridos, não obtiveram nenhum cruzamento superior a esse híbrido comercial. Por outro lado, Carvalho et al. (2003) constataram que 17% dos híbridos $S_2 \times S_2$ foram superiores ao híbrido DKB333B, demonstrando o grande potencial da utilização desse tipo de germoplasma para síntese de híbridos.

No que se refere aos híbridos intrapopulacionais, apesar de alguns cruzamentos apresentarem desempenho acima da média geral, eles não conseguiram superar os híbridos genitores. Uma hipótese a ser levantada para explicar este fato é que foram avaliados apenas 33 híbridos intrapopulacionais, colaborando para o baixo desempenho desses genótipos, devido à maior probabilidade de se encontrar combinações endogâmicas. Porém, levando em consideração também o fato de que foram avaliados apenas 27 híbridos interpopulacionais e que grande parte desses híbridos apresentou alta performance, é razoável afirmar que esse tipo de cruzamento deve ser prioritário para essas populações.

Deve-se destacar que, na Tabela 2, apresentaram-se os BLUPs da produtividade e não as médias fenotípicas. Assim, testes de comparações múltiplas não são recomendados para valores genotípicos, pois o BLUP é tido como estimador do tipo “shrinkage”, que penaliza os genótipos que foram mais influenciados pela avaliação genética, conduzindo a valores próximos à media geral.

No que se refere à CEC e CGC e suas respectivas variâncias, alguns comentários merecem destaque. Segundo Sprague & Tatum (1942), a CGC

reflete o comportamento médio de um genitor em um conjunto de cruzamentos. Além disso, de acordo com Boer & Hoeschele (1993), cruzamentos endogâmicos podem influenciar as estimativas dos componentes de variância, fato esse que pode ser observado em dialelos completos, nos quais ocorrem cruzamentos de indivíduos aparentados e não aparentados. Neste estudo, esse resultado não foi diferente, ou seja, a estimativa do componente da variância da CGC que reflete a variabilidade do desempenho dos genitores nos cruzamentos foi muito baixa e próxima de zero.

Considerando o conceito de CGC e cruzamentos endogâmicos e não endogâmicos avaliados em conjunto em dialelos completos, observa-se, pelos dados da Tabela 2, que as linhagens que apresentaram bom desempenho em cruzamentos interpopulacionais (não endogâmicos), normalmente, apresentaram baixa produtividade em cruzamentos intrapopulacionais (endogâmicos). Esse resultado pode ter contribuído para subestimação dos valores de CGC que foram próximos a zero e, conseqüentemente, variância da CGC também próxima de zero (Tabela 3). Assim, pode-se inferir que o conjunto de híbridos avaliados neste estudo não foi fortemente influenciado pela CGC dos genitores nos cruzamentos, convergindo em valores da variância aditiva e não aditiva referentes a CGC e CEC bastantes distintos do que seria esperado para esses componentes da variância na cultura do milho (Hallauer & Miranda Filho, 1981).

Outro comentário a ser realizado seria a razão de não se utilizar o modelo de dialelo parcial em vez do dialelo completo desbalanceado como utilizado neste estudo para análise dialélica. Como já comentado no tópico Material e métodos, os cruzamentos foram realizados em esquema 11x6, o que levaria à utilização de esquemas de dialelos parciais para avaliação da CGC e CEC. No entanto, esses cruzamentos foram obtidos com o objetivo de recuperar combinações mais e menos divergentes, o que acarretou na alocação de

indivíduos aparentados fora dos grupos, ou seja, não necessariamente ocorreram grupos distintos.

Quando se utiliza um esquema de dialelo parcial, parte da informação de parentesco é perdida na matriz A, ao passo que, quando se considera um dialelo completo, mesmo altamente desbalanceado, essa informação pode ser recuperada, acarretando em maior acurácia na predição dos valores de CGC e CEC. Esse fato fica evidente na Figura 1, em que os seis últimos parentais, quando alocados em um grupo distintos dos demais (11x6), obtiveram menor acurácia nos BLUPs da CGC em relação ao modelo de dialelo completo desbalanceado e, conseqüentemente, menores valores de acurácia da CEC, que também foram influenciadas pela perda de informação de parentesco (Figura 1).

Devido a isso, pode-se justificar a preferência, neste estudo, pela utilização do modelo de dialelo completo desbalanceado em vez de dialelo parcial. Deve-se ressaltar que, na abordagem tradicional de cruzamentos dialélicos por meio do método dos quadrados mínimos ordinários (OLS) ou até mesmo a utilização de modelos mistos sem a informação de parentesco, essa recuperação de informação não seria possível, pois não assume a covariância entre os genótipos. Com isso, é possível inferir que, quando não se dispõe de qualquer informação de parentesco a respeito dos genótipos que estão sendo avaliados, a informação de marcadores pode contribuir para a acurácia de predição dos valores genéticos.

No entanto, alguns autores (Bauer et al., 2006; Balestre et al., 2008c) têm sugerido a utilização da similaridade no estado para a predição dos valores genéticos utilizando a metodologia da melhor predição linear não viesado (BLUP). Porém, observando-se os valores de acurácia para as predições da CEC e CGC (Figura 1), verifica-se que o coeficiente de LR proporcionou valores de acurácia de maior magnitude quando comparado com a DMR.

Esse maior erro de predição ocasionado pela DMR nos valores de CEC e CGC já era esperado devido à recuperação da informação genética que leva em consideração apenas o aspecto evolutivo (similaridade no estado) dos indivíduos e não seu parentesco. Assim, em maior ou menor peso, todos os valores de covariância genética são recuperados no preditor linear (BLUP) de cada valor de CEC e CGC, inclusive daqueles indivíduos que não partilham alelos herdados de genitores em comum. No entanto, esse viés pode ser eliminado quando se utiliza o coeficiente de Lynch & Ritland (1999), pois ele, apesar de utilizar dados moleculares, leva em consideração a probabilidade condicional de que dois indivíduos possuam alelos herdados de genitores em comum. Assim, pode-se inferir que a utilização da similaridade no estado para construção da matriz A pode acarretar em erro nas predições dos valores genéticos.

As distâncias genéticas obtidas por meio de marcadores de microsatélites apresentaram moderada eficiência na pré-seleção dos híbridos avaliados neste estudo. Os doze híbridos mais divergentes, em média, foram superiores em produtividade, CEC e heterose, em relação aos menos divergentes, apresentando eficiência de seleção de 20% em relação aos maiores valores observados de produtividade. Entretanto, o híbrido de maior desempenho não foi alocado dentro do grupo de híbridos selecionados, demonstrando que pode ocorrer perda de combinações promissoras quando se utiliza a pré-seleção por distância genética (Tabela 4).

Os resultados obtidos de correlação entre DMR e produtividade (0,546) foram praticamente idênticos aos obtidos por Amorim et al. (2006) para progênies S_0 procedentes dessas mesmas populações (0,55). Isto permite inferir que existe repetibilidade na associação entre DMR e produtividade para as populações envolvidas nesse estudo. Esse valor de correlação pode ser considerado de média magnitude, porém, não permite confiabilidade de predição. Isso porque, quando se realizou a separação entre híbridos intra e

interpopulacionais, não houve associação entre DMR e produtividade, ou seja, embora seja de média eficiência na predição conjunta de híbridos inter e intrapopulacionais, a DMR proporcionou pouca eficácia preditiva em cruzamentos intra e interpopulacionais.

Esse resultado discorda dos resultados obtidos por Reif et al. (2003), Amorim et al. (2006) e Balestre et al. (2008c), que encontraram grandes diferenças nos valores de correlação quando separados grupos intra e interpopulacionais. Porém, esses autores discordam na magnitude da associação encontrada entre a DMR com produtividade em híbridos intra e intergrupos.

Assim, pode-se inferir que, embora os resultados de correlação sejam aparentemente satisfatórios, a DMR só apresentou uma eficiência de seleção de média magnitude, quando ocorreu alta variabilidade nos valores de distância genética e da produtividade. Quando essa variabilidade diminuiu, no caso da separação de híbridos inter e intrapopulacionais, essa moderada eficiência esvaeceu, o que já era esperado pela expressão de correlação de Pearson. As mesmas inferências podem ser realizadas para a CEC e heterose, apesar da correlação entre a DMR e heterose (0,661) apresentar maior valor quando comparada com produção e CEC.

Ainda com relação ao valor de correlação entre DMR e produtividade, podem-se utilizar os resultados da simulação Monte Carlo, apenas para caráter ilustrativo, com a finalidade de demonstrar sua aplicabilidade em estudos de correlação.

Focando a atenção no trabalho realizado por Amorim et al. (2006), nota-se que esses autores utilizaram apenas 24 híbridos para o estudo de correlação de um conjunto de 163 híbridos avaliados e apenas metade desses híbridos (12) era intra e a outra metade, interpopulacional. Esses autores observaram que o alto valor de correlação entre DMR e produtividade só foi obtido quando houve separação entre híbridos inter e intrapopulacionais, tendo apenas os híbridos

interpopulacionais apresentado alta correlação (0,84), enquanto que com os intrapopulacionais obtiveram-se baixos valores de correlação (0,18).

Considerando que, neste estudo, foram utilizadas as mesmas populações e que as progênies utilizadas nos cruzamentos são derivadas das mesmas obtidas por Amorim et al. (2006) e, ainda, que o coeficiente de correlação obtido neste estudo foi praticamente o mesmo obtido por estes autores, é possível inferir que eles podem ter obtido um alto valor de correlação entre DMR e produtividade nos híbridos interpopulacionais (0,84) por mero acaso, devido ao tamanho da amostra (12) retirada do conjunto de dados com valor de correlação de 0,55 e tamanho igual a 24. O método de simulação Monte Carlo dá respaldo para essa inferência (Figura 1).

