

**ESTRATÉGIAS DE PREDIÇÃO DE  
CRUZAMENTOS NÃO REALIZADOS EM  
ANÁLISE DIALÉLICA, UTILIZANDO  
INFORMAÇÕES DE MARCADORES  
MOLECULARES**

**JOÃO LUÍS DA SILVA FILHO**

**2004**

**JOÃO LUÍS DA SILVA FILHO**

**ESTRATÉGIAS DE PREDIÇÃO DE CRUZAMENTOS NÃO  
REALIZADOS EM ANÁLISE DIALÉLICA, UTILIZANDO  
INFORMAÇÕES DE MARCADORES MOLECULARES**

Tese 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 "Doutor".

Orientador

Prof. Dr. Eduardo Bearzoti

LAVRAS  
MINAS GERAIS - BRASIL

2004

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

Silva Filho, João Luís da

Estratégias de predição de cruzamentos não realizados em análise dialéctica,  
utilizando informações de marcadores moleculares / João Luís da Silva Filho. --

Lavras : UFLA, 2003.

70 p. : il.

Orientador: Eduardo Bearzoti.

Tese (Doutorado) – UFLA.

Bibliografia.

1. Feijão. 2. Cruzamento. 3. Marcador molecular. I. Universidade Federal de  
Lavras. II. Título.

CDD-635.65223

**JOÃO LUÍS DA SILVA FILHO**

**ESTRATÉGIAS DE PREDIÇÃO DE CRUZAMENTOS NÃO  
REALIZADOS EM ANÁLISE DIALÉLICA, UTILIZANDO  
INFORMAÇÕES DE MARCADORES MOLECULARES**

Tese 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 “Doutor”.

APROVADA em 13 de maio de 2004

Dr. Antônio Augusto Franco Garcia	ESALQ/USP
Dr. Daniel Furtado Ferreira	UFLA
Dr. João Batista Duarte	UFG
Dr. Júlio Sílvio de Souza Bueno Filho	UFLA

Prof. Dr. Eduardo Bearzoti  
UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS - BRASIL

Ao *Senhor Jesus*, de *Sabedoria* e *Bondade* inimagináveis aos cérebros dos  
homens mais sábios, por mais essa conquista na minha vida,

OFEREÇO.

Aos meus pais, João Luiz e Antonieta.

Às minhas irmãs, Nágida e Naíla.

Ao pequeno Ahas, meu sobrinho.

Aos meus avós, Antonio Ferreira (*in memoriam*) e Maria Ferreira.

Aos meus padrinhos, Chiquinho e Maria de Lourdes (*in memoriam*).

Aos meus tios, primos, parentes e amigos.

DEDICO.

## AGRADECIMENTOS

A *DEUS*, pela dádiva da minha existência.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Biologia, pela oportunidade da realização da pós-graduação, tanto o Mestrado como o Doutorado.

Ao Fundo de Amparo à Pesquisa e Extensão do Estado de Minas Gerais (FAPEMIG) e ao Conselho de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro para a realização do meu curso.

À Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), que permitiu que me ausentasse do trabalho em momentos críticos do desenvolvimento desta tese.

Ao professor Eduardo Bearzoti, meu muito obrigado pela orientação segura e valiosos ensinamentos transmitidos.

Ao professor Magno Antonio Patto Ramalho, pelos valiosos ensinamentos transmitidos e a certeza de que sua carreira de pesquisador dedicado, competente e prestativo será uma referência para a minha, que se inicia.

Aos demais professores e funcionários do Departamento de Biologia, em especial aos professores João Bosco e César Brasil e a professora Lisete Davide, pelos ensinamentos e cordial convivência.

Aos Doutores Antonio Augusto Franco Garcia, João Batista Duarte, Júlio Silvío de Souza Bueno Filho e Daniel Furtado Ferreira, pela disponibilidade em participarem da banca de defesa e valiosas contribuições e sugestões. Um agradecimento especial ao professor Júlio, pelas valiosas sugestões durante a elaboração deste trabalho.

Um agradecimento muito especial a Dra. Ângela de Fátima Barbosa Abreu, exímia pesquisadora, que gentilmente cedeu os dados experimentais.

Aos colegas do Curso de Genética e Melhoramento de Plantas, pelo convívio e amizade. Um agradecimento hiper-especial àqueles que tiveram a paciência de compartilhar comigo a “República dos Perdidos”: Airton Nunes (Airtinho de Açúcar), Alex Teixeira (Alésquio), Dudu Lambert (Congonhalense), e Vanderlei (Vandertupa). Agradeço também àqueles que, por último, tive a oportunidade conviver: Breno, Cristovão e Kaesel. Não poderia deixar de citar a D. Maria, nossa secretária e uma amiga de todas as horas. Muito obrigado pelo carinho e dedicação.

Agradeço ainda aos amigos Francislei Vitti Raposo e Jefferson Matavelli, pela amizade e convivência agradável e também aos “embrapianos” Ana Claudia e Arnaldo Rocha.

A todos aqueles que não foram citados, mas que de alguma forma contribuíram para o êxito deste trabalho.

## SUMÁRIO

	<b>Página</b>
RESUMO .....	i
ABSTRACT .....	iii
1 INTRODUÇÃO .....	1
2 REFERENCIAL TEÓRICO .....	4
2.1 A análise dialélica no melhoramento de plantas .....	4
2.2 Divergência genética e capacidade combinatória dos genitores .....	10
2.3 Covariância genética entre parentes e predição de cruzamentos .....	13
2.4 Utilização das EMM no melhoramento genético de plantas .....	15
2.5 Procedimentos de estimação de parâmetros genéticos .....	19
2.5.1 Quadrados mínimos ordinários (OLS) .....	20
2.5.2 Quadrados mínimos generalizados (GLS) .....	21
2.5.3 Máxima verossimilhança (ML) .....	21
2.5.4 Equações de modelos mistos (EMM) .....	22
3 MATERIAL E MÉTODOS .....	25
3.1 Material genético e delineamento experimental .....	25
3.2 Caracterização molecular dos genitores .....	26
3.3 Cálculo de parentesco e coeficiente de dominância, via marcadores moleculares, de acordo com Lynch e Ritland (1999) .....	27
3.4 Análise de similaridade genética entre os genitores .....	29
3.5 Modelos utilizados na predição de cruzamentos dialélicos .....	29
3.5.1 Modelo aditivo .....	30
3.5.2 Modelo de distância .....	31
3.5.3 Modelos de regressão fatorial .....	32
3.5.4 Equações de modelos mistos (EMM) .....	35
3.6 Procedimentos de simulação de perdas e estatísticas de avaliação das predições dos cruzamentos com perda simulada .....	39
4 RESULTADOS .....	41
4.1 Divergência genética e parentesco via Lynch e Ritland (1999) .....	41
4.2 Predição de cruzamentos não realizados pelos diferentes modelos .....	44
4.2.1 Modelo aditivo e modelo de distância .....	47
4.2.2 Modelos de regressão fatorial .....	48
4.2.3 Modelos EMM .....	51
5 DISCUSSÃO .....	53
5.1 Considerações sobre a metodologia de Lynch e Ritland (1999) .....	53
5.2 Considerações gerais sobre os modelos de predição de cruzamentos .....	55
5.3 Modelo aditivo e modelo de distância .....	56
5.4 Modelos de regressão fatorial .....	59
5.5 Equações de modelos mistos (EMM) .....	60



6 CONCLUSÕES.....	64
REFERÊNCIAS BIBLIOGRÁFICAS.....	65

## RESUMO

SILVA FILHO, João Luís da. **Estratégias de predição de cruzamentos não realizados em análise dialélica, utilizando informações de marcadores moleculares.** Lavras: UFLA, 2004. 70p. (Tese - Doutorado em Agronomia/Genética e Melhoramento de Plantas).\*

Os cruzamentos dialélicos estão entre as estratégias mais utilizadas pelos melhoristas para a escolha de genitores e ou populações segregantes. Uma de suas principais limitações, contudo, é o trabalho demandado na realização de cruzamentos, que cresce geometricamente à medida que o número de genitores que se deseja inter cruzar aumenta. O presente trabalho teve por objetivo apresentar diferentes estratégias de predição de cruzamentos não realizados em análise dialélica, utilizando dados de marcadores moleculares. Foram utilizados dados de experimentos com a cultura do feijoeiro, envolvendo 120 cruzamentos derivados de 22 genitores (de quatro diferentes raças), avaliados em três locais do estado de Minas Gerais. A disponibilidade de uma matriz de padrões moleculares dos genitores (matriz de “0’s” e “1’s”), constituída por 136 produtos polimórficos de marcadores RAPD, permitiu a aplicação de metodologias que utilizam diferentes princípios para a predição de cruzamentos não realizados, as quais são caracterizadas a seguir: (i) modelo aditivo, em que a predição dos cruzamentos não avaliados foi realizada com base nas capacidades gerais de combinação dos genitores, estimadas a partir dos cruzamentos avaliados em que eles participaram; (ii) modelo de distância, similar ao modelo aditivo, contudo apenas a informação adicional de uma covariável, cujo valor atribuído a cada cruzamento foi a distância genética entre os genitores, calculada a partir dos marcadores moleculares; (iii) modelos de regressão fatorial, em número de quatro, que também fazem uso de covariáveis, só que calculadas a partir dos escores dos quatro primeiros componentes principais da matriz de “0” e “1”, em que cada produto de amplificação foi tomado como uma variável; (iv) metodologia de modelos mistos, considerando duas situações: a primeira delas, desconsiderando a endogamia existente nos cruzamentos, dado que os genitores são aparentados e a segunda, considerando a endogamia dos cruzamentos, numa tentativa de minimizar sua influência sobre as predições dos cruzamentos não avaliados. Perdas de 12, 24, 36, 48 e 60 cruzamentos foram simuladas 1000 vezes cada uma para os modelos aditivo, de distância e de regressão fatorial, e 250 vezes para as equações de modelos mistos, sendo, então, seus valores

---

\* Comitê Orientador: Prof. Dr. Eduardo Bearzoti (Orientador) – UFLA.

preditos com base no desempenho dos cruzamentos avaliados. Em cada simulação, estimavam-se as correlações de Pearson e de Spearman entre os valores preditos dos cruzamentos e os valores fenotípicos observados. Ao longo das simulações obtidas para cada modelo em cada tamanho de perda, foram estimadas as correlações médias, mínimas, máximas e desvios padrões. Em todos os modelos, houve uma tendência de se obter maiores correlações quando se tinha um maior número de cruzamentos preditores (menor número de perdas simuladas). Os modelos de regressão fatorial apresentaram melhores estimativas de correlações entre os valores preditos e observados. As equações de modelos mistos e o modelo de distância apresentaram resultados similares. A inclusão da covariável nas equações de modelos mistos não trouxe nenhum benefício às predições, provavelmente devido ao fato das estimativas das endogamias dos cruzamentos, tomadas como covariáveis, terem sido muito próximas entre si.

## ABSTRACT

SILVA FILHO, João Luís da. **Alternatives for predicting effects of missing crosses in diallels, with information from molecular markers** Lavras: UFLA, 2004. 70p. (Thesis – Doctoral in Genetics and Plant Breeding) \*

Diallel crosses are largely used by breeders to assist choosing parents and/or populations. One constraint of such designs is the effort required to obtain crosses, the amount of which grows exponentially as the number of parents increases. This work aimed at evaluation some approaches to predict effects of missing crosses in diallel analysis, comprising 120 crosses among 22 common bean parents (from four different races), carried out in three sites of Minas Gerais State in Brazil. A matrix of 0's and 1's was available, with respect to 136 RAPD loci, aiming the application of methodologies with different principles to predict missing crosses. Such methodologies were: (i) an additive model, where prediction in made with the general combining abilities of the parents (molecular information not used); (ii) a distance model, similar to (i), but with additional covariate, corresponding to the genetic distance (calculate with molecular data) among parents of each cross; (iii) factor regression models (four variants were considered) using covariates that were subsets of the first four principal components relative to the matrix of 0's and 1's; (iv) a mixed models (best linear unbiased prediction) with two variants, considering and non-considering the inbreeding that arose in crosses where parents were related. In each model, losses of 12, 24, 36, 48 and 60 crosses were simulated one thousand times in each situation for the additive, the genetic distance and factor models. For the mixed models, in each situation 250 simulations were performed. The effects of such crosses were predicted considering the remaining data. In each simulated condition, Pearson's and Spearman's coefficient of correlation were estimated among predictions and observed phenotypic values of missing crosses. Across simulations, averages, maximum and minimum values, and standard deviations of those coefficients were calculated. With all models, there was a trend of increasing the coefficients of correlation with less missing crosses. Factor models showed the highest coefficients of correlation among predicted and observed values. Best linear unbiased prediction and the distance model showed similar results. Inclusion of a covariate in the mixed model, to account for inbreeding, brought no further improvement on the predictions. This may

---

\* Guidance committee: Dr. Eduardo Bearzoti (Adviser) – UFLA.

have been a result of the quite similar values of inbreeding estimated in the set of crosses.

## 1 INTRODUÇÃO

Um dos objetivos de um programa de melhoramento é aumentar a frequência de alelos favoráveis em uma população pela escolha e posterior intercruzamento de indivíduos julgados como superiores. Assim, a realização de hibridações é uma prática necessária e indispensável no melhoramento de qualquer cultura.

Uma das estratégias mais consagradas para a avaliação de linhagens ou populações em combinações híbridas são os chamados cruzamentos dialélicos. Um conjunto de cruzamentos dialélicos é constituído por todas as combinações duas a duas envolvendo  $n$  genitores. Vários são os modelos existentes para análise dialélica e sua utilização varia de acordo com os objetivos do estudo (Jinks e Hayman, 1953; Griffing, 1956; Gardner e Eberhart, 1966). Os modelos de Griffing (1956), por exemplo, decompõem o valor do cruzamento em efeitos das capacidades de combinação geral (CGC) e específica (CEC) dos genitores.

Uma das limitações dos cruzamentos dialélicos é a realização de cruzamentos, cujo número cresce geometricamente à medida que o número de genitores que se deseja inter cruzar aumenta. Variações na estrutura das realizações de cruzamentos têm sido propostas, como os dialelos circulantes (Kempthorne e Curnow, 1961).

Veiga (1998) verificou, tanto via simulação como com dados experimentais, que os dialelos circulantes são eficientes para a estimação da CGC. Entretanto, uma de suas limitações é a não estimação da CEC de cruzamentos não realizados.

Quando um dado cruzamento foi perdido ou não realizado, uma alternativa é predizê-lo com base nas estimativas das CGC dos genitores. Essa

aproximação é melhor quando a magnitude da variância aditiva é relativamente superior à não-aditiva (Cruz e Regazzi, 1998).

Mesmo em plantas autógamas, como o feijoeiro, em que o objetivo final é a obtenção de linhagens homozigóticas, não só efeitos significativos da CEC têm sido obtidos (Mendonça, 2001; Abreu, 1997), como também podem ser de suma importância na seleção de populações segregantes. Isso porque a significância desse parâmetro depende da heterose que, de acordo com Falconer (1987), ocorre em função da dominância e de locos complementares entre os genitores. Uma maior CEC, portanto, pode ser um indicio de um maior número de locos segregando.

Algumas estratégias para a aproximação da CEC de cruzamentos não realizados têm sido propostas. Entre elas estão as estimativas de distância genética, calculadas a partir de características morfológicas ou de marcadores moleculares. No caso do feijoeiro, essa estratégia foi aplicada em algumas oportunidades, sem resultados muito animadores (Machado, 1999; Duarte, 1998). Como a heterose depende da divergência genética nos locos que controlam o caráter, assume-se que genitores mais divergentes, seja em caracteres morfológicos ou em padrões de amplificação de marcadores moleculares, tenham maior possibilidade de apresentar heterose para o caráter desejado, o que nem sempre se verifica.

Outra alternativa que permite estimar o valor genético de cruzamentos, ainda que não realizados, é utilizar as Equações de Modelos Mistos (EMM), propostas por Henderson, em 1950 (Mrode, 1996), em que as predições dos efeitos aleatórios são denominadas *Best Linear Unbiased Prediction*, ou apenas BLUP. Em tal procedimento, estimativas de parentesco entre os genitores podem ser consideradas na análise, sejam essas calculadas com base na genealogia ou, na sua ausência, por meio de marcadores moleculares (Bernardo, 1994).

É preciso estar ciente de que cruzamentos entre indivíduos aparentados geram uma descendência endogâmica. Sabe-se que a endogamia, quando presente, influencia as estimativas de componentes de média e variância (Boer e Hoeschele, 1993). No caso de dialelos parciais envolvendo linhagens de grupos heteróticos a descendência é não endogâmica. Já no caso de dialelos completos envolvendo genitores aparentados, esse fato se verifica.

