



JHONATHAN PEDROSO RIGAL DOS SANTOS

**INCLUSÃO DE EFEITOS DE DOMINÂNCIA NO
MODELO GBLUP MULTIVARIADO PARA
PREDIÇÃO DE HÍBRIDOS SIMPLES
DE MILHO**

**LAVRAS-MG
2015**

JHONATHAN PEDROSO RIGAL DOS SANTOS

**INCLUSÃO DE EFEITOS DE DOMINÂNCIA NO MODELO GBLUP
MULTIVARIADO PARA PREDIÇÃO DE HÍBRIDOS SIMPLES
DE MILHO**

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

Orientador
Dr. Renzo Garcia Von Pinho

Coorientador
Dr. Marcio Balestre

**LAVRAS- MG
2015**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo (a) próprio (a) autor (a).

dos Santos, Jhonathan Pedroso Rigal.

Inclusão de efeitos de dominância no modelo GBLUP
multivariado para predição de híbridos simples de milho /
Jhonathan Pedroso Rigal dos Santos. – Lavras: UFLA, 2015.
91 p. : il.

Dissertação (mestrado acadêmico) – Universidade Federal de
Lavras, 2015.

Orientador (a): Renzo Garcia Von Pinho.

Bibliografia.

1. Modelo GBLUP. 2. Análise Univariada. 3. Análise
Multivariada. 4. Efeitos não aditivos. I. Universidade Federal de
Lavras. II. Título.

JHONATHAN PEDROSO RIGAL DOS SANTOS

**INCLUSÃO DE EFEITOS DE DOMINÂNCIA NO MODELO GBLUP
MULTIVARIADO PARA PREDIÇÃO DE HÍBRIDOS SIMPLES
DE MILHO**

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

APROVADA em 01 de abril de 2015.

Dr. Renzo Garcia Von Pinho	UFLA
Dr. Marcio Balestre	UFLA
Dr. Roberto Fritsche-Neto	ESALQ

Dr. Renzo Garcia Von Pinho
Orientador

Dr. Marcio Balestre
Coorientador

**LAVRAS - MG
2015**

AGRADECIMENTOS

A Deus, pela saúde, disposição e paciência concedida para a realização deste trabalho.

Aos meus Pais, Maria Elenir Rigal dos Santos e Álvaro Cesar dos Santos pelo dom da vida, amor incondicional e todo apoio. Ao meu irmão Gustavo Eugênio Rigal dos Santos pela amizade, incentivo e apoio. A minha namorada Thaise Caroline Moço Zerbini por todo carinho, paciência e amor.

A Universidade Federal de Lavras, ao programa de pós-graduação em Genética e Melhoramento de Plantas e a FAPEMIG pela oportunidade e financiamento à execução desse trabalho.

Ao professor Renzo Garcia Von Pinho, pela orientação, conselhos, paciência, amizade e ensinamentos transmitidos ao longo do curso.

Ao professor Marcio Balestre, pela amizade, orientação, dedicação, paciência e longas horas de escritório dedicadas ao meu crescimento profissional, técnico e pessoal.

Aos meus orientadores de ontem e hoje Gilberto Pozar, Justino Mario, Adriana Gonela, Maria Celeste Gonçalves Vidigal e Pedro Soares Vidigal Filho, pelos ensinamentos transmitidos.

Aos amigos Luiz Paulo Miranda Pires e Renato Coelho de Castro Vasconcellos pelas contribuições acadêmicas e pessoais.

A todos que contribuíram direta ou indiretamente para a realização desse trabalho.

RESUMO GERAL

Novas propostas de modelos e aplicações de processos de predição com informações de marcadores moleculares podem contribuir na redução de gastos financeiros e identificações de genótipos superiores em programas de melhoramento de milho. Estudos avaliando modelos GBLUP com a inclusão de efeitos de dominância não foram realizados no contexto univariado e multivariado na análise de dados dessa cultura. Assim, neste estudo foi conduzido um procedimento de construção de híbridos simples com dados fenotípicos e de marcadores moleculares reais de 4.091 linhagens de milho do banco de dados público Panzea. Nesse processo, obteve-se 400 híbridos simples que foram analisados com o modelo GBLUP univariado e multivariado, considerando apenas efeitos aditivos, e em outra configuração com a inclusão de efeitos de dominância. Utilizou-se cenários de herdabilidade histórica de cinco caracteres e em outras condições de arquitetura genética para investigação comparativa entre os modelos, avaliando a capacidade preditiva e a de decomposição de componentes de variância. Diferenças expressivas não foram detectadas entre os modelos multivariados e univariados. As principais justificativas para essa pequena discrepância entre os modelos são as correlações de magnitude baixa à moderada entre os caracteres estudados e as moderadas herdabilidades observadas. Essa condição não favorece as vantagens da análise multivariada. A inclusão de efeitos de dominância nos modelos mostrou ser uma estratégia eficiente para melhorar a capacidade preditiva e a qualidade da decomposição dos componentes de variância.

Palavras-chave: Modelo GBLUP. Análise Univariada. Análise Multivariada. Efeitos não aditivos.

GENERAL ABSTRACT

New proposals of models and application of prediction processes using molecular markers information can contribute to the reduction of financial expenses and identification of superior genotypes in maize breeding programs. Studies evaluating GBLUP model with the inclusion of dominance effects were not made in the univariate and multivariate context in the analysis of maize data. In this study, it was conducted a construction procedure of simple hybrids with phenotypic data and real molecular markers of 4091 maize lines of the public database Panzea. In this process, 400 simple hybrids were obtained and then analyzed using univariate and multivariate GBLUP model considering only additive effects, and in another configuration with the inclusion of dominance effects. Historical heritability scenario of five maize characters and in other conditions of genetic architecture for comparative research among the models, assessing their predictive capacity and decomposition of variance components. Significant differences were not detected between the multivariate and univariate models. The main reason for this small discrepancy between the models is the low to moderate magnitude correlations among the characters studied and the moderate heritability observed. This condition does not favor the advantages of multivariate analysis. The inclusion of dominance effects in models is an efficient strategy to improve the predictive capacity and the quality of the variance components decomposition.

Keywords: GBLUP Model. Univariate analysis. Multivariate Analysis.
Non-additive effects.

LISTA DE TABELAS

SEGUNDA PARTE - ARTIGO	
Tabela 1	Contagens de coincidência de QTLs identificados no processo de ordenamento dos 100 maiores efeitos aditivos (a) e dos 100 maiores efeitos de dominância (d) em relação aos efeitos genéticos totais obtidos na análise com o modelo BayesB entre os cinco caracteres..... 47
Tabela 2	Grupos heteróticos de linhagens definidas por análise de componentes principais..... 48
Tabela 3	Herdabilidades médias históricas de cinco caracteres de milho..... 50
Tabela 4	Valores de correlações genéticas paramétricas (P) e estimados entre cinco caracteres pelos modelos GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD em herdabilidades históricas..... 56
Tabela 5	Componentes de variância genética estimados e paramétricos e soma do quadrado dos erros de predição dos efeitos genéticos estimados pelos modelos GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD em herdabilidades históricas..... 58
TABELAS SUPLEMENTARES	
Tabela S1	Correlações aditiva entre caracteres estimadas pelo modelo GBLUP-MV-A..... 80
Tabela S2	Correlações aditiva, dominante e genética total entre caracteres estimada pelo GBLUP-MV-AD..... 81
Tabela S3	Componentes de variância paramétricos dos valores genéticos aditivos, dominância e genético total dos híbridos simples..... 82
Tabela S4	Componentes de variância genética e soma do quadrado dos erros de predição dos efeitos genéticos estimados pelo modelo GBLUP-UV-A, GBLUP-UV-AD em faixas de herdabilidade (em parêntesis) fixas (em parêntesis) em 0,3; 0,5 e 0,7 para os cinco caracteres, respectivamente..... 83
Tabela S5	Componentes de variância genética e soma do quadrado dos erros de predição dos efeitos genéticos estimados pelo modelo GBLUP-MV-A, GBLUP-MV-AD em faixas de herdabilidade (em parêntesis) fixas em 0,3; 0,5 e 0,7 para os cinco caracteres, respectivamente, e por ordem determinada por sorteio..... 84

LISTA DE FIGURAS

Figura 1	Agrupamento genético de 4.091 linhagens utilizando análise de componentes principais (PCA). As linhagens são pertencentes ao banco de germoplasma do departamento de agricultura dos Estados Unidos (USDA). O eixo X representa o primeiro componente principal (pc1) e o Y ao segundo componente principal (pc2). As cores foram atribuídas de acordo com o grupo genético, em que azul equivale as 19 linhagens próximas da B73 e vermelho as 19 linhagens próximas da Mo17.....	54
Figura 2	Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidades históricas. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG).(A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.....	60
Figura 3	Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade histórica. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG).(A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.....	61
FIGURAS SUPLEMENTARES		
Figura S1	Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,3. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG).(A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.....	86

Figura S2	Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,5. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.....	87
Figura S3	Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,7. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.....	88
Figura S4	Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade sorteada. Modelos utilizados: GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.....	89
Figura S5	Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixa de 0,3. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.....	89

Figura S6	<p>Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixada de 0,5. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.....</p>	90
Figura S7	<p>Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixada de 0,7. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.....</p>	91
Figura S8	<p>Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade sorteada. Modelos utilizados: GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.....</p>	91

SUMÁRIO

	PRIMEIRA PARTE.....	13
1	INTRODUÇÃO GERAL.....	13
2	REFERENCIAL TEÓRICO.....	15
2.1	Predição do desempenho de híbridos de milho baseada em covariâncias entre parentes e distâncias genéticas.....	15
2.2	Modelos lineares mistos.....	17
2.3	Modelos lineares mistos em processos de análise genética.....	19
2.4	Predição do desempenho de híbridos simples de milho utilizando modelos lineares mistos com informação de parentesco.....	22
2.5	Análise genética de componentes de produção de milho.....	26
3	CONCLUSÕES.....	30
	REFERÊNCIAS.....	31
	SEGUNDA PARTE – ARTIGO.....	38
	INCLUSÃO DE EFEITOS DE DOMINÂNCIA NO MODELO GBLUP MULTIVARIADO PARA PREDIÇÃO DE HÍBRIDOS SIMPLES DE MILHO	38
	INTRODUÇÃO.....	41
	MATERIAL E MÉTODOS.....	44
	RESULTADOS.....	54
	DISCUSSÕES.....	63
	CONCLUSÕES.....	70
	REFERÊNCIAS.....	71
	INFORMAÇÕES SUPLEMENTARES.....	75
	TEXTO SUPLEMENTAR.....	78
	TABELAS SUPLEMENTARES.....	80
	FIGURAS SUPLEMENTARES.....	86

PRIMEIRA PARTE

1 INTRODUÇÃO GERAL

Entre as dificuldades dos programas de melhoramento de milho, uma das principais é o teste de todos os híbridos simples sintetizados por todas as combinações de linhagens derivadas de populações em processo de melhoramento. Isso decorre do elevado custo financeiro e as dificuldades operacionais em condições de ensaios com várias repetições e instalados em número excessivo de locais. Nesse contexto, a disponibilidade de técnicas acuradas de predição das melhores combinações, entre linhagens, pode ser um ponto-chave para a maximização do aproveitamento de recursos financeiros em programas de melhoramento.

Recentemente, um grande interesse científico foi despertado com a proposta de novos modelos de regressão, aplicados em procedimentos de seleção e predição de forma integrada em um processo conhecido popularmente por seleção genômica (MEUWISSEN; HAYES; GODDARD, 2001). A principal premissa desse método é cobrir amplamente o genoma com marcadores moleculares, procurando encontrar relações entre os marcadores e variantes fenotípicos de uma população, sendo ou não possível esses marcadores serem pertencentes à regiões genômicas causais (DE LOS CAMPOS et al., 2013; STRANGER; STAHL; RAJ, 2011). Essa técnica permite a predição de grupo de híbridos simples não testados, devido ao parentesco genético capturado por marcadores moleculares, com outro grupo avaliado fenotipicamente em condição de campo.

Vários trabalhos relatam a melhoria na capacidade preditiva utilizando modelos que consideram informações de caracteres secundários de alta herdabilidade, em conjunto com o caráter-alvo (BALESTRE et al., 2013; HENDERSON; QUAAS, 1976; PIEPHO et al., 2008). Esses modelos são

conhecidos como múltiplos caracteres ou multivariados. Praticamente não há relatos sobre uso dessa classe de modelo em trabalhos utilizando espécies cultiváveis anuais (PIEPHO et al., 2008). Relatos de utilização de modelos multivariados que consideram relacionamento de parentesco estimados por marcadores moleculares são menos ainda frequentes (CALUS; VEERKAMP, 2011; JIA; JANNINK, 2012; MALOSETTI et al., 2007).

Assim, os objetivos com este trabalho são: (i) elaborar a construção de híbridos simples de milho com informações fenotípicas e genotípicas reais de linhagens de milho, de modo que seja conservada a arquitetura genética de caracteres; (ii) propor a inclusão do efeito de dominância no modelo GBLUP multivariado (GBLUP-MV-AD); (iii) comparar o modelo GBLUP-MV-AD com os modelos GBLUP multivariado aditivo (GBLUP-MV-A); GBLUP univariado aditivo (GBLUP-UV-A); e GBLUP univariado com a inclusão do efeito de dominância (GBLUP-UV-AD).

2 REFERENCIAL TEÓRICO

2.1 Predição do desempenho de híbridos de milho baseada em covariâncias entre parentes e distâncias genéticas

Entre as dificuldades dos programas de melhoramento, uma das principais é obter todas as combinações entre linhagens elites e testar todos os descendentes na forma de híbridos simples. Isso decorre do elevado custo financeiro e das dificuldades operacionais em testes situados em vários locais. Por exemplo, a combinação entre 50 linhagens de um grupo heterótico e 50 linhagens de outro grupo heterótico permite a obtenção de 1225 combinações de híbridos simples. Assim, a disponibilidade de técnicas acuradas de predição das melhores combinações, entre linhagens, é um ponto-chave quando almeja-se maximizar o aproveitamento de recursos financeiros em programas de melhoramento.