Nota-se que, mesmo quando duas populações apresentam correlação de 0,55, retirando-se inúmeras amostras de tamanho 10 dessas mesmas populações, a probabilidade de se observar valores de correlação 0,55 é praticamente a mesma de observar 0,80 (Figura 2), ou seja, ocorre erro de amostragem na estimativa de correlação.

No entanto, quando se utilizam amostras maiores (60 e 100), essa paridade desaparece, prevalecendo o coeficiente populacional (Figura 2). Resultado semelhante a este foi observado por Millsap (1988), em estudos de correlação via simulação de Monte Carlo. Este autor concluiu que amostras menores que 60 podem apresentar viés devido a erro de amostragem, o que também observado neste estudo, no qual a amostra de tamanho 60 obteve praticamente o mesmo resultado da amostra de tamanho 100. Além disso, os valores de correlações obtidos nas análises individuais estão dentro do intervalo de confiança para a amostra de tamanho 60 (0,402 e 0,603).

Assim, pelo método de estudo de correlação via simulação Monte Carlo, é possível testar a significância do coeficiente de correlação para dados assimétricos (DMR), como também obter as informações necessárias para

dimensionar o tamanho da amostra para não correr o risco de observar valores de correlações meramente por acaso, conforme enfatizado por Fan et al., (2002).

Com base nos resultados deste estudo, pode-se concluir que a utilização de linhagens parcialmente endogâmicas é uma alternativa viável na síntese de híbridos de alta performance e que as distâncias genéticas apresentaram moderada capacidade preditiva da produtividade, capacidade de combinação e heterose desses híbridos.

A simulação por Monte Carlo apresenta-se como alternativa viável para se testar a correlação entre distâncias genética e produtividade, além de poder ser utilizada para dimensionar o tamanho adequado de amostra em estudos de associação entre essas duas variáveis.

A estimativa de parentesco por marcadores moleculares por meio do coeficiente de Lynch e Ritland demonstra ser uma metodologia promissora na predição de valores genéticos em modelos lineares mistos (BLUP), principalmente quando não se dispõe de nenhuma informação de pedigree.

1.6 Agradecimentos

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pelo auxílio concedido para a realização deste trabalho.

2 REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, E.P.; AMORIM, U.B.O.; SANTOS, J.B. dos. Genetic distance based on SSR and grain yield of inter an intrapopulational maize single cross hybrids. **Maydica**, Bergamo, v.51, p. 507-513, 2006.

ARAÚJO, P.M. **Dialelo parcial circulante interpopulacional e cruzamentos “top cross” na avaliação de linhagens parcialmente endogâmicas de milho.** 2000. 170p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Universidade de São Paulo, Piracicaba, SP.

BALESTRE, M.; MACHADO, J.C.; LIMA, J.L.; SOUZA, J.C.; FILHO, L.N. Genetics distances estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids **Genetics and Molecular Research**, Ribeirão Preto, v.7, n. 1, p. 65-73, 2008a.

BALESTRE, M.; VON PINHO, R.G.; SOUZA, J.C.; LIMA, J.L. Comparison of maize similarity and dissimilarity genetic coefficients based on microsatellites markers. **Genetics and Molecular Research**, Ribeirão Preto v.7, n. 3, p. 695-705, 2008b.

BALESTRE, M.; VON PINHO, R.G.; SOUZA, J.C.; MACHADO J.C. Potential of maize single-cross hybrids for extraction of inbred lines using the mean components and mixed models with microsatellite marker information **Genetics and Molecular Research**, Ribeirão Preto v.7, n. 4, p.1106-1118, 2008c.

BAUER, A.M.; REETZ, T.C.; LEON, J. Estimation of Breeding Values of Inbred Lines using Best Linear Unbiased Prediction(BLUP) and Genetic Similarities. **Crop Science**, Madison, v. 46, n. 4, 2695-2691, 2006.

BERNARDO, R. Correlation between tests cross performance of lines early and last selfing generation. **Theoretical and Applied Genetics**, Berlin, v. 82, n. 1, p. 17-21, 1991.

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. Relationship between single cross performance and molecular marker heterozygosity. **Theoretical and Applied Genetics**, Berlin, v.83, p. 628-634, 1992.

BOER, I.J.M. de.; HOESCHELE, I. Genetic evaluation methods for populations with dominance and inbreeding. **Theoretical and Applied Genetics**, Berlin, v. 86, n. 2/3, p. 245-258, April. 1993.

CARLONE, M. R.; RUSSEL, W. A. Evaluation of S2 maize lines reproduced from several generations by random mating within lines: II. Comparisons of testcross performance of original and advanced S2 and S8 lines. **Crop Science**, Madison, v. 28, n. 6, p. 916-920, Nov./Dec. 1988.

CARVALHO, A. D. F.; SOUZA, J. C.; RIBEIRO, P. H. Desempenho de híbridos de linhagens parcialmente endogâmicas de milho em regiões dos Estados de Roraima e Minas Gerais. **Ciência e Agrotecnologia**, Lavras, v. 27, n. 8, p. 985-990, set./out. 2003.

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, 2004.

FALCONER, D.S.; MACKAY, T.F. **Introduction to quantitative genetics** 4.Ed. Essex, England: Ed. Longman,1996. 463 p.

FAN, X.O.; FELSÓVÁLY,I A.; SIVO, S.A.; KEENAN, S.C. **SAS® for Monte Carlo Studies: A Guide for Quantitative Researchers** Cary, NC: SAS Institute Inc.2002. 251 p.

FLEISHMAN, A.I. A method for simulating non-normal distributions. **Psychometrika**, Greensboro, v. 43, p. 521-532, 1978.

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

GUIMARÃES, O.S.; PATERNIANI, M.E.A.G.Z.; LUDERS, R.R.; SOUZA, A.P.; LABORTA, P.R.; OLIVEIRA, Q.M. Correlação da heterose de híbridos de milho com divergência genética entre linhagens. **Pesquisa. Agropecuária Brasileira**, Brasília v. 42, p.811-816, 2007.

HALLAUER, A. R.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. Ames: Iowa State University Press, Ames: Iowa State University Press, 1981. 468 p.

HALLAUER, A.R.; RUSSEL, W.A.; LAMKEY, K.R. Corn breeding. In: SPRAGE, G.F.; DUDLEY, J.W. (Ed.) **Corn and Corn Improvement**, 3rd., Madison: Iowa State University Press, 1988. p. 453-464.

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

KAISER, H.F.; DICKMAN, K. Sample and Population Score Matrices and Sample Correlation Matrices from an Arbitrary Population Correlation Matrix. **Psychometrika**, Greensboro ,v. 27, p. 179-182, 1962.

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

MANLY, B.F.J. **Randomization and Monte Carlo Methods in Biology**. Chapman & Hall, London, 1991, 281 p.

MANTEL, N . The detection of disease clustering and a generalized regression approach **Cancer Research**, Philadelphia, v. 27, p. 209-220, 1967

MAY, E.C.; SPOTTISWOODE, S.J.P.; FAITH, L.V. A methodological issue in the study of correlation between psychophysiological variables. Proceedings of Presented Papers. In **The Parapsychological Association 44th Annual Convention**, p. 166-178, 2001.

MILLSAP, R.E. Sampling variance in attenuated correlation coefficients: A monte carlo study. **Journal of Applied Psychology**, Washington ,v. 73, p. 316-319, 1988.

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.; HOISINGTON, D. A.; VASAL, S. K.; SRINIVASAN, G.; BOHN, M.; FRISCH, M Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. **Crop Science**, Madison, v. 43, n. 4, p. 1275-1282. 2003.

RESENDE, M.D.V. (2002) **Genética Biométrica e Estatística no Melhoramento de Plantas Perenes**. Brasília, Embrapa Informação Tecnológica, 2002, 975 p.

SAS INSTITUTE. **SAS language and procedures: Usage**. Version 8. 1. Cary, NC: SAS: SAS Institute, 2000. 1CD-ROM.

SOUZA JR., C. L. Interpopulation genetic variances and hybrid breeding programs. **Brazilian Journal of Genetics**, Ribeirão Preto, v. 15, n. 3, p.643-656, Sept. 1992.

SPRAGUE, G.F.; TATUM, L.A . General vs specific combining ability in single crosses of corn. **Journal of American Society of Agronomy**, Washington v. 34, n. 3, p. 923-932, Mar. 1942

TROYER, A.F. Breeding early corn. In HALLAUER, A.R. **Speciality corns**: Press. Boca Raton, Florida, p.342-396, 1994.

VAN VLECK, L.D. **Selection index in introduction to mixed models methods for genetic improvement of animals. The green book**, Press. Boca Raton, Florida. 1993. 481 p.

3 TABELAS

TABELA 1 Momentos referentes aos dados de distância genética e produtividade de 60 híbridos de milho e conjunto de dados moldados e simulados pelo método de Monte Carlo, com base nos dados amostrais.

Dados amostrais	n	média	Desvio padrão	Assimetria	Curtose
DMR	60	0,169	0,101	-0,332	-1,098
Produção	60	8,962	1,529	*	*
Correlação	0,546				
Dados simulados	n	média	Desvio padrão	Assimetria	Curtose
DMR	2000000	0,169	0,102	-0,345	-0,975
Produção	2000000	8,958	1,528	*	*
Correlação	0,547				

*dados com distribuição normal, pelo teste de Shapiro-Wilk.

TABELA 2 BLUPS da produtividade média de espigas despalhadas ($t\ ha^{-1}$) e heterose dos híbridos intra e interpopulacionais avaliados em dois locais, na safra 2007-2008. Lavras, UFLA, 2009.