O presente trabalho tem por objetivos apresentar e comparar estratégias para a predição de cruzamentos não avaliados, na cultura do feijoeiro, que fazem uso de diferentes propriedades de estimação:

(i) o modelo aditivo que, para a predição de cruzamentos não avaliados faz uso apenas dos efeitos das CGC dos genitores, obtidos a partir dos cruzamentos avaliados;

(ii) o modelo de distância, que visa uma aproximação para a CEC do cruzamento pela introdução da distância genética entre os genitores como covariável;

(iii) os modelos de regressão fatorial, que fazem uso de covariáveis, como o modelo de distância, só que, em vez de distância genética para a construção dessas, utilizam os escores dos genitores obtidos a partir de uma análise de componentes principais dos padrões moleculares, tomando-se cada banda polimórfica como uma variável;

(iv) aplicação das EMM em duas circunstâncias: a primeira delas, sem a inclusão da endogamia dos cruzamentos como covariável, admitindo que sua magnitude não traz prejuízo à predição e a Segunda, a inclusão da endogamia de cada cruzamento como covariável numa tentativa de minimizar seus efeitos sobre as estimativas dos componentes do modelo.



## 2 REFERENCIAL TEÓRICO

### 2.1 A análise dialélica no melhoramento de plantas

No melhoramento vegetal, a hibridação tem por objetivo combinar, em um mesmo indivíduo, alelos que conferem fenótipos desejáveis e que se encontram em indivíduos distintos (Ramalho, Santos e Zimmermann 1993).

Em geral, os melhoristas de qualquer espécie dispõem de grande quantidade de germoplasma potencialmente utilizável em programas de melhoramento. Como consequência, seja do ponto de vista prático ou do econômico, a realização e a condução de todas as hibridações desejadas são inviáveis. O desenvolvimento de metodologias que permitam escolher com maior segurança as combinações que resultarão em populações mais promissoras para a seleção de novas cultivares é, então, de grande relevância.

Entre as metodologias mais consagradas da literatura para essa finalidade estão os cruzamentos dialélicos que, em seu sentido mais amplo, são definidos como o conjunto de todas as combinações duas a duas possíveis envolvendo  $n$  genitores, sendo:  $n(n-1)/2$  combinações híbridas,  $n(n-1)/2$  combinações híbridas recíprocas e os  $n$  genitores (Ramalho, Santos e Zimmermann, 1993; Cruz e Regazzi, 1998).

Existem vários modelos biométricos para a análise dialélica. A escolha de um deles dependerá dos objetivos do estudo. Entre os modelos mais difundidos estão os de Jinks e Hayman (1953), adequado para estudos de herança, o de Griffing (1956), utilizado em estudos de capacidade de combinação e o de Gardner e Eberhart (1966), usado em estudos de heterose.

O modelo proposto por Jinks e Hayman (1953) baseia-se na suficiência de um modelo aditivo-dominante no controle genético do caráter. Ele permite

estimar os componentes de variância genética e parâmetros genéticos como herdabilidades no sentido amplo e restrito, grau médio de dominância, dentre outros.

Santos, Vencovsky e Ramalho (1985) estudaram o controle genético da produção de grãos do feijoeiro e seus componentes primários pelo método proposto por Jinks e Hayman (1953). Foram utilizados sete genitores e as respectivas gerações  $F_2$  dos seus intercruzamentos. Os autores concluíram que a ação gênica aditiva foi predominante em relação à de dominância na expressão de todos os caracteres. Com base nos resultados, sugerem que o comportamento médio dos genitores e das populações segregantes seria um bom indicador do potencial de cada uma.

Os modelos de Griffing (1956) estimam a CGC e a CEC dos genitores. Os conceitos de CGC e CEC foram preliminarmente definidos por Sprague e Tatum (1942): a primeira é o comportamento médio de uma linhagem, cultivar ou material genético em um conjunto de cruzamentos; a CEC é o desvio observado de uma dada combinação híbrida em relação ao que seria esperado somente pelas CGC dos genitores.

Griffing (1956) propôs quatro modelos que variam conforme a utilização das diferentes partições da tabela dialélica: híbridos, cruzamentos recíprocos e genitores. No modelo I, a tabela dialélica é utilizada na íntegra, incluindo, portanto, genitores, híbridos e os cruzamentos recíprocos. No modelo II, os cruzamentos recíprocos são excluídos do quadro de análise, enquanto que, no modelo III, os genitores é que são excluídos. Por fim, no modelo IV, são excluídos os genitores e os cruzamentos recíprocos. A utilização do modelo I de Griffing (1956) justifica-se quando o objetivo da pesquisa é verificar a presença de efeito materno.

Os modelos propostos por Griffing (1956) têm sido amplamente usados em estudos de melhoramento com a cultura do feijoeiro. Machado (1999),

utilizando o modelo IV na análise dialélica, envolvendo 12 cultivares de feijão, verificou que as diferenças na produtividade de grãos das gerações  $F_2$  foram explicadas pelas CEC e CGC. Da mesma forma, Abreu (1997), avaliando a capacidade de combinação de cinco cultivares do grupo Mesoamérica, também pelo modelo IV, verificou que as fontes de variação CGC e CEC apresentaram magnitudes semelhantes de quadrados médios, indicando, segundo a autora, que elas contribuem igualmente para explicar a variação constatada no dialelo. Ressalte-se, entretanto, que Abreu (1997) utilizou em seus estudos a geração  $F_4$  dos cruzamentos, enquanto Machado (1999) utilizou gerações  $F_2$ .

Vencovsky e Barriga (1992) comentam que os tipos de capacidade combinatória diferenciam-se pela natureza da ação gênica envolvida. A CGC é influenciada pela ação aditiva dos genes, além de efeitos de dominância. Os efeitos constituintes da CEC, por sua vez, são devidos, principalmente, a efeitos de dominância.

A modelagem proposta por Gardner e Eberhart (1966) permite também o uso da tabela dialélica completa ou da meia tabela (híbridos e genitores). O modelo faz a decomposição da heterose em três níveis: heterose média atribuída a todos os cruzamentos; as heteroses inerentes a cada um dos genitores e a heterose específica que é intrínseca a cada um dos cruzamentos. Diferentemente dos modelos de Griffing (1956), este pressupõe que os genitores sejam de uma população em equilíbrio de Hardy-Weinberg.

Existem alguns trabalhos disponíveis na literatura, comparando as diferentes modelagens citadas. Singh e Singh (1984) estudaram as inter-relações existentes entre os modelos I e IV de Griffing (1956), o modelo de Morley-Jones, o modelo de Walters e Morton e o modelo de Gardner e Eberhart (1966). Para tanto, foi utilizado um dialelo 8x8 de *Vigna radiata*, em que se analisou a variável peso de 100 sementes. De acordo com os autores, todos os modelos mostraram-se inter-relacionados. Contudo, o modelo de Gardner e Eberhart

(1966), por decompor a heterose em diferentes componentes, forneceu informações mais detalhadas sobre a capacidade de combinação.

Cruz e Vencovsky (1989) também compararam, teoricamente e via simulação, os modelos II e IV de Griffing (1956) e o de Gardner e Eberhart (1966), nos quais foram considerados dialelos intervarietais, supondo genitores em equilíbrio de Hardy-Weinberg. Verificaram que tais modelos são equivalentes quanto às informações genéticas fornecidas. Com relação à CEC de um genitor com ele mesmo no modelo II de Griffing (1956), concluíram ser um parâmetro fundamental para indicar a direção dos desvios de dominância. Concluíram, ainda, que a CGC continua sendo o melhor critério para a seleção de genitores.

Extensões dos modelos dialélicos para o caso multivariado também são relatadas na literatura. Ledo (2002) desenvolveu estimadores para a análise de variância multivariada do modelo de Gardner e Eberhart (1966). Segundo o autor, a análise de variância multivariada para cruzamentos dialélicos pode ser realizada para melhorar a heterose em várias características, simultaneamente.

Como já comentado, uma das ressalvas à utilização dos modelos dialélicos tradicionais é que o número de cruzamentos aumenta consideravelmente com o aumento do número de genitores. Alternativas para superar essa limitação podem ser citadas, como, por exemplo, os cruzamentos dialélicos parciais, em que dois grupos de genitores são inter cruzados. Entre os modelos disponíveis na literatura estão: o de Griffing (1956) adaptado, semelhante a um modelo fatorial, quando não são utilizados os genitores; o de Gardner e Eberhart (1966), adaptado a dialelos parciais, sugerido por Miranda Filho e Geraldi (1984); o proposto por Geraldi e Miranda Filho (1988) adaptado ao modelo de Griffing (1956), em que são utilizados os cruzamentos entre grupos e os dois grupos de genitores.

De acordo com Ramalho, Santos e Zimmermann (1993), um aspecto favorável dos dialelos parciais é que nem sempre o melhorista está interessado em obter todos os cruzamentos possíveis a partir de um conjunto de genitores, os quais podem, então, ser separados em dois grupos, tomando-se apenas os cruzamentos entre os grupos. Nesse sentido, Abreu (1997) estudou o potencial de cruzamentos inter-raciais em feijoeiro. Considerando-se um dialelo parcial envolvendo genitores da raça Mesoamérica com outros das raças Durango e Jalisco, apenas as fontes de variação CGC para ambos os grupos foram significativas; no dialelo envolvendo as raças Mesoamérica e Nova Granada, apenas a fonte de variação CGC do grupo Nova Granada foi significativa.

Uma outra alternativa que exige menos cruzamentos do que a meia tabela tradicional ou, sob outro enfoque, que permite avaliar mais genitores com o mesmo número de cruzamentos obtido em um dialelo completo, são os chamados dialelos circulantes, propostos inicialmente por Kempthorne e Curnow (1961). De acordo com esse procedimento, cada um dos  $n$  genitores envolvidos é cruzado com um número “ $s$ ”, tal que “ $s$ ” seja menor que  $n-1$ . Os modelos de análise desses dialelos são idênticos aos modelos II e IV de Griffing (1956), conforme apresentado por Cruz e Regazzi (1998). A adaptação desse procedimento para o caso interpopulacional foi proposta por Miranda Filho e Vencovsky (1999).

Alguns trabalhos na literatura validam a utilização dos dialelos circulantes em programas de melhoramento. Oliveira Júnior (1995) avaliou a concordância das análises de dialelos circulantes nas gerações F1 e F3. Considerando-se as características número de vagens por planta e peso de vagens por planta, foi detectada maior correlação entre os valores preditos dessas características na geração F1 (por meio das estimativas da CGC) e valores fenotípicos da geração F3 do que entre os valores fenotípicos das duas gerações.

Veiga (1998) avaliou a eficiência dos dialelos circulantes na cultura do milho, tanto via simulação como via dados experimentais e verificou coerência dos resultados nos dois tipos de análises. Ao simular dialelos circulantes a partir de dois dialelos completos com 18 e 28 genitores, este autor constatou, nas duas situações, correlações médias acima de 0,80 entre a CGC, ou CEC, estimadas nos dialelos completos e as respectivas estimativas nos dialelos circulantes quando o número de cruzamentos por genitor, “s”, foi, aproximadamente, 50% dos genitores. Deve ser ressaltado que as herdabilidades dos conjuntos de cruzamentos originais foram, respectivamente, 28% e 63% para os dialelos com 28 e 18 genitores. Resultados similares foram obtidos por Ferreira (2003), também com a cultura do milho. Simulando dialelos circulantes a partir de um dialelo completo de 28 genitores, este autor verificou que, com  $s=19$ , não houve prejuízos no que diz respeito à escolha dos melhores genitores e das melhores combinações híbridas.

Aplicando o método de Miranda Filho e Vencovsky (1999) a um dialelo interpopulacional de linhagens S2 de milho, Araújo (2000) verificou que esse método foi mais eficiente na seleção do que o uso de “top crosses”, devido a uma melhor precisão das estimativas de CGC e as informações da CEC, essa última não disponível quando se realiza “top crosses”.

Mendonça (2001) avaliou um dialelo circulante com 12 genitores, sendo o número de cruzamentos por genitor igual a cinco. Este autor concluiu que, no feijoeiro, existe efeito de dominância no sentido de aumentar a produção de grãos, sendo as estimativas de CEC e  $d$  (componente genético de média que representa o desvio do heterozigoto em relação à média dos genitores) úteis para identificar as populações segregantes mais promissoras.

Considerando-se o método tradicional de estimação de parâmetros, quadrados mínimos, a estimação do desempenho dos cruzamentos não avaliados

é, então, feita com base nas CGC dos genitores. Quando ocorre predominância de efeitos aditivos no controle do caráter, essa pode ser uma boa aproximação.

Baseado na premissa da estimação de um cruzamento não avaliado a partir das capacidades combinatórias dos genitores, Ferreira (2003), na cultura do milho, obteve correlação de 0,69 entre os valores preditos dos cruzamentos não avaliados e os valores fenotípicos observados. Esses resultados foram obtidos a partir de um dialelo circulante de  $s=19$  de um total de 28 genitores.

## 2.2 Divergência genética e capacidade combinatória dos genitores

No item anterior, comentou-se que a predição dos híbridos não avaliados com base nas CGC dos genitores fornece boa aproximação quando, no controle do caráter, atuam genes de ação predominantemente aditiva. De fato, na ausência de dominância, a média da descendência corresponde à média dos genitores. Mesmo em tal circunstância há um outro contratempo: a impossibilidade de se predizer a variabilidade genética dos cruzamentos.

Na realidade, como apresentado em Falconer (1987), a heterose para uma característica quantitativa, pode ser expressa por:  $h = \Sigma dy^2$ , em que  $y$  é a diferença nas frequências alélicas entre os dois genitores e  $d$  é o desvio do heterozigoto em relação à média do homozigoto nos locos que controlam tal característica. Dessa forma, caso os genitores não difiram em frequências alélicas, ou não houver efeito de dominância, não haverá heterose.

Não obstante, a heterose está relacionada com a variabilidade genética liberada em um cruzamento. Do cruzamento entre dois genitores puros e contrastantes para um dado loco, as variâncias genéticas aditiva e de dominância presentes na geração  $F_2$  são:  $\sigma_a^2 = \frac{1}{2}a^2$  e  $\sigma_d^2 = \frac{1}{4}d^2$  (Ramalho, Santos e Zimmermann, 1993). O componente “ $a$ ” indica o desvio do homozigoto em

relação à média do loco e “ $d$ ” o desvio do heterozigoto. Para vários locos tem-se que  $\sigma_a^2 = \frac{1}{2} \Sigma a^2$  e  $\sigma_d^2 = \frac{1}{4} \Sigma d^2$ . Dessa forma, pode-se admitir que, quanto maior a heterose, maior será a quantidade de locos segregantes e, conseqüentemente, maior a variabilidade genética, tomando-se como verdadeiro o modelo aditivo-dominante. Deve ser ressaltado que ausência de heterose não implica em ausência de variabilidade, devido à possibilidade de ocorrência de direções opostas dos desvios de dominância ao longo dos locos que controlam o caráter.

Muito esforço em pesquisa tem sido demandado na busca de procedimentos que possibilitem identificar, a priori, os melhores cruzamentos a serem realizados, como por exemplo, o uso de distâncias genéticas. Inferências a respeito dessa grandeza podem ser feitas pela análise de genealogia dos genitores potenciais ou, ainda, por meio de técnicas multivariadas que caracterizam a divergência genética pelas diferenças morfológicas e, mais recentemente, por meio de técnicas moleculares.

Os métodos que utilizam a genealogia pressupõem que genitores com menor grau de parentesco têm maiores chances de apresentarem locos complementares, diferindo, portanto, em frequências alélicas. A maior limitação desse método é que, em muitas ocasiões, não há disponibilidade de genealogia. Abreu (1997) verificou que cruzamentos entre cultivares de feijão de diferentes raças, portanto não aparentados, foram promissores para a seleção visando ao aumento da produtividade de grãos. Contudo, devido a fenótipos indesejáveis para outros caracteres, tais como cor, tipo de grão e hábito de crescimento, a chance de obtenção de linhagens que associem produtividade e aceitação por agricultores e consumidores é reduzida.

A aplicação de técnicas multivariadas para a escolha de genitores também é relatada na literatura. Abreu (1997) utilizou 14 características para



estimar a distância genética de Mahalanobis entre cultivares de diferentes raças de feijão. Esta autora verificou que as correlações entre a distância genética dos genitores e a CEC e entre a primeira e a heterose dos cruzamentos na geração  $F_4$ , foram  $-0,02$  e  $-0,35$ , respectivamente. Contudo, ela atribuiu essas baixas correlações à baixa variabilidade observada na CEC das gerações  $F_4$ , concluindo, ainda, que a divergência não foi eficiente na escolha de genitores, pois, entre os escolhidos, estavam cultivares não adaptadas.