Um dos trabalhos pioneiros na predição de desempenho de híbridos de milho foi o realizado por Jenkins (1934). Nesse trabalho foram propostos quatro métodos de predição de desempenho de híbridos duplos, utilizando dados de desempenho de híbridos simples. O método mais efetivo foi chamado de método B. Nesse método a predição era baseada no desempenho médio de quatro híbridos simples não parentais. Nessa época, a metodologia teve grande aceitação e reduziu significativamente a quantidade de híbridos duplos testados, o que proporcionou marcante aumento na eficiência dos programas de melhoramento da época (HALLAUER et al., 2010).

Após alguns anos, Ebenhart (1964) dessecou os métodos propostos por Jenkins (1934) e sugeriu uma nova metodologia de predição de híbridos duplos, utilizando valores médios de produtividade de híbridos simples e triplos. Nesse novo método proposto, era considerado efeito de epistasia dominante-dominante, ao contrário dos métodos sugeridos por Jenkins (1934), que assumiam ausência de efeitos de epistasia. Cockerham (1967) também propôs uma nova metodologia de

predição de híbridos duplos. Nesta proposta, foi considerado tanto efeitos genéticos quanto ambientais, e também foram utilizados valores de desempenho de híbridos simples nas predições de híbridos duplos.

Com o predomínio de vendas de híbridos simples no mercado de sementes após a década de 60, os trabalhos relacionados à predição de híbridos duplos se desalinham com o interesse comercial dos programas de melhoramento.

A descoberta do DNA e a intensificação de estudos com marcadores moleculares despertaram grande interesse de desenvolvimento de novos processos de predição. Vários estudos foram desenvolvidos explorando as informações de marcadores moleculares em procedimentos preditivos de desempenho de híbridos simples de milho (BALESTRE et al., 2008; LANZA et al., 1997; LEE et al., 1989).

A principal justificativa no uso das marcas nas predições de híbridos simples foi oriunda da própria teoria da genética quantitativa, em que híbridos com locos com maior dominância direcional positiva, obtidos de linhagens com maior divergência genética apresentam maior heterose (FALCONER; MACKAY, 1996). Portanto, acreditava-se que a divergência genética, estimada por marcadores moleculares na forma de distâncias genéticas, poderia predizer as melhores combinações entre as linhagens na síntese dos melhores híbridos simples (LEE et al., 1989).

Entretanto, predições baseadas apenas nas distâncias genéticas não foram eficientes para prever o desempenho de híbridos (MELCHINGER, 1999). De acordo com Reif, Hahne Melchinger (2012) as principais razões foram: (1) pequena associação entre a heterozigosidade dos locos - que controlam caracteres quantitativos - e a heterozigosidade estimada pelas marcas; (2) pobre associação entre heterozigosidade e a heterose; (3) epistasia.

2.2 Modelos lineares mistos

Após a década de 50, várias contribuições científicas foram realizadas por C. R. Henderson ao melhoramento de animais e plantas. A principal contribuição deste pesquisador foi a percepção de que seria conveniente na construção de modelos genéticos tanto a inclusão de efeitos fixos, fatores ambientais (β), quanto aleatórios, valores genéticos (g), amostrados de uma distribuição de probabilidade (HENDERSON, 1948). Neste trabalho foi proposta pela primeira vez a metodologia BLUP – *Best Linear Unbiased Predictors* (HILL, 2014). Após essa proposta inicial, Henderson e outros pesquisadores aprimoraram essa abordagem. Uma obra descritiva deste método de análise pode ser encontrada em Henderson (1984). Porém, para melhor compreensão dessa metodologia, uma descrição geral do BLUP será realizada na sequência.

O BLUP é um procedimento estatístico que prediz efeitos aleatórios em modelos mistos (PIEPHO et al., 2008). Os modelos mistos são modelos estatísticos que apresentam efeitos fixos e aleatórios (SEARLE et al., 2006). De acordo com Henderson (1975), um modelo misto com várias aplicações genéticas, pode ser descrito matricialmente por:

$$y = X\beta + Zg + e \quad (1)$$

em que y é um vetor de observações $n \times 1$, X é uma matriz de incidência dos efeitos fixos $n \times p$, β é vetor de efeitos fixos, Z é uma matriz de incidência dos efeitos aleatórios $n \times q$, g e e são vetores de efeitos aleatórios genéticos e ambientais. As distribuições dos efeitos aleatórios são consideradas como $g \sim MVN(0, G)$ e $e \sim MVN(0, R)$.

As observações são assumidas $\mathbf{y} \sim MVN(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$. A matriz \mathbf{V} é estruturada por:

$$\text{Var}(\mathbf{y}) = \mathbf{ZGZ}' + \mathbf{R} = \mathbf{V} \quad (2)$$

em que: \mathbf{G} é uma matriz de variância-covariância genética, e \mathbf{R} é uma matriz de variância-covariância residual.

Os efeitos fixos são estimados via BLUE – *Best Linear Unbiased Estimators* – e os aleatórios, como supracitado, pelo BLUP. As letras das siglas BLUP e BLUE significam: (1) *Best*: o quadrado médio do erro $E(\hat{\mathbf{g}} - \mathbf{g})^2$ é minimizado, quando dentro da classe dos estimadores não viesados; (2) *Linear*: $\hat{\mathbf{g}}$ são funções lineares de \mathbf{y} em situações de componentes de variância dados como conhecidos; (3) *Unbiased*: $\hat{\mathbf{g}}$ é um estimador não viesado de \mathbf{g} , ou seja, $E(\hat{\mathbf{g}} - \mathbf{g}) = 0$ (THOMPSON, 2008).

De acordo com Henderson (1975) o melhor preditor linear (BLP) de $k' \boldsymbol{\beta}^* + m' \mathbf{g}$, quando $\boldsymbol{\beta}$, \mathbf{G} , e \mathbf{R} são conhecidos, é dado por:

$$k' \boldsymbol{\beta} + m' \mathbf{GZ}' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (3)$$

em que \mathbf{V}^{-1} é a inversa da matriz de variância-covariância fenotípica (\mathbf{V}).

Entretanto, no modelo (1) $\boldsymbol{\beta}$ é desconhecido. Uma alternativa é utilizar o $\hat{\boldsymbol{\beta}}$ estimado pelo método dos quadrados mínimos generalizados:

$$\mathbf{XV}^{-1} \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \quad (4)$$

Com o novo $\hat{\beta}$ obtém-se o BLUP de $k' \beta^* + m'g$:

$$k' \hat{\beta} + m'GZ'V^{-1}(y - X \hat{\beta}) \quad (5)$$

Como V tem uma dimensão muito grande para ser invertida diretamente, é atrativo computacionalmente utilizar equações de modelos mistos (EMM) descritas em Henderson (1975), que são dadas por:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (6)$$

A resolução da EMM é obtida pela inversão direta do sistema de equações (6). Os componentes de variância podem ser estimados via resoluções numéricas, com os componentes das matrizes de variância-covariância R e G obtidos por métodos matemáticos iterativos, que maximizam de forma direta a função de máxima verossimilhança residual ou REML (Residual Maximum Likelihood) de Paterson e Thompson (1971). Por exemplo, o método simplex proposto por Nelder e Mead (1965). Outros algoritmos que podem ser utilizados para a estimação dos componentes de variância são: (i) Expectation-Maximization (EM) (DEMPSTER, 1977); (ii) Average Information (AI) (GILMOUR et al., 1995) e (iii) Newton/Raphson (MRODE; THOMPSON, 2005).

2.3 Modelos lineares mistos em processos de análise genética

A possibilidade de utilização de informações geradas por *pedigrees* complexos em programas de melhoramento ou por marcadores moleculares são características importantes do método BLUP. As informações de parentesco são recuperadas devido às correlações entre os indivíduos de uma população original, relacionados pelo *pedigree*, até a uma população de referência, da qual os componentes de variância são inerentes (PIEPHO et al., 2008). Várias

estimativas úteis em programas de melhoramento, como por exemplo: o coeficiente de herdabilidade, o ganho esperado com a seleção e a correlação entre caracteres, são funções dos componentes de variância (RAMALHO, 2012).

As informações de parentesco são recuperadas utilizando a matriz de parentesco genético (estruturada pelo parentesco médio de Malecot) ou genômico (estimado por marcadores moleculares). Cada elemento da matriz \mathbf{G} pode ser definido como resultado de uma “contribuição gamética” de parentais ao longo das gerações, estruturada pela informação de parentesco. A variância genética total teoricamente é dada por:

$$\hat{\sigma}_g^2 = \hat{\sigma}_u^2 + \hat{\sigma}_\delta^2 + \hat{\sigma}_{uu}^2 + \dots \quad (10)$$

De acordo com Henderson (1985), o valor genotípico ($\hat{\mathbf{g}}$) e a matriz \mathbf{G} podem ser decompostos em:

$$\hat{\mathbf{g}} = \hat{\mathbf{u}} + \hat{\boldsymbol{\delta}} + \mathbf{u}\mathbf{u} + \dots \quad (10)$$

$$\mathbf{G} = \sum_1^i \mathbf{G}_i = \mathbf{G}_u + \mathbf{G}_\delta + \mathbf{G}_{uu} + \dots = \mathbf{A}\hat{\sigma}_u^2 + \mathbf{D}\hat{\sigma}_\delta^2 + \mathbf{A}\#\mathbf{A}\hat{\sigma}_{uu}^2 + \dots \quad (11)$$

O valor $\hat{\sigma}_u^2$ é a variância genética aditiva, que corresponde à variância dos valores genéticos ou *breeding values* ($\hat{\mathbf{u}}$), $\hat{\sigma}_\delta^2$ é a variância genética de dominância, à variância relacionada aos desvios de dominância ($\hat{\boldsymbol{\delta}}$), $\hat{\sigma}_{uu}^2$ é a variância genética aditiva-aditiva epistática, à variância relacionada aos efeitos de desvio aditivo-aditivo epistático ($\mathbf{u}\mathbf{u}$), que é um exemplo possível de efeito de epistasia. Todas essas variâncias são estimativas de parâmetros inerentes a uma população de referência, de onde os indivíduos foram derivados.

O valor \hat{u} é definido como o valor de um indivíduo, julgado pelo valor médio de sua progênie. O valor $\hat{\delta}$ é definido como o efeito oriundo da interação entre alelos, ou seja, o valor resultante de dois alelos juntos, especificamente, na constituição de um dado genótipo. Por fim, o valor de epistasia (no caso do exemplo do tipo $\square uu$) pode ser definido como o desvio adicional devido à combinação de dois ou mais genes, ou seja, a falta de aditividade dos valores genotípicos de dois genes diferentes (FALCONER; MACKAY, 1996).

A matriz G_u equivale a matriz de variância-covariância genética aditiva, a G_δ a matriz de variância-covariância genética de dominância. Por fim, a matriz A , a matriz de parentesco aditiva e a D , a matriz de parentesco de dominância, ambas estimadas com informação de *pedigree*. A matriz A informa nas EMM os parentescos relacionados aos valores aditivos dos indivíduos. Cada elemento não diagonal de A corresponde a duas vezes o coeficiente de parentesco (f), que pode ser definido como a probabilidade de um alelo tomado ao acaso de um indivíduo, seja idêntico por descendência a outro alelo, do mesmo loco, tomado ao acaso de um segundo indivíduo. Em relação aos elementos da diagonal de A , cada elemento tem equivalência à $(1 + F)$, em que F corresponde ao coeficiente de endogamia, que é definido como a probabilidade de dois alelos de um mesmo loco tomados ao acaso de um indivíduo sejam idênticos por descendência (SCHAEFFER, 2010). Sobre a matriz D , cada elemento dessa matriz representa a probabilidade de que dois indivíduos parentes apresentem o mesmo par de alelos idênticos por descendência (MRODE; THOMPSON, 2005). Por fim, a matriz $A \# A$, em que $(\#)$ equivale ao produto matricial de Hadamard, é a matriz que relaciona os parentescos dos indivíduos em relação ao efeito $\square uu$ (HENDERSON, 1985; MRODE; THOMPSON, 2005).

As mesmas equações de modelos mistos (EMM) considerando a decomposição de G e g sugerido por Henderson (1985) podem ser observadas abaixo:

$$\begin{bmatrix} X'X & X'Z & X'Z & X'Z & \dots \\ Z'X & Z'Z + G_u^{-1} \hat{\sigma}_e^2 & Z'Z & Z'Z & \dots \\ Z'X & Z'Z & Z'Z + G_\delta^{-1} \hat{\sigma}_e^2 & Z'Z & \dots \\ Z'X & Z'Z & Z'Z & Z'Z + G_{uu}^{-1} \hat{\sigma}_e^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \\ \hat{\delta} \\ \hat{uu} \\ \vdots \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \\ Z'y \\ \vdots \end{bmatrix} \quad (12)$$

Por fim, os componentes de variância podem ser estimados via EM-REML por:

$$\hat{\sigma}_e^2 = (e'e - \text{tr}(WC^{-1}W')) / n$$

$$\hat{\sigma}_u^2 = (\hat{u}'A^{-1}\hat{u} + \text{tr}(C_{22}^{-1})\hat{\sigma}_e^2) / q_u$$

$$\hat{\sigma}_\delta^2 = (\hat{\delta}'D^{-1}\hat{\delta} + \text{tr}(C_{33}^{-1})\hat{\sigma}_e^2) / q_\delta$$

$$\hat{\sigma}_{uu}^2 = (\hat{uu}'(A\#A)^{-1}\hat{uu} + \text{tr}(C_{44}^{-1})\hat{\sigma}_e^2) / q_{uu}$$

em que q_u é o número de estimativas de \hat{u} ; q_δ é o número de estimativas de $\hat{\delta}$;

q_{uu} é o número de estimativas \hat{uu} ; $e = y - X\hat{\beta} - Z\hat{g}$, o resíduo; C é a resolução do sistema de equações (12); e $W = [X | Z]$ é a matriz X concatenada com a matriz Z .