Genótipo	Produtividade	Heterose	Genótipo	Produtividade	Heterose
1BX28C	12,375	2,998	19CX25C	8,713	-1,710
14CX29A	11,222	1,300	15CX7C	8,712	-1,711
30BX28C	10,813	1,436	12CX25C	8,707	-1,716
*DKB199	10,715	-	19CX23C	8,695	-1,728
30BX23C	10,690	1,313	22CX23C	8,583	-1,840
1BX7C	10,649	1,272	30BX29A	8,579	-0,296
29CX28B	10,628	1,251	12CX28C	8,560	-1,863
*DKB333B	10,423	-	15CX29A	8,545	-1,377
30BX7C	10,412	1,035	*DOW657	8,331	-
12CX28B	10,264	0,887	19CX7C	8,511	-1,912
15CX28B	10,194	0,817	12CX23C	8,467	-1,956
5CX29A	10,029	0,107	14CX23C	8,464	-1,959
27BX28C	9,942	0,565	1CX29A	8,434	-1,488
27BX23C	9,928	0,551	5CX25C	8,301	-2,122
22CX29A	9,894	-0,027	5CX7C	8,238	-2,185
5CX28B	9,894	0,517	12CX7C	8,167	-2,256
14CX28B	9,879	0,502	14CX28C	8,152	-2,271
19CX28B	9,745	0,368	19CX28C	8,142	-2,281
27BX7C	9,653	0,276	5CX23C	8,134	-2,289
1BX23C	9,631	0,254	1BX25C	8,091	-1,286
22CX28C	9,566	-0,857	29CX25C	8,087	-2,336
1CX28C	9,558	-0,865	22CX25C	8,084	-2,339
1BX29A	9,467	0,591	29CX23C	7,866	-2,557
5CX28C	9,424	-0,999	29CX7C	7,835	-2,588
*P30F45	9,420	-	1CX7C	7,797	-2,626
19CX29A	9,417	-0,504	14CX7C	7,741	-2,682
1CX28B	9,273	-0,104	15CX25C	7,480	-2,943
15CX23C	9,268	-1,155	14CX25C	7,407	-3,016
27BX25C	9,115	-0,262	1CX23C	7,398	-3,025
15CX28C	9,113	-1,310	1BX28B	7,310	-1,021
27BX29A	9,064	0,189	22CX7C	7,050	-3,373
1CX25C	8,761	-1,662	27BX28B	5,627	-2,704
Media	9,010				

* Testemunhas

A=População proveniente do híbrido simples P30F45.

B= População proveniente do híbrido simples Dow 657.

C=População proveniente do híbrido simples DKB 333B.

TABELA 3 BLUPs da capacidade geral (CGC) e específica de combinação (CEC) de 17 linhagens S_{0,2}, considerando o modelo de dialelo completo desbalanceado. Lavras, UFLA, 2009.

	23C	25C	28B	7C	28C	29A	CGC
30B	1,661	-	-	1,386	1,776	-0,457	0,005
27B	0,905	0,098	-3,399	0,633	0,911	0,034	-0,001
19C	-0,328	-0,304	0,718	-0,508	-0,889	0,387	-0,001
22C	-0,438	-0,931	-	-1,968	0,537	0,866	-0,002
1B	0,604	-0,930	-1,721	1,625	3,339	0,432	0,004
1C	-1,623	-0,253	0,248	-1,219	0,530	-0,594	-0,003
14C	-0,557	-1,608	0,855	-1,277	-0,877	2,194	-0,003
29C	-1,156	-0,928	1,602	-1,183	-	-	-0,002
5C	-0,889	-0,715	0,867	-0,781	0,393	0,999	-0,001
12C	-0,553	-0,306	1,240	-0,849	-0,468	-	-0,004
15C	0,247	-1,535	1,169	-0,306	0,084	-0,484	-0,002
CGC	-0,003	-0,009	0,001	-0,006	0,005	0,004	
σ_{CGC}^2	0,002						
σ_{CEC}^2	1,622						
σ_e^2	1,137						
h_a^2	0,741						

TABELA 4 Seleção dos doze híbridos mais divergentes e comparação com os doze híbridos menos divergentes, considerando os valores de produtividade, heterose e capacidade específica de combinação (CEC). Lavras, UFLA, 2009.

Dez cruzamentos mais divergentes					
Ordem	Cruzamento	Distância	Produção	CEC	heterose
1	1BX25C	0,337	8,091	-0,926	-1,286
2	27BX25C	0,313	9,115	0,110	-0,262
3	14CX28B	0,306	9,879	0,854	0,502
4	14CX29A	0,306	11,222	2,190	1,300
5	22CX29A	0,283	9,894	0,860	-0,027
6	19CX28B	0,274	9,745	0,716	0,368
7	1BX23C	0,272	9,631	0,598	0,254
8	30BX23C	0,271	10,690	1,656	1,313
9	30BX28C	0,271	10,813	1,759	1,436
10	1CX28B	0,265	9,273	0,248	-0,104
11	5CX28B	0,265	9,894	0,865	0,517
12	15CX28B	0,265	10,194	1,169	0,817
Média		0,285	9,087	0,842	0,402
Dez cruzamentos menos divergentes					
Ordem	Cruzamento	Distancia	Produção	CEC	heterose
1	27BX28B	0,074	5,627	-3,4059	-2,704
2	1CX28C	0,061	9,558	0,524	-0,865
3	12CX28C	0,058	8,560	-0,473	-1,863
4	19CX7C	0,030	8,511	-0,499	-1,912
5	22CX23C	0,030	8,583	-0,433	-1,840
6	22CX7C	0,030	7,050	-1,958	-3,373
7	1CX7C	0,030	7,797	-1,207	-2,626
8	14CX7C	0,030	7,741	-1,265	-2,682
9	29CX7C	0,030	7,835	-1,172	-2,588
10	5CX7C	0,030	8,238	-0,771	-2,185
11	12CX7C	0,030	8,167	-0,836	-2,256
12	15CX7C	0,030	8,712	-0,294	-1,711
Média		0,039	8,032	-0,982	-2,217

A=População proveniente do híbrido simples P30F45.

B= População proveniente do híbrido simples Dow 657.

C=População proveniente do híbrido simples DKB 333B.

TABELA 5 Correlação entre as distâncias modificada de Roger (DMR) com produtividade, capacidade específica de combinação (CEC) e heterose referentes aos 60 híbridos avaliados em dois locais, na safra 2007/2008. Lavras, UFLA, 2009.

Distância	produção	CEC	heterose
DMR	0,546 ⁺⁺	0,567 ⁺⁺	0,661 ⁺⁺
DMR INTRA	0,107	0,141	0,153
DMR INTER	0,031	0,015	0,047

⁺⁺ Significativo, a 1% de probabilidade, pelo teste de mantel baseado em 10.000 simulações

4 FIGURAS

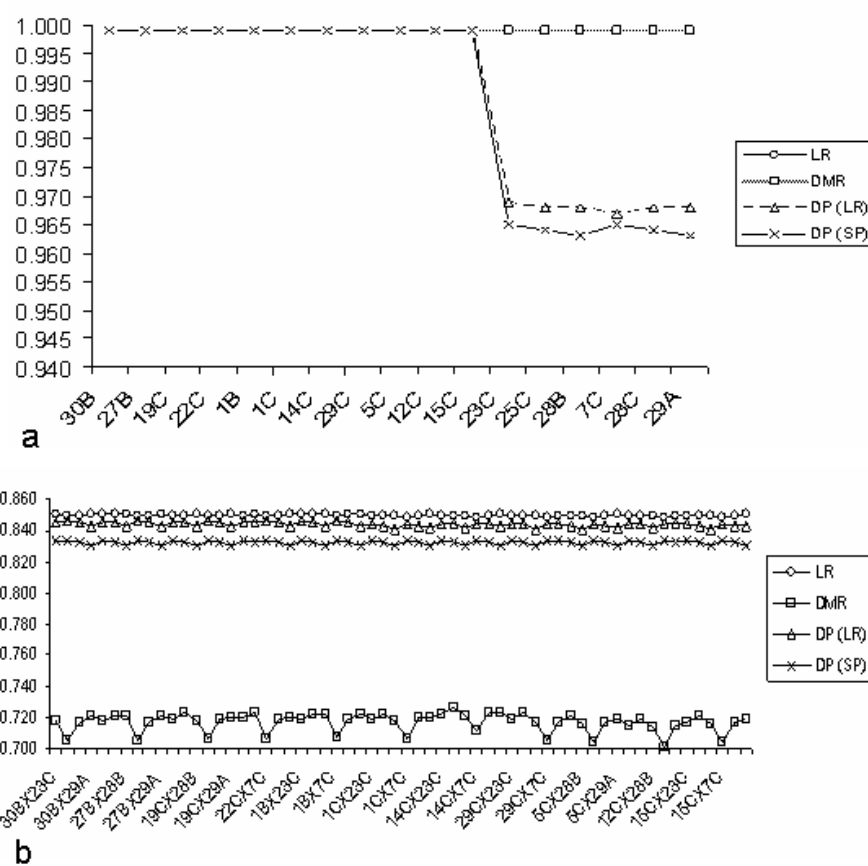


FIGURA 1 Acurácia dos BLUPs da CGC (a) e CEC (b), considerando o modelo de dialeto completo desbalanceado com parentesco de Lynch e Ritland (LR), distância modificada de Roger (DMR), dialeto parcial com parentesco de Lynch e Ritland (DP_{LR}) e dialeto parcial sem informação de marcadores (DP_{SP}).