Duarte (1998), utilizando os cruzamentos realizados por Abreu (1997), verificou que a correlação obtida entre a distância genética dos genitores, via marcadores RAPD e a CEC dos cruzamentos foi de  $-0,02$  e da primeira com a heterose de  $-0,39$ . Tais resultados foram similares àqueles obtidos quando a distância genética foi calculada por características morfológicas. A correlação entre as distâncias genéticas calculadas via marcadores moleculares e características morfológicas foi de  $0,89$ .

Resultados semelhantes foram obtidos por Machado (1999). A autora constatou que as distâncias genéticas obtidas por marcadores morfológicos e por marcadores RAPD foram altamente correlacionadas. No entanto, não foram observadas correlações entre a distância de Mahalanobis e as estimativas dos parâmetros do dialelo relativo à produção de grãos e, em particular, à CEC.

Mendonça (2001), trabalhando com um dialelo circulante em feijão, verificou que as menores distâncias obtidas via marcadores moleculares RAPD estavam associadas a genitores geneticamente mais semelhantes. As correlações obtidas entre as distâncias moleculares e as CEC, nas gerações  $F_3$  e  $F_4$ , foram  $0,006$  e  $-0,04$ , respectivamente.

Nas análises de divergência genética por meio de marcadores moleculares, linhagens são caracterizadas por um número muito grande de padrões de amplificação ou bandas. Alguns trabalhos na literatura evidenciam que o uso de alguns componentes principais desses padrões de amplificação

constitui-se em uma relevante sumariação da caracterização molecular de genótipos (Dubreuil et al., 1996; Lee et al., 1994; Burstin et al., 1989).

### **2.3 Covariância genética entre parentes e predição de cruzamentos**

Existem outras alternativas que, em vez da divergência genética, fazem uso da similaridade como parte da estratégia da predição do desempenho de cruzamentos não avaliados. Como exemplo, citam-se as EMM, desenvolvida por Henderson em 1949 (Mrode, 1996). Como soluções, são obtidos os BLUP dos efeitos aleatórios e o *Best Linear Unbiased Estimation*, ou BLUE, dos efeitos fixos.

Nesse procedimento pode-se utilizar o parentesco existente entre os indivíduos com o objetivo de otimizar a análise, considerando, na predição do desempenho de um dado indivíduo, as observações fenotípicas dos seus parentes. A predição dos indivíduos não avaliados é feita com base no desempenho dos indivíduos avaliados e na similaridade genética entre os primeiros e os últimos.

A matriz de similaridade genética entre indivíduos ou populações pode ser construída com base no coeficiente de coancestria de Malecot, desde que informações de genealogia estejam disponíveis. Essa era uma das limitações para sua utilização, visto que nem sempre informações de genealogia estão disponíveis. O advento dos marcadores moleculares supriu parte dessa limitação e diversas estratégias têm sido sugeridas para se estimar parentesco a partir de marcas moleculares, inclusive para análise dialélica.

A análise dialélica sob enfoque de EMM e cálculo de parentesco via marcadores moleculares, conforme apresentado em Bernardo (1994), é um exemplo do seu uso potencial no melhoramento de plantas.

A abordagem genética-estatística tratada pelo autor se aplica perfeitamente a dialelos parciais entre grupos heteróticos. Bernardo (1994) considera o cruzamento entre indivíduos de dois grupos, em que os indivíduos dentro de cada grupo são aparentados, mas não há parentesco entre indivíduos de grupos diferentes. Portanto, nessas circunstâncias, a descendência é não endogâmica. Assim, tomando-se dois genitores do grupo X, x e x', e dois indivíduos do grupo Y, y e y', a covariância entre os híbridos (x,y) e (x',y') pode ser expressa por:

$$Cov[(x, y), (x', y')] = f_{xx'}V_{(A/X)} + f_{yy'}V_{(A/Y)} + f_{xx'}f_{yy'}V_{(D/XY)},$$

em que:

$f_{xx'}$  é o coeficiente de coancestria entre as linhagens do grupo X;  $f_{yy'}$  é o coeficiente de coancestria entre as linhagens do grupo Y;  $V_{(A/X)}$  é a variância aditiva entre as linhagens do grupo X;  $V_{(A/Y)}$  é a variância aditiva entre as linhagens do grupo Y;  $V_{(D/XY)}$  é a variância de dominância.

O coeficiente  $f_{xx'}$  (ou  $f_{yy'}$ ) pode ser calculado por:

$$f_{xx'} = \frac{\left[ S_{xx'} - \frac{1}{2}(S_{xy} + S_{x'y}) \right]}{\left[ 1 - \frac{1}{2}(S_{xy} + S_{x'y}) \right]},$$

$f_{xx'}$  é o coeficiente de coancestria estimado por marcadores moleculares entre as linhagens x e x';  $S_{xx'}$  a proporção de locos marcadores com variantes comuns entre x e x';  $S_{xy}$  (ou  $S_{x'y}$ ) é a proporção média de variantes comuns para x (ou para x') em uma série de linhagens do grupo heterótico Y.

Contudo, essa abordagem não contempla a análise de dialelos completos, pois o cruzamento entre indivíduos do mesmo grupo gera endogamia, uma vez que esses são aparentados. É sabido que a endogamia altera os componentes genéticos de variância e tal efeito não está considerado no referido modelo. Boer e Hoeschele (1993) mostram os efeitos da endogamia sobre a média e a variância genética para caracteres quantitativos em uma população com efeitos aditivos e de dominância, sendo que a complexidade do modelo cresce sobremaneira.

Além disso, tomando-se um dialelo completo, não faz sentido considerar dois grupos de indivíduos, mas um único grupo. Assim, a covariância entre os indivíduos  $(x,y)$  e  $(x',y')$  teria que contemplar também as similaridades entre  $x$  e  $y'$  e entre  $x'$  e  $y$ , e não apenas entre  $x$  e  $x'$  e  $y$  e  $y'$ .

Uma alternativa à proposta de Bernardo (1994) é a de Lynch e Ritland (1999). Estes autores apresentam um procedimento de estimação conjunto dos coeficientes de parentesco e de dominância via marcadores moleculares co-dominantes, em uma população de acasalamentos ao acaso. Tal método foi denominado de estimadores de regressão e seus passos principais estão descritos no item 3.3.

## **2.4 Utilização das EMM no melhoramento genético de plantas**

Como salientado anteriormente, o diferencial desse modelo é a possibilidade de ele considerar em sua estrutura de análise a matriz de variâncias e covariâncias entre os níveis dos fatores aleatórios, considerando na predição do desempenho de um indivíduo, o desempenho de seus parentes. Além disso, permite a predição de indivíduos não avaliados.

Embora as EMM tenham sido originalmente propostas para programas de melhoramento genético animal, seu potencial de utilização no melhoramento

genético de plantas tem sido verificado em várias culturas, como eucalipto (Bueno Filho, 1997), soja (Panter e Allen, 1995), pêssago (Souza, Byrne e Taylor, 1998a, 1998b, 2000), dendê (Soh, 1994) e milho (Bernardo, 1994, 1995, 1996a, 1996b, 1996c, 1996d). Estudos via simulação também são encontrados na literatura (André, 1999).

Bernardo (1994) comparou, em milho, a predição de cruzamentos usando informações de parentesco calculadas via marcadores moleculares ou de genealogia. De um total de 54 cruzamentos foram usados como preditores tamanhos de 10, 15, 20, 25 e 30 cruzamentos. As correlações entre os valores preditos e observados para produção variaram de 0,654 a 0,800. Maiores correlações foram obtidas quando as estimativas de parentesco foram determinadas via RFLP, em vez de dados de genealogia. À medida em que aumentou o número de híbridos preditores, maiores correlações foram obtidas.

Bernardo (1995) avaliou 67 cruzamentos entre dois grupos de linhagens, sem que houvesse cruzamentos dentro de cada grupo. O estudo tinha por objetivo avaliar diferentes modelos genéticos utilizando-se as EMM na predição de híbridos simples de milho. Conjuntos de dados contendo 10, 15, 20, 25 e 30 híbridos foram escolhidos aleatoriamente para predição dos demais híbridos com perda simulada. Os modelos com inclusão de componentes epistáticos aditivo x aditivo não melhoraram as predições quando comparados aos modelos genéticos baseados apenas em efeitos aditivos e de dominância apenas. As correlações variaram de 0,583 a 0,749. Em todos os modelos, maiores correlações foram obtidas quando se aumentou o número de cruzamentos preditores. Considerando-se as situações em que os híbridos preditos (aqueles com perda simulada) e os preditores não tinham linhagem parental em comum as correlações variaram de 0,388 a 0,493.

Em soja, Panter e Allen (1995) compararam as predições BLUP e Quadrados Mínimos Ordinários, ou OLS, na predição de cruzamentos.

Utilizando dados de séries históricas de experimentos avaliados em vários anos e locais, estes autores assumiram como predição BLUP dos cruzamentos a média dos BLUP dos genitores obtida ao longo dos ensaios e, como predição OLS, a média de quadrados mínimos. As performances dos bulks  $F_4$  e  $F_6$ , além de famílias  $F_{5,6}$  de cada um dos cruzamentos, foram utilizadas como estimativas do desempenho de cada um dos 24 cruzamentos. Para a produtividade de grãos, as predições BLUP, tanto em situações de dados balanceados como de dados desbalanceados, apresentaram maiores correlações e menores erros-padrão.

Bernardo (1996a) avaliou o potencial das EMM na predição de cruzamentos não avaliados em milho, para diversas características, envolvendo linhagens de nove grupos heteróticos. Foram utilizadas 16 combinações de cruzamentos entre os nove grupos. As matrizes de parentesco foram calculadas a partir de informações de pedigree. Para cada intercruzamento entre dois grupos, um híbrido por vez era retirado do quadro de análise e, então, sua performance era predita pelos demais. As correlações entre os valores preditos e observados para produção de grãos, nas 16 diferentes combinações de cruzamentos, variaram de 0,136 a 0,762. As maiores estimativas de correlações foram obtidas nos intercruzamentos com maior número de híbridos preditores.

A vantagem do uso de modelos com presença de efeitos aditivos e dominância em relação àqueles que utilizam apenas efeitos aditivos tem sido o objetivo de alguns trabalhos. Bernardo (1996d) verificou que a análise dialélica sob enfoque de modelos mistos foi mais eficiente na predição de cruzamentos não avaliados quando foi utilizado um modelo incluindo efeitos aditivo e de dominância. As correlações entre os valores preditos e observados para a produção de grãos com esse modelo variaram de 0,50 a 0,66; para o modelo somente aditivo, foram um pouco menores, com redução variando de 0,02 a 0,06. Segundo o autor, o modelo aditivo pode ser usado se informações a priori

indicarem que a variância de dominância para a característica em questão tiver pouca contribuição para a variância genética total.

Em muitas ocasiões, as informações de parentesco disponíveis não são fidedignas. Bernardo (1996b) investigou a robustez das predições BLUP para estimativas viesadas de parentesco, com 10% de estimativas desconhecidas ou 25% das estimativas de parentesco entre as linhagens desconhecidas. As estimativas de parentesco foram calculadas a partir de dados de pedigree. A metodologia mostrou-se robusta nessas situações uma vez que as correlações das predições foram pouco influenciadas pelas perdas das informações de parentesco.

Bernardo (1996c) estudou o uso das EMM na predição de cruzamentos em que um ou ambos os genitores não participavam dos cruzamentos presentes na análise. Foram utilizadas 16 combinações de cruzamentos entre linhagens pertencentes a nove grupos heteróticos. Maiores correlações foram obtidas quando ambos os genitores participavam dos cruzamentos utilizados como preditores. A predição de cruzamentos em que nenhum dos genitores participava de cruzamentos do quadro de análise não apresentou resultados satisfatórios. Nas situações em que apenas um dos genitores participava de cruzamentos do quadro de análise, as correlações foram satisfatórias quando o número de cruzamentos envolvidos era grande.

Charcosset et al. (1998) compararam diferentes modelos de predição de cruzamentos não realizados, na cultura do milho, em duas situações: (i) todos os cruzamentos disponíveis, independentemente da genealogia das linhagens, eram utilizados; e (ii) considerando-se apenas os cruzamentos entre linhagens não aparentadas. Na primeira situação, o modelo de distância (formado pelos efeitos da CGC e uma covariável constituída pela distância genética dos genitores), estimado por OLS, foi superior ao modelo aditivo (formado exclusivamente pelos efeitos da CGC), às predições BLUP e ao modelo de regressão fatorial

(formado pelos efeitos aditivos e covariáveis obtidas a partir dos escores dos componentes principais dos genitores, tendo como variáveis bandas polimórficas), estimado via OLS. Na segunda situação, as predições BLUP foram mais eficientes. Para todos os modelos, nas duas situações, maiores estimativas de  $r^2$  entre os valores preditos e observados foram obtidas quando um maior número de híbridos preditores foram utilizados.

André (1999) comparou, via simulação, predições BLUP e OLS de efeitos de cruzamentos e de capacidade de combinação. Verificou que a metodologia BLUP foi superior em diferentes condições de herdabilidades e graus médios de dominância, independentemente de se utilizar os métodos II ou IV de Griffing (1956).

## **2.5 Procedimentos de estimação de parâmetros genéticos**

Nos modelos biométricos, um dos interesses dos melhoristas é explicar quanto do valor fenotípico de um indivíduo é devido ao seu potencial genético. A correta estimação dos parâmetros que compõem o valor genético de um indivíduo é condição básica para o sucesso da seleção.

Searle (1971) apresenta três procedimentos de estimação, considerando-se um modelo linear  $Y = X\beta + \varepsilon$  que, sob determinadas pressuposições, geram os mesmos estimadores. Esses procedimentos são: quadrados mínimos ordinários (**OLS**), quadrados mínimos generalizados (**GLS**) e máxima verossimilhança (**ML**). Contudo, nenhuma deles considera em seus pressupostos as matrizes correlações genéticas entre os indivíduos avaliados e, portanto, são mais apropriados para se estimar os efeitos do modelo considerados como fixos.

Admitir que os níveis de um dado fator são aleatórios é assumir que os mesmos advêm de uma mesma população de referência e, portanto, inter-relações entre eles existem. A modelagem de EMM, que trata os efeitos



aleatórios de forma mais conveniente, uma vez que considera em seu procedimento de análise a estrutura de covariância entre os níveis de um dado fator, é também apresentada.

### 2.5.1 Quadrados mínimos ordinários (OLS)

Seja um modelo linear que possa ser expresso matricialmente por  $Y = X\beta + \varepsilon$ , em que  $Y$  é o vetor de dados observados,  $X$  a matriz de incidência dos efeitos do modelo (tomados como fixos),  $\beta$  o vetor dos parâmetros a ser predito e  $\varepsilon$  o vetor de erros aleatórios. O princípio do método é encontrar soluções para o vetor  $\beta$  tal que a soma de quadrados residual seja mínima.

Desde que  $\varepsilon = Y - X\beta$ , então  $\varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$ . Tomando-se a derivada dessa função e igualando-a a zero, tem-se que o estimador de  $\beta$  calculado pelo método **OLS** pode ser expresso por como solução de:

$$(X'X)\hat{\beta} = X'Y$$

Na metodologia **OLS**, nenhuma pressuposição a respeito da distribuição de probabilidade dos erros aleatórios do modelo, representados por  $\varepsilon$ , é assumida. Contudo, admite-se que esses não são correlacionados.

Em adição, o posto coluna da matriz  $X$ , em muitos modelos biométricos, pode não ser completo e assim,  $X'X$  não teria inversa única. Todavia, essas situações podem ser contornadas utilizando-se restrições no conjunto solução ou restrições paramétricas diretamente feitas na matriz  $X$  ou, ainda, utilização de uma inversa generalizada.

### 2.5.2 Quadrados mínimos generalizados (GLS)

Assume-se que existem correlações entre os erros (matriz  $R$ ), mas nenhuma informação a respeito da distribuição de probabilidade deles é assumida. O estimador de  $\beta$  pode ser expresso como solução de:

$$(X' R^{-1} X) \hat{\beta} = X' R^{-1} Y$$

Dessa forma, quando  $R=I\sigma^2$ , ou seja, quando os erros são não correlacionados, o estimador **GLS** é idêntico ao **OLS**. As ressalvas feitas no item anterior sobre o posto coluna de  $X$  também se aplicam aqui.

### 2.5.3 Máxima verossimilhança (ML)

Diferentemente dos procedimentos OLS e GLS, o estimador de ML de  $\beta$  pressupõe que os erros têm uma distribuição de probabilidade. Por esse método o melhor estimador para  $\beta$  é aquele que maximiza a função de verossimilhança para esse parâmetro. Assumindo-se que os erros sejam normalmente distribuídos, com média zero e matriz de variância e covariância  $R$ , a função de verossimilhança é dada por:

$$L = (2\pi)^{(-1/2)N} |R|^{-1/2} \exp\left\{-1/2(Y - X\beta)' R^{-1}(Y - X\beta)\right\}$$

e maximizada em relação a  $\beta$ , tem-se:

$$(X' R^{-1} X) \hat{\beta} = X' R^{-1} Y$$

Assim, os procedimentos e GLS e OLS são casos particulares do método da ML, quando determinadas pressuposições são atendidas.