2.4 Predição do desempenho de híbridos simples de milho utilizando modelos lineares mistos com informação de parentesco

Após o insucesso dos métodos de predição de desempenho híbridos de milho baseados em distâncias genéticas, estimadas por marcadores moleculares, Bernardo (1994) propôs uma nova metodologia de predição utilizando modelos

lineares mistos. Esse novo método era baseado no modelo genético sugerido por Henderson (1985), porém com algumas particularidades. Entre as diferenças, um ponto importante entre o modelo de Bernardo (1994) e Henderson (1985) era o processo de construção das matrizes de parentesco. A matriz A de Henderson (1985) era gerada por informações de *pedigree*, já a de Bernardo (1994) era estimada por informações oriundas de marcadores moleculares. No modelo proposto por Henderson (1985) eram estimados os efeitos genéticos aditivos, de dominância e desvios epistáticos. Já no sugerido por Bernardo (1994), as estimativas obtidas eram a capacidade geral de combinação (CGC) e capacidade específica de combinação (CEC) de linhagens de milho. Em outro trabalho, Bernardo (1994) realizou a validação do método com um maior volume de dados. Este autor conseguiu encontrar correlações de ordem de 0,5 a 0,81 entre valores preditos e observados, para o caráter produtividade de híbridos simples de milho.

No começo do século XXI, Meuwissen, Hayes e Goddard (2001) propuseram um novo método preditivo conhecido na atualidade como Seleção Genômica Ampla (SGA) – do termo em inglês *Genome-Wide Selection*. O método era baseado na regressão de valores fenotípicos, utilizando marcadores moleculares (uma marca a cada 1 cM) espalhados densamente pelo genoma como covariáveis em uma função de regressão (DE LOS CAMPOS et al., 2013). Meuwissen, Hayes e Goddard (2001) utilizaram quatro métodos de análise genômica em um estudo de simulação, os métodos eram: (i) métodos dos quadrados mínimos (MQM); (ii) BLUP; (iii) Bayes A; (iv) Bayes B. Meuwissen, Hayes e Goddard (2001) conseguiram encontrar correlações entre os valores estimados e observados de 0,318 (MQM), 0,732 (BLUP), 0,798 (Bayes A) e 0,848 (Bayes B) na simulação executada no trabalho. Detalhe sobre particularidades dos métodos podem ser encontrados no trabalho original. Uma crítica detalhada sobre esses métodos em Gianola (2013) e uma revisão recente em de los Campos et al. (2013). Resultados e comentários de estudos práticos e

de simulação dessa metodologia em grandes culturas podem ser encontrados em Jannink, Lorenz e Iwata (2010) e em específico na cultura do milho em Albrecht et al. (2011) e Habier, Fernando e Garrick (2013).

Os modelos bayesianos BayesA e BayesB utilizados no trabalho de Meuwissen, Hayes e Goddard (2001) permitem a incorporação de informações *a priori*, modelagem de heterogeneidade da variância de locos e a seleção e encolhimento de marcas de pequeno efeito, atribuindo, desta forma, o relaxamento de fortes pressupostos e maior versatilidade de modelagem. A principal desvantagem dos modelos bayesianos de variância específica relaciona-se a sua deficiência em descrever a variância genética de forma efetiva e devido à baixa entropia presente nas distribuições *prioris*, pouco se aprende com os dados observados (GIANOLA, 2013).

Após a proposta de Meuwissen, Hayes e Goddard (2001), especialmente em relação ao uso do método BLUP, é muito frequente em trabalhos com SGA a utilização de um modelo conhecido por rBLUP (*Ridge Regression BLUP*). Neste modelo são estimados efeitos genéticos de marcadores moleculares espalhados no genoma de um grupo de indivíduos. A soma dos efeitos das marcas de cada indivíduo recupera o valor genético de cada indivíduo. A principal desvantagem desse método é a realização de uma regressão direta nas marcas (identidade por estado), sem considerar uma estrutura de parentesco comum entre os indivíduos.

Uma alternativa para melhorar esse modelo foi considerar um relacionamento genético entre os indivíduos, via identidade por descendência, assumindo uma população base de passado comum. Nesse sentido, Powell, Visscher e Goddard (2010) sugeriram um processo de estimação da matriz A por marcadores moleculares. A proposta encontra-se em harmonia com a teoria da coalescência, que estabelece que todos os indivíduos compartilham um ancestral

comum. O método é baseado em considerar a população “atual” como a “base”. O estimador de Powell, Visscher e Goddard (2010) é dado por:

$$A = \begin{cases} \frac{1}{N} \sum_k \frac{(x_{ik} - 2p_k)(x_{jk} - 2p_k)}{2p_k q_k}, i \neq j \\ 1 + \frac{1}{N} \sum_k \frac{x_{ik}^2 - (1 + 2p_k)x_{ik} + 2p_k^2}{2p_k q_k}, i = j \end{cases} \quad (13)$$

em que A é a matriz de parentesco unificada de Powell, x é a soma de alelos (marcas) de um dado loco de um indivíduo i ou j em um loco k , com cada alelo favorável codificado por 1 e cada desfavorável por 0, p é a frequência do alelo favorável e q frequência do alelo desfavorável, N é o número de locos.

Um ajuste empírico, visando eliminar o viés e torná-lo um estimador não viesado do parentesco entre indivíduos, de acordo com Powell, Visscher e Goddard (2010), pode ser encontrado em Yang et al. (2010). Porém, em algumas situações, é possível que a matriz A não seja positiva semidefinida (DE LOS CAMPOS et al., 2013). Nessa situação, Meuwissen, Luan e Woolliams(2011) recomendam não calcular a diagonal separadamente, como proposto originalmente por Powell, Visscher e Goddard (2010), e adicionar uma constante 10^{-4} a matriz A .

Assim, com a utilização da matriz A estimada pelos marcadores é possível estimar o valor \hat{u} e a $\hat{\sigma}_u^2$ por um novo método chamado *Genomic BLUP* ou *GBLUP* (VANRADEN, 2008). Em trabalhos envolvendo SG é comum observar-se a utilização da matriz de parentesco A proposta por VanRaden (2008). Foi exatamente nesse trabalho em que foi proposto pela primeira vez o modelo GBLUP. O modelo GBLUP, estruturalmente, é o mesmo modelo usado por muito tempo no melhoramento animal, conhecido popularmente como modelo animal, porém, com a estrutura de variância-

covariância genética (G) entre os indivíduos, estimada por marcadores moleculares. A principal limitação da matriz de VanRaden (2008) é que a parametrização da matriz de desvios não é realizada loco a loco ou em *pairwise*. Também, conforme salientado por Meuwissen, Luan e Woolliams (2011), o uso dessa matriz com dados reais acaba subestimando a participação de alelos em baixa frequência, que na maioria das vezes refletem com intensidade na expressão fenotípica. Por exemplo, genes que conferem resistência à doenças de herança monogênica.

Por fim, em um estudo prático, Massman et al. (2013) compararam o método preditivo BLUP proposto por Bernardo (1994) e o método rrBLUP sugerido por Meuwissen, Hayes e Goddard (2001). Nesse trabalho, o autor concluiu que o método BLUP e rrBLUP se equivalem quando não existe desequilíbrio de ligação (LD) entre as marcas - o que é esperado na teoria - e que em condições de maior LD o rrBLUP tende a tornar-se um método preditivo mais eficiente. O maior valor de acurácia média entre os valores genéticos preditos e observados no processo de validação cruzada foi obtido quando as informações relativas aos dois parentais dos híbridos não testados eram utilizadas na análise.

2.5 Análise genética de componentes de produção de milho

O aumento da produtividade de grãos é o principal objetivo de programas de melhoramento de milho, seja de forma direta ou indireta (HALLAUER et al., 2010). Por exemplo, mesmo quando é direcionado alguns cruzamentos visando ao aumento da resistência a doenças, indiretamente, aumenta-se o desempenho produtivo em híbridos.

Sobre a arquitetura genética do caráter produção de grãos, as principais características são: (i) baixa herdabilidade; (ii) controlado por vários genes de pequeno efeito; (iii) poucos genes de grande efeito; (iv) genes de efeito

epistático; (v) e genes com efeito pleiotrópico (ALESTRE et al., 2012; BERNARDO, 2009; HALLAUER et al., 2010; MA et al., 2007; PATERNIANI; AJUSTAR, 1999).

Devido à alta complexidade genética da produção de grãos, vários estudos foram realizados buscando estratégias de maximização do ganho com a seleção de genótipos superiores na expressão deste caráter. Uma das estratégias foi a seleção indireta de caracteres secundários de alta herdabilidade e correlacionados com a produção de grãos em milho (BENTO; RAMALHO; SOUZA, 2003; CARENA; SANTIAGO, 1998; LONNQUIST, 1967; SINGH; KHERRA; DHILLON, 1986). Uma tabela resumindo o resultado médio de 390 estudos, com parâmetros de correlação genética entre caracteres em nível de planta e em nível de espiga, em relação à produção de grãos na cultura do milho, pode ser encontrada em Hallauer et al. (2010).

De acordo com Bernardo (2010), o que se observou após vários anos de pesquisa com a seleção indireta avaliando componentes de produção, não só para o milho, mas também para outras culturas, é que a seleção direta é mais efetiva no aumento de produtividade de grãos. Falconer e Mackay (1996) afirmam que a principal justificativa é que à medida em que é praticada, a seleção indireta, e aumenta-se favoravelmente a expressão de dois caracteres, ao mesmo tempo, outros caracteres, que também são componentes de produção, e que são negativamente correlacionados aos que foram selecionados, acabam reduzindo a produtividade média final na maioria das vezes após o ciclo de melhoramento.

Entretanto, vários trabalhos relatam a melhoria na capacidade preditiva utilizando modelos que consideram informações de caracteres secundários de alta herdabilidade, em conjunto com dados de produção de grãos (HENDERSON; QUAAS, 1976; MALOSETTI et al., 2007; MRODE; THOMPSON, 2005; PIEPHO et al., 2008). Esses modelos são conhecidos por

“multipletraits” ou múltiplos caracteres ou multivariados. Praticamente não há relatos sobre uso dessa classe de modelo em trabalhos utilizando espécies cultiváveis anuais (MALOSETTI et al., 2007; PIEPHO et al., 2008).

Malosetti (2007) propôs um modelo com o uso de informações de vários caracteres em ambientes diversos. O objetivo principal desse trabalho foi identificar QTLs que explicavam a variação genética de componentes de produção da cultura do milho. Esse autor analisou e identificou QTLs com efeitos pleiotrópicos, ligados e também conseguiu quantificar efeitos de interação QTLs x ambientes. Por fim, foi concluído no final do trabalho que informações obtidas por ensaios em vários locais podem ser úteis no entendimento da arquitetura genética de caracteres de importância econômica.

Balestre (2013) também realizou uma proposta de um modelo de múltiplos caracteres, porém, adicionalmente, comparou este modelo com outros modelos considerando os caracteres separadamente. O principal objetivo do trabalho foi verificar se os modelos de múltiplos caracteres proporcionavam estimativas com maior precisão, do que os modelos considerando cada caráter separadamente. As análises foram executadas tanto na abordagem de modelos mistos quanto na bayesiana. Este autor concluiu que quando não existem informações *a priori* sobre os parâmetros, tanto os métodos bayesianos quanto os modelos mistos se equivalem, o que, conforme afirmado pelo próprio autor, é esperado na teoria. Em relação aos modelos de múltiplos caracteres, relatou que, em geral, quando são utilizados dados fenotípicos de componentes de produção, em conjunto com dados de produtividade, a análise é mais eficiente do que quando se analisa cada caráter em separado, principalmente, quando os componentes de produção apresentam alta correlação entre si. Outro ponto mencionado, é que no geral, a análise é mais eficiente quando são utilizados componentes com herdabilidade mediana à baixa.

Um estudo comparativo entre o método de seleção indireta com caracteres secundários, com o método da SGA, especialmente para tolerância à seca, foi conduzido por Ziyomo e Bernardo (2013). Neste trabalho, foi estimado com resultados fenotípicos de ensaio de campo as variâncias dos caracteres secundários, as covariâncias e as correlações entre os caracteres usando as duas metodologias. A principal conclusão desses autores foi que, no geral, a SGA é 24% mais eficiente do que a seleção indireta baseada apenas em dados fenotípicos. Também relataram que praticando três ciclos de seleção (em 1 ano) com a SGA, a eficiência predita é cerca de 3,7 vezes mais alta, em termos de ganho genético, do que praticar um ciclo de avaliação fenotípica indireta por resultados de *test crosses* (em 2 anos). Porém, alertam a possibilidade de superestimação, baseado em resultados de outros estudos práticos (MASSMAN et al., 2013). Nesse último estudo, não foi utilizado um modelo multivariado, seja no uso do método de seleção indireta fenotípica, ou na utilização do método de SGA.

3 CONCLUSÕES

Processos de predição de híbridos simples apresentam alto potencial para promover a redução de gastos financeiros em programas de melhoramento genético de milho.

Modelos mistos apresentam propriedades de grande relevância para execução de predições. As pressuposições de normalidade, propriedades estatísticas e versatilidade de modelagem o torna atrativo.

Estimações de matrizes de parentesco com informações de marcadores moleculares são alternativas eficientes para capturar relacionamentos genéticos entre indivíduos.

O modelo GBLUP multivariado pode recuperar informações genéticas aditivas e não aditivas entre caracteres e indivíduos. Essa característica pode torná-lo mais eficiente do que os modelos atuais.

REFERÊNCIAS

ALBRECHT, T. et al. Genome-based prediction of testcross values in maize. **Theoretical and Applied Genetics**, Berlin, v. 123, n. 2, p. 339–350, July 2011.

BALESTRE, M. et al. Applications of multi-trait selection in common bean using real and simulated experiments. **Euphytica**, Wageningen, v. 189, n. 2, p. 225–238, Sept. 2013.

BALESTRE, M. et al. Bayesian mapping of multiple traits in maize: the importance of pleiotropic effects in studying the inheritance of quantitative traits. **Theoretical and Applied Genetics**, Berlin, v. 125, n. 3, p. 479–493, Aug. 2012.

BALESTRE, M. et al. Genetic distance estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids. **Genetics and Molecular Research**, Ribeirão Preto, v. 7, n. 1, p. 65–73, 2008.