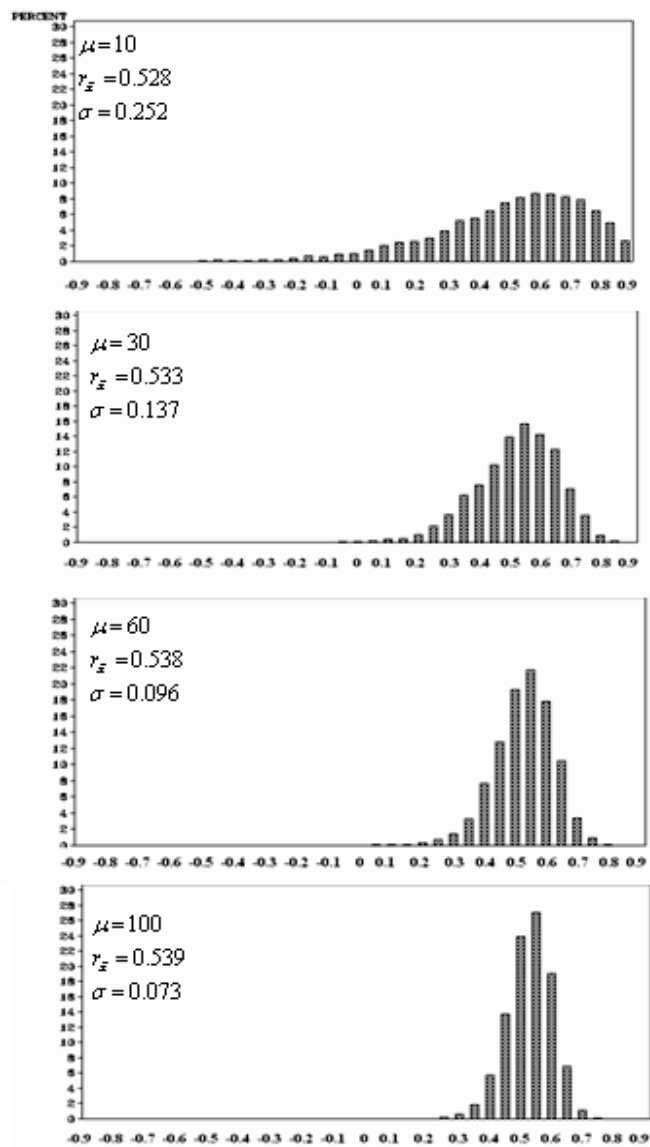


FIGURA 2 Tamanho da amostra (μ), desvio padrão (σ) e correlação média ($r_{\bar{x}}$) de dados obtidos via simulação Monte Carlo moldados de acordo com os valores de distância genética e produtividade dos 60 híbridos provenientes de linhagens parcialmente endogâmicas.

CAPÍTULO 2

O capítulo número 2 será transcrito em formato de artigo e encaminhado para submissão do Periódico Científico **Australian Journal of Agriculture Research** com o seguinte título:

Predição do desempenho de híbridos de milho utilizando modelos lineares mistos com informação de marcadores de microssatélites

Marcio Balestre¹, Renzo Garcia Von Pinho², João Candido de Souza¹, Julio Silvio de Bueno Filho³.

Corresponding author:

E-mail: marciobalestre@hotmail.com

¹Departamento de Biologia, Universidade Federal de Lavras. CP 3037. Lavras. MG, Brazil.

²Departamento de Agricultura, Universidade Federal de Lavras. CP 3037. Lavras. MG. Brazil.

³Departamento de Ciências Exatas, Universidade Federal de Lavras. CP 3037. Lavras. MG. Brazil.

1.1 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 do parentesco de Lynch & Ritland nas predições de híbridos não testados. Foram utilizadas 90 progênies $S_{0,2}$, provenientes de três populações oriundas de três híbridos simples comerciais. Todas as 90 progênies foram

genotipadas com 25 marcadores de microssatélites (SSR), sendo 9 dessas marcas ligadas à QTL para produtividade de grãos. Com base nas similaridades genéticas, foram selecionadas 17 linhagens, que foram intercruzadas em um sistema de dialelo parcial, com o objetivo de recuperar os cruzamentos mais e menos divergentes. Os resultados de similaridade e parentesco foram utilizados para a construção da matriz genética aditiva e de dominância que, posteriormente, foram empregadas em associação com o BLUP para predição dos valores genotípicos e de CEC dos híbridos não avaliados. Para isso, foram simulados cinco graus de desbalanceamento (5, 10, 20, 30 e 40 híbridos retirados). Os valores de correlações entre os valores genotípicos preditos e as médias observadas variaram de 0,55 a 0,70, dependendo do grau de desbalanceamento. Resultado semelhante foi observado para as predições da CEC, ou seja, variaram de 0,61 a 0,70. Além disso, constatou-se que o coeficiente de parentesco de Lynch e Ritland associado ao BLUP proporcionou predições mais acuradas quando comparado às predições obtidas pela similaridade no estado, estimada pela distância modificada de Roger. Assim, foi possível constatar que o método (BLUP) demonstrou ser uma alternativa viável na predição de cruzamentos não realizados, mesmo com cruzamentos com alta influência de efeitos não aditivos, como observado neste estudo.

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

1.2 Introdução

A utilização de cruzamentos dialélicos em programas de melhoramento de milho para a seleção de genótipos é de fundamental importância. Por meio dessas combinações é possível decompor os valores dos cruzamentos em capacidade geral de combinação (CGC) e capacidade específica de combinação

(CEC) (Griffing, 1956). Os valores de CGC e CEC auxiliam o melhorista na escolha de genitores e seleção de linhagens, principalmente por estarem ligados aos efeitos aditivo-dominante e dominante, respectivamente (Vencovsky & Barriga, 1992).

Porém, uma das limitações ao uso de cruzamentos dialélicos completos é a grande quantidade de genótipos que podem ser obtidos com a utilização de poucos genitores. Uma alternativa seria a realização de dialelos parciais e dialelos parciais circulantes (Kempthorne & Curnow, 1961; Ramalho et al., 1993). Entretanto, a grande restrição ao uso de dialelos parciais e parciais circulantes é a não possibilidade de predição de CEC dos cruzamentos não realizados.

Para contornar esse problema, uma estratégia normalmente empregada é a predição dos cruzamentos apenas com base na informação da CGC dos genitores. Porém, quando os efeitos de dominância são significativos ou maiores que os efeitos aditivos, esse método torna-se ineficiente (Cruz & Regazzi, 1997).

Levando em consideração esse fato, algumas metodologias têm sido propostas para a predição do desempenho de cruzamentos não avaliados. Como exemplo, pode-se citar o emprego da distância genética obtida por meio de marcadores moleculares na predição da CEC e heterose (Melchinger, 1999; Reif et al., 2003; Balestre et al., 2008a). Contudo, em relação ao seu potencial preditivo, poucos resultados animadores têm sido encontrados na literatura (Dias et al 2004).

Uma alternativa seria a utilização da metodologia dos modelos lineares mistos (BLUP) 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 o emprego da técnica.

Para contornar essa limitação, alguns autores têm sugerido a utilização da informação de marcadores moleculares (similaridade alélica no estado) como opção, quando não se dispõem de dados de pedigree (Bernardo, 1994; 1995; André, 1999; Balestre et al., 2008c). Segundo Nejavati-Javameri (1997), o emprego da similaridade no estado possui a “vantagem” de se utilizar toda a semelhança alélica existente entre os indivíduos avaliados.

No entanto, Bernardo (1994) propõe o uso de um coeficiente de coancestria, que utiliza a informação da similaridade genética, no intuito de corrigir possíveis vieses da similaridade genética no estado, visto que esta tende a superestimar o coeficiente de parentesco. Entretanto, essa metodologia somente é válida para dialelos parciais nos quais não ocorrem cruzamentos endogâmicos, ou seja, pode ser utilizada apenas quando já se possui prévia informação a respeito de grupos heteróticos.

Como alternativa a essa restrição Lynch & Ritland (1999) propõem um procedimento não viesado para estimação conjunta de parentesco por meio de marcadores moleculares, que pode ser utilizado em dialelos completos e parciais. Esse coeficiente tem sido o mais recomendado, pois apresenta menor variância de amostragem quando comparado aos demais coeficientes sugeridos na literatura (Lynch & Ritland, 1999; Bouvet et al., 2008).

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 de Lynch & Ritland (1999) nas predições de híbridos não testados.

1.3 Materiais e métodos

Descrição dos materiais

Foram utilizadas 90 progênies $S_{0:2}$, provenientes de três populações oriundas de três híbridos simples comerciais, sendo 30 progênies da população A (híbrido P30F45), 31 da população B (híbrido Dow 657) e 29 da população C (híbrido DKB333B).

Todas as 90 progênies foram genotipadas com 25 marcadores de microssatélites (SSR), sendo 9 dessas marcas ligadas à QTL para produtividade de grãos e as demais marcas foram selecionadas de acordo com resultado de análise “bootstrap” para essas mesmas populações (Amorim et al., 2006). A extração de DNA, o preparo das reações e as estimativas das distâncias genéticas foram descritas no trabalho desenvolvido por Balestre et al. (2008b). Com base nos resultados desses autores, utilizaram-se as distâncias modificadas de Roger (DMR) para a seleção de 17 linhagens, com o objetivo de recuperar as cinco combinações mais e menos divergentes. A distância modificada de Roger é a expressão mais recomendada para esse tipo de análise (Reif et al., 2005), além de possuir valores aceitáveis de distorção e estresse para essas mesmas populações (Balestre et al., 2008b).

A partir das 17 linhagens selecionadas, realizou-se um dialelo parcial (11 x 6), tendo, dentro de cada grupo, as linhagens sido alocadas de forma a recuperar os cruzamentos entre as linhagens mais e menos divergentes.