Considerando-se modelos em que dois ou mais fatores estão incluídos, as equações normais são idênticas. Supondo um delineamento em blocos ao acaso, expresso pelo modelo  $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ , pode-se imaginar que a matriz  $X$ , considerada anteriormente, é composta por duas submatrizes  $X_1$  e  $X_2$ , enquanto o vetor  $\beta$  pode ser particionado em  $\beta_1$  e  $\beta_2$ . Sendo,  $X_1$  e  $\beta_1$  a matriz de incidência e o vetor dos parâmetros dos efeitos associados ao intercepto e aos blocos, enquanto  $X_2$  e  $\beta_2$  a matriz de incidência e o vetor dos parâmetros dos tratamentos, tem-se o seguinte sistema de equações considerando-se normalidade e ausência de correlação entre os erros:

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}$$

#### 2.5.4 Equações de modelos mistos (EMM)

No item 2.1 foi comentado que a abordagem tradicional OLS para a estimação de parâmetros genéticos nos modelos dialélicos não permite a estimação da CEC dos cruzamentos ausentes. O potencial genético de tais cruzamentos é então avaliado com base na CGC dos genitores.

Como abordado anteriormente, ao se assumir que os níveis de um dado fator são aleatórios, a estrutura de covariância existente pode ser utilizada para enriquecer a informação sobre a população. Assim, cada nível desse fator é, na realidade, uma realização da variável aleatória em questão.

Seja o modelo linear misto expresso de acordo com Mrode (1996):

$$y = Xb + Za + e, \text{ em que:}$$

$y$ : vetor  $n \times 1$  de observações;

$b$ : vetor  $p \times 1$  de efeitos fixos,  $p$  sendo o número de níveis;

$a$ : vetor  $q \times 1$  de efeitos aleatórios,  $q$  sendo o número de níveis;

$e$ : vetor  $n \times 1$  de efeitos aleatórios residuais;

$X$ : matriz de incidência de ordem  $n \times p$ , associada aos efeitos fixos;

$Z$ : matriz de incidência de ordem  $n \times q$ , associada aos efeitos aleatórios.

Os BLUP de  $a$ , solução dos efeitos aleatórios, podem ser expressos por  $\hat{a} = BLUP(a) = GZV^{-1}(y - X\hat{b})$ , enquanto que a solução dos efeitos fixos  $b$  é o estimador de quadrados mínimos generalizados, ou seja,  $b = (X'V^{-1}X)^{-1}X'V^{-1}y$ . Nesse caso,  $G$  é uma matriz de variâncias e covariâncias genéticas entre os níveis dos efeitos aleatórios, enquanto  $V$  é uma matriz de variâncias e covariâncias fenotípicas entre as observações.

Estimar os efeitos de  $a$  e  $b$  por essas expressões requer a inversão de  $V$ , que é de difícil inversão devido à sua ordem. Henderson, em 1950, apresentou as equações de modelos mistos para estimar  $a$  e  $b$ , sem a necessidade da inversão de  $V$  (Mrode, 1996). Matricialmente, pode-se apresentar as equações de modelos mistos da seguinte forma:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{bmatrix}$$

Assumindo que os erros sejam não correlacionados ( $R=I\sigma^2$ ), as equações de modelos mistos dependem apenas da inversão da matriz de variâncias e covariâncias genéticas  $G$ . Sendo  $G=A\sigma_a^2$ , em que  $A$  é a matriz de similaridade genética aditiva entre os níveis dos efeitos aleatórios e  $\sigma_a^2$  é a variância genética

aditiva, pode-se reescrever as equações de modelos mistos da seguinte forma (Mrode, 1996):

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1} \frac{\sigma^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \text{ ou}$$

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1} \frac{(1-h^2)}{h^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

### 3 MATERIAL E MÉTODOS

#### 3.1 Material genético e delineamento experimental

No presente trabalho, utilizou-se um conjunto de dados obtidos por Abreu (1997), envolvendo 22 cultivares de feijoeiro cruzados em esquema dialélico. A caracterização molecular dessas linhagens foi obtida por Duarte (1998) via marcadores RAPD.

A disponibilidade de uma matriz binária, elementos 0 e 1, obtida dos padrões de amplificação dos marcadores moleculares de cada genitor, permitiu a construção de matrizes de parentesco e de coeficientes de dominância entre os genitores e ou cruzamentos, de acordo com Lynch e Ritland (1999), conforme descrito no item 3.3 e a aplicação dos modelos de Charcosset et al. (1998), mediante análise prévia de componentes principais.

Os 22 genitores utilizados eram proveninetes de quatro diferentes raças (Tabela 1). Foram obtidas 120 das 231 combinações de cruzamentos possíveis, avançadas pelo método massal até a geração  $F_4$ .

As 120 populações, juntamente com os 22 genitores e mais duas outras cultivares, Ouro e A-77, foram avaliados em delineamento látice 12x12 com duas repetições em três locais do estado de Minas Gerais (Lavras, Lambari e Patos de Minas), conduzidos na época da seca do ano de 1994. As observações referentes às cultivares Ouro e A-77 foram descartadas, uma vez que não participaram como genitores dos cruzamentos avaliados. Detalhes adicionais da condução dos experimentos são apresentados em Abreu (1997).

**TABELA 1** Cultivares de feijão utilizadas na obtenção dos cruzamentos dialélicos e suas respectivas raças preconizadas.

Raça	Cultivares
Mesoamérica	Sanilac, Brasil-2*, A-62*, Porrilo Sintético, Jamapa, Milionário, IPA-7419, A-175*, Rojo de Seda, A-140, Carioca*, Compuestro Negro Chimaltenango*, Puebla 152.
Durango	Great Northern*, Ojo de Liebre*, A-114*
Jalisco	Flor de Mayo*, Apetito Blanco*
Nova Granada	Mont Calm*, Jalo, Diacol Andino*, Perry Marrow*

\* Cultivares seguidas por asteriscos foram as utilizadas por Abreu (1997) em seu estudo de capacidade combinatória, envolvendo 3 dialelos: (i) dentro da raça Mesoamérica; (ii) entre as cultivares das raças Mesoamérica e Nova Granada; (iii) entre cultivares da raça Mesoamérica e as raças Durango e Jalisco.

Todos os cruzamentos obtidos foram considerados como um único dialelo, analisado sob os vários modelos descritos no item 3.5. Perdas de cruzamentos eram simuladas no quadro de dados original e a predição desses realizada por meio dos modelos adotados.

### 3.2 Caracterização molecular dos genitores

Considerando os genitores, 136 bandas polimórficas foram de 33 iniciadores RAPD (Duarte, 1998). Cada banda foi tratada como um caráter único, com dois alelos, sendo sua presença em um dado indivíduo designada por 1 (um) e sua ausência por 0 (zero), obtendo-se assim uma matriz binária. As possíveis combinações para dois genitores, para um dado loco, podem ser expressas conforme o seguinte diagrama:

		<i>Cultivar j</i>	
		1	0
<i>Cultivar i</i>	1	(1,1)	(1,0)
	0	(0,1)	(0,0)

O tratamento algébrico da matriz binária variou de acordo com o modelo, exceto para o modelo aditivo em que não foi utilizada nenhuma informação de similaridade genética.

### 3.3 Cálculo de parentesco e coeficiente de dominância, via marcadores moleculares, de acordo com Lynch e Ritland (1999)

Sejam dois indivíduos  $X$  e  $Y$ , em que  $X$ , oriundo do cruzamento das linhagens  $i$  e  $j$ , possui alelos  $a$  e  $b$ , e  $Y$ , oriundo do cruzamento entre as linhagens  $m$  e  $n$ , possui alelos  $c$  e  $d$ , em um loco  $k$  qualquer. Sejam também  $\hat{r}_{xy}$  e  $\hat{\Delta}_{xy}$  os coeficientes de parentesco e de dominância, respectivamente, entre os indivíduos  $X$  e  $Y$ . Tomando-se  $X$  como referência, então,  $\hat{r}_{xy}$  e  $\hat{\Delta}_{xy}$  são expressos por:

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

$$\hat{\Delta}_{xy(k)} = \frac{2p_a p_b - p_a(S_{bc} + S_{bd}) - p_b(S_{ac} + S_{ad}) + (S_{ac} S_{bd}) + (S_{ad} S_{bc})}{(1 + S_{ab})(1 - p_a - p_b) + 2p_a p_b},$$

em que:

$S_{ab}$ : igual a 1 (um) se  $a$  e  $b$  forem idênticos, e igual 0 (zero) caso contrário;

$S_{ac}$ : igual a 1 (um) se  $a$  e  $c$  forem idênticos, e igual 0 (zero) caso contrário;

$S_{ad}$ : igual a 1 (um) se  $a$  e  $d$  forem idênticos, e igual 0 (zero) caso contrário;

$S_{bc}$ : igual a 1 (um) se  $b$  e  $c$  forem idênticos, e igual 0 (zero) caso contrário;

$S_{bd}$ : igual a 1 (um) se  $b$  e  $d$  forem idênticos, e igual 0 (zero) caso contrário;

$p_a$  e  $p_b$ : são as freqüências dos alelos  $a$  e  $b$  ao longo das linhagens para um dado loco  $k$ .



Normalmente, as estimativas de parentesco são obtidas ao longo de muitos locos. Um peso dado à estimativa de parentesco, para um dado loco  $k$ , considerando-se o indivíduo  $X$  como referência, é apresentado pelos autores.

Sejam  $w_{r,x}$  e  $w_{\Delta,x}$  as ponderações para o  $k$ -ésimo loco das estimativas  $\hat{r}_{xy}$  e  $\hat{\Delta}_{xy}$  e  $W_{r,x}$  e  $W_{\Delta,x}$  os somatórios das ponderações atribuídas aos vários locos envolvidos. As estimativas de parentesco entre  $X$  e  $Y$  são, então, dadas por:

$$\hat{r}_{x,y} = \frac{1}{W_{r,x}} \sum_1^L w_{r,x(k)} \hat{r}_{x,y(k)}$$

$$\hat{\Delta}_{x,y} = \frac{1}{W_{\Delta,x}} \sum_1^L w_{\Delta,x(k)} \hat{\Delta}_{x,y(k)},$$

em que:

$$w_{r,x(k)} = \frac{1}{\text{Var}[\hat{r}_{x,y(k)}]} = \frac{(1 + S_{ab})(p_a + p_b) - 4p_a p_b}{2p_a p_b}$$

$$w_{\Delta,x(k)} = \frac{1}{\text{Var}[\hat{\Delta}_{x,y(k)}]} = \frac{(1 + S_{ab})(1 - p_a - p_b) + 2p_a p_b}{2p_a p_b}$$

Dado que a escolha dos indivíduos  $X$  ou  $Y$  é meramente arbitrária, Lynch e Ritland (1999) sugerem calcular, para cada loco, os parentescos e as ponderações, considerando, em um primeiro momento, o indivíduo  $X$  como referência e, em seguida, considerando o indivíduo  $Y$  como referência. Então, as expressões finais para o cálculo do parentesco aditivo e do parentesco de dominância são dadas por:

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

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

### 3.4 Análise de similaridade genética entre os genitores

A partir dos parentescos entre os genitores,  $\hat{r}_{xy}$ , calcula pelo procedimento apresentado no item anterior, foi obtida a matriz de dissimilaridade entre eles. A medida de dissimilaridade, ou de distância genética, entre dois genitores quaisquer foi assumida como  $1 - \hat{r}_{xy}$ .

A representação simplificada das distâncias genéticas foi feita por meio de um dendrograma, obtido pelo método hierárquico aglomerativo da média aritmética entre pares não ponderados (UPGMA) utilizando-se o programa NTSYS-PC 2.0 (Rohlf, 1992).

### 3.5 Modelos utilizados na predição de cruzamentos dialélicos

No total, foram utilizados oito modelos: (i) o modelo aditivo, sem nenhuma aproximação para a CEC e que não fez uso das informações de marcadores moleculares para predição dos cruzamentos faltantes; (ii) modelo de distância genética, que é similar ao modelo aditivo, sendo apenas adicionada uma covariável de distância genética calculada a partir da matriz de elementos 0 e 1; (iii) quatro modelos de regressão fatorial, em que as covariáveis são obtidas a partir de uma análise de componentes principais da matriz de elementos 0 e 1; (iv) análise sob enfoque de modelos mistos, em duas situações, na presença e na ausência da endogamia do cruzamento como covariável.

Uma análise preliminar, sem considerar simulação de perdas, foi realizada para os modelos aditivo, de distância e os de regressão fatorial. O

objetivo dessa análise foi verificar quanto da variação da CEC é explicada pela co-variável, ou co-variáveis, quando incorporadas a um modelo composto previamente apenas pelos efeitos da CGC (modelo aditivo).

### 3.5.1 Modelo aditivo

O uso desse modelo na predição de cruzamentos não inclui nenhuma aproximação para a CEC. Foi utilizado o modelo II de Griffing (1956), na forma reduzida, conforme apresentado em Cruz e Regazzi (1998), sendo a estimação dos parâmetros obtida pelo método dos quadrados mínimos. Como valor fenotípico dos cruzamentos, foi utilizada a média ajustada de cada um deles obtida pela análise intrablocos conjunta dos três locais.

A performance do cruzamento  $Y_{ij}$  entre as linhagens  $i$  e  $j$  é dada por:

$$Y_{ij} = \mu + g_i + g_j + e_{ij},$$

em que:

$Y_{ij}$ : é a performance média observada do cruzamento entre as linhagens  $i$  e  $j$ ;

$\mu$ : é a performance média dos híbridos avaliados;

$g_i$ : é o efeito da CGC da linhagem  $i$ ;

$g_j$ : é o efeito da CGC da linhagem  $j$ ;

$e_{ij}$ : é o desvio entre o valor médio observado e o valor predito pelo modelo.

A predição de um dado cruzamento entre as linhagens  $k$  e  $l$  com perda simulada pode ser feita pela seguinte expressão de predição:

$$\hat{Y}_{kl} = \hat{\mu} + \hat{g}_k + \hat{g}_l,$$

em que:

$\hat{Y}_{kl}$ : o valor predito do cruzamento entre as linhagens  $k$  e  $l$ , com perda simulada;

$\hat{\mu}$  : é a estimativa da performance média dos cruzamentos avaliados;

$\hat{g}_k$  : é a estimativa da CGC da linhagem  $k$ , considerando os cruzamentos avaliados de que ela participou;

$\hat{g}_l$  : é a estimativa da CGC da linhagem  $l$ , considerando os cruzamentos avaliados de que ela participou.

### 3.5.2 Modelo de distância

Da mesma forma como realizado para o **modelo aditivo**, no **modelo de distância** foi utilizada como valor fenotípico do cruzamento entre as linhagens  $i$  e  $j$ , a média ajustada obtida da análise intrablocos conjunta.

Charcosset et al. (1998) propõem, como uma primeira aproximação para o componente da CEC, assumir que o termo  $s_{ij}$  está relacionado com a distância genética  $D_{ij}$  entre as linhagens parentais do cruzamento.

De acordo com essa hipótese, a interação entre as linhagens  $i$  e  $j$  pode ser aproximada por uma função linear de  $D_{ij}$ . O modelo para os cruzamentos avaliados, com estimação de parâmetros pelo método dos quadrados mínimos, inclui o valor de  $D_{ij}$  como uma covariável e a estimação de um parâmetro de regressão, resultando no seguinte modelo:

$$Y_{ij} = \mu + g_i + g_j + \rho D_{ij} + e_{ij},$$

em que  $\mu$ ,  $g_i$ ,  $g_j$  e  $e_{ij}$  já foram definidos anteriormente e  $\rho$  é o coeficiente de regressão estimado atribuído a  $D_{ij}$  no modelo. Observa-se que esse modelo necessita apenas de um grau de liberdade a mais que o modelo aditivo anterior.

A predição de um cruzamento com perda simulada entre as linhagens  $k$  e  $l$  é feita, então, pela seguinte expressão de predição:

$$\hat{Y}_{kl} = \hat{\mu} + \hat{g}_k + \hat{g}_l + \hat{\rho}D_{kl},$$

em que:

$\hat{\mu}$  : é a estimativa da performance média dos cruzamentos avaliados;

$\hat{g}_k$  e  $\hat{g}_l$  são as estimativas das CGC das linhagens  $k$  e  $l$ , considerando os cruzamentos avaliados de que cada linhagem participou;

$\hat{\rho}$  é a estimativa do coeficiente de regressão estimado com base nos cruzamentos avaliados;

$D_{kl}$  a distância genética estimada entre as linhagens  $k$  e  $l$ .