BENTO, D. A. V.; RAMALHO, M. A. P.; SOUZA, J. C. Seleção massal para prolificidade em milho na época normal e na “safrinha”. **Revista Brasileira de Milho e Sorgo**, Sete Lagoas, v. 2, n. 3, p. 78–87, 2003.

BERNARDO, R. **Breeding for quantitative traits in plants**. 2nd ed. Woodbury: Stemma, 2010. 390 p.

BERNARDO, R. Genomewide Selection for Rapid Introgression of Exotic Germplasm in Maize. **Crop Science**, Madison, v. 49, n. 2, p. 419, 2009.

BERNARDO, R. Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. **Crop Science**, Madison, v. 34, n. 1, p. 20–25, 1994.

CALUS, M. P. L.; VEERKAMP, R. F. Accuracy of multi-trait genomic selection using different methods. **Genetics, Selection, Evolution**, Paris, v. 43, p. 26, Jan. 2011.

CARENA, M. J. I.; SANTIAGO, A. O. Direct and correlated response to recurrent selection for prolificacy in maize at two planting densities. **Maydica**, Bergamo, v. 43, p. 95–102, 1998.

COCKERHAM, C. C. Prediction of double crosses from single crosses. **Theoretical and Applied Genetics**, Berlin, v. 37, n. 4, p. 160–169, 1967.

DE LOS CAMPOS, G. et al. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, Austin, v. 193, n. 2, p. 327–345, Feb. 2013.

DEMPSTER, A. P. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society**, London, v. 39, n. 1, p. 1–38, 1977.

EBERHART, S. A. **Theoretical relations among single, three-way, and doubled cross hybrids**. 1964. Disponível em: <<http://www.jstor.org/>>. Acesso em: 5 mar. 2014.

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

GIANOLA, D. Priors in whole-genome regression: the bayesian alphabet returns. **Genetics**, Austin, v. 194, n. 3, p. 573–596, July 2013.

GILMOUR, A. R. et al. Linear mixed models algorithm for average information REML: an efficient in linear mixed models variance parameter estimation. **Biometrics**, Washington, v. 51, n. 4, p. 1440–1450, 1995.

HABIER, D.; FERNANDO, R. L.; GARRICK, D. J. Genomic BLUP decoded: a look into the black box of genomic prediction. **Genetics**, Austin, v. 194, n. 3, p. 597–607, July 2013.

HALLAUER, A.R. et al. **Quantitative genetics in maize breeding**. 2nd ed. Ames: Iowa State University, 2010. 468p.

HENDERSON, C. R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 384p.

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

HENDERSON, C. R. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. **Journal of Animal Science**, Champaign, v. 60, n. 1, p. 111–117, 1985.

HENDERSON, C. R. **Estimation of general, specific and maternal combining abilities in crosses among inbred lines of swine**. 1948. Thesis (PhD) - Iowa State University, Ames, 1948.

HENDERSON, C. R.; QUAAS, R. L. Multiple trait evaluation using relatives' records. **Journal of Animal Science**, Champaign, v. 43, n. 6, p. 1188–1197, 1976.

HILL, W. G. Applications of population genetics to animal breeding, from wright, fisher and lush to genomic prediction. **Genetics**, Austin, v. 196, n. 1, p. 1–16, Jan. 2014.

JANNINK, J. L.; LORENZ, A. J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**, London, v. 9, n. 2, p. 166–177, Mar. 2010.

JENKINS, M. T. Methods of estimating the performance of double crosses in corn. **Journal American Society of Agronomy**, Madison, v. 26, p. 199–204, 1934.

JIA, Y.; JANNINK, J. L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. **Genetics**, Austin, v. 192, n. 4, p. 1513–22, Dec. 2012.

LANZA, L. L. B. et al. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. **Theoretical and Applied Genetics**, Berlin, v. 94, n. 8, p. 1023–1030, 27 Juny 1997.

LEE, M. Association of restriction length polymorphism among maize inbreds with agronomic performance of their crosses. **Crop Science**, Madison, v. 29, p. 1067–1071, 1989.

LEE, J. H. Mass selection for prolificacy in corn. **Der Züchter**, Berlin, v. 37, p. 185–188, 1967.

MALOSETTI, M. et al. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). **Euphytica**, Wageningen, v. 161, n. 1/2, p. 241–257, Dec. 2007.

MASSMAN, J. M. et al. Genomewide predictions from maize single-cross data. **Theoretical and applied genetics**, Berlin, v. 126, n. 1, p. 13–22, Jan. 2013.

MA, X. Q. et al. Epistatic interaction is an important genetic basis of grain yield and its components in maize. **Molecular Breeding**, Dordrecht, v. 20, n. 1, p. 41–51, Mar. 2007.

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

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Austin, v. 157, n. 4, p. 1819–1829, Apr. 2001.

MEUWISSEN, T. H. E.; LUAN, T.; WOOLLIAMS, J. A. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. **Journal of Animal Breeding and Genetics**, Berlin, v. 128, n. 6, p. 429–439, Dec. 2011.

MRODE, R. A.; THOMPSON, R. **Linear models for the prediction of animal breeding values**. 2nd. ed. Cambridge: CABI, 2005. 344p.

NELDER, J. A.; MEAD, R. A simplex method for function minimization. **Computer Journal**, London, v.7, n.4, p. 308-313, 1965.

PATERNIANI, E.; AJUSTAR, B. Melhoramento do milho. In: BORÉM, A. **Melhoramento de espécies cultivadas**. Viçosa, MG: UFV, 1999. p. 491–547.

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

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, Wageningen, v. 161, n. 1-2, p. 209–228, July 2008.

POWELL, J. E.; VISSCHER, P. M.; GODDARD, M. E. Reconciling the analysis of IBD and IBS in complex trait studies. **Nature reviews. Genetics**, London, v. 11, n. 1, p. 800–805, Nov. 2010.

RAMALHO, M. A. P. **Experimentação em genética e melhoramento de plantas**. 3. ed. Lavras: UFLA. 2012. 328 p.

REIF, J. C.; HAHN, V.; MELCHINGER, A. E. Genetic basis of heterosis and prediction of hybrid performance. **Helia**, Novi Sad, v. 35, n. 57, p. 1–7, 2012.

SCHAEFFER, L. R. **Linear models and animal breeding**. Guelph: University of Guelph, 2010. 314 p.

SEARLE, S. R. et al. **Variance components**. 2nd ed. New Jersey: J. Wiley & Sons, 2006. 501 p.

SINGH, M. A. S.; KHERRA, A. S.; DHILLON, B. S. Direct and correlated response to recurrent full-sib selection for prolificacy in maize. **Crop Science**, Madison, v. 26, p. 275–278, 1986.

STRANGER, B. E.; STAHL, E. A.; RAJ, T. Progress and promise of genome-wide association studies for human complex trait genetics. **Genetics**, Austin, v. 187, n. 2, p. 367–83, Feb. 2011.

THOMPSON, R.; BROTHERSTONE, S.; WHITE, I. M. S. Estimation of quantitative genetic parameters. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, London, v. 360, n. 1459, p. 1469–1477, July 2005.

THOMPSON, R. Estimation of quantitative genetic parameters. **Proceedings. Biological sciences / The Royal Society**, Edinburg, v. 275, n. 1635, p. 679–686, Mar. 2008.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, Champaign, v. 91, n. 11, p. 4414–4423, Nov. 2008.

YANG, J. et al. Common SNPs explain a large proportion of the heritability for human height. **Nature Genetics**, New York, v. 42, n. 7, p. 565–571, July 2010.

ZIYOMO, C.; BERNARDO, R. Drought tolerance in maize: indirect selection through secondary traits versus genomewide selection. **Crop Science**, Madison, v. 53, n. 4, p. 1269–1275, 2013.

SEGUNDA PARTE - ARTIGO

**ARTIGO: INCLUSÃO DE EFEITOS DE DOMINÂNCIA NO MODELO
GBLUP MULTIVARIADO PARA PREDIÇÃO DE HÍBRIDOS SIMPLES
DE MILHO**

Normas da Revista Científica Plos One

**Inclusão de Efeitos de Dominância no Modelo GBLUP Multivariado para
Predição de Híbridos Simples de Milho**

Jhonathan Pedroso Rigal dos Santos^{1¶}, Renato Coelho de Castro Vasconcellos^{1&}, Luiz Paulo Miranda Pires^{3&}, Marcio Balestre^{2¶*}, Renzo Garcia Von Pinho^{3&}

¹ Departamento de biologia (DBI), Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil.

² Departamento de ciências exatas (DEX), Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil.

³ Departamento de agricultura (DAG), Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil.

* Corresponding author

E-mail: marciobalestre@dex.ufla.br (MB)

¶ These authors contributed equally to this work.

& These authors also contributed equally to this work.

RESUMO

Novas propostas de modelos e aplicações de processos de predição com informações de marcadores moleculares podem contribuir na redução de gastos financeiros e identificações de genótipos superiores em programas de melhoramento de milho. Estudos avaliando modelos GBLUP com a inclusão de efeitos de dominância não foram realizados no contexto univariado e multivariado na análise de dados dessa cultura. Assim, neste estudo foi conduzido um procedimento de construção de híbridos simples com dados fenotípicos e de marcadores moleculares reais de 4.091 linhagens de milho do banco de dados público Panzea. Nesse processo, obteve-se 400 híbridos simples que foram analisados com o modelo GBLUP univariado e multivariado considerando apenas efeitos aditivos, e em outra configuração com a inclusão de efeitos de dominância. Utilizou-se cenários de herdabilidade histórica de cinco caracteres e em outras condições de arquitetura genética para investigação comparativa entre os modelos, avaliando a capacidade preditiva e a de decomposição de componentes de variância. Diferenças expressivas não foram detectadas entre os modelos multivariados e univariados. As principais justificativas para essa pequena discrepância entre os modelos são as correlações de magnitude baixa à moderada entre os caracteres estudados e as moderadas herdabilidades observadas. Essa condição não favorece as vantagens da análise multivariada. A inclusão de efeitos de dominância nos modelos mostrou ser uma estratégia eficiente para melhorar a capacidade preditiva e a qualidade da decomposição dos componentes de variância.

Palavras-chave: Modelo GBLUP. Análise Univariada. Análise Multivariada. Efeitos de dominância.

INTRODUÇÃO

Entre as dificuldades dos programas de melhoramento de milho, uma das principais é o teste de todos os híbridos simples, sintetizados por todas as combinações de linhagens derivadas de populações em processo de melhoramento. Isso decorre do elevado custo financeiro e as dificuldades operacionais em condições de ensaios com várias repetições e instalados em número excessivo de locais. Por exemplo, a combinação entre 50 linhagens de um grupo heterótico e 50 linhagens de outro grupo heterótico permite a obtenção de 1225 combinações de híbridos simples. Nesse contexto, a disponibilidade de técnicas acuradas de predição das melhores combinações, entre linhagens, pode ser um ponto-chave para a maximização do aproveitamento de recursos financeiros em programas de melhoramento.

Nos últimos anos, devido ao desenvolvimento de técnicas de genotipagem em larga escala, por exemplo, *Genotyping-by-sequencing* (GBS) [1], e a diminuição do custo de genotipagem, associado à forte tendência de aumento de gastos com fenotipagem, há um interesse crescente de programas de melhoramento na exploração de processos de predição com informações de marcadores moleculares [2].

Recentemente, novos modelos de regressão foram propostos para análise de dados fenotípicos e de marcadores moleculares de programas de melhoramento, aplicados em procedimentos de seleção e predição de forma integrada em um processo conhecido popularmente por seleção genômica (SG) [3]. A principal premissa desse método é cobrir amplamente o genoma com marcadores moleculares, procurando encontrar relações entre os marcadores e variantes fenotípicos de uma população, sendo ou não possível esses marcadores serem pertencentes à regiões genômicas causais [4,5]. Essa metodologia do ponto de vista prático é vantajosa, pois não necessita a adoção de delineamentos

e estruturas populacionais específicas, quando comparado às outras metodologias tradicionais que utilizam marcadores, por exemplo, o processo de mapeamento de QTLs (*Quantitative Trait Nucleotide*).

Na SG geralmente são utilizados com maior frequência modelos de regressão múltipla, *best linear unbiased prediction* (BLUP), *ridge regression*, métodos semi-paramétricos e modelos bayesianos [6]. Após vários trabalhos, poucas diferenças foram observadas entre esses métodos na estimação de efeitos genéticos aditivos, com diferenças marginais dependendo da espécie e da arquitetura genética do caráter avaliado [7–10]. Modelos bayesianos permitem estimações de variâncias genéticas específicas dos efeitos genéticos dos marcadores e incorporações de informações *a priori* sobre as distribuições de probabilidade dos efeitos genéticos. O BLUP pode ser empregado por regressão direta nas marcas pelo método *ridge regression* BLUP (rrBLUP) ou utilizando o método *Genomic* BLUP (GBLUP). Este último método recupera informações genéticas entre indivíduos por matrizes de parentesco construídas por marcadores moleculares. Essas matrizes elaboradas por marcadores têm evidenciado maior vantagem em relação às estimadas por *pedigrees*, principalmente por recuperar informações tanto em função de amostragem mendeliana quanto em relação a parentes próximos, permitindo a exploração de relacionamentos genéticos desde gerações recentes até as mais distantes [2].

Vários trabalhos relatam a melhoria na capacidade preditiva utilizando modelos que consideram informações de caracteres secundários de alta herdabilidade, em conjunto com o caráter-alvo [11–13]. Esses modelos são conhecidos como múltiplos caracteres ou multivariados. Praticamente não há relatos sobre uso dessa classe de modelo em trabalhos utilizando espécies cultiváveis anuais [13]. Relatos de utilização de modelos multivariados que consideram relacionamento de parentesco estimados por marcadores moleculares são ainda menos frequentes [14–16].