Delineamento experimental e correção das médias via modelo misto

Devido à pequena quantidade de sementes, dos 66 cruzamentos possíveis no dialelo, apenas 60 foram avaliados, juntamente com quatro testemunhas, em dois locais no município de Lavras (câmpus da Universidade

Federal de Lavras e Fazenda Vitorinha), na safra 2007/2008. Das quatro testemunhas utilizadas, uma foi o híbrido DKB199 e as três restantes, os híbridos simples que originaram as populações A (P30F45), população B (Dow 657) e população C (DKB 333B). O delineamento experimental utilizado foi um látice 8x8, com três repetições. Cada parcela foi constituída de 2 linhas de 3m com aplicação de 450 kg.ha⁻¹ de NPK (8-28-16) na semeadura e 100 kg.ha⁻¹ de nitrogênio de cobertura.

O modelo linear misto, na forma matricial, utilizado para análise dos dados levando em consideração a recuperação da informação interblocos, seguiu a seguinte expressão:

$$y = X\beta + Z_1a + Z_2g + \delta$$

em que y é o vetor de dados observados, β é o vetor dos efeitos fixos (médias de repetições), a é o vetor aleatório dos valores genotípicos, g é o vetor aleatório dos efeitos de blocos dentro de repetições, δ é o erro aleatório e X , Z_1 e Z_2 são as matrizes de incidência relacionadas a β , a e g , respectivamente.

As médias dos tratamentos ajustadas para cada local foram estimadas pelo procedimento BLUP, conforme Resende (2002), por meio da seguinte expressão:

$$y = \mu + \hat{a} / h^2$$

em que: μ é a média geral do experimento, \hat{a} é o valor genotípico predito + valor referente à recuperação interblocos e h^2 é a herdabilidade do caráter produtividade de grãos.

Segundo Resende (2002), esse estimador é equivalente às médias corrigidas na recuperação interblocos via análise de variância.

Análise dialélica via modelos mistos

A partir dos dados corrigidos para cada local realizou-se a análise dialélica conjunta utilizando o método IV, proposto por Griffing (1956). As análises foram realizadas utilizando-se o software SAS[®] PROC IML System (SAS Institute, 2000). As estimativas dos efeitos fixos e BLUPs das capacidades de combinação (CEC e CGC) foram obtidas de maneira semelhante à apresentada por Bernardo (1995), proposta para dialelos parciais. No entanto, neste estudo, optou-se por utilizar o modelo de dialelo completo, considerando todos os genótipos dentro de um mesmo grupo.

O modelo linear considerado foi dado por:

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

em que y é o vetor das observações; β é o vetor dos efeitos fixos (média geral, e locais); g é o vetor de efeitos da capacidade geral de combinação; s é o vetor da capacidade específica de combinação; 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 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 $\gamma_1 = \frac{\sigma_e^2}{\sigma_{CGC}^2}$ $\gamma_2 = \frac{\sigma_e^2}{\sigma_{CEC}^2}$, em que: A_l é a matriz de similaridade genética

aditiva entre as linhagens. A matriz de parentesco aditiva foi obtida, conforme Lynch & Ritland (1999), pela seguinte expressão:

$$\hat{r}_{xy(k)} = \frac{p_a(S_{bc} + S_{bd}) + p_b(S_{ac} + S_{sd}) - 4p_a p_b}{(1 + S_{ab})(p_a + p_b) + 2p_a p_b}$$

em que:

$\hat{r}_{xy(k)}$ é o estimador do parentesco 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 dos alelos a e b ao longo das linhagens para um dado loco k ;

Considerando que as estimativas de parentescos são obtidas ao longo de muitos locos, é sugerida a utilização de pesos às estimativas, além da alternância dos indivíduos tomados como referência, ou seja, ora utiliza o indivíduo X como referência e ora o indivíduo Y (Lynch & Ritland, 1999). A expressão final, levando em consideração todos os locos, os pesos atribuídos a cada loco e o parentesco estimado para cada loco, é dada por:

$$\hat{r}_{xy} = \frac{\sum_1^L w_{r,x(k)} \hat{r}_{x,y(k)} + \sum_1^L w_{r,y(k)} \hat{r}_{y,x(k)}}{W_{r,x} + W_{r,y}}$$

em que $w_{r,x(k)}$ e $w_{r,y(k)}$, são os pesos para o k -ésimo loco das estimativas de $\hat{r}_{xy(k)}$ e $W_{r,x}$ e $W_{r,y}$ os somatório dos pesos atribuídos a todos os locos envolvidos.

A matriz de parentesco de dominância (A_2) foi construída como sugerido por Henderson (1984) e Van Vleck (1993).

Também foram utilizadas as distâncias de Roger para a construção das matrizes A_1 e A_2 pela expressão $s_{ij} = 1 - d_{ij}$, em que d_{ij} é o valor da distância modificada de Roger. Com isso, foram comparados os coeficientes de Lynch e Ritland (LR) e a distância modificada de Roger (DMR) na eficiência de predição de híbridos não testados.

Predição de híbridos não avaliados

As predições dos valores genotípicos e das CEC dos cruzamentos não avaliados foram realizadas com base em perdas simuladas. Assim, de um total de 60 híbridos avaliados, simularam-se cinco situações considerando desbalançamentos de 5, 10, 20, 30 e 40 híbridos. Para cada situação, o processo foi repetido por 100 vezes. O modelo utilizado foi semelhante ao apresentado por Bernardo (1994).

Considerando que \hat{y}_p sejam as médias dos híbridos avaliados corrigidos para os efeitos de β , tem-se:

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

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

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

em que: \hat{y}_{nt} é o vetor do desempenho 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 variância e covariâncias fenotípicas entre os cruzamentos avaliados.

De modo semelhante, os valores da CEC foram preditos por meio da seguinte expressão:

$$d_{nt} = SV^{-1}\hat{s}_p$$

em que: d_{nt} é o vetor da CEC dos cruzamentos com perda simulada, S é a matriz de covariância genética entre os cruzamentos com perda simulada e cruzamentos avaliados; V é a matriz de variância e covariâncias entre os cruzamentos avaliados, \hat{s}_p vetor das CEC dos híbridos avaliados.

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 e Zimmermann (1986), comparando os doze maiores valores preditos de CEC e valores genotípicos, com os doze 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 neste estudo.

Comparação das predições obtidas pelo coeficiente de Lynch e Ritland e da distância modificada de Roger

Para a predição dos híbridos não avaliados, utilizou-se a informação da similaridade no estado (DMR) e parentesco estimada pelo coeficiente de Lynch & Ritland (1999). O objetivo foi verificar a influência do uso da similaridade no estado sobre a predição de híbridos não avaliados. Para isso, utilizaram-se as correlações entre os valores observados de CEC e as médias fenotípicas com os valores preditos dos híbridos não avaliados.

Também foi aplicado o erro médio padrão de predição, utilizando-se a seguinte expressão:

$$EMP_p = \frac{1}{I} \sum_{i=1}^I \frac{1}{J_i} \sum_{j=1}^{J_i} |\hat{\theta}_{ij} - \theta_i|$$

em que θ_i é o valor médio predito (efeito da CEC ou de cruzamento) do i -ésimo híbrido não testado; I é o número de parâmetros a serem preditos; J_i é o número de vezes que o parâmetro i foi predito e $\hat{\theta}_{ij}$ é a j -ésima predição de θ_i . Essa expressão mede o erro padrão das predições ou a estabilidade de predição dos híbridos não testados em relação ao seu valor médio predito.

Além disso, foi estimado o desvio médio das predições dos híbridos não testados em relação aos valores observados da CEC e do valor genotípico dos cruzamentos. O erro médio de predições foi estimado por meio da seguinte expressão:

$$DM_p = \frac{1}{I} \sum_{i=1}^I \frac{1}{J_i} \sum_{j=1}^{J_i} |\hat{\theta}_{ij} - \theta_i|$$

em que θ_i se refere ao BLUP da CEC ou valor genotípico do i -ésimo cruzamento observado. Essa expressão mede o desvio médio de predição dos híbridos não testados em relação ao valor observado.

1.4 Resultados

O conjunto de genótipos avaliados apresentou valor médio de similaridade genética de 0,817 e valor médio de parentesco de 0,077 (Tabela 1). Esse valor médio de parentesco, apesar de relativamente baixo, quando comparado à estimativa de similaridade genética, pode ser considerado superestimado, visto que os valores negativos obtidos pelo coeficiente de Lynch e Ritland foram considerados iguais a zero.

Pode-se observar que não foram observados valores de parentesco diferentes de zero entre genótipos 30B, 27B, 1B, 28B e 29A, em relação aos demais genótipos (Tabela 1). Com isso, é possível inferir que o coeficiente de Lynch & Ritland (1999) proporcionou uma estimativa coerente de parentesco entre os genótipos avaliados, visto que não apresentou qualquer estimativa de

parentesco entre as progênies da população B e C, embora tenha apresentado um pequeno valor de parentesco entre a progênie 30B e a 29A, as quais pertencem a diferentes populações.

Na Figura 1, pode-se observar que a distância modificada de Roger (DMR), que mede a similaridade genética, agrupou os genótipos de maneira similar ao coeficiente de parentesco de Lynch & Ritland (1999). Apesar de terem ocorrido pequenas modificações dentro dos grupos, fica evidente a formação de dois grupos principais que representam as populações B e C. Deve-se destacar que, embora o genótipo 29A pertença à população A, este foi alocado dentro do grupo de genótipos provenientes da população B, devido ao parentesco estimado somente com o genótipo 30B.

A média fenotípica referente aos sessenta híbridos avaliados foi de 8,962 t ha⁻¹. O valor de herdabilidade obtido neste estudo foi de 0,741. A variância da CGC foi de baixa magnitude (0,002) em relação à variância da CEC (1,622).