De acordo com Charcosset et al. (1998), as estimativas de  $g_i$  e de  $g_j$  diferem daquelas que seriam obtidas pelo modelo aditivo desde que esses efeitos não sejam ortogonais para as distâncias entre os híbridos parentais.

No presente trabalho, a medida de distância utilizada como aproximação da CEC corresponde à seguinte expressão:  $1 - \hat{r}_{xy}$ , em que  $\hat{r}_{xy}$  é o coeficiente de parentesco entre as linhagens  $x$  e  $y$ , conforme apresentado em Lynch e Ritland (1999).

### 3.5.3 Modelos de regressão fatorial

Da mesma forma como comentado para o modelo aditivo e o modelo de distância, utilizou-se como observação fenotípica de um dado cruzamento ou genitor, a média ajustada de cada um deles da análise intrablocos conjunta.

Para os modelos apresentados nesse item, o tratamento algébrico das matrizes binárias foi diferente daquele empregado no **modelo distância**, embora, em ambos os casos, coeficientes de regressão tenham sido estimados pelo método OLS.

Inicialmente, procedeu-se a uma análise de componentes principais considerando cada uma das 136 bandas polimórficas entre os 22 genitores como uma variável. De acordo com Johnson e Wichern (1988) nenhuma pressuposição a respeito da variável em estudo é necessária para a utilização das chamadas componentes principais.

A análise consiste em transformar um conjunto original de variáveis, no presente caso 136 padrões polimórficos de RAPD, em um outro conjunto de dimensão equivalente, mas com propriedades distintas. Cada componente principal é uma combinação linear das variáveis originais, sendo ortogonais entre si (covariância zero) e estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos da variância total contida nos dados originais (Cruz e Regazzi, 1997).

Charcosset et. al. (1998) consideram que um pequeno número de componentes principais fornece uma sumariação satisfatória da caracterização molecular das linhagens. Os autores apresentam, em seu trabalho, 10 covariáveis, as quais são utilizadas em sete modelos diferentes. No presente estudo, apenas quatro covariáveis foram consideradas, a partir das quais foram construídos os quatro modelos aqui utilizados.

Sejam  $C_{ip}$  e  $C_{jp}$ , respectivamente, os escores das linhagens  $i$  e  $j$  do  $p$ -ésimo componente principal, gerados a partir da matriz de variância e covariância dos elementos 0 e 1. Considerando o cruzamento entre as linhagens  $i$  e  $j$ , se quatro componentes principais forem considerados, as seguintes covariáveis podem ser calculadas:

$$A1_{ij} = C_{i1}C_{j1}$$

$$A2_{ij} = C_{i2}C_{j2}$$

$$A3_{ij} = C_{i3}C_{j3}$$

$$A4_{ij} = C_{i4}C_{j4}$$

Diferentes subconjuntos de covariáveis podem ser introduzidos em um modelo de regressão fatorial. A adição seqüencial das covariáveis supracitadas ao modelo aditivo resulta nos seguintes modelos:

$$(1) Y_{ij} = \mu + g_i + g_j + \rho_1 A1_{ij} + e_{ij} \text{ (Charcosset 1)}$$

$$(2) Y_{ij} = \mu + g_i + g_j + \rho_1 A1_{ij} + \rho_2 A2_{ij} + e_{ij} \text{ (Charcosset 2)}$$

$$(3) Y_{ij} = \mu + g_i + g_j + \rho_1 A1_{ij} + \rho_2 A2_{ij} + \rho_3 A3_{ij} + e_{ij} \text{ (Charcosset 3)}$$

$$(4) Y_{ij} = \mu + g_i + g_j + \rho_1 A1_{ij} + \rho_2 A2_{ij} + \rho_3 A3_{ij} + \rho_4 A4_{ij} + e_{ij} \text{ (Charcosset 4)}$$

em que  $\mu$ ,  $g_i$ , e  $g_j$ , já foram definidos anteriormente e  $\rho_1$ ,  $\rho_2$ ,  $\rho_3$ ,  $\rho_4$ , os coeficientes de regressão associados às covariáveis  $A1_{ij}$ ,  $A2_{ij}$ ,  $A3_{ij}$ ,  $A4_{ij}$ , de acordo com cada modelo. Da mesma forma que no modelo de distância, os parâmetros foram estimados pelo método OLS considerando as covariáveis de efeito fixo.

A predição do desempenho de um cruzamento com perda simulada entre as linhagens  $k$  e  $l$  foi obtida mediante procedimentos análogos aos apresentados para a predição dos cruzamentos no modelo de distância, resultando nas seguintes expressões de predição:

$$(1) \hat{Y}_{kl} = \hat{\mu} + \hat{g}_k + \hat{g}_l + \hat{\rho}_1 A1_{kl}$$

$$(2) \hat{Y}_{kl} = \hat{\mu} + \hat{g}_k + \hat{g}_l + \hat{\rho}_1 A1_{kl} + \hat{\rho}_2 A2_{kl}$$

$$(3) \hat{Y}_{kl} = \hat{\mu} + \hat{g}_k + \hat{g}_l + \hat{\rho}_1 A1_{kl} + \hat{\rho}_2 A2_{kl} + \hat{\rho}_3 A3_{kl}$$

$$(4) \hat{Y}_{kl} = \hat{\mu} + \hat{g}_k + \hat{g}_l + \hat{\rho}_1 A1_{kl} + \hat{\rho}_2 A2_{kl} + \hat{\rho}_3 A3_{kl} + \hat{\rho}_4 A4_{kl},$$

em que:

$\hat{g}_k$  e  $\hat{g}_l$  são as estimativas das CGC das linhagens  $k$  e  $l$ , considerando os cruzamentos avaliados em que cada linhagem participou em um dado modelo;  
 $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4$  são as estimativas dos coeficientes de regressão estimado a partir dos cruzamentos avaliados em cada um dos modelos;  
 $A1_{kl}, A2_{kl}, A3_{kl}, A4_{kl}$  os valores das covariáveis calculadas a partir dos escores das linhagens  $k$  e  $l$ .

### 3.5.4 Equações de modelos mistos (EMM)

Na análise de EMM, a matriz de incidência de efeitos fixos foi constituída pelos blocos dos látices dos três locais, enquanto os efeitos da CGC e da CEC, modelados conforme o modelo II de Griffing (1956), foram tomados como aleatórios.

Dois enfoques foram considerados: o primeiro, um modelo sem covariável; o segundo, um modelo em que a endogamia foi assumida como covariável, uma vez que esta influência nas estimativas dos parâmetros genéticos normalmente é ignorada na modelagem. Considerou-se como estimativa da endogamia da geração  $F_1$  do cruzamento entre as linhagens  $i$  e  $j$ , o coeficiente de parentesco entre elas. Porém, como foram analisadas as gerações  $F_4$ , foi acrescido o efeito das autofecundações.

As estimativas dos efeitos fixos e predição dos efeitos aleatórios, capacidades geral e específica de combinação, foram obtidas de forma semelhante ao apresentado por Bernardo (1994), em cruzamentos dialélicos parciais. O modelo linear considerado pode ser expresso por:

$$Y = X\beta + Z_1g + Z_2s + e,$$

em que:

$Y$ : é o vetor de observações;



$\mathbf{X}$ : é a matriz de incidência dos efeitos fixos, no caso da ausência de covariável formada pelos blocos dos látices; no caso da presença de covariável com uma coluna adicional, com valores da endogamia dos cruzamentos.

$\beta$ : é o vetor de efeitos fixos; no caso de ausência de covariável, constituída pelos efeitos dos blocos dos látices; no caso de presença de covariável, um parâmetro a mais, o de regressão;

$\mathbf{Z}_1$ : é a matriz de incidência dos efeitos aleatórios da CGC;

$\mathbf{g}$ : é o vetor dos efeitos aleatórios das CGC;

$\mathbf{Z}_2$ : é a matriz de incidência dos efeitos aleatórios da CEC;

$\mathbf{s}$ : é o vetor dos efeitos aleatórios das CEC;

$\mathbf{e}$ : é o vetor de resíduos, com variância  $\sigma^2$ .

Dado que os efeitos de CGC e de CEC são aleatórios, a estrutura de covariância associada a cada um desses fatores pode ser incluída na análise. As matrizes de variância e covariância genética entre os efeitos das CGC e entre os efeitos das CEC podem ser expressas por:

$\mathbf{G}_1 = \mathbf{A}_1 \sigma_g^2$ , matriz de variâncias e covariâncias genéticas associada aos efeitos do vetor  $\mathbf{g}$ , em que  $\mathbf{A}_1$  é a matriz de parentesco entre as linhagens e  $\sigma_g^2$  é o componente de variância da CGC;

$\mathbf{G}_2 = \mathbf{A}_2 \sigma_s^2$ , matriz de variância e covariância genéticas associadas aos efeitos do vetor  $\mathbf{s}$ , em que  $\mathbf{A}_2$  é a matriz coeficientes de dominância entre os cruzamentos e  $\sigma_s^2$  é o componente de variância da CEC.

A matriz de parentesco,  $\mathbf{A}_1$ , associada aos efeitos das CGC, e a matriz de coeficientes de dominância,  $\mathbf{A}_2$ , associada aos efeitos das CEC, foram calculadas de acordo com Lynch e Ritland (1999) (item 3.3).

Assumindo que os efeitos não genéticos são não correlacionados,  $\beta$ ,  $\mathbf{g}$ , e  $\mathbf{s}$  podem ser estimadas pelas seguintes equações de modelos mistos:

$$\begin{bmatrix} \beta^0 \\ \hat{\mathbf{g}} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 & \mathbf{X}'\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}_1^{-1}\gamma_1 & \mathbf{Z}_1'\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{X} & \mathbf{Z}_2'\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{A}_2^{-1}\gamma_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}_1'\mathbf{Y} \\ \mathbf{Z}_2'\mathbf{Y} \end{bmatrix}, \text{ ou renomeado:}$$

$$\begin{bmatrix} \beta^0 \\ \hat{\mathbf{g}} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}_1'\mathbf{Y} \\ \mathbf{Z}_2'\mathbf{Y} \end{bmatrix}, \text{ em que:}$$

$$\gamma_1 = \frac{\sigma^2}{\sigma_g^2} \text{ e } \gamma_2 = \frac{\sigma^2}{\sigma_s^2} .$$

As estimativas dos componentes de variância foram obtidas pelo método da máxima verossimilhança restrita (REML), utilizando o algoritmo de maximização de esperança (EM-REML). A cada quarenta iterações o acelerador de convergência de Aitken foi utilizado, conforme descrito em Conte (1971).

As estimativas dos componentes de variância, em uma dada iteração, foram obtidas por:

$$\hat{\sigma}^2 = \frac{Y'Y - \beta^0 X'Y - \hat{\mathbf{g}}' \mathbf{Z}_1' Y - \hat{\mathbf{s}}' \mathbf{Z}_2' Y}{N - r(X)}$$

$$\hat{\sigma}_g^2 = \frac{\hat{\mathbf{g}}' \mathbf{A}_1^{-1} \hat{\mathbf{g}} + tr(\mathbf{A}_1^{-1} \mathbf{C}_{11}) \sigma^2}{q_1}$$

$$\hat{\sigma}_s^2 = \frac{\hat{s}' A_2^{-1} \hat{s} + tr(A_2^{-1} C_{22}) \sigma^2}{q_2},$$

sendo  $N$  o número de observações,  $r(X)$  o posto da matriz  $X$ ,  $tr(\ )$  o operador traço matricial,  $q_1$  o número de linhagens e  $q_2$  o número de cruzamentos avaliados.

Para solução do sistema, inicialmente foram atribuídos valores arbitrários para os componentes de variância de modo a se obter a primeira solução. Com essa solução, novos componentes de variância foram calculados e então utilizados para a obtenção de novas soluções. O processo iterativo foi então repetido até que houvesse convergência satisfatória dos componentes de variância.

A predição dos cruzamentos não avaliados, com perda simulada, foi obtida com base no desempenho dos cruzamentos avaliados. Seja  $Y_p$  o vetor do desempenho médio dos cruzamentos avaliados corrigidos para os efeitos de  $\beta$ .

$$\hat{Y}_p = (Z_2' Z_2)^{-1} Z_2' (Y - X\beta)$$

O desempenho dos cruzamentos com perda simulada pode ser expresso por (Bernardo, 1994):

$$\hat{Y}_M = CV^{-1} \hat{Y}_p,$$

em que:

$Y_M$ : vetor do desempenho dos cruzamentos com perda simulada;

$C$ : matriz de covariâncias genéticas entre os cruzamentos com perda simulada e os cruzamentos avaliados;

$V$ : matriz de variâncias e covariâncias fenotípicas entre os cruzamentos avaliados.

As matrizes C e V podem ser expressas por:

$$C = Z_{1NA}(G_1)Z'_{1A} + Z_{2NA}(G_2)Z'_{2A}$$
$$V = Z_{1A}(G_1)Z'_{1A} + Z_{2A}(G_2)Z'_{2A} + inv(Z'_2Z_2)\sigma^2$$

As matrizes  $Z_{1NA}$  e  $Z_{1A}$  são, respectivamente, matrizes de incidência dos cruzamentos não avaliados e avaliados referentes às CGC (contendo um só registro de cada cruzamento) enquanto as matrizes  $Z_{2NA}$  e  $Z_{2A}$  são, respectivamente, as matrizes de incidência das CEC dos cruzamentos não avaliados e avaliados (contendo um só registro de cada cruzamento).

### **3.6 Procedimentos de simulação de perdas e estatísticas de avaliação das predições dos cruzamentos com perda simulada**

Em todos os oito modelos considerados, a partir dos 142 tratamentos originais, 120 populações e 22 genitores, perdas simuladas de 12, 24, 36, 48 e 60 tratamentos foram feitas, sendo, então, as análises realizadas com os cruzamentos remanescentes. Para os modelos aditivo, de distância, Charcosset 1, Charcosset 2, Charcosset 3, Charcosset 4, foram realizadas mil simulações para cada tamanho de perda, enquanto que para os modelos EMM com covariável e EMM sem covariável, foram feitas 250 simulações. As simulações de perda foram realizadas de tal forma que, nos tratamentos remanescentes, cada genitor participasse de pelo menos um cruzamento ou, ele próprio, estivesse no quadro de análise.

Ao final de cada simulação, as correlações de Pearson e Spearman entre os valores preditos e os valores observados dos cruzamentos com perdas simuladas eram obtidas. Estatísticas descritivas das correlações, de Pearson e de Spearman, ao longo das simulações, para cada tamanho de perda em cada um

dos modelos, foram obtidas: correlações mínima, média e máxima, além do desvio-padrão.

## 4 RESULTADOS

### 4.1 Divergência genética e parentesco via Lynch e Ritland (1999)

Na Tabela 2, estão apresentadas as estimativas do coeficiente de parentesco entre os 22 genitores ( $\hat{r}_{xy}$ ), que corresponde à matriz de similaridade  $A_I$ , associada aos efeitos da CGC, utilizada nas EMM. Em negrito, estão os cruzamentos obtidos por Abreu (1997), 120 ao todo.

Como já comentado anteriormente, os complementos em relação à unidade ( $1-\hat{r}_{xy}$ ), foram tomados como medidas de dissimilaridade. Essas medidas foram usadas como covariável no modelo de distância. O dendrograma elaborado a partir dessas dissimilaridades encontra-se na Figura 1.

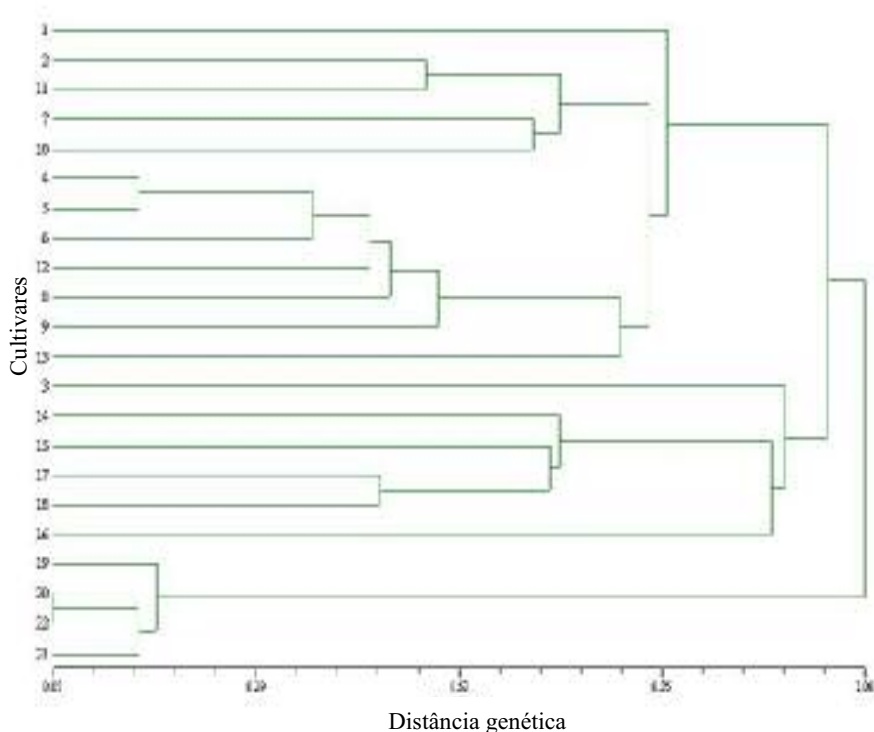
Observando-se os valores em negrito, vê-se que, em 76 dos 120 cruzamentos avaliados, o parentesco entre os genitores foi igual a zero. Na realidade, as estimativas foram negativas. Uma vez que não faz sentido considerar parentesco negativo, atribuíram-se então, valores nulos. Os genitores numerados de 1 a 13 são da raça Mesoamérica; 15 e 18, da raça Jalisco; 14, 16 e 17 da raça Durango; 19, 20, 21 e 22 da raça Nova Granada.