Durante o melhoramento de espécies em que é possível a exploração do fenômeno da heterose, como o caso da cultura do milho, efeitos de desvio de dominância, oriundos de interação intraloco, podem ter impacto positivo na estimação de efeitos genéticos obtidos em procedimentos de seleção e predição. O entendimento da predominância de efeitos aditivos e de dominância de populações em estágio de melhoramento são preponderantes em algumas tomadas de decisão. Por exemplo, a escolha de parentais em sistemas de cruzamentos, e identificação de populações segregantes promissoras. No contexto de SG, inclusão de efeitos de dominância pode ser realizada no modelo GBLUP com a utilização da matriz de parentesco de dominância estimada por marcadores moleculares [17].

O modelo GBLUP multivariado (GBLUP-MV) [11] permite a exploração de informação entre caracteres secundários, por meio da covariância entre esses caracteres, e entre indivíduos [15,16]. Especialmente no caso do milho, há carências de trabalhos procurando avaliar a utilidade desses modelos. Inclusões de efeitos de dominância no contexto de SG com modelos multivariados ainda não foram reportados na literatura. Assim, os objetivos com este trabalho são: (i) elaborar a construção de híbridos simples de milho com informações fenotípicas e genotípicas reais de linhagens de milho, de modo que seja conservada a arquitetura genética de caracteres; (ii) propor a inclusão do efeito de dominância no modelo GBLUP multivariado (GBLUP-MV-AD); (iii) comparar o modelo GBLUP-MV-AD com os modelos GBLUP multivariado aditivo (GBLUP-MV-A); GBLUP univariado aditivo (GBLUP-UV-A); e GBLUP univariado com a inclusão do efeito de dominância (GBLUP-UV-AD).

MATERIAL E MÉTODOS

Banco de dados Panzea de linhagens de milho

Neste estudo foram utilizados dados de genotipagem SNPs/GBS e fenotipagem das linhagens derivadas da *Nested Association Mapping population* (NAM) do banco de germoplasma do Departamento de Agricultura dos Estados Unidos (USDA-ARS), caracterizado geneticamente por [18]. Esse banco de dados está disponível publicamente pela plataforma Panzea (<http://www.panzea.org/>). A população NAM foi obtida a partir dos cruzamentos entre a linhagem B73 com outras 25 linhagens endogâmicas de alta diversidade genética. Esses cruzamentos foram planejados visando à representação do *pool* gênico mundial da cultura do milho.

Amostragem sistemática de 27,000 marcadores distribuídos uniformemente ao longo dos 10 grupos de ligação do milho foi promovida. A seleção dos marcadores foi realizada pelo sorteio da primeira marca na faixa entre a 1° e a 25° marcas, e as demais foram selecionadas linearmente a cada 25 marcas até a cobertura completa de todos grupos de ligação (2,700 marcas selecionadas por cromossomo). As marcas perdidas foram imputadas pela função *A.mat*, método *meando* pacote *rrBLUP* [19] do software R (R core team, 2014).

No total foram estudados cinco caracteres: (i) altura de planta – AP (4,065 linhagens); (ii) altura de inserção de espiga – AIE (4,061 linhagens); (iii) comprimento de espiga – CE (3,178 linhagens); (iv) número de fileiras por espiga – NFE (3,084 linhagens); (v) peso de grãos - PG (3,313 linhagens). Os dados fenotípicos foram obtidos em ensaios conduzidos no ano de 2006, na safra de verão, na cidade de Aurora no estado de Nova York na latitude 42,75° N e longitude 76,70° W. O processo de manipulação de dados do projeto Panzea foi realizado no software TASSEL [21] e software R.

Análise genética das linhagens

O método BayesB foi utilizado para estimação dos efeitos genéticos dos 27.000 marcadores moleculares, sendo realizadas análises individuais para cada um dos cinco caracteres das linhagens reportados anteriormente. O modelo genético utilizado para análise com o BB segue abaixo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}_A \mathbf{a} + \mathbf{W}_D \mathbf{d} + \mathbf{e}$$

em que \mathbf{y} é um vetor de observações $n \times 1$, em que n é o número de observações; \mathbf{X} é uma matriz de incidência dos efeitos fixos $n \times p$, em que p é o número de efeitos fixos; $\boldsymbol{\beta}$ é vetor de efeitos fixos (média); \mathbf{W}_A e \mathbf{W}_D são matrizes de desvios de substituição alélica e de efeitos de dominância $n \times q$, respectivamente, em que q é o número de efeitos aleatórios; \mathbf{a} , \mathbf{d} são vetores de efeitos aleatórios de substituição alélica e dominantes das marcas $q \times 1$, respectivamente, e \mathbf{e} é o vetor dos efeitos residuais $n \times 1$.

As matrizes de incidência \mathbf{W}_A e \mathbf{W}_D foram construídas seguindo a métrica de Cockerham [22,23]:

$$\mathbf{W}_A \begin{cases} 2 - 2p_k \\ 1 - 2p_k \text{ para os genótipos} \\ -2p_k \end{cases} \begin{cases} \mathbf{A}^1 \mathbf{A}^1 \\ \mathbf{A}^1 \mathbf{A}^2 \\ \mathbf{A}^2 \mathbf{A}^2 \end{cases}$$

$$\mathbf{W}_D \begin{cases} -2(1 - p_k)^2 \\ 2p(1 - p_k) \text{ para os genótipos} \\ -2p_k^2 \end{cases} \begin{cases} \mathbf{A}^1 \mathbf{A}^1 \\ \mathbf{A}^1 \mathbf{A}^2 \\ \mathbf{A}^2 \mathbf{A}^2 \end{cases}$$

em que: p é a frequência do alelo favorável no loco k .

Uma descrição completa sobre as especificações das distribuições de probabilidade dos efeitos aleatórios e dos parâmetros do modelo Bayes B pode ser encontrada em [24] e [5]. Todas as análises bayesianas foram realizadas no pacote BGLR [24] do software R, com a função *BGLR*, ajustada para 10,000 iterações com os primeiros 1,000 ciclos descartados como *burn in* como apresentado em [3]. O parâmetro de mistura π (100/27,000) do modelo Bayes B foi introduzido para o ajuste de 100 efeitos de substituição alélica e de 100 efeitos de dominância das marcas. Todas as demais configurações da função BGLR foram mantidas no modo padrão do pacote.

Construção do genoma de referência

Após o processo de análise, foram ordenados os 100 maiores efeitos de substituição alélica entre os 27.000 efeitos de substituição alélica estimados pelo método BayesB. O mesmo procedimento foi executado para os 100 maiores efeitos de dominância. Esses 200 efeitos genéticos foram reparametrizados utilizando a equação disponível no Texto S1. O objetivo dessa reparametrização foi inflacionar os efeitos genéticos das marcas, que são consequência de um procedimento de encolhimento característico do método Bayes B. Esse processo permite obter valores genéticos que serão considerados como efeitos de regiões genômicas causais para construção do genoma de referência dos híbridos simples. Executamos esse procedimento procurando mitigar alterações na arquitetura genética dos efeitos de substituição alélica e de dominância das marcas dos caracteres obtidos pela análise genética das linhagens.

Após a obtenção e reparametrização dos 200 efeitos genéticos foi construído um genoma de referência com a distribuição desses efeitos ao longo dos 27.000 marcadores de acordo com a posição identificada antes do processo de ordenamento. Os 100 maiores efeitos de substituição alélica foram utilizados para a construção de um vetor dos efeitos de substituição alélica do genoma construído artificialmente (\mathbf{a}^*)

(27.000×1), os QTLs que não eram relacionados a esses efeitos foram zerados durante a construção desse vetor. O mesmo procedimento foi executado para construção do vetor de efeitos de dominância (\mathbf{d}^*) (27.000×1), utilizando os 100 maiores efeitos de dominância. Alguns QTLs com efeitos pleiotrópicos foram identificados no processo de análise genética. As contagens de coincidência dos efeitos genéticos mapeamentos por QTLs com efeitos pleiotrópicos podem ser observadas na Tabela (1).

Tabela 1 Contagens de coincidência de QTLs identificados no processo de ordenamento dos 100 maiores efeitos aditivos (**a**) e dos 100 maiores efeitos de dominância (**d**) em relação aos efeitos genéticos totais obtidos na análise com o modelo BayesB entre os cinco caracteres

		AIE	CE	NFE	PG
a	AP	17	3	1	0
	AIE	-	4	4	2
	CE	-	-	6	2
	NFE	-	-	-	1
d	AP	20	1	0	2
	AIE	-	1	0	0
	CE	-	-	0	3
	NFE	-	-	-	1

Caracteres: altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE), peso de grãos (PG).

Esses QTLs com efeitos pleiotrópicos, identificados naturalmente no processo de análise, e QTLs ligados não conhecidos, participaram como fonte de informação para gerar as correlações genéticas entre os caracteres após o processo de construção dos híbridos.

Por fim, cada caráter teve o seguinte número de QTLs (considerando QTLs com ambos efeitos aditivos e de dominância): 186 (AP); 179 (AIE); 200 (CE); 196 (NFE); e 197 (PG).

Construção dos híbridos simples

Alguns cruzamentos artificiais foram elaborados no software R para obtenção de 400 híbridos simples em um dialelo parcial 20 (grupo heterótico 1) x 20 (grupo heterótico 2) sem os recíprocos. Os dois grupos heteróticos foram definidos graficamente por análise de componentes principais (Tabela 2). Um procedimento de decomposição espectral foi realizado na matriz de parentesco aditiva (A) entre as 4,065 linhagens, e em seguida foram plotados os dois primeiros componentes principais (detalhes sobre a matriz de parentesco serão dados adiante). O Grupo 1 foi definido graficamente pelas linhagens mais próximas da B73, e da mesma forma o Grupo 2 pelas linhagens próximas da Mo17 (Tabela 2). Essas duas linhagens também foram integradas em cada um dos grupos heteróticos. Houve a opção das linhagens B73 e Mo17 como de referência pelo alto padrão heterótico bem conhecido pelos híbridos obtidos em cruzamentos utilizando essas linhagens.

Tabela 2 Grupos heteróticos de linhagens definidas por análise de componentes principais

Grupo1		Grupo2	
B73	GEMS-0086	Mo17	EZ18
NC328	PI601004	Mo44	NC44
Ames22753	PI559382	R177	Va17
PI539927	PI539923	B97	GEMN-0081
Ames27151	NSL438033	GA224	IDS91
Ames27218	GEMS-0223	PI601685	PI601416
PI538009	N192	A682	Ames19008
NSL437913	PI546485	Ames10261	Ames19287
A679	Ames30797	Ames27178	Ames26764
B109	PI550473	NSL437903	PI542778

O estado genotípico de cada loco dos híbridos foi definido pela esperança da contribuição alélica de cada linhagem parental. A esperança foi computada por:

$$E(m_{ij}) = p_{\lambda_j} p_{\Omega_j} 2 + p_{\lambda_j} (1 - p_{\Omega_j}) 1 + p_{\Omega_j} (1 - p_{\lambda_j}) 1$$

em que $E(m_{ij})$ é a esperança do valor genotípico no loco j do híbrido i oriundo do cruzamento entre uma linhagem λ e Ω ; p_{λ} é a frequência alélica favorável da linhagem λ para o loco j ; p_{Ω} é a frequência alélica favorável da linhagem Ω no loco j .

Conforme relatado, os dados perdidos da matriz de marcadores das linhagens foram imputados, após esse processo houve locos com valores entre os intervalos de 0~1 e 1~2. Interpolações numéricas foram realizadas durante a estimação da frequência dos alelos favoráveis para esses locos perdidos. No caso em que as duas linhagens parentais apresentaram o mesmo loco como perdido antes do procedimento de imputação, o estado genotípico do loco foi definido pelo mesmo procedimento de imputação descrito anteriormente, porém de forma direta na matriz \mathbf{M} (matriz de marcadores) dos híbridos.

O valor fenotípico dos híbridos para cada um dos cinco caracteres foi obtido por:

$$\mathbf{y}_{HS} = \mu + \mathbf{W}_A \mathbf{a}^* + \mathbf{W}_D \mathbf{d}^* + \mathbf{e}$$

em que: \mathbf{y}_{HS} é um vetor de observações 400×1 dos híbridos simples; μ é a média amostral obtida pelo BB (BayesB) usando os dados fenotípicos reais das linhagens; \mathbf{a}^* é o vetor de efeitos de substituição alélica do genoma construído artificialmente (27.000×1); \mathbf{d}^* é o vetor dos efeitos de dominância do genoma construído artificialmente (27.000×1); \mathbf{e} é o vetor de resíduos amostrados de

uma distribuição normal univariada ajustando as observações para diferentes cenários de herdabilidade, que serão discutidos subsequentemente; \mathbf{W}_A e \mathbf{W}_D são as matrizes de desvios aditivos e de dominância construída pela matriz \mathbf{M} dos híbridos.

Os valores fenotípicos dos híbridos foram ajustados para faixas de herdabilidade 0,3; 0,5; 0,7 e histórica para cada caráter por resíduos amostrados de uma distribuição normal com variância residual ajustada em função da herdabilidade:

$$\sigma_e^2 = \frac{(1-h^2)}{h^2 \sigma_g^2}$$

em que h^2 é a herdabilidade no sentido amplo desejada para ajuste; σ_g^2 é a variância genética total do caráter.

As herdabilidades (h^2) históricas foram baseadas em um levantamento [25] que pode ser observado de forma adaptada na Tabela 3.

Tabela 3 Herdabilidades médias históricas de cinco caracteres de milho

Caracteres	Herdabilidades (h^2)	Número de Estudos
AP	0,569	45
AIE	0,662	52
CE	0,381	36
NFE	0,57	18
PG	0,418	11

Altura de planta (AP); Altura de inserção de espiga (AIE); Comprimento de espiga (AP); Número de fileiras por espiga (NFE); Peso de grãos (PG). Adaptado de [8].

Modelo GBLUP geral

Para finalidade de comparação entre o modelo GBLUP-MV com o GBLUP-UV, foram analisados todos os híbridos simples simulados nas seguintes configurações de herdabilidade: (i) os cinco caracteres conjuntamente

em h^2 de 0,3; 0,5; 0,7; respectivamente - GBLUP-MV-A e GBLUP-MV-AD; (ii) com os caracteres AP, AIE, CE, NFE, PG, em h^2 de 0,3; 0,5; 0,7; 0,7; 0,3 respectivamente, em ordem determinada por sorteio - GBLUP-MV-A e GBLUP-MV-AD; (iii) os cinco caracteres conjuntamente considerando h^2 histórica (Tabela 2)- GBLUP-MV-A e GBLUP-MV-AD; (vi) considerando todos os caracteres individualmente para as herdabilidades 0,3; 0,5; 0,7 e históricas - GBLUP-UV-A e GBLUP-UV -AD.