As correlações entre os valores preditos e os valores fenotípicos encontram-se na Tabela 2. É possível observar que os valores de correlações, dependendo do número de híbridos retirados, variaram de 0,555 a 0,704 quando se utilizou o coeficiente de LR (Lynch & Ritland 1999). Esse valor pode ser considerado de moderada magnitude, visto que a acurácia obtida neste experimento foi de 0,86.

No entanto, considerando a DMR, os valores de correlações foram inferiores aos obtidos pelo coeficiente de LR, ou seja, variaram de 0,351 a 0,60 (Tabela 2). Também foi observado que os valores de correlação foram maiores quando o número de híbridos preditores utilizado também foi grande, fato este também observado em estudos anteriores (Bernardo, 1994; 1995).

O desvio padrão referente às estimativas de correlações diminuiu à medida que o número de híbridos preditores também diminuiu, ou seja, o desvio padrão dos valores de correlações foi menor quando o desbalanceamento foi

maior (Tabela 2). Além disso, foi observado que o desvio padrão das correlações utilizando a DMR foi maior que o observado no coeficiente LR, em todas as situações de desbalanceamento apresentada.

As mesmas inferências podem ser feitas para os valores de correlações entre a CEC dos híbridos não testados e os seus valores observados (Tabela 3). Os valores de correlações variaram de 0,617 a 0,704, quando se utilizou o coeficiente LR e de 0,507 a 0,632, quando se utilizou a DMR. Também foi observado que quanto maior o grau de desbalanceamento menor a correlação entre o valor predito e observado.

Pelos dados apresentados na Tabela 4, pode-se verificar que a DMR apresentou maior erro médio padrão e maior desvio médio de predição em relação ao coeficiente de parentesco de LR. Ainda na mesma tabela, pode-se constatar que, quando se utilizou o coeficiente LR, a correlação entre os valores observados e os valores médios preditos foi de maior magnitude, quando comparadas as medidas de similaridade (DMR). Assim, é possível inferir que os valores de predição referentes ao uso da DMR comparados aos valores obtidos com o coeficiente de LR, apresentaram maior instabilidade de predição e maior desvio de predição em relação aos valores observados.

Os doze maiores valores médios de predição dos híbridos não testados estão apresentados na Tabela 5. Pode-se observar que os valores fenotípicos correspondentes a esses doze valores preditos, em sua grande maioria, foram de alta magnitude ($10,218 \text{ t ha}^{-1}$), com vantagem de $1,26 \text{ t ha}^{-1}$ em relação à média geral. A eficiência de seleção, levando em consideração os doze maiores valores fenotípicos observados, foi de 40% e a porcentagem de coincidência de seleção foi de 50% (Tabela 5). No entanto, esse resultado somente foi verificado quando se utilizou o coeficiente de LR.

No que se refere às predições utilizando a DMR, houve baixa coincidência de genótipos selecionados e o índice de eficiência de seleção igual

a zero (Tabela 5). Embora houvesse moderada correlação entre os valores observados e valores preditos, os doze maiores valores preditos médios não coincidiram com os melhores híbridos testados.

Na Tabela 6, na qual estão as predições referentes às CEC, pode-se observar que a utilização do melhor preditor linear não viesado, associado ao coeficiente de LR, proporcionou a seleção de indivíduos com altos valores de CEC (0,968) e eficiência de seleção de 52%. Além disso, houve 58% de coincidência ao considerar os doze maiores valores de CEC dos híbridos não testados e os doze maiores valores observados.

O mesmo não se aplica às predições com a DMR, ou seja, a porcentagem de coincidência foi baixa e com baixa eficiência de seleção (Tabela 6). Porém, a DMR proporcionou a seleção do maior valor de CEC observado, o que não foi alcançado com o uso do coeficiente LR (Tabela 6). Porém, em média, o coeficiente LR foi superior a DMR na seleção dos melhores indivíduos.

Com isso, é possível inferir que a DMR, que é um coeficiente de distância genética que utiliza a similaridade no estado, apresentou menor eficiência de seleção de genótipos superiores, quando comparado ao coeficiente de LR.

1.5 Discussões

A utilização do melhor preditor linear não viesado (BLUP) na predição de híbridos não avaliados tem sido recomendado por vários autores (Bernardo, 1994; 1995; 1997; Charcosset et al., 1998). Em todos esses trabalhos, foram verificadas correlações de moderada magnitude entre os híbridos não avaliados e o seu valor fenotípico observado. Esses resultados também foram verificados neste estudo, no qual os valores de correlações entre os híbridos não avaliados e seus valores observados foram de média magnitude, demonstrando que essa

metodologia apresenta moderada eficiência preditiva dos valores genotípicos dos híbridos não avaliados. Essa eficiência pode ser quantificada levando-se em consideração a raiz quadrada da herdabilidade (0,861), ou seja, o valor máximo de correlação a ser obtido nas predições (Bernardo 1994). Nesse caso, as magnitudes das correlações capturaram de 64,5% a 81,4% do valor de correlação limite delimitado pela raiz quadrada da herdabilidade.

Segundo Bernardo (1996), esses valores de correlações podem ser considerados suficientemente altos, visto que proporcionam alta probabilidade ($\pm 80\%$) de se selecionar o melhor híbrido em um conjunto de genótipos selecionados com índice de seleção de 20%. No entanto, pode-se observar que o melhor genótipo não foi alocado dentro do grupo de híbridos selecionados, mesmo havendo coincidência de 50% e eficiência de seleção de 40% entre os híbridos selecionados e as maiores médias observadas. Deve-se lembrar que esses valores genotípicos médios dos híbridos não avaliados foram obtidos levando-se em consideração todas as predições, que foram, em média, de 150 para cada genótipo. Assim, a confiança desse valor médio deve ser medida pelo erro médio padrão associado a cada valor genotípico (Tabela 5).

No que se refere aos valores de correlação, considerando apenas a CEC dos híbridos não testados, pode-se verificar que estas foram de média magnitude (0,617 a 0,704), porém, maiores que as reportadas por Bernardo (1994, 1995), André (1999) e Iemma (2003). Bernardo (1994; 1995) atribuiu a baixa correlação entre a CEC dos híbridos não testados com a CEC observada ($r \leq 0,15$) devido a viés na estimativa da variância da CEC. Este autor observou que a variância da CEC foi extremamente baixa quando comparada à variância da CGC. Sendo assim, os cruzamentos obtidos por ele foram fortemente influenciados pela CGC, convergindo em predições mais acuradas e mais fáceis de serem obtidas, visto que não houve CEC significativa. No trabalho desenvolvido por Bernardo (1994), as predições considerando o modelo aditivo

apresentaram praticamente o mesmo resultado do modelo completo, demonstrando baixa participação da CEC nas predições dos híbridos não avaliados.

Neste estudo, a variância da CEC (V_{CEC}) foi bem maior que a variância da CGC, captando quase 99% da variância genética total. Assim, pode-se inferir que as predições obtidas nesse estudo são de grande relevância, considerando que as combinações híbridas foram altamente influenciadas pela CEC. Deve-se levar em consideração a hipótese de possível viés na estimativa V_{CEC} , principalmente devido a combinações endogâmicas e não endogâmicas obtidas no dialelo e valores de variância aditiva e de dominância normalmente reportados na cultura do milho (Boer & Hoeschele 1993; Hallauer & Miranda Filho, 1981). Sendo assim, esse resultado pode ser considerado restrito ao conjunto de híbridos avaliados neste estudo.

Considerando a probabilidade de selecionar as melhores combinações específicas (CEC), pode-se verificar, com base nas predições dos híbridos não testados, que a eficiência de seleção entre os doze maiores valores de CEC observados e doze maiores valores de CEC preditos foi de 50%, com coincidência de 58%. Embora os valores de CEC fossem obtidos ao longo de várias predições com diferentes combinações e número de preditores, verificasse, pelo erro médio de predição, que grande parte dos doze maiores valores de predições da CEC foi positiva, independentemente do grau de desbalanceamento e combinação de preditores (Tabela 6). Esse fato demonstrou a eficiência do BLUP na predição das CEC dos híbridos não testados ao longo de todas as 500 diferentes combinações obtidas, considerando os cinco níveis de desbalanceamento.

Na comparação entre a utilização da similaridade no estado não corrigida (DMR) com o coeficiente LR, pode-se verificar que, em todas as análises comparativas, o uso da DMR apresentou resultados inferiores ao coeficiente LR.

Além dos menores valores de correlações e maior desvio padrão desse estimador, a DMR também apresentou maior erro padrão de predição, demonstrando que as predições obtidas por meio do uso desse coeficiente foram menos estáveis em relação ao coeficiente de LR. Além disso, considerando os valores genotípicos observados e os valores de CEC, constata-se que o uso da DMR acarretou em valores de predição mais distantes do valor observado. Essa inferioridade também pode ser observada considerando a eficiência de seleção das melhores combinações híbridas.

Esse maior erro de predição ocasionado pela DMR nos valores preditos já era esperado, devido à recuperação da informação genética que leva em consideração apenas o aspecto evolutivo (similaridade no estado) dos indivíduos e não seu parentesco. Assim, em maior ou menor peso, todos os valores de covariância genética são recuperados no preditor linear (BLUP) de cada valor de CEC e CGC, inclusive daqueles indivíduos que não partilham alelos herdados de genitores em comum. Esse fato convergiu para predições menos estáveis, visto que, independentemente do conjunto de híbridos utilizados como preditores, todos esses poderão contribuir de certa forma na predição do híbrido não avaliado.

No entanto, esse viés pode ser eliminado quando se utiliza o coeficiente de Lynch & Ritland (1999), pois esse coeficiente, apesar de utilizar dados moleculares, leva em consideração a probabilidade condicional de que dois indivíduos possuam alelos herdados de genitores em comum. Assim, somente indivíduos com probabilidade de possuírem alelos em comum participam da predição dos híbridos não avaliados. Porém, aqueles com parentescos iguais à zero não contribuirão na predição, concorrendo em resultados mais coerentes.