É perceptível a eficiência da análise utilizada no agrupamento de cultivares, de acordo com as raças preconizadas, desde que seja assumido que cultivares de raças diferentes sejam menos aparentadas. Dentro da raça Mesoamérica, houve apenas três estimativas de parentesco iguais a zero, todas elas envolvendo o genitor 3: 1 e 3; 3 e 6; 3 e 11. Aliás, essas foram as únicas estimativas iguais a zero envolvendo genitores de uma mesma raça.

**TABELA 2** Matriz dos coeficientes de parentesco ( $r_{xy}$ ) entre os genitores, obtidos de acordo com Lynch e Ritland (1999). Em negrito estão os cruzamentos avaliados por Abreu (1997).

Pais*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
1	<b>1</b>	0,20	0,00	0,18	0,29	0,32	0,11	0,16	0,33	0,13	0,37	0,22	0,22	<b>0,00</b>	<b>0,00</b>	<b>0,02</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
2		<b>1</b>	<b>0,04</b>	0,19	0,17	0,14	0,32	<b>0,15</b>	0,06	0,25	<b>0,51</b>	<b>0,22</b>	0,13	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
3			<b>1</b>	0,05	0,08	0,00	0,13	<b>0,27</b>	0,09	0,13	<b>0,00</b>	<b>0,11</b>	0,22	<b>0,03</b>	<b>0,16</b>	<b>0,04</b>	<b>0,06</b>	<b>0,17</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
4				<b>1</b>	0,85	0,58	0,26	0,61	0,58	0,24	0,35	0,62	0,38	0,02	<b>0,00</b>	<b>0,09</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
5					<b>1</b>	0,71	0,25	0,61	0,64	0,28	0,34	0,62	0,32	<b>0,05</b>	<b>0,00</b>	<b>0,12</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
6						<b>1</b>	0,27	0,48	0,47	0,19	0,30	0,49	0,12	<b>0,02</b>	<b>0,00</b>	<b>0,20</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
7							<b>1</b>	0,38	0,24	0,39	0,45	0,52	0,22	<b>0,00</b>	<b>0,00</b>	<b>0,21</b>	<b>0,07</b>	<b>0,04</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
8								<b>1</b>	0,44	0,39	<b>0,27</b>	<b>0,52</b>	0,26	<b>0,09</b>	<b>0,01</b>	<b>0,15</b>	<b>0,05</b>	<b>0,07</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
9									<b>1</b>	0,17	0,33	0,36	0,37	<b>0,07</b>	<b>0,00</b>	<b>0,17</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
10										<b>1</b>	0,40	0,19	<b>0,00</b>	<b>0,00</b>	<b>0,07</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
11											<b>1</b>	0,22	0,17	<b>0,00</b>	<b>0,00</b>	<b>0,05</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
12												<b>1</b>	0,26	<b>0,00</b>	<b>0,01</b>	<b>0,16</b>	<b>0,05</b>	<b>0,00</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
13													<b>1</b>	<b>0,03</b>	<b>0,03</b>	<b>0,00</b>	<b>0,06</b>	<b>0,08</b>	0,00	0,00	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
14														<b>1</b>	0,34	0,15	0,33	0,39	0,00	0,00	0,00	0,00	0,00
15															<b>1</b>	0,13	0,28	0,45	0,00	0,00	0,00	0,00	0,00
16																<b>1</b>	0,07	0,08	0,00	0,00	0,00	0,00	0,00
17																	<b>1</b>	0,57	0,00	0,00	0,00	0,00	0,00
18																		<b>1</b>	<b>0,00</b>	0,00	<b>0,00</b>	<b>0,00</b>	0,00
19																			<b>1</b>	0,82	<b>0,82</b>	<b>0,84</b>	<b>0,84</b>
20																				<b>1</b>	<b>0,86</b>	<b>0,95</b>	<b>0,95</b>
21																					<b>1</b>	<b>0,84</b>	<b>0,84</b>
22																						<b>1</b>	<b>0,84</b>

1-Sanilac; 2-Brasil 2; 3-A 62; 4-Porrilo Sintético; 5-Jamapa; 6-Milionário; 7- IPA 7419; 8-A 175; 9- Rojo de Seda; 10- A 140; 11-Carioca; 12- Compuesto Negro Chimaltenango; 13-Puebla 152; 14- Great Northern; 15- Flor de Mayo; 16- A 114; 17- Ojo de Liebre; 18-Apetito Blanco; 19- Mont Calm; 20- Jalo; 21- Diacol Andino; 22- Perry Marrow.



**FIGURA 1** Dendrograma de distâncias genéticas ( $1-r_{xy}$ ) entre as 22 cultivares de feijão, obtidas pelo critério de agrupamento UPGMA. A numeração dos cultivares está de acordo com a Tabela 2.

Por outro lado, percebe-se que todas as estimativas de parentesco entre as cultivares das raças Mesoamérica, Durango e Jalisco, pertencentes ao centro de domesticação Mesoamérica, com as da raça Nova Granada, do centro de domesticação Sul do Andes, foram iguais a zero. Fica evidenciada também a diferenciação entre as raças dentro do centro de domesticação Mesoamérica. Dos 65 cruzamentos possíveis entre as cultivares da raça Mesoamérica com os das raças Durango/Jalisco, em 34 deles a estimativa do parentesco foi igual a zero e 57 foram inferiores a 0,10.

Os comentários tecidos a respeito das estimativas de parentesco são mais bem visualizados pela análise gráfica do dendrograma (Figura 1). Vê-se o



agrupamento das cultivares de acordo com os respectivos centros de origem, e, dentro do Centro de Domesticação Mesoamérica, uma distinção entre as cultivares da raça Mesoamérica e os das raças Durago/Jalisco. A única exceção nessa tendência de agrupamento foi a cultivar 3 que, pela metodologia empregada, foi considerada mais próxima das cultivares da raças Durango/Jalisco do que das cultivares da raça Mesoamérica.

#### **4.2 Predição de cruzamentos não realizados pelos diferentes modelos**

Os resumos das estimativas das correlações de Pearson e de Spearman, entre os valores preditos e os observados dos cruzamentos, ao longo das simulações realizadas para cada modelo, são apresentados nas Tabelas 3 e 4.

Pelas correlações de Pearson (Tabela 3), bem como as de Spearman (Tabela 4), observa-se, como esperado, uma tendência de redução na estimativa de correlação média quando menos cruzamentos preditores são avaliados (maior número de perdas simuladas). As menores estimativas de correlações médias de Pearson e de Spearman, para todos os modelos, foram obtidas quando 60 perdas foram simuladas, sempre inferiores a 0,50.

Percebe-se, ainda, uma redução gradual na estimativa do desvio padrão quando o número de cruzamentos a serem preditos (número de perdas simuladas) é maior. Esse fato pode ser justificado, uma vez que correlações obtidas a partir de um menor número de pontos estão sujeitas a maiores flutuações das estimativas. Coerentemente, em todos os modelos, foram encontradas menores amplitudes de estimativas de correlações (diferença entre as estimativas de correlações máximas e mínimas), quando o número de cruzamentos a ser predito foi maior.

**TABELA 3** Valores médios, mínimos e máximos e o desvio padrão das correlações de Pearson, obtidos ao longo de 1000 simulações para os modelos aditivo, de distância, Charcosset 1, Charcosset 2, Charcosset 3 e Charcosset 4, e de 250 simulações para EMM com e sem covariável, em cada um dos tamanhos de perda.

Número	Parâmetros	Modelos							
		Aditivo	Distância	Charcos. 1	Charcos. 2	Charcos. 3	Charcos. 4	EMM s/cov	EMM c/cov
12	Mínima	-0,534	-0,517	-0,580	-0,313	-0,669	-0,457	-0,479	-0,055
	Média	<b>0,453</b>	<b>0,495</b>	<b>0,526</b>	<b>0,525</b>	<b>0,529</b>	<b>0,512</b>	<b>0,504</b>	<b>0,504</b>
	Máxima	0,948	0,882	0,968	0,948	0,917	0,932	0,909	0,887
	D. Padrão	0,242	0,216	0,238	0,233	0,225	0,234	0,226	0,208
24	Mínima	-0,162	-0,100	-0,035	-0,176	-0,135	-0,087	0,018	0,093
	Média	<b>0,467</b>	<b>0,491</b>	<b>0,526</b>	<b>0,523</b>	<b>0,528</b>	<b>0,508</b>	<b>0,507</b>	<b>0,493</b>
	Máxima	0,804	0,852	0,847	0,859	0,852	0,860	0,790	0,786
	D. Padrão	0,154	0,146	0,150	0,155	0,150	0,153	0,141	0,144
36	Mínima	0,032	0,030	0,112	-0,152	0,000	0,023	0,193	0,083
	Média	<b>0,443</b>	<b>0,473</b>	<b>0,517</b>	<b>0,509</b>	<b>0,509</b>	<b>0,496</b>	<b>0,490</b>	<b>0,480</b>
	Máxima	0,760	0,748	0,825	0,821	0,834	0,777	0,792	0,736
	D. Padrão	0,120	0,112	0,117	0,122	0,114	0,118	0,110	0,114
48	Mínima	-0,001	0,112	0,155	0,151	0,058	0,133	0,182	0,235
	Média	<b>0,440</b>	<b>0,463</b>	<b>0,505</b>	<b>0,502</b>	<b>0,498</b>	<b>0,480</b>	<b>0,485</b>	<b>0,482</b>
	Máxima	0,676	0,718	0,769	0,771	0,716	0,753	0,716	0,713
	D. Padrão	0,100	0,096	0,093	0,098	0,096	0,098	0,096	0,090
60	Mínima	0,166	0,046	0,183	0,165	0,094	0,103	0,219	0,283
	Média	<b>0,418</b>	<b>0,443</b>	<b>0,487</b>	<b>0,477</b>	<b>0,476</b>	<b>0,465</b>	<b>0,469</b>	<b>0,468</b>
	Máxima	0,633	0,657	0,727	0,697	0,736	0,683	0,664	0,632
	D. Padrão	0,083	0,083	0,082	0,084	0,084	0,087	0,078	0,077
Média Geral		<b>0,444</b>	<b>0,472</b>	<b>0,512</b>	<b>0,507</b>	<b>0,508</b>	<b>0,492</b>	<b>0,491</b>	<b>0,485</b>

**TABELA 4** Valores médios, mínimos e máximos e o desvio padrão das correlações de Spearman, obtidos ao longo de 1000 simulações para os modelos aditivo, de distância, Charcosset 1, Charcosset 2, Charcosset 3 e Charcosset 4, e de 250 simulações para EMM com e sem covariável, em cada um dos tamanhos de perda.

Número Perdas	Parâmetros	Modelos							
		Aditivo	Distância	Charcos. 1	Charcos. 2	Charcos. 3	Charcos. 4	EMM s/cov	EMM c/cov
12	Mínima	-0,413	-0,455	-0,364	-0,315	-0,692	-0,455	-0,378	-0,126
	Média	<b>0,434</b>	<b>0,474</b>	<b>0,509</b>	<b>0,505</b>	<b>0,514</b>	<b>0,499</b>	<b>0,490</b>	<b>0,480</b>
	Máxima	0,944	0,930	0,965	0,958	0,944	0,972	0,916	0,916
	D. Padrão	0,249	0,228	0,236	0,234	0,229	0,237	0,229	0,223
24	Mínima	-0,051	-0,105	-0,019	-0,183	-0,241	-0,032	-0,038	0,090
	Média	<b>0,458</b>	<b>0,485</b>	<b>0,516</b>	<b>0,522</b>	<b>0,531</b>	<b>0,510</b>	<b>0,494</b>	<b>0,479</b>
	Máxima	0,820	0,849	0,837	0,843	0,857	0,883	0,809	0,806
	D. Padrão	0,155	0,153	0,152	0,157	0,150	0,152	0,154	0,148
36	Mínima	0,042	0,062	0,146	-0,111	0,129	0,071	0,148	0,131
	Média	<b>0,440</b>	<b>0,474</b>	<b>0,516</b>	<b>0,514</b>	<b>0,518</b>	<b>0,502</b>	<b>0,483</b>	<b>0,473</b>
	Máxima	0,767	0,790	0,841	0,828	0,860	0,792	0,829	0,761
	D. Padrão	0,122	0,118	0,114	0,121	0,114	0,120	0,111	0,115
48	Mínima	0,032	0,030	0,163	0,180	0,097	0,169	0,204	0,158
	Média	<b>0,437</b>	<b>0,466</b>	<b>0,508</b>	<b>0,509</b>	<b>0,510</b>	<b>0,491</b>	<b>0,486</b>	<b>0,483</b>
	Máxima	0,702	0,720	0,766	0,746	0,752	0,754	0,721	0,714
	D. Padrão	0,103	0,101	0,093	0,097	0,097	0,098	0,097	0,100
60	Mínima	0,096	0,002	0,151	0,175	0,113	0,104	0,240	0,264
	Média	<b>0,419</b>	<b>0,449</b>	<b>0,492</b>	<b>0,486</b>	<b>0,490</b>	<b>0,480</b>	<b>0,474</b>	<b>0,469</b>
	Máxima	0,653	0,663	0,705	0,718	0,719	0,724	0,693	0,648
	D. Padrão	0,086	0,087	0,080	0,084	0,083	0,087	0,082	0,078
Média Geral		<b>0,438</b>	<b>0,470</b>	<b>0,508</b>	<b>0,507</b>	<b>0,513</b>	<b>0,496</b>	<b>0,485</b>	<b>0,477</b>

**TABELA 5** Relação SQP/SQT para os modelos aditivo, de distância, Charcosset 1, Charcosset 2, Charcosset 3 e Charcosset 4 em relação à soma de quadrados total, na ausência de perdas simuladas.

	Modelos					
	Aditivo	Distância	Charcos.1	Charcos.2	Charcos.3	Charcos.4
SQP/SQT	0,410	0,443	0,476	0,480	0,490	0,490

A proporção da variação explicada pela parte preditiva (SQP) dos modelos aditivo, de distância e de regressão fatorial, ou seja, a parte do modelo composta pelos parâmetros de predição, em relação à variação total explicada por cada um deles (SQT), quando uma análise sem nenhuma perda foi realizada é apresentada na Tabela 5. Observa-se que, nos modelos em que se utilizou covariável, as relações SQP/SQT foram maiores que no modelo aditivo, indicando que parte da variação atribuída a CEC foi captada pela variável regressora ou variáveis regressoras inerentes a cada um dos modelos.

Além disso, as proporções obtidas pelos modelos propostos por Charcosset et al. (1998) foram superiores à obtida pelo modelo de distância, corroborando com os resultados das correlações de Pearson e de Spearman (Tabelas 3 e 4). Dado que essas duas formas de análise fizeram uso do mesmo conjunto de informações, isso é um indício de que a simples escolha do procedimento de análise pode influenciar no sucesso da pesquisa.

#### **4.2.1 Modelo aditivo e modelo de distância**

##### **a) Modelo aditivo**

O modelo aditivo foi considerado como modelo base de comparação para a predição de cruzamentos. Nesse modelo, apenas os efeitos médios dos genitores em um conjunto de cruzamentos, CGC, são usados na predição de

cruzamentos não avaliados. Não há nenhuma aproximação para a CEC do cruzamento a ser predito. Quando comparado aos demais modelos, este sempre apresentou as menores estimativas de correlações médias, de Pearson e de Spearman.

A correlação média de Pearson ao longo dos tamanhos de perdas foi de 0,444 e a de Spearman 0,438. A amplitude entre a maior e a menor estimativa de correlação média de Pearson, obtida com 24 e 60 perdas, respectivamente, foi de 0,05, enquanto para a de Spearman, obtida também com 24 e 60 perdas, foi 0,04.