O modelo GBLUP geral [11], adaptado para inclusão de efeito de dominância é dado por:

$$y_i = X_i\beta_i + Z_i\alpha_i + Z_i\delta_i + e_i$$

em que y_i é vetor de observações fenotípicas para o caráter i ($i = 1, 2, 3, 4, 5$) de dimensão $n_i \times 1$, n_i é o número total de observações para o caráter i ; X_i é a matriz de incidência de efeitos fixos (média amostral) $n_i \times 1$ para o caráter i ; Z_i é matriz de incidência de efeitos aleatórios (valores genéticos) $n_i \times n_i$ para o caráter i e $\beta_i, \alpha_i, \delta_i$, são vetores dos efeitos relacionados a X_i, Z_i, Z_i , respectivamente; e e_i são efeitos residuais $n_i \times 1$.

Durante a utilização dos modelos aditivos o componente $Z_i\delta_i$ foi desconsiderado; α_i é o efeito genético aditivo dos híbridos simples $n_i \times 1$ e δ_i é o efeito genético de dominância dos híbridos simples $n_i \times 1$. No caso do modelo univariado foi analisado cada caráter isoladamente.

As matrizes de parentesco aditiva (**A**) e de dominância (**D**) estimadas por marcadores moleculares [17] utilizadas foram:

$$\mathbf{A} = \frac{\mathbf{W}_A \mathbf{W}_A'}{2 \sum_k \mathbf{p}_k (1 - \mathbf{p}_k)}$$

em que \mathbf{p}_k é a frequência do alelo favorável no loco \mathbf{k} ;

$$\mathbf{D} = \frac{\mathbf{W}_D \mathbf{W}_D'}{4 \sum_k (\mathbf{p}_k (1 - \mathbf{p}_k))^2}$$

As matrizes de variância-covariância residual (\mathbf{R}) (5x5), variância-covariância genética aditiva entre os cinco caracteres (\mathbf{G}_α) (5x5), e variância-covariância genética de dominância entre os cinco caracteres (\mathbf{G}_δ) (5x5) foram consideradas como não estruturadas.

As matrizes obtidas pelo produto de Kronecker (\otimes) - $\mathbf{R} \otimes \mathbf{I}_{n_i \times n_i}$; $\mathbf{G}_\alpha \otimes \mathbf{A}$; $\mathbf{G}_\delta \otimes \mathbf{D}$ - foram inseridas nas equações de modelos mistos (EMM) de Henderson [26], de acordo com as especificações de cada classe de modelo. As soluções foram obtidas pela inversão direta das EMM após a estimação dos componentes de variância pela maximização da função de Máxima Verossimilhança Residual (REML) pelo algoritmo *Expectation-Maximization* (EM) [27].

Decomposição espectral

As duas matrizes de parentesco \mathbf{A} e \mathbf{D} caracterizaram-se por ser negativa definida, devido a isso, foi realizado um procedimento de decomposição espectral [28,29]. O procedimento foi basicamente a decomposição das matrizes de parentesco em matrizes de autovalores (Diagonal) e de autovetores. Na matriz diagonal dos autovalores, em que os elementos eram negativos, foram realizadas substituições por uma pequena constante de (10^{-4}) magnitude decrescente ao longo de cada elemento diagonal da matriz. Após a reconstrução, obteve-se uma matriz positiva definida, que caracteriza o processo gaussiano necessário para o modelo misto adotado neste estudo.

Medidas de avaliação de modelos

A eficiência dos três modelos foi avaliada por três estatísticas genéticas. Uma delas foi a correlação de Pearson $r_{\hat{g}y}$ entre os valores genéticos paramétricos com os valores genéticos estimados pelos diferentes modelos.

A outra estatística foi o coeficiente de herdabilidade, estimado por:

$$h_{\hat{g}y}^2 = \frac{cov(\hat{g}, y_i)}{var(y_i)}$$

em que \hat{g} é apenas α_i (modelo aditivo) ou $\alpha_i + \delta_i$ (modelo aditivo-dominante).

A última estatística empregada foi a soma de quadrados dos resíduos preditos (*SQRP*):

$$SQRP = \sum_{i=1}^{n_i} (\hat{\theta}_i - \theta_i)^2$$

em que $\hat{\theta}_i$ é valor estimado das variáveis α_i e δ_i – computado individualmente para cada variável - e θ_i os paramétricos.

RESULTADOS

Definição de grupos heteróticos por PCA

Uma distinção clara entre os 2 grupos genéticos formados pelas linhagens próximas da B73 e Mo17 estão demonstradas na Figura. 1.

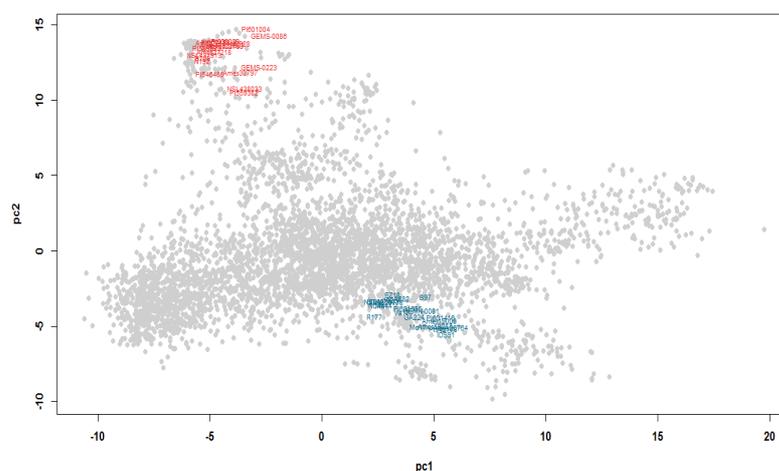


Figura 1 Agrupamento genético de 4.091 linhagens utilizando análise de componentes principais (PCA). As linhagens são pertencentes ao banco de germoplasma do departamento de agricultura dos Estados Unidos (USDA). O eixo X representa o primeiro componente principal (pc1) e o Y ao segundo componente principal (pc2). As cores foram atribuídas de acordo com o grupo genético, em que azul equivale as 19 linhagens próximas da B73 e vermelho as 19 linhagens próximas da Mo17.

O grupo genético definido pelas linhagens próximas da B73 ficou localizado no canto superior esquerdo do gráfico e o da Mo17 no canto inferior central. Algumas linhagens foram observadas em réplica no banco de dados, principalmente as de maior importância econômica. Por exemplo, a linhagem B73 apresentava 35 réplicas e a Mo17, 7 réplicas. Conforme esperado, essas linhagens ficaram repetidas de maneira próxima no gráfico.

Os grupos genéticos foram definidos nessas regiões onde se concentrava um maior número dessas duas linhagens de referência.

Correlações entre caracteres

Com o objetivo de comparar as correlações entre os caracteres obtidas pelos modelos univariados e multivariados foram computadas as correlações entre os efeitos genéticos dos caracteres. Correlações genéticas aditivas, de dominância e genética total entre caracteres e valores paramétricos podem ser observadas na Tabela 4 e Tabelas S1 e S2. O procedimento de construção de híbridos utilizando as estimativas de efeitos genéticos aditivos e de dominância das linhagens dos acessos do USDA evidenciou ser efetivo para o objetivo deste trabalho, com valores de correlação que refletem o que é observado na realidade. Por exemplo, a correlação genética aditiva considerada como paramétrica (com efeitos genéticos estimados nas linhagens) entre os caracteres AP e AIE foi de 0,5501, valor elevado que é esperado na realidade, já que plantas altas normalmente apresentam alta AIE. Um comportamento similar foi observado em relação aos outros caracteres. Não houve uma predominância clara de superioridade entre os modelos durante a estimação das correlações entre caracteres. Apesar da pequena distinção entre os modelos, o uso do modelo GBLUP-UV-AD evidenciou estimativas de correlação em maior proximidade com os valores paramétricos considerados no processo de construção dos híbridos.

Tabela 4 Valores de correlações genéticas paramétricas (P) e estimados entre cinco caracteres pelos modelos GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD em herdabilidades históricas.

Caracteres	Modelo	Efeitos	AIE (0,662)	CE (0,381)	NFE (0,381)	PG (0,418)
AP	P	α	0,5501	-0,1310	0,0865	0,2448
		δ	0,5528	0,3302	0,0875	0,3089
		g	0,5477	-0,014	0,0792	0,2665
AIE	P	α	-	0,1720	0,5319	0,5583
		δ	-	0,3154	0,0941	0,2364
		g	-	0,2005	0,3400	0,4225
CE	P	α	-	-	0,0550	0,4083
		δ	-	-	0,3040	0,5831
		g	-	-	0,1641	0,4245
NFE	P	α	-	-	-	0,0898
		δ	-	-	-	0,5959
		g	-	-	-	0,2486
AP	UV-A		0,6266	-0,2848	0,112	0,2229
AIE		α	-	-0,0191	0,5607	0,5092
CE			-	-	0,2064	0,2898
NFE			-	-	-	0,2898
AP	UV-AD	α	0,6787	-0,2502	0,1562	0,2741
		δ	0,4928	0,3554	0,078	0,3459
		g	0,5731	-0,1066	0,0977	0,233
AIE	UV-AD	α	-	-0,0017	0,4497	0,568
		δ	-	0,2299	0,0736	0,2457
		g	-	0,0439	0,3376	0,41
CE	UV-AD	α	-	-	0,2064	0,2898
		δ	-	-	0,2345	0,2938
		g	-	-	0,2189	0,3372
NFE	UV-AD	α	-	-	-	0,2898
		δ	-	-	-	0,1071
		g	-	-	-	0,6239
AP	MV-A		0,7602	-0,2442	0,0971	0,4061
AIE		α	-	-0,0347	0,4422	0,4534
CE			-	-	0,0418	0,4250
NFE			-	-	-	-0,0474
AP	MV-AD	α	0,7880	-0,2209	0,1902	0,4349
		δ	0,3339	0,3377	-0,0275	0,2769
		g	0,5868	-0,0547	0,1025	0,3692

Continuação...

AIE	α	-	0,0323	0,3593	0,5447
	δ	-	-0,0991	-0,0217	-0,1155
	g	-	-0,0069	0,2025	0,3545
CE	α	-	-	0,0817	0,4432
	δ	-	-	-0,0854	0,8243
	g	-	-	0,0364	0,4954
NFE	α	-	-	-	-0,0467
	δ	-	-	-	0,3700
	g	-	-	-	0,0401

Caracteres: altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE), peso de grãos (PG). Os valores de herdabilidades históricas encontram-se em parênteses. Os valores de correlações genéticas entre os caracteres obtidos pelo processo de análise com o modelo GBLUP-UV foram determinados pela correlação de Pearson entre os efeitos genéticos estimados nas análises individuais de cada caráter.

Componentes de variância e PRESS

Os valores de componente de variância e PRESS estimados pelos modelos univariados e multivariados podem ser constatados na Tabela 5 e Tabelas S4 e S5. Diferenças expressivas não foram detectadas em todas as configurações de herdabilidade na eficiência dos modelos GBLUP-UV-A e GBLUP-MV-A na estimação variância genética aditiva, e entre os modelos GBLUP-UV-AD e GBLUP-MV-AD na estimação da variância genética aditiva (σ_a^2), de dominância (σ_δ^2) e genética total (σ_g^2). Pequenas flutuações de superioridade foram observadas entre os modelos dependendo do caráter, porém, devido à pequena magnitude, essas variações não foram informativas para sugerir superioridade entre os modelos. Os valores semelhantes de componentes de variância genética aditiva obtidos pelos modelos aditivos e aditivo-dominante sugerem ortogonalidade na decomposição dos componentes de variância em boa parte das arquiteturas genéticas estudadas. Entretanto, confundimento da variância genética aditiva com a de dominância foi observado nos modelos aditivos em cenários de alta herdabilidade. Por exemplo, a σ_a^2 estimado pelo modelo GBLUP-UV-A usando os dados do caráter AP ajustado em

herdabilidade 0,3, 0,5 e 0,7, apresentaram valores de 30,99, 43,13, e 51,45, respectivamente, ao passo que o valor paramétrico foi de 23,36. Esse resultado evidência a maior absorção da variância genética de dominância no componente de variância genética aditiva em cenários de herdabilidade superior com a utilização de modelos que consideram apenas efeitos aditivos.

Tabela 5 Componentes de variância genética estimados e paramétricos e soma do quadrado dos erros de predição dos efeitos genéticos estimados pelos modelos GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD em herdabilidades históricas

	Modelo	AP (0,569)	AIE (0,662)	CE (0,381)	NFE (0,57)	PG (0,418)
σ_a^2		23,36	74,65	112,10	0,80	0,35
σ_δ^2	P	19,07	60,55	33,0	1,00	0,06
σ_g^2		42,44	135,20	145,10	1,80	0,41
σ_a^2		34,86	80,39	159,39	1,23	0,29
$SQRP_\alpha$	UV-A	986,80	1.476,40	6.544,0	20,0	12,0
$SQRP_g$		7.667,70	24.961,30	18.144,0	388,0	32,0
σ_a^2		29,30	86,00	159,22	1,03	0,29
σ_δ^2		23,87	67,99	34,92	0,59	0,05
σ_g^2	UV-AD	53,17	154,00	194,14	1,62	0,35
$SQRP_\alpha$		1.117,80	2.008,30	6.802,1	20,7	10,2
$SQRP_\delta$		2.461,90	4.251,90	4.473,7	66,7	7,9
$SQRP_g$		2.724,30	5.185,50	9.796,5	70,6	18,8
σ_a^2		34,47	81,75	154,49	1,20	0,32
$SQRP_\alpha$	MV-A	1.092,69	1.610,73	6.752,66	23,35	9,89
$SQRP_g$		7.754,87	25.127,55	18.464,66	392,2	33,38
σ_a^2		30,11	80,86	153,57	1,02	0,31
σ_δ^2	MV-AD	23,16	66,43	34,54	0,60	0,04
σ_g^2		53,27	147,29	188,11	1,61	0,35
$SQRP_\alpha$		1.322,95	2.406,62	7.158,72	22,62	9,76

Continuação...