A utilização da similaridade no estado não se restringe apenas a DMR, mas também a coeficientes que levam em consideração a coocorrência de bandas (0 e 1). Com isso, Bernardo (1994) sugeriu uma expressão para corrigir

esse viés. No entanto, como já comentado, em dialelos completos, esse modelo não se aplica, pois não contempla todas as similaridades possíveis entre os genótipos em avaliação. Por outro lado, o coeficiente de LR pode ser utilizado em dialelos completos, pois as medidas de parentesco obtidas por esse estimador levam em consideração todos os indivíduos em avaliação.

Assim, pode-se inferir que a utilização da similaridade no estado, mesmo utilizando qualquer outro coeficiente que considere todas as covariâncias genéticas, sem ponderar a probabilidade de que determinado alelo seja herdado de genitores em comum, pode apresentar menor eficiência na predição em híbridos não avaliados.

Outro comentário a ser realizado seria a razão de se utilizar o modelo de dialelo completo desbalanceado em vez de dialelo parcial. Como já comentado no tópico Material e métodos, os cruzamentos foram realizados em esquema 11 x 6, o que levaria à utilização de esquemas de dialelos parciais para avaliação da CGC e CEC. No entanto, esses cruzamentos foram obtidos com o objetivo de recuperar combinações mais e menos divergentes, o que acarretou na alocação de indivíduos aparentados fora dos grupos, ou seja, não necessariamente ocorreram grupos distintos.

Quando se utiliza um esquema de dialelo parcial, parte da informação de parentesco é perdida na matriz A, ao passo que, quando se considera um dialelo completo, mesmo altamente desbalanceado, essa informação pode ser recuperada acarretando em maior acurácia na predição dos valores da CGC e CEC. Esse fato ficou evidente ao se analisar o valor médio de acurácia de predição da CGC e CEC considerando o modelo de dialelo completo desbalanceado (0,99 e 0,85 para CGC e CEC, respectivamente) e dialelo parcial com e sem a informação de parentesco (0,988 e 0,844; 0,987 e 0,833). Devido a isso, é possível justificar a preferência, neste estudo, pela utilização do modelo de dialelo completo desbalanceado em vez do dialelo parcial.

Conclui-se que o BLUP é uma metodologia eficaz para predição e seleção e cruzamentos não realizados, mesmo quando ocorrem predominância de efeitos não aditivos e alto grau de desbalanceamento.

Além disso, constatou-se que o coeficiente de parentesco de Lynch e Ritland associado ao BLUP proporcionou predições mais acuradas quando comparado à similaridade no estado, obtida pelas distâncias modificadas de Roger.

1.6 Agradecimentos

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), pelos auxílios concedidos para a realização deste trabalho.

2 REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, E.P.; AMORIM, U.B.O.; SANTOS, J.B. dos. Genetic distance based on SSR and grain yield of inter an intrapopulational maize single cross hybrids. **Maydica**, Bergamo, v.51, p. 507-513, 2006.

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, MG.

BALESTRE, M.; MACHADO, J.C.; LIMA, J.L.; SOUZA, J.C.; NOBREGA FILHO, L. Genetics distances estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids **Genetics and Molecular Research**, Ribeirão Preto, v.7, n. 1, p. 65-73, 2008a.

BALESTRE, M.; VON PINHO, R.G.; SOUZA, J.C.; LIMA, J.L. Comparison of maize similarity and dissimilarity genetic coefficients based on microsatellites markers. **Genetics and Molecular Research**, Ribeirão Preto, v.7, n. 3, p. 695-705, 2008b.

BALESTRE, M.; VON PINHO, R.G.; SOUZA, J.C.; MACHADO J.C. Potential of maize single-cross hybrids for extraction of inbred lines using the mean components and mixed models with microsatellite marker information **Genetics and Molecular Research**, Ribeirão Preto, v.7, n. 4, p.1106-1118, 2008c.

BERNARDO, R. Best linear unbiased prediction of maize single cross performance. **Crop Science**, Madison, v. 36, n. 1, p. 50-56, Jan./Feb. 1996.

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. 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. RFLP markers and predicted testcross performance of maize sister inbreds. **Theoretical and Applied Genetics**, Berlin, v.95, p. 655-659, 1997.

BOER, I.J.M. de.; HOESCHELE, I. Genetic evaluation methods for populations with dominance and inbreeding. **Theoretical and Applied Genetics**, Berlin, v. 86, n. 2/3, p. 245-258, April. 1993.

BOUVET, J.M.; KELLY, B.; SANOU, H.; ALLAL, F. Comparison of marker and pedigree based methods for estimating heritability in an agroforestry population of *Vitellaria paradoxa* C.F. Gaertn. (shea tree) **Genet Resour Crop Evol**, 2008, (In Press).

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**, v.38, n. 1, p. 38-44, 1998.

CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético** Viçosa, UFV. 1997. 390 p.

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, 2004.

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

HALLAUER, A. R.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. Ames: Iowa State University Press, Ames: Iowa State University Press, 1981. 468 p.

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

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

IEMMA, M. **Uso do melhor preditor linear não viesado (BLUP) em análise dialélica e predição de híbridos**. 2003. 81p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP.

KEMPTHORNE, O.; CURNOW, R.N. The partial diallel cross. **Biometrics**, North Caroline, v.17, n. 2, p.229-250, June 1961.

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. p. 99-118, 1999.

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**, v. 75, n. 7, p. 1738-1735, July., 1997.

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

REIF, J. C.; MELCHINGER, A. E.; XIA, X. C.; WARBURTON, M. L.; HOISINGTON, D. A.; VASAL, S. K.; SRINIVASAN, G.; BOHN, M.; FRISCH, M Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. **Crop Science**, Madison, v. 43, n. 4, p. 1275-1282. 2003.

RESENDE, M.D.V. (2002) **Genética Biométrica e Estatística no Melhoramento de Plantas Perenes**. Brasília, Embrapa Informação Tecnológica, 2002, 975 p.

VAN VLECK, L.D. **Selection index in introduction to mixed models methods for genetic improvement of animals. The green book**, Press. Boca Raton, Florida. 1993. 481 p.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 496 p.

3 TABELAS

TABELA 1 Estimativas da similaridade genética no estado (acima da diagonal), obtidas por meio da distância modificada de Roger (DMR) e estimativa de parentesco, obtidas por meio do coeficiente de Lynch e Ritland (LR), em dezessete progênies $S_{0.2}$ de milho.

		<i>SIMILARIDADE GENÉTICA NO ESTADO (DMR)</i>																
		30B	27B	19C	22C	1B	1C	14C	29C	5C	12C	15C	23C	25C	28B	7C	28C	29A
PARENTESCO (RL)	30B	1,00	0,93	0,73	0,73	0,90	0,73	0,69	0,74	0,74	0,76	0,74	0,73	0,68	1,00	0,80	0,73	0,76
	27B	0,21	1,00	0,74	0,74	0,88	0,74	0,70	0,75	0,74	0,76	0,74	0,74	0,69	0,93	0,80	0,74	0,74
	19C	0,00	0,00	1,00	0,84	0,73	0,88	0,84	0,90	0,88	0,93	0,88	0,84	0,81	0,73	0,97	0,88	0,74
	22C	0,00	0,00	0,08	1,00	0,73	0,88	0,78	0,90	0,85	0,88	0,85	0,97	0,80	0,73	0,97	0,90	0,72
	1B	0,10	0,12	0,00	0,00	1,00	0,76	0,71	0,72	0,73	0,75	0,73	0,73	0,66	0,90	0,80	0,75	0,74
	1C	0,00	0,00	0,04	0,07	0,00	1,00	0,85	0,86	0,90	0,98	0,90	0,88	0,77	0,73	0,97	0,94	0,74
	14C	0,00	0,00	0,00	0,03	0,00	0,18	1,00	0,81	0,88	0,85	0,88	0,78	0,78	0,69	0,97	0,86	0,69
	29C	0,00	0,00	0,09	0,00	0,00	0,20	0,09	1,00	0,88	0,94	0,88	0,90	0,84	0,74	0,97	0,88	0,73
	5C	0,00	0,00	0,13	0,04	0,00	0,29	0,20	0,12	1,00	0,90	0,97	0,85	0,85	0,74	0,97	0,88	0,77
	12C	0,00	0,00	0,00	0,01	0,00	0,09	0,14	0,14	0,15	1,00	0,90	0,88	0,88	0,76	0,97	0,94	0,77
	15C	0,00	0,00	0,07	0,00	0,00	0,21	0,18	0,21	0,18	0,29	1,00	0,85	0,85	0,74	0,97	0,88	0,77
	23C	0,00	0,00	0,07	0,24	0,00	0,09	0,00	0,15	0,17	0,00	0,02	1,00	0,80	0,73	0,97	0,90	0,72
	25C	0,00	0,00	0,00	0,00	0,00	0,01	0,13	0,33	0,07	0,28	0,16	0,04	1,00	0,68	0,97	0,78	0,67
	28B	0,22	0,27	0,00	0,00	0,13	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,80	0,73	0,76
	7C	0,00	0,00	0,07	0,03	0,00	0,50	0,50	0,10	0,37	0,39	0,39	0,24	0,36	0,00	1,00	0,97	0,80
	28C	0,00	0,00	0,05	0,00	0,00	0,25	0,12	0,37	0,14	0,04	0,15	0,06	0,12	0,00	0,24	1,00	0,71
	29A	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

TABELA 2 Valores mínimos médios e máximos e desvio padrão de correlação entre o valor genético observado e o valor predito ao longo de 100 simulações, considerando o parentesco estimado pelo coeficiente de Lynch e Ritland (LR) e a similaridade genética obtida por meio da distância modificada de Roger (DMR), em dezessete progênies $S_{0,2}$ de milho.