#### **b) Modelo de distância**

As estimativas de correlações obtidas com a inclusão da covariável de distância genética não só tiveram a mesma tendência de comportamento, como também foram próximas daquelas obtidas com o modelo aditivo, embora sempre maiores (Tabelas 3 e 4).

A correlação média de Pearson ao longo dos tamanhos de perdas para o modelo de distância foi de 0,472, enquanto que a de Spearman foi de 0,470. Vê-se, ainda, que, tanto para o modelo aditivo como para o modelo de distância, todas as estimativas de correlações médias, de Pearson e de Spearman, obtidas nos diferentes tamanhos de perda simulada de cruzamentos, foram iguais ou inferiores a 0,5.

### **4.2.2 Modelos de regressão fatorial**

#### **a) Modelo Charcosset 1**

De modo geral, a introdução da covariável formada a partir dos escores das linhagens do primeiro componente principal (da matriz de elementos 0 e 1) proporcionou estimativas de correlações médias, de Pearson e de Spearman,

mais elevadas que as obtidas nos modelos anteriores (Tabelas 3 e 4). Comparando-as com as do modelo aditivo, o incremento nas correlações médias, de Pearson e de Spearman, foi, aproximadamente, de 0,07. A correlação média de Pearson, ao longo das simulações nos diferentes tamanhos de perda, foi de 0,512 e a de Spearman, 0,508. Em termos absolutos, essa foi a maior estimativa de correlação média de Pearson dentre todos os modelos. Excetuando-se a situação em que 60 perdas foram simuladas, todas as correlações nos demais tamanhos de perdas foram superiores a 0,5.

#### **b) Modelo Charcosset 2**

A inclusão no modelo da covariável obtida a partir dos escores do segundo componente principal das linhagens não alterou de forma substancial as estimativas das estatísticas, quando comparadas com as obtidas pelo modelo anterior (Tabela 3 e 4).

As estimativas das correlações médias ao longo das perdas foram similares às obtidas pelo modelo anterior. Para ambas as correlações, de Pearson e de Spearman, a estimativa da correlação média ao longo dos tamanhos de perda foi de 0,507.

De forma similar ao modelo anterior, a exceção da situação em que 60 perdas foram simuladas, nos demais tamanhos de perda, tanto as correlações médias de Pearson como as de Spearman foram superiores a 0,50.

#### **c) Modelo Charcosset 3**

O modelo preditivo composto por efeitos aditivos e três covariáveis, formadas a partir dos escores dos três primeiros componentes principais dos genitores, proporcionou estimativas de correlações médias similares às obtidas

quando uma ou duas covariáveis foram utilizadas, tanto para as estatísticas de Pearson como para as de Spearman (Tabelas 3 e 4).

A correlação média de Pearson, ao longo das simulações, foi de 0,508, enquanto a estimativa média das correlações de Spearman foi de 0,513 que, em valores absolutos, foi a maior estimativa dentre todos os modelos.

#### **d) Modelo Charcosset 4**

A inclusão da quarta covariável não adicionou ganhos em relação aos modelos Charcosset 1, Charcosset 2 e Charcosset 3, mas apresenta resultados satisfatórios quando comparados aos demais modelos, especialmente o ativo (Tabelas 3 e 4). Dentre os modelos propostos por Charcosset (1998) utilizados nesse estudo, este foi o que proporcionou, ao longo dos tamanhos de perdas, as menores estimativas de correlações médias de Pearson e de Spearman, 0,492 e 0,496, respectivamente, embora seja o modelo que inclua o maior número de covariáveis.

O desempenho dos diferentes modelos que fazem uso de covariáveis com relação às correlações de Pearson e de Spearman entre os valores preditos e os observados dos cruzamentos é corroborado pelos resultados apresentados na Tabela 5. Nas análises sem perdas, vê-se que a inclusão sucessiva de covariáveis para os modelos de Charcosset et al. (1998) pouco influenciou na relação SQP/SQT do modelo quando comparado ao modelo contendo apenas uma covariável.

### **4.2.3 Modelos EMM**

#### **a) EMM sem covariável**

Inicialmente, vale lembrar que, nos modelos anteriores, à exceção do modelo aditivo, a tentativa de recuperar parte da informação da CEC dos cruzamentos foi realizada com inclusão de covariáveis no modelo e a respectiva estimação dos parâmetros de regressão pelo método dos quadrados mínimos. No procedimento EMM, como já comentado, a estimação dos parâmetros genéticos considera a estrutura de parentesco entre os diferentes níveis de um fator aleatório, sem a necessidade de estimação de parâmetros extras. A estrutura de parentesco, contudo, é influenciada pela endogamia e sua consideração ou não na análise poderia conduzir a diferentes resultados.

Observa-se, pelas Tabelas 3 e 4, que as estimativas das correlações médias de Pearson e de Spearman foram intermediárias entre as obtidas pelo modelo de distância e as obtidas pelos modelos propostos por Charcosset et al. (1998), sendo, respectivamente, 0,491 e 0,485.

#### **b) EMM com covariável**

A inclusão da endogamia como covariável não trouxe benefício algum quando comparado ao modelo sem ela. Na média das estimativas ao longo dos tamanhos de perdas, os valores das correlações foram inferiores aos obtidos pelo modelo EMM sem covariável e bem próximos dos obtidos pelo modelo de distância (Tabelas 3 e 4).

Para esse modelo, todas as correlações médias, de Pearson e de Spearman, obtidas em cada tamanho de perda, foram inferiores a 0,5, exceto para a correlação média de Pearson quando 12 perdas foram simuladas. A média



das correlações de Pearson ao longo das perdas foi de 0,485, enquanto a de Spearman foi de 0,477.

## 5 DISCUSSÃO

Como comentado inicialmente, o presente trabalho utilizou os dados obtidos por Abreu (1997). Todavia, o enfoque aqui apresentado é diferente do discutido pela autora em seu estudo. Abreu (1997) avaliou o potencial de cruzamentos inter-raciais para a obtenção de populações segregantes. Entretanto, muitos dos cruzamentos do dialelo, aproximadamente a metade, não foram obtidos.

O propósito do presente trabalho tenta preencher essa lacuna, discutindo e apresentando alternativas de predição de cruzamentos perdidos ou não realizados.

### 5.1 Considerações sobre a metodologia de Lynch e Ritland (1999)

O estabelecimento de medidas de similaridade entre dois indivíduos, ou grupos de indivíduos, é o ponto de partida para a caracterização da divergência genética entre eles. Medidas de similaridade (ou de dissimilaridade) podem ser obtidas, tanto por características morfológicas quanto por marcadores moleculares. Marcadores do tipo RAPD são dicotômicos e dominantes, ou seja, variáveis binárias em que cada indivíduo apresenta apenas um dos padrões de bandas para um dado loco.

O agrupamento obtido pela medida de dissimilaridade aqui empregada (ver Figura 1) é similar aos obtidos por Duarte (1998), usando o coeficiente de Dice e por Abreu (1997), usando características morfológicas. A tendência de agrupamento obedeceu, primeiramente, aos centros de domesticação e, em seguida, às raças dentro do centro de domesticação Mesoamérica, como já comentado. As cultivares do centro de domesticação Sul dos Andes, que

possuem grãos grandes, formaram um grupo a parte. Dentro do centro de domesticação Mesoamérica, as raças Durango e Jalisco, cujos tamanhos de grãos são classificados como médios, formaram um grupo distinto das cultivares da raça Mesoamérica, caracterizados por grãos pequenos.

O agrupamento das cultivares dentro da raça Mesoamérica corrobora mais uma vez os resultados obtidos por Duarte (1998), evidenciando assim uma proximidade entre o coeficiente de Dice e o estimador de parentesco de Lynch e Ritland (1999). Os subgrupos obtidos indicam uma tendência de separação das cultivares de acordo com a cor do grão; um formado por cultivares de cor bege (Brasil, IPA 7419, A-140, Carioca) e outro formado por cultivares de grãos pretos (Jamapa, Porrilo Sintético, Milionário, Compuesto Negro Chimaltenango, Puebla), além da cultivar Rojo de Seda, de cor roxa.

Uma diferença obtida entre o agrupamento aqui realizado e o obtido por Duarte (1998) diz respeito às cultivares A-114 (número 16), raça Durango e A-62 (número 3), raça Mesoamérica. No trabalho do referido autor, a primeira cultivar foi agrupada juntamente com as cultivares da raça Mesoamérica. No presente trabalho, a cultivar A-62 é que foi agrupada juntamente com as cultivares das raças Durango/Jalisco.

É importante frisar que o coeficiente de Dice é uma medida de co-ocorrência, limitada entre “0” e “1”. As estatísticas de Lynch e Ritland (1999) são estimadores de parentesco e, portanto, estimativas negativas podem ser obtidas. Uma vez que parentesco negativo não faz sentido do ponto de vista genético, em tais situações é mais conveniente considerar as estimativas como zero. Essa situação ocorreu para os cruzamentos entre linhagens de centros de origens diferentes. Uma outra particularidade desse estimador é considerar, em cada loco, a frequência média dos padrões de amplificação (“0” ou “1”) ao longo das linhagens, o que não acontece no coeficiente de Dice.

Um último comentário, contudo, deve ser feito. O estimador de parentesco de Lynch e Ritland (1999) foi proposto para marcadores de natureza co-dominante; sabe-se também que marcadores RAPD não possuem essa característica. Assim, ao ser assumido que a ausência de uma dada banda em duas linhagens corresponde ao mesmo “produto de amplificação”, vieses de estimativas podem ter ocorrido.

## **5.2 Considerações gerais sobre os modelos de predição de cruzamentos**

De acordo com os resultados obtidos é possível verificar que a capacidade preditiva de todos os modelos foi influenciada pelo número de cruzamentos com perdas simuladas. Quanto maior o número de cruzamentos preditores, melhor a correspondência entre o valor predito e o valor fenotípico. Esse mesmo fato foi verificado por Bernardo (1994, 1995), na cultura do milho, avaliando o potencial das EMM na predição de cruzamentos não realizados. Maiores estimativas de correlações entre os valores preditos e observados sempre foram obtidas quando se aumentou o número de híbridos preditores. Charcosset et al. (1998), comparando diferentes modelos preditivos em cruzamentos de milho, verificaram também que as estimativas de coeficientes de determinação ( $r^2$ ) entre valores preditos e observados aumentaram com o aumento no número de híbridos preditores.

As estimativas dos desvios-padrões, obtidos ao longo das simulações, também apresentaram uma tendência comum em todos os modelos. Tais estimativas sempre foram maiores em menores tamanhos de perdas. Esses resultados diferem dos obtidos por Charcosset et al. (1998) que verificaram ocorrência de menores estimativas de desvios- padrões para as estimativas  $r^2$ , quando um número intermediário de híbridos preditores foi utilizado. O desvio-padrão obtido para o coeficiente  $r^2$ , quando se usou a maior quantidade de

cruzamentos preditores, foi superior ao obtido quando se usou a menor dentre aquelas utilizadas pelos autores.

Algumas suposições podem ser feitas para explicar tais resultados no presente estudo. Em menores tamanhos de perdas (menor quantidade de cruzamentos a ser predito), por questão de amostragem, apenas cruzamentos com predição destoante do valor fenotípico observado foram amostrados; ou, por outro lado, apenas foram amostrados os cruzamentos com predições mais próximas dos valores observados. Dessa forma, estimativas de correlações extremas, mínimas e máximas, teriam ocorrido e, por conseguinte, inflacionado as estimativas dos desvios-padrões.

De modo geral, observa-se que, com 36 perdas (tanto para correlações de Pearson como para as de Spearman), as estimativas dos desvios-padrão foram aproximadamente metade das obtidas quando 12 perdas foram simuladas e já bem próximas das obtidas com 60 perdas.

### **5.3 Modelo aditivo e modelo de distância**

O modelo aditivo, em todas as situações, foi o que proporcionou as menores estimativas de correlações médias, de Pearson e de Spearman. Esse fato já era esperado, uma vez que as predições dos cruzamentos ausentes são realizados com base nas CGC dos genitores e, portanto, nenhuma aproximação para a CEC do cruzamento é realizada. Esses resultados estão de acordo com os obtidos por Charcosset (1998), que observaram menores estimativas de  $r^2$  para o modelo aditivo quando comparadas às dos modelos de distância, regressão fatorial e predições BLUP, na cultura do milho.

Sabe-se que o sucesso de predição de cruzamentos com base nas CGC dos genitores depende da predominância dos efeitos aditivos no controle do caráter. Isto sugere predominância dos componentes da CGC sobre os da CEC.

No caso do feijoeiro, em algumas situações, ambos os componentes têm se mostrado importantes (Mendonça, 2001; Machado, 1999).

No presente caso, as análises preliminarmente realizadas por Abreu (1997), houve predominância da CGC. A autora, estratificando a análise dialéctica, verificou que, nos cruzamentos dentro da raça Mesoamérica, houve efeito significativo para as fontes de variação da CEC e CGC. Por outro lado, na análise do dialeto parcial envolvendo genitores da raça Mesoamérica, em um grupo, com genitores das raças Durango/Jalisco, em outro, apenas as CGC para ambos os grupos foram significativas. Por fim, no dialeto parcial envolvendo os genitores da raça Mesoamérica, em um grupo, com os da raça Nova Granada, em outro, apenas a CGC referente ao grupo Nova Granada foi significativa.

Pelo exposto, pode-se inferir que os genitores da raça Mesoamérica possuem propriedades de capacidade combinatória variadas, que dependem da raça com a qual esses genitores são cruzados. Assim, o padrão combinatório exibido em cruzamento com uma dada raça não pode ser estendido a cruzamentos com outra raça. Esse fato justificaria, pelo menos em parte, o pior desempenho dessa metodologia na predição de cruzamentos.

Vale ressaltar que, dos 120 cruzamentos avaliados, 113 deles envolviam cultivares da raça Mesoamérica, sendo 10 desses dentro da raça, 64 com as raças Durango/Jalisco e 39 com a raça Nova Granada.

A incorporação da covariável de distância genética ao modelo aditivo proporcionou, para todos os tamanhos de perdas simuladas, um pequeno aumento nas estimativas de correlações médias de Pearson e de Spearman (Tabelas 3 e 4).

Esses resultados estão de acordo com os obtidos por Charcosset et al. (1998). O autor verificou que o modelo de distância foi superior ao modelo aditivo na predição de cruzamentos não realizados quando, entre os híbridos preditores, estavam incluídos os oriundos de cruzamentos entre linhagens

aparentadas e os oriundos de cruzamentos entre linhagens não aparentadas. Contudo, quando dentre os híbridos preditores estavam apenas incluídos os híbridos oriundos de linhagens não aparentadas, os dois modelos tiveram eficiências equivalentes. No primeiro caso, os coeficientes de determinação para o modelo aditivo variaram de 54,1% a 63,4%; para o modelo de distância esses valores variaram de 65,7% a 73,6%, dependendo do número de cruzamentos usados como preditores. No segundo caso, as estimativas dos coeficientes de determinação variaram de 60,5% a 69,8% para o modelo aditivo e de 60,5% a 70,4% para o modelo de distância, dependendo do número de cruzamentos usados como preditores.

Não obstante, o sucesso de uma variável regressora para captar a tendência do comportamento de uma variável resposta depende da correlação entre elas; no presente caso, distância genética e CEC. Duarte (1998) já havia verificado que as correlações entre distâncias genéticas obtidas por marcadores e as CEC dos cruzamentos avaliados por Abreu (1997) foram de  $-0,02$ . No presente trabalho, 76 dos 120 cruzamentos tiveram estimativa de parentesco igual a zero. Assim, tais cruzamentos apresentaram o mesmo valor para a variável regressora na matriz de incidência, no caso, o valor "1". Porém, para que haja correlação é necessário variabilidade em ambas as variáveis.

Outros trabalhos realizados com a cultura do feijoeiro também têm relatado baixas correlações entre CEC e a distância estimada por marcadores moleculares. Mendonça (2001) verificou correlações praticamente nulas entre distância genética obtida por marcadores moleculares e as CEC das gerações  $F_2$  e  $F_3$  do feijoeiro. Em ambas as gerações do dialelo, houve efeito significativo tanto para a CGC como para a CEC.

#### **5.4 Modelos de regressão fatorial**

De início, convém ressaltar que a principal diferença entre as covariáveis utilizadas nos modelos de Charcosset et al. (1998) e a utilizada no modelo de distância está no tratamento algébrico dado à matriz binária. No modelo de distância, o complemento do parentesco estimado conforme Lynch e Ritland (1999) foi utilizado como covariável. Nos modelos de regressão fatorial, as covariáveis são formadas a partir dos escores de cada genitor, obtidos de uma análise de componentes principais, tomando-se cada padrão de amplificação como uma variável. O modelo, então, baseia-se no princípio de que poucos componentes principais consigam “captar” a variabilidade existente nos genótipos e “fornecer” um padrão de comportamento da CEC.