$SQRP_{\delta}$	2.585,86	4.130,27	4.783,29	69,59	10,17
$SQRP_g$	2.810,18	4.937,93	10.354,14	71,52	20,98

Variância aditiva (σ_{α}^2), de dominância (σ_{δ}^2) e genética total (σ_g^2) e a soma do quadrado dos resíduos preditos aditivos ($SQRP_{\alpha}$), de dominância ($SQRP_{\delta}$) e genético total ($SQRP_g$) para os caracteres altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE), peso de grãos (PG). Os valores de herdabilidades históricas encontram-se em parênteses.

Apesar desse último resultado aparentar ser paradoxal, já que normalmente espera-se maior eficiência na análise de caracteres em condições de alta herdabilidade, sabemos pela teoria da genética quantitativa, que ambas as variâncias genética aditiva e de dominância são funções do efeito de dominância dos marcadores:

$$\sigma_{\alpha}^2 = \sum_k 2p_k(1-p_k)[a_k + d_k(q_k - p_k)]^2 ; \sigma_{\delta}^2 = \sum_k (2p_k(1-p_k)d_k)^2$$

Nesse cenário, é esperado teoricamente, que ambos os componentes de variância genética aditiva e de dominância apresentem forte confundimento e que percam a ortogonalidade.

Valores de soma de quadrados dos resíduos preditos ($SQRP$) dos efeitos genéticos totais obtidos pelo modelo GBLUP-MV-A foram superiores quando comparados aos resultantes do modelo GBLUP-UV-A, e marginais entre os modelos GBLUP-MV-AD e GBLUP-UV-AD. Reduções expressivas do $SQRP$ foram observadas com a inclusão de efeitos de dominância no modelo. A redução de maior relevância foi observada na análise do caráter AIE no cenário de herdabilidade histórica com o modelo GBLUP-MV-AD, em que houve a redução de 25.127,55 para 4.937,93 do $SQRP$ dos efeitos genéticos totais com a inclusão do efeito de dominância no modelo (Tabela 4). Reduções

do *SQRP* dos efeitos genéticos totais e pequenas oscilações sem padrão do *SQRP* dos efeitos genéticos aditivos foram observadas em todos os cenários de análise com a inclusão do efeito de dominância. Esses resultados evidenciam a importância da inclusão de efeitos de dominância em modelos durante a análise de caracteres de híbridos de milho.

Capacidade preditiva

As correlações entre os valores genéticos paramétricos com os estimados na análise dos dados ajustados em cenários de herdabilidade histórica podem ser observadas na Figura 2.

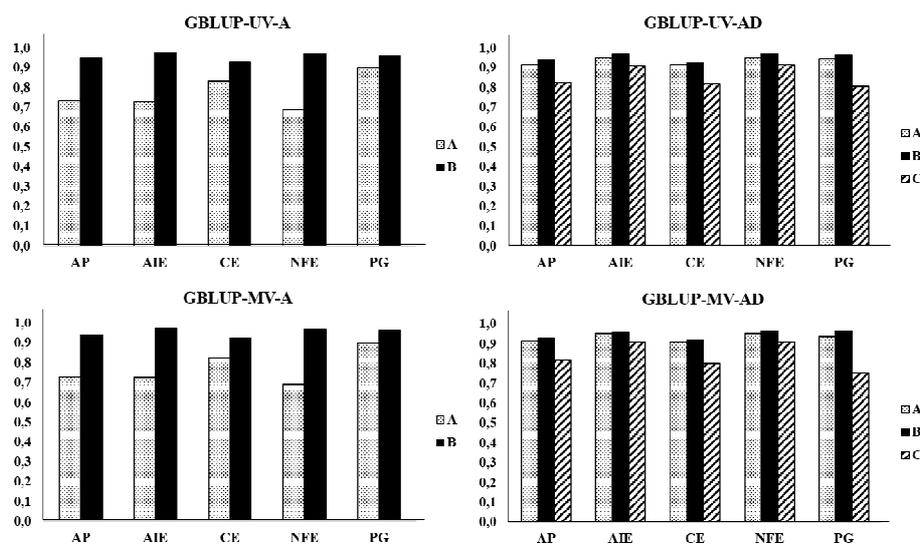


Figura 2 Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidades históricas. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

Uma clara distinção da capacidade preditiva entre os modelos univariados e multivariados não foram observadas em todos os cenários estudados (Figura 2, Figuras S1-S4). Entretanto, diferenças expressivas foram observadas entre o modelo puramente aditivo e o modelo aditivo-dominante. Por exemplo, na análise do caráter AIE ajustado em herdabilidade histórica, com o modelo GBLUP-UV-AD foi obtida uma correlação genética total superior em 38,91% em relação à obtida pelo modelo GBLUP-UV-A, e no modelo GBLUP-MV-AD superior em 37,94% quando comparado ao GBLUP-MV-A. As correlações foram praticamente idênticas entre os efeitos genéticos aditivos paramétricos e estimados em todos modelos utilizados. Esse mesmo comportamento de correlações foi observado para os efeitos genéticos de dominância.

Coefficientes de herdabilidade

As estimativas dos coeficientes de herdabilidade obtidos por todos os modelos em todos os cenários de análise podem ser observadas nas Figuras S5-S8 e Figura. 3.

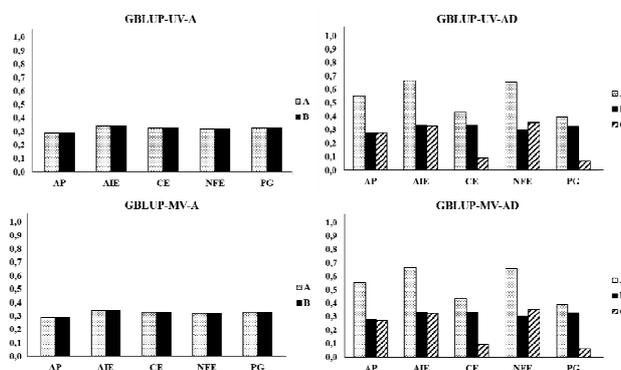


Figura 3 Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade histórica. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.

O mesmo padrão de ausência de diferença de desempenho entre modelos multivariados e univariados foi sugerido pelas estimativas de herdabilidade. A superioridade de modelos que incluem efeito de dominância também é clara pela quantidade de informação genética total recuperada. Em cenários históricos de herdabilidade o menor impacto da inclusão do efeito de dominância foi observado com o modelo GBLUP-MV-AD, na análise do caráter PG, em que houve aumento de 17,85% da herdabilidade no sentido amplo, em relação ao modelo GBLUP-MV-A. Porém, o modelo GBLUP-MV-AD foi o que teve a maior capacidade de recuperar informações genéticas com a inclusão do efeito de dominância. Nesse modelo houve incremento de 0,3228 para 0,6584 da herdabilidade no sentido amplo na análise do caráter NFE (Figura 3). A inclusão do efeito de dominância fez com que dobrasse a quantidade de informação genética total recuperada na análise desse caráter. Novamente, esses resultados confirmam a superioridade de modelos que incluem efeitos genéticos de dominância em recuperações de informações genéticas, e a pequena discrepância de capacidade preditiva entre os modelos univariados e multivariados nos cenários de arquitetura genética abordados.

DISCUSSÕES

Neste estudo, vários cenários de arquitetura genética com variações de condições de herdabilidade foram criados procurando avaliar a utilidade de modelos multivariados em programas de melhoramento genético de milho, principalmente buscando contemplar condições de arquitetura genética histórica já reportadas em outras publicações científicas. Especialmente na cultura do milho, esse é o primeiro trabalho que avalia a utilidade do modelo GBLUP multivariado em análises de componentes de produção. Estudos avaliando o desempenho de modelos multivariados com informações de marcadores moleculares já foram reportados, porém aqui é, pela primeira vez, relatado o impacto da inclusão do efeito de dominância em processos de predição.

A utilização das informações genéticas da população NAM (4.091 linhagens endogâmicas) permitiu a exploração de um *pool* genético dificilmente encontrado em um programa de melhoramento genético restrito a uma empresa ou universidade. A alta variabilidade genética dessa população auxiliou na estimação de parâmetros genéticos que são representativos de várias fontes de germoplasma. Cenários de parâmetros genéticos baseados em resultados históricos também foram adotados por [12], com a análise de dados reais e simulados de caracteres da cultura do feijoeiro comum com modelos multivariados. Entretanto, esse estudo não avaliou o impacto da incorporação de informações de parentesco estimado por marcadores moleculares e o da inclusão de efeitos de dominância. A maioria dos estudos com simulações que consideram informações de parentesco assumem cenários extremos de que podem não ser encontrados na realidade. Por exemplo, em trabalho anterior [16] foram adotadas condições de herdabilidade de 0,1, 0,5 e 0,8 e avaliado apenas duas características conjuntamente. Devido a isso, seguindo valores históricos [25], procuramos avaliar condições de herdabilidade e correlação intermediárias que são observadas com frequência durante o melhoramento genético da cultura do milho.

GBLUP univariado *versus* GBLUP multivariado

Durante o emprego de modelos multivariados com informações de parentesco os principais fatores já relatados que podem influenciar o desempenho preditivo são: (i) disponibilidade de caracteres fenotípicos bem correlacionados com o caráter-alvo; (ii) boa cobertura do genoma com marcadores moleculares; (iii) baixa/moderada herdabilidade do caráter-alvo; (iv) alta herdabilidade dos caracteres correlacionados ao alvo; e (v) moderada/alta correlação entre os caracteres [12,15].

Nossos resultados divergiram parcialmente desses relatos. Nas análises com modelos multivariados para os caracteres AP, AIE, CE, NFE e PG, ajustados em herdabilidade 0,3; 0,5; 0,7; 0,7 e 0,3 respectivamente, não houve melhorias na capacidade preditiva entre AP e PG. Esse resultado não era esperado, já que esses dois caracteres apresentam baixa herdabilidade e são moderadamente correlacionados em relação aos demais.

Os resultados das análises dos dados gerados em cenários de herdabilidade histórica também não evidenciaram diferenças claras entre os modelos GBLUP multivariado e univariado. Esse fato é concordante com os reportados por outros estudos, em que análises conjuntas de caracteres com modelos multivariados tendem a ter desempenho preditivo semelhante a de modelos univariados em condições intermediárias de herdabilidade e correlação [12,15,16]. Assim, sugerimos que modelos univariados são os melhores modelos para análise dos caracteres investigados no nosso estudo. A principal razão dessa recomendação é que os resultados obtidos pela análise conjunta dos caracteres não justificam a adoção desse modelo, já que não existiram diferenças na capacidade preditiva entre modelos univariados e multivariados em todos os cenários de arquitetura genética estudados, e principalmente, porque o modelo

GBLUP univariado é menos intensivo computacionalmente que o GBLUP multivariado.

Entre as condições que podem ter impacto prejudicial no desempenho da análise multivariada, destaca-se situações em que todos ou boa parte dos caracteres apresentam baixa herdabilidade [12]. Na Figura S1 podemos observar que a análise dos cinco caracteres ajustados em herdabilidade 0,3 não implicou em diferenças de capacidade preditiva entre os modelos univariados e multivariados. Nossos resultados sugerem que a análise multivariada de caracteres em condições de baixa/moderada correlação com herdabilidade de até 0,3 não são prejudiciais em relação às univariadas. Outra observação importante é que não houve benefício da análise multivariada com todos os cinco caracteres ajustados em herdabilidade fixa de 0,7 (Figura S3). Esse resultado sugere que a análise conjunta de todos os caracteres em alta herdabilidade também não oferece benefício em relação a análise individual em condições de baixa/moderada correlação entre caracteres.

As correlações dos efeitos genéticos totais entre caracteres variaram de -0,014 entre os caracteres AP e CE à 0,5477 entre os caracteres AP e AIE na análise dos dados das 4091 linhagens endogâmicas do banco de dados Panzea. Em um estudo de simulação [15] foi observado que a análise multivariada é vantajosa apenas em cenários de correlação entre caracteres superiores a 0,5 em alguns *designs* específicos de arquitetura genética. Essa constatação pode ser uma justificativa plausível para ausência de vantagem do uso do modelo GBLUP multivariado investigado em nosso estudo, já que foram observadas correlações baixas/moderadas entre os valores genéticos dos caracteres analisados das linhagens, que serviram como parâmetros durante a construção dos híbridos, com a maioria apresentando correlações entre si inferiores a 0,5.

Em processos de seleção indireta na cultura de milho, é bem conhecida a desvantagem da seleção indireta de caracteres secundários em relação à direta

para produção de grãos. Estudos anteriores evidenciaram que apesar de existir caracteres bem correlacionados com a produção de grãos, na maioria da vezes esses caracteres secundários não apresentam alta herdabilidade que incentive a adoção da seleção indireta para aumentar o ganho com a seleção, e quando há alta herdabilidade, não há alta correlação [25,30]. Acreditamos que essa pode ser a mesma justificativa pela pequena discrepância dos resultados obtidos com os modelos univariados e multivariados na análise dos cinco caracteres de milho estudados, já que observamos apenas baixa/moderada correlação e moderada herdabilidade entre os efeitos genéticos dos caracteres.

Inclusão de efeitos de dominância

A primeira proposta de utilização de modelos multivariados com a abordagem de modelos mistos foi realizada por [11]. A proposta de inclusão de efeitos não aditivos em modelos univariados com informações de parentesco estimados por *pedigrees* também foi proposta por [31]. Nesse trabalho C. R. Henderson já havia previsto a possibilidade da inclusão de efeitos não aditivos em modelos multivariados. Trabalhos explorando esses modelos não foram desenvolvidos de forma imediata devido a restrições computacionais que existiam naquele momento.