LR					
Número de híbridos retirados					
	5	10	20	30	40
Mínimo	0,219	-0,029	0,007	0,006	0,013
Média	0,702	0,643	0,601	0,560	0,555
Máximo	0,961	0,945	0,825	0,763	0,722
Desvio padrão	0,208	0,203	0,173	0,173	0,112
DMR					
Número de híbridos retirados					
	5	10	20	30	40
Mínimo	-0,702	-0,025	0,070	0,012	-0,299
Média	0,600	0,562	0,503	0,459	0,351
Máxima	0,983	0,919	0,783	0,686	0,632
Desvio padrão	0,382	0,261	0,189	0,175	0,172

TABELA 3 Valores mínimos médios e máximos e desvio padrão de correlação entre a capacidade específica de combinação observada (CEC) e o valor predito ao longo de 100 simulações, considerando o parentesco estimado pelo coeficiente de Lynch e Ritland (LR) e a similaridade genética obtida por meio da distância modificada de Roger (DMR), em dezessete progênies $S_{0,2}$ de milho.

	LR				
	Número de híbridos retirados				
	5	10	20	30	40
Mínimo	-0,093	0,402	0,398	0,528	0,347
Média	0,704	0,701	0,647	0,644	0,617
Máximo	0,967	0,958	0,831	0,839	0,750
Desvio padrão	0,300	0,136	0,108	0,082	0,090

	DMR				
	Número de híbridos retirados				
	5	10	20	30	40
Mínimo	-0,698	-0,103	-0,233	0,227	-0,253
Média	0,632	0,573	0,525	0,524	0,507
Máximo	0,975	0,949	0,851	0,711	0,761
Desvio padrão	0,381	0,269	0,233	0,129	0,201

TABELA 4 Erro médio padrão de predição (EMP), desvio médio (DM) e correlação dos valores médios preditos com os valores observados, obtidos ao longo de 500 simulações de desbalanceamento, considerando a estimativa de parentesco de Lynch e Ritland (LR) e similaridade obtida pela distância modificada de Roger (DMR), em dezessete progênies $S_{0,2}$ de milho.

	LR		DMR	
	Valor genotípico	CEC	Valor genotípico	CEC
EMP	0,469	0,171	0,691	0,391
DM	0,911	0,839	1,126	0,978
Correlação	0,537	0,665	0,350	0,467

TABELA 5 Eficiência de seleção (ES%) e porcentagem de coincidência (C%) dos doze maiores valores genotípicos preditos (VG) e suas respectivas produtividades, considerando 500 simulações, em relação às doze maiores produtividade obtidas na avaliação de sessenta híbridos de milho.

LR				
Ordem	Cruzamento	VG	Produção	EMP ¹
1	30BX7C	9,644	10,777	0,388
2	27BX7C	9,636	9,762	0,407
3	27BX28C	9,615	10,162	0,476
4	1CX28B	9,613	9,239	0,401
5	1BX25C	9,582	7,667	0,485
6	30BX28C	9,580	11,347	0,509
7	5CX28B	9,548	10,108	0,370
8	29CX28B	9,510	11,110	0,536
9	15CX28B	9,505	10,504	0,446
10	30BX23C	9,456	11,211	0,616
11	14CX28B	9,413	10,102	0,327
12	12CX28B	9,356	10,629	0,342
Média		9,538	10,218	0,442
C%	50,00			
ES %	40,00			
DMR				
Ordem	Cruzamento	VG	Produção	EMP
1	27BX28B	10,949	4,457	0,902
2	1BX28B	10,295	6,752	1,436
3	1BX29A	10,204	9,622	1,519
4	30BX29A	10,201	8,415	1,244
5	30BX28C	9,932	11,347	0,601
6	1CX29A	9,836	8,169	1,094
7	27BX29A	9,784	9,080	0,785
8	1CX28B	9,769	9,239	0,825
9	27BX28C	9,718	10,162	0,596
10	1BX23C	9,681	9,771	0,619
11	30BX23C	9,635	10,108	0,486
12	5CX28B	9,621	10,777	0,609
Média		9,969	9,027	0,901
C%	16,70			
ES%	0,00			

¹Erro médio padrão associado a cada predição considerando 500 simulações de desbalanceamento

TABELA 6 Eficiência de seleção (ES%) e porcentagem de coincidência (C%) dos doze maiores valores preditos de capacidade específica de combinação (CEC¹) e seus respectivos valores observados (CEC²), considerando 500 simulações, em relação aos doze maiores valores de capacidade específica obtidas na avaliação de sessenta híbridos de milho.

LR				
Ordem	Cruzamento	CEC ¹	CEC ²	EMP ¹
1	30BX7C	0,285	1,386	0,167
2	1CX28B	0,278	0,248	0,167
3	27BX7C	0,267	0,633	0,160
4	30BX28C	0,254	1,776	0,146
5	1BX25C	0,235	-0,930	0,326
6	15CX28B	0,232	1,169	0,154
7	1BX7C	0,231	1,625	0,184
8	5CX28B	0,218	0,867	0,131
9	30BX23C	0,211	1,661	0,323
10	12CX28B	0,209	1,240	0,220
11	27BX28C	0,197	0,910	0,196
12	1BX23C	0,188	0,603	0,344
Média		0,234	0,933	0,194
C%	58,00			
ES	50,00			
DMR				
Ordem	Cruzamento	CEC ¹	CEC ²	EMP
1	27BX28B	0,388	-3,399	0,519
2	30BX29A	0,297	-0,457	0,297
3	30BX28C	0,227	1,776	0,254
4	1BX28B	0,180	-1,721	0,500
5	1BX29A	0,168	0,432	0,512
6	1BX23C	0,120	0,604	0,344
7	1CX28B	0,097	0,248	0,466
8	30BX23C	0,088	1,661	0,348
9	1BX28C	0,087	3,339	0,251
10	27BX28C	0,078	0,911	0,196
11	27BX29A	0,074	0,034	0,766
12	12CX28B	0,068	1,240	0,315
Média		0,156	0,389	0,397
C%	33,33			
ES	20,00			

¹Erro médio padrão associado a cada predição considerando 500 simulações de desbalanceamento

4 FIGURA

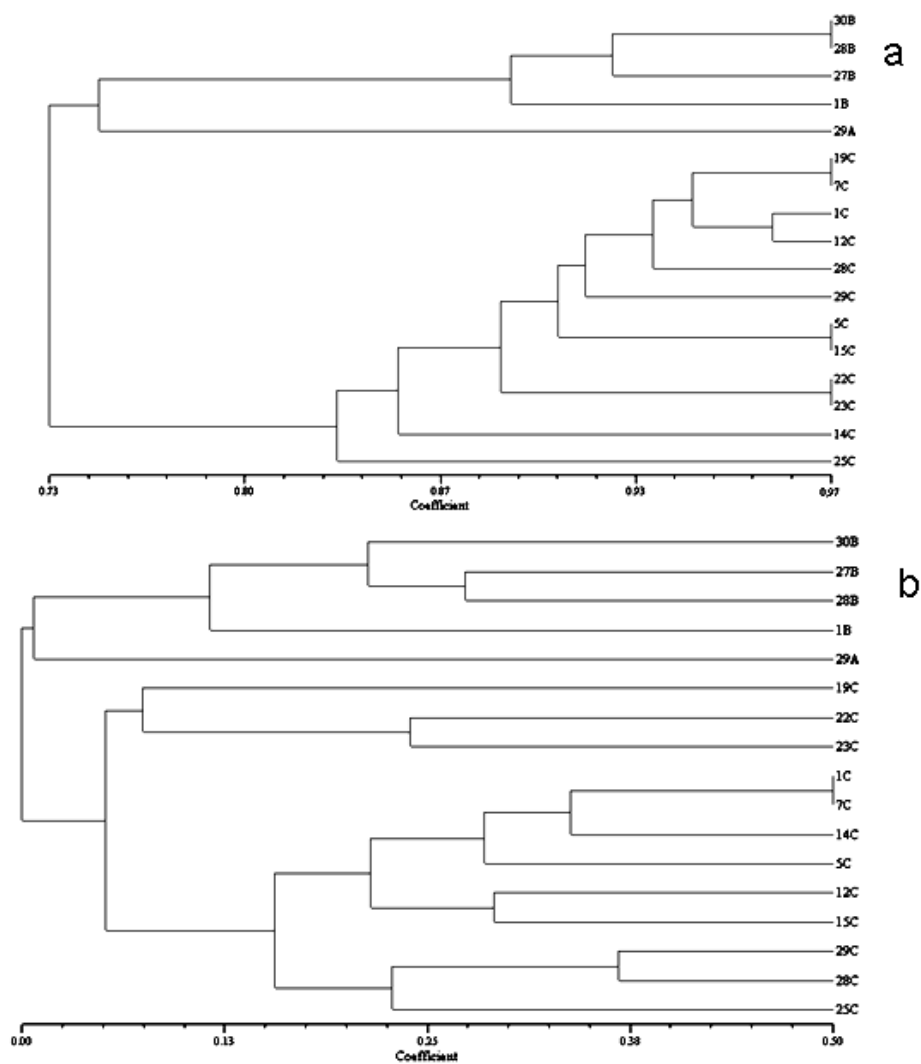


FIGURA 1 Dendrograma da similaridade genética (a) e do parentesco (b) estimado em dezessete progênies $S_{0.2}$ de milho. A progênie com código A é proveniente do híbrido P30F45 e as progênies B e C são oriundas dos híbridos Dow657 e DBB333B, respectivamente.