As correlações médias, de Pearson e de Spearman, obtidas nos quatro modelos utilizados, foram sempre superiores àquelas obtidas nos demais modelos. As correlações obtidas com uma, duas ou três covariáveis foram sempre superiores a 0,50.

Esses resultados destoam um pouco dos obtidos por Charcosset et al. (1998). Ao comparar diferentes modelos preditivos, estes autores verificaram que o modelo de distância superou as predições BLUP, aditivo e de regressão fatorial, quando entre os híbridos preditores estavam cruzamentos envolvendo linhagens aparentadas (com ancestrais comuns) e não aparentadas. Quando, entre os híbridos preditores, estavam incluídos apenas cruzamentos entre linhagens não aparentadas, a predição BLUP foi a melhor, tendo alguns modelos de regressão fatorial sido superiores ao modelo de distância. O modelo aditivo sempre apresentou menor eficiência.

É interessante comentar que, como as covariáveis foram calculadas a partir dos escores dos componentes principais, é provável que essas não sejam correlacionadas. Isso implica em não haver informação redundante entre elas.



Contudo, a eficiência de uma variável regressora depende da correlação que ela tem com a variável dependente e, a princípio, não há razão para que os primeiros componentes principais, embora respondam pela maior parte da variação, sejam os mais adequados em qualquer circunstância. De acordo com van Eeuwijk et al. (1996), um dos pontos cruciais para a aplicação dos modelos de regressão fatorial é a escolha de bons conjuntos de covariáveis.

Modelos de regressão fatorial já são comumente usados no estudo da interação genótipos por ambientes, sendo o modelo mais consagrado, ou, pelo menos, mais divulgado, o modelo AMMI (Dias e Krzanowski, 2003; Duarte e Vencovsky, 1999; Piepho, 1994). A aplicabilidade da modelagem AMMI ao estudo da interação, CEC, em cruzamentos dialélicos também já tem sido demonstrada na literatura (Ortiz et al., 2001; Duarte e Pinto, 2002). Contudo, seu potencial como modelo preditivo de cruzamentos não realizados ainda tem sido pouco explorado.

### **5.5 Equações de modelos mistos (EMM)**

De acordo com os resultados obtidos pelas EMM, na presença ou ausência da endogamia como covariável, infere-se que o poder preditivo desse modelo foi, na média, inferior aos modelos de regressão fatorial e muito próximo do obtido pelo modelo de distância. A inclusão da endogamia como covariável proporcionou, em algumas situações, menores estimativas dos parâmetros avaliados que o modelo de distância.

Esperava-se que o uso da covariável melhorasse as estimativas das predições. Tal correção, a princípio, parecia necessária uma vez que os genitores foram assumidos como uma amostra de uma população de referência e, assim, no cruzamento entre indivíduos aparentados, a descendência seria endogâmica, o que influenciaria as estimativas dos componentes de média e variância.

Em princípio, a correção para a endogamia pouco deve ter influenciado o poder preditivo da metodologia por duas razões: (i) a primeira delas, é que, salvo os cruzamentos realizados dentro de uma mesma raça, o parentesco entre os genitores obtidos pela metodologia de Lynch e Ritland (1999), coerentemente, foi muito baixo. Como mostrado, em 76 dos 120 cruzamentos avaliados, as estimativas de parentesco aditivo entre os genitores, endogamia da primeira geração filial, foi igual a zero e, em 99 deles, inferior a 0,1; (ii) os cruzamentos foram avaliados na geração  $F_4$  e, nesse caso, mesmo com endogamia inicial zero, a endogamia na geração de avaliação seria 0,875. Dessa forma, devido tanto à proximidade entre os valores para as covariáveis dos cruzamentos, como ao número elevado de cruzamentos com o mesmo valor para a covariável, a presença da endogamia do cruzamento como variável regressora pouco teria contribuído para melhorar o desempenho preditivo do modelo.

O grande número de parentescos iguais a zero talvez tenha sido uma das principais razões para o baixo desempenho na predição de cruzamentos não realizados. Isso porque a Tabela 2 corresponde à matriz de parentesco A1, usada como matriz dos coeficientes de variância e covariâncias associadas aos efeitos da CGC. Em tais circunstâncias, pouca informação é recuperada entre os níveis dos efeitos aditivos do modelo e, por esse motivo, parte da informação recuperada pela metodologia talvez seja devido ao coeficiente de dominância, associada aos efeitos da CEC. Bernardo (1996d), em milho, verificou que EMM na análise dialéctica contendo efeitos aditivos e de dominância foi mais eficiente na predição de cruzamentos não realizados em relação a modelos exclusivamente aditivos. Ainda com relação à ausência de parentesco, Bernardo (1996b) verificou que a predição BLUP é robusta quando são desconhecidos cerca de 25%, ou quando existem vieses nas estimativas. Todavia, para os dados em questão, mais de 60% dos cruzamentos avaliados apresentam coeficiente de parentesco igual a zero.

Outra questão a ser considerada diz respeito ao grande desbalanceamento inicial existente nos dados utilizados. Além do limitado poder de recuperação intergenotípica comentado para a matriz de parentesco, a amostragem completamente aleatória dos cruzamentos a serem preditos poderia ter contribuído para o reduzido poder preditivo. Cruzamentos cujos genitores participassem inicialmente em um menor número de hibridações poderiam ter sido prejudicados na predição. Charcosset et al. (1998) verificaram aumento no poder preditivo de diferentes modelos quando se aumentou o número de híbridos preditores. Em adição, Bernardo (1996c) verificou, na cultura do milho, que a eficiência das EMM na predição de cruzamentos dialélicos não avaliados é maior quando os genitores participam de hibridações no conjunto de cruzamentos preditores.

No presente trabalho, em cada simulação, tomou-se o cuidado para que cada genitor participasse de pelo menos um cruzamento dentre os preditores. Contudo, é provável que uma amostragem mais sistemática, que mantivesse uma participação mais equitativa dos genitores dentre os cruzamentos preditores, talvez conduzisse a resultados mais satisfatórios. Como já apresentado, as estimativas das correlações médias, de Pearson e de Spearman, para as predições BLUP nos diferentes tamanhos de perdas ficaram abaixo de 0,51. Essas estimativas são bem inferiores às obtidas por Bernardo (1994, 1995), na cultura do milho. No primeiro caso, as correlações variaram de 0,66 a 0,80 e, no segundo, de 0,59 a 0,68, dependendo do modelo e do número de perdas simuladas.

Assim, em uma última análise, o baixo desempenho desses modelos deve-se em parte às particularidades dos dados em estudo e da estratégia utilizada para simulação e predição dos cruzamentos excluídos do quadro de análise e não, exclusivamente, a um limitado poder preditivo de cruzamentos não realizados.

Uma última consideração que deve ser feita, válida para todos os modelos, diz respeito à precisão dos valores fenotípicos observados, uma vez que esses foram tomados como referência para os valores preditos. Tais valores são estimativas dos valores paramétricos e, portanto, existe erro associado.

Dessa forma, as correlações efetuadas entre os valores preditos e os valores fenotípicos são apenas uma aproximação das correlações entre os primeiros e os valores paramétricos que, sob condições experimentais, não podem ser conhecidos.

De acordo com Bernardo (1994), a correlação teórica entre os valores fenotípicos e os valores genéticos é igual à raiz quadrada da herdabilidade. Portanto, esse é o valor máximo teórico da correlação entre os valores preditos e os valores fenotípicos. O valor da herdabilidade obtido quando todos os cruzamentos foram considerados na análise foi de 0,607, sendo, portanto, o limite teórico dessa correlação 0,78. Em nenhuma das simulações realizadas, a estimativa da correlação média de Pearson, ou de Spearman, foi superior a esse valor (Tabelas 3 e 4).

## 6 CONCLUSÕES

Os estimadores de parentesco e de coeficiente de dominância, propostos por Lynch e Ritland (1999), proporcionaram estimativas coerentes com o esperado, tendo em vista a origem preconizada dos diferentes genitores.

Os modelos de regressão fatorial, dentre os utilizados, foram os que proporcionaram as melhores predições de cruzamentos não avaliados.

O fato de considerar a endogamia como covariável na análise dialética via EMM não trouxe benefício algum para as predições de cruzamentos não avaliados. A princípio, as baixas estimativas de parentesco entre os genitores devem ter contribuído para o baixo desempenho preditivo.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, A. de F. B. **Predição do potencial genético de populações segregantes do feijoeiro utilizando genitores inter-raciais**. 1997. 80 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG.
- ANDRÉ, C. M. G. **Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores moleculares na análise dialélica**. 1999, 101 p. Dissertação – (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras, Lavras, MG.
- ARAÚJO, P. M. de. **Dialelo parcial circulante interpopulacional e cruzamentos “top cross” na avaliação de linhagens parcialmente endogâmicas de milho (*Zea mays* L)**. 2000. 170 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- BERNARDO, R. Best linear unbiased prediction of maize single-cross performance. **Crop Science**, Madison, v. 36, n. 1, p. 50-56, Jan./Feb. 1996a.
- BERNARDO, R. Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationship. **Crop Science**, Madison, v. 36, n. 4, p. 862-866, July/Aug. 1996b.
- BERNARDO, R. Best linear unbiased prediction of the performance of crosses between untested maize inbreds. **Crop Science**, Madison, v. 36, n. 4, p. 872-876, July/Aug. 1996c.
- BERNARDO, R. Genetic models for prediction maize single-cross performance in unbalanced yield trial data. **Crop Science**, Madison, v. 35, n. 1, p. 141-147, Jan./Feb. 1995.
- BERNARDO, R. Predicting of maize single cross performance using RFLP and information from related hybrid. **Crop Science**, Madison, v. 34, n. 1, p. 20-25, Jan./Feb. 1994.
- BERNARDO, R. Test cross additive and dominance effects in best linear unbiased prediction of maize single-cross performance. **Theoretical and Applied Genetics**, Berlin, v. 93, n. 7, p. 1098-1102, Nov. 1996d.

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, Apr. 1993.

BUENO FILHO, J. S. de S. **Modelos mistos na predição de valores genéticos aditivos em testes de progênies florestais**. 1997. 118 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura de

BURSTIN, J.; VIENNE, D de.; DUBREUIL, P.; DAMERVAL, C. Molecular markers and protein quantities as genetic descriptors in maize. I. Genetic diversity among 21 inbred lines. **Theoretical and Applied Genetics**, Berlin, v. 89, p. 943-950, 1994.

CHARCOSSET, A.; BONNISSEAU, B.; TOUCHEBEUF, O. et al. Prediction of maize hybrids silage performance use markers data: comparison of several models for specific combining ability. **Crop Science**, Madison, v. 38, n. 1, Jan./Feb. 1998.

CONTE, S. D. **Elementos de análise numérica**. 3. ed. Porto Alegre: Globo, 1971. 331 p.

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

CRUZ, C. D.; VENCOVSKY, R. Comparação de alguns métodos de análise dialélica. **Revista Brasileira de Genética**, Ribeirão Preto, v. 12, n. 2, p. 432-438, jan. 1989.

DIAS, C. T. dos S.; KRZANOWSKI, W. J. Model selection and cross validation in additive main effect and multiplicative interaction models. **Crop Science**, Madison, v. 43, n. 3, p. 865-873, May/June 2003.

DUARTE, J. B.; VENCOVSKY, R. **Interação genótipos x ambientes: uma introdução à análise “AMMI”**. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60 p. (Série Monografias, 9).

DUARTE, J. M. **Estudo da divergência genética entre raças de feijão por meio de marcadores RAPD**. 1998. 78 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras, Lavras, MG.

DUBREUIL, P.; DUFOUR, P.; KREJCI, E. et al. Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. **Crop Science**, Madison, v. 36, n. 3, p. 790-799, May/June 1996.

FALCONER, D. S. **Introdução à genética quantitativa**. Viçosa: UFV, 1987. 279 p.

FERREIRA, F. M. **Eficiência de dialelos circulantes por reamostragem de um dialelo completo**. 2003. 76 p. Dissertação (Mestrado) - Universidade Federal de Viçosa, Viçosa, MG.

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

GERALDI, I. O.; MIRANDA FILHO, J. B. de. Adapted models for the analysis of combining ability of varieties in partial diallel crosses. **Revista Brasileira de Genética**, Ribeirão Preto, v. 11, n. 2, p. 419-430, jun. 1988.

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

JINKS, J. L.; HAYMAN, B. I. The analysis of diallel crosses. **Maize Genetics Corporation Newsletter**, Ithaca, v. 27, n. 1, p. 48-54, 1953.

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

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

LEDO, C. A. da SILVA. **Análise de variância multivariada para cruzamentos dialélicos**. 2002. 79 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras, Lavras, MG.

LEE, M.; GOLDSHALK, E. B.; LAMKEY, K. R.; WOODMAN, W. W. Association of restriction fragment length polymorphisms among maize inbred with agronomic performance of their crosses. **Crop Science**, Madison, v. 29, p. 1067-1071, June/Aug. 1989.



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

MACHADO, C. de F. **Procedimentos para a escolha de genitores de feijão**. 1999. 118 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras, Lavras, MG.

MEDONÇA, H. A. de. **Escolha de populações segregantes de feijoeiro utilizando parâmetros genéticos fenotípicos e marcadores RAPD**. 2001. 100 p. Tese (Doutorado em Genética e Melhoramento de Plantas) Universidade Federal de Lavras, Lavras, MG.

MIRANDA FILHO, J. B de.; GERALDI, I. O. An adapted model for the analysis of partial diallel crosses. **Revista Brasileira de Genética**, Ribeirão Preto, v. 7, n. 4, p. 677-688, dez. 1984.

MIRANDA FILHO, J. B de.; VENCOSKY, R. The partial circulant diallel cross at the interpopulation level. **Genetics and Molecular Biology**, Ribeirão Preto, v. 22, n. 2, p. 249-255, June 1999.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. Guildford: Biddles, 1996. 187 p.

OLIVEIRA JÚNIOR, A. **Metodologias de dialelos circulante e de meia-tabela desbalanceados e correlações inter e intrapopulacionais: exemplo com a cultura do feijão (*Phaseolus vulgaris* L.)**. 1995. 136 p. Dissertação (Mestrado) - Universidade Federal de Viçosa, Viçosa, MG.

ORTIZ, R.; MADSEN, S.; WAGOIRE, W. W. et al. Additive main effect and multiplicative interaction model for diallel- cross analysis. **Theoretical and applied genetics**, Berlin, v. 102, n. 6/7, p. 1103-1106, May 2001.

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

PIEPHO, H. P. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison for additive main effects and multiplicative interaction (AMMI) analysis. **Theoretical and applied genetics**, Berlin, v. 89, n. 5, p. 647-654, Nov. 1994.

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

ROHLF, F. J. **Numerical taxonomy and multivariate analysis system**. New York, 1992. 470 p.

SANTOS, J. B. dos.; VENCOVSKY, R.; RAMALHO, M. A. P. Controle genético da produção de grãos e seus componentes primários em feijoeiro. **Pesquisa Agropecuária Brasileira**, Brasília, v. 20, n. 7, p. 1203-1211, jul. 1985.

SEARLE, S. R. **Linear models**. New York: John Wiley, 1971. 532 p.

SINGH, M.; SINGH, R. K. A comparison of different methods of half-diallel analysis. **Theoretical and Applied Genetics**, Berlin, v.67, n. 4, p. 323-326, 1984.

SOH, A. C. Ranking parents by best linear unbiased prediction (BLUP) breeding values in oil palm. **Euphytica**, Wageningen, v. 76, n. 1/2, p. 13-21, 1994.

SOUZA, V. A. B de.; BYRNE, D. H.; TAYLOR, J. F. Heritability, genetic and phenotypic correlations, and predicted selection response of quantitative traits in peach: I. An analysis of several fruit traits. **Journal of American Society Horticultural Science**, Alexandria, v. 123, n. 4, p. 598-603, June 1998a.

SOUZA, V. A. B de.; BYRNE, D. H.; TAYLOR, J. F. Heritability, genetic and phenotypic correlations, and predicted selection response of quantitative traits in peach: II. An analysis of several reproductive traits. **Journal of American Society Horticultural Science**, Alexandria, v. 123, n. 4, p. 604-611, June 1998b.

SOUZA, V. A. B de.; BYRNE, D. H.; TAYLOR, J. F. Predicted breeding values for nine plant and fruit characteristics of 28 peach genotypes. **Journal of American Society Horticultural Science**, Alexandria, v. 125, n. 4, p. 460-465, June 2000.

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.

VAN EEUWIJK, F. A.; DENIS, J. B.; KANG, M. S. Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In: KANG, M. S.; GAUCH Jr., H. G. (Ed.). **Genotype-by-environment interaction**. Boca Raton, Florida: CRC Press, 1996. 416 p.

VEIGA, R. D. **Eficiência de dialelos circulantes na escolha de genitores, avaliada com simulação de dados**. 1998. 95 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1982. 486 p.