Vários trabalhos aplicados de predição com a inclusão de efeitos não aditivos foram desenvolvidos. Entre esses estudos, [32] investigaram a habilidade de decomposição de componentes de variância genética considerando apenas informação de marcadores moleculares e utilizando isoladamente informações de *pedigrees* de *Pinus taeda* com modelos univariados. A principal conclusão foi que marcadores moleculares são mais informativos para estimação de matrizes de parentesco de efeitos aditivos e não aditivos e permitem a decomposição ortogonal dos componentes de variância. Esses autores também afirmam que matrizes de parentesco de efeitos não aditivos construídas por

pedigree apresentam sérias restrições que podem comprometer processos de inferência. A justificativa é que todos os elementos dessas matrizes são funções dos elementos da matriz de parentesco aditiva, e devido a isso, são altamente correlacionados com os elementos das outras matrizes de parentesco, e conseqüentemente não permitem a decomposição efetiva de componentes de variância. Devido a essa constatação não procuramos comparar informações de *pedigrees versus* marcadores moleculares. Utilizando apenas marcadores moleculares concluímos que tanto o modelo GBLUP-UV-AD e GBLUP-MV-AD são efetivos em decompor componentes de variância genética total em componentes aditivos e de dominância, principalmente em cenários de baixa à moderada herdabilidade.

Em um estudo anterior [33] foi reportado que em cenários de baixa herdabilidade a diferença de capacidade preditiva entre modelos aditivos e aditivos-dominantes é expressiva e que desaparece em condições de alta herdabilidade. Apesar de não observarmos essa mesma tendência na capacidade preditiva, foi constatado que em condições de alta herdabilidade tende a ocorrer maior absorção do componente de variância de dominância no de variância aditiva. Por isso, recomendamos cautela em processos de inferência em condições de alta herdabilidade.

Confundimento de componentes de variância podem incentivar conclusões equivocadas que levando a decisões prejudiciais em programas de melhoramento. Por exemplo, durante a seleção de populações efeitos aditivos confundidos com de dominância podem levar a escolha de populações julgadas como geneticamente superiores que podem ter efeitos de dominância perdidos ao longo de gerações, já que esse efeito desvanece após recombinações gaméticas sucessivas. Entre as matrizes de parentesco disponíveis, utilizamos a proposta por [17]. Essa matriz no ponto de vista teórico apresenta maior conformidade com a formulação proposta por Cockerham[22]. Essa métrica

permite maior ortogonalidade na decomposição de componentes de variância genética [23,34,35]. Não avaliamos a inclusão de efeitos epistáticos devido à pequena capacidade de modelos mistos em capturar sinais genéticos epistáticos já bem relatados, tanto no ponto de vista de predição quanto de estimação de componentes de variância [36]. Novos estudos procurando procedimentos alternativos da estimação de efeitos epistáticos e a avaliação do impacto da inclusão dessas covariáveis em modelos multivariados são necessários.

Aumento expressivo de correlações de efeitos genéticos totais paramétricos e simulados foi observado com a inclusão de efeitos de dominância nos modelos GBLUP-UV e GBLUP-MV. Esse impacto já foi reportado em outros estudos com o modelo GBLUP-UV [17,32,33,37]. Além de aprimorar a capacidade preditiva desses modelos, a inclusão de efeitos de dominância melhora a qualidade da estimação de componentes de variância genética aditiva, dado o confundimento já relatado entre esses componentes em modelos puramente aditivos em estudo de modelos de seleção genômica [37].

Assim, concluímos que modelos que consideram efeitos de dominância estimados por marcadores moleculares são recomendados para a análise de banco de dados de programas de melhoramento, uma vez que há benefícios tanto em processos de predição quanto de inferência durante o emprego da seleção genômica.

Perspectivas para o modelo GBLUP multivariado

Apesar de não ter sido observada uma vantagem clara da análise com o modelo GBLUP multivariado em nosso estudo entre caracteres correlacionados da cultura do milho, outros benefícios da análise multivariada podem ser explorados. O modelo GBLUP multivariado permite a obtenção dos efeitos genéticos dos SNPs e dissecação em nível genômico da contribuição de efeitos aditivos e de dominância no controle genético de caracteres e identificação e discriminação de efeitos de QTNs (*Quantitative Trait*

Nucleotide) de ação pleiotrópica e de QTNs em estado de ligação. A identificação de QTNs de ação pleiotrópica pode gerar informações relevantes para o melhoramento simultâneo de vários caracteres [38]. A dissecação dos efeitos genéticos aditivos e de dominância podem auxiliar a elaboração de novos delineamentos de cruzamento para a maximização do ganho com a seleção por ciclo de melhoramento. A principal vantagem da utilização do modelo GBLUP multivariado para identificação de QTNs quando comparado a métodos tradicionais de mapeamento de QTLs (*Quantitative Trait Loci*) é sua simplicidade, ser menos intensivo computacionalmente e lidar bem com o problema “ $p \gg n$ ”. Entretanto, esse modelo ainda apresenta a forte desvantagem de encolher na mesma intensidade os efeitos genéticos dos marcadores. Novos estudos investigando a utilização do modelo GBLUP multivariado para identificação de QTNs são necessários.

CONCLUSÕES

A nossa estratégia de construção de híbridos simples por meio de dados reais de linhagens de milho evidenciou ser efetiva para a conservação da arquitetura genética dos caracteres avaliados. Não houve benefícios em aumento de capacidade preditiva e de qualidade de decomposição de componentes de variância na análise dos caracteres altura de planta, altura de inserção de espiga, comprimento de espiga, número de fileiras por espiga e peso de grãos de milho com a adoção do GBLUP multivariado quando comparado ao univariado. Há ganhos expressivos de capacidade preditiva e de qualidade de decomposição de componentes de variância com a inclusão de efeitos de dominância tanto na análise univariada quanto multivariada com o modelo GBLUP para os caracteres estudados.

REFERÊNCIAS

1. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011; 6: e19379. Available: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019379>
2. Heslot N, Mark ES, Jannink, JL. Perspectives for genomic selection applications and research in plants. *Crop Science* 2014; In press.
3. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001; 157: 1819–29.
4. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*. 2011; 187: 367–383.
5. De Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 2013; 193: 327–345.
6. Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R. Additive genetic variability and the Bayesian alphabet. *Genetics*. 2009; 183: 347–363.
7. Morota G, Boddhireddy P, Vukasinovic N, Gianola D, Denise S. Kernel-based variance component estimation and whole-genome prediction of pre-corrected phenotypes and progeny tests for dairy cow health traits. *Front Genet*. 2014; 5: 1-9.
8. Heslot N, Yang HP, Sorrells ME, Jannink JL. Genomic selection in plant breeding: a comparison of models. *Crop Science*. 2012; 52: 146-160.
9. Gianola D, Morota G, Crossa J. Genome-enabled prediction of complex traits with kernel methods: what have we learned? *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*. 2014. p. 6.
10. Ober U, Erbe M, Long N, Porcu E, Schlather M, Simianer H. Predicting genetic values: a kernel-based best linear unbiased prediction with genomic data. *Genetics*. 2011; 188: 695–708.

11. Henderson CR, Quaas RL. Multiple trait evaluation using relatives' records. *J Anim Sci.* 1976; 43: 1188–1197.
12. Balestre M, Torga PP, Pinho RG, Santos JB. Applications of multi-trait selection in common bean using real and simulated experiments. *Euphytica.* 2013; 189: 225–238.
13. Piepho HP, Möhring J, Melchinger AE, Büchse A. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica.* 2008; 161: 209–228.
14. Malosetti M, Ribaut JM, Vargas M, Crossa J, Eeuwijk FA. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica.* 2007; 161: 241–257.
15. Calus MPL, Veerkamp RF. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol.* 2011; 43: 26.
16. Jia Y, Jannink J-L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics.* 2012; 192: 1513–1522.
17. Vitezica ZG, Varona L, Legarra A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics.* 2013; 195: 1223–1230.
18. Romay MC et al. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology;* 14: 1–18.
19. Endelman JB. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome.* 2011; 4: 250–255.
20. Team R core. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; Available: <http://www.R-project.org/>. 2014.
21. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007; 23: 2633–2635.
22. Cockerham CC. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics.* 1954; 39: 859–82.

23. Zeng Z-B, Wang T, Zou W. Modeling quantitative trait Loci and interpretation of models. *Genetics*. 2005; 169: 1711–1725.
24. Pérez P, De Los Campos G. Genome-wide regression & prediction with the BGLR statistical package. In: *Genetics* [Internet]. 2014; Available: <http://www.ncbi.nlm.nih.gov/pubmed/25009151>
25. Hallauer AR et al. *Quantitative genetics in maize breeding*. 2nd ed. Ames: Iowa State University; 2010.
26. Henderson CR. *Applications of linear models in animal breeding models*. 1984.
27. Dempster AP. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc*. 1977; 39: 1–38.
28. Colvin JA, Dykstra RL. Maximum likelihood estimation of a set of covariance matrices under lower order restrictions with applications to balanced multivariate variance components models. *Ann Stat*. 1991; 19: 850–69.
29. Calinski T, Czajka S, Kaczmarek Z, Krajewski P, Pilarczyk W. Analyzing multi-environment variety trials using randomization-derived mixed models. *Biometrics*. 2005; 61: 448–55.
30. Falconer DS, Mackay TF. *Introduction to quantitative genetics*. 4th ed. London: Longman group Ltd; 1996.
31. Henderson CR. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *J Anim Sci*. 1985; 60: 111–117.
32. Muñoz PR et al. Unraveling additive from non-additive effects using genomic relationship matrices. *Genetics*. 2014; 198: 1759-1768.
33. Da Y, Wang C, Wang S, Hu G. Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. *PLoS One*. 2014; 9: e87666. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3907568&tool=pmcentrez&rendertype=abstract>
34. Yang R-C. Epistasis of quantitative trait loci under different gene action models. *Genetics*. 2004; 167: 1493–1505.

35. Kao C-H, Zeng Z-B. Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics*. 2002; 160: 1243–1261.
36. Gianola D, Fernando RL, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*. 2006; 173: 1761–76.
37. Nishio M, Satoh M. Including dominance effects in the genomic BLUP method for genomic evaluation. *PLoS One*. 2014; 9: e85792. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3885721&tool=pmcentrez&rendertype=abstract>
38. Balestre M., Von Pinho RG., de Souza CL., Bueno Filho JSS. Bayesian mapping of multiple traits in maize: the importance of pleiotropic effects in studying the inheritance of quantitative traits. *Theor Appl Genet*. 2012; 125: 479–493.

INFORMAÇÕES SUPLEMENTARES

Texto S1. Reparametrização dos valores genéticos.

Tabela S1. Correlações aditiva entre caracteres estimadas pelo modelo GBLUP-MV-A.

Tabela S2. Correlações aditiva, dominante e genética total entre caracteres estimada pelo GBLUP-MV-AD

Tabela S3. Componentes de variância paramétricos dos valores genéticos aditivos, dominância e genético total dos híbridos simples.

Tabela S4. Componentes de variância genética e soma do quadrado dos erros de predição dos efeitos genéticos estimados pelo modelo GBLUP-UV-A, GBLUP-UV-AD em faixas de herdabilidade fixas (em parêntesis) em 0,3; 0,5 e 0,7 para os cinco caracteres, respectivamente.

Tabela S5. Componentes de variância genética e soma do quadrado dos erros de predição dos efeitos genéticos estimados pelo modelo GBLUP-MV-A, GBLUP-MV-AD em faixas de herdabilidade fixas (em parêntesis) em 0,3; 0,5 e 0,7 para os cinco caracteres, respectivamente, e por ordem determinada por sorteio.

Figura S1. Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,3. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

Figura S2. Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,5. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

Figura S3. Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,7. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

Figura S4. Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade sorteada. Modelos utilizados: GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

Figura S5. Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixa de 0,3. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.

Figura S6. Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixa de 0,5. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.

Figura S7. Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixa de 0,7. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.

Figura S8. Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade sorteada. Modelos utilizados: GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.

TEXTO SUPLEMENTAR

Texto S1: Reparametrização dos efeitos de substituição alélica e de dominância das marcas

As reparametrizações dos efeitos genéticos das 100 marcas de maior efeito estimadas pela análise das linhagens endogâmicas pelo método BayesB seguem abaixo:

$$a_{rk} = h_A^2 \left[\frac{a_k}{0,7 \left(1 - \frac{h_{Ak}^{2a}}{h_A^2} \right)} \right], \quad \text{em que} \quad h_A^2 = \frac{\sum_{i=1}^{100} \sigma_a^2 + \sum_{i=1}^{100} \sigma_d^2}{\sum_{i=1}^{100} \sigma_a^2 + \sum_{i=1}^{100} \sigma_d^2 + \sigma_e^2}$$

$$\text{e } h_{Ak}^{2a} = \frac{\sum_{i=1}^{99} \sigma_{a \neq k}^2 + \sum_{i=1}^{100} \sigma_d^2}{\sum_{i=1}^{100} \sigma_{a \neq i}^2 + \sum_{i=1}^{100} \sigma_d^2 + \sigma_e^2}$$

em que: a_{rk} é o efeito de substituição alélica da marca k reparametrizado; h_A^2 é a herdabilidade no sentido amplo considerando a variância individual das 100 marcas de maior efeito; h_{Ak}^{2a} é a herdabilidade no sentido amplo considerando a variância aditiva individual das 99 marcas de maior efeito, desconsiderando a variância aditiva individual da marca k ; a_k é o efeito de substituição alélica da marca k entre os 100 maiores efeitos de substituição alélica das marcas sem a reparametrização; $\sum_{i=1}^{100} \sigma_a^2$ é o somatório da variância aditiva individual das 100 marcas de maior efeito; $\sum_{i=1}^{100} \sigma_d^2$ é o somatório da variância de dominância individual das 100 marcas de maior efeito; $\sum_{i=1}^{99} \sigma_{a \neq k}^2$ é o somatório da variância aditiva individual das 99 marcas de maior efeito, desconsiderando a variância aditiva individual da marca k ; σ_e^2 é a variância residual estimada com o modelo BayesB.

$$d_{rk} = h_A^2 \left[\frac{d_k}{0,7 \left(1 - \frac{h_{Ak}^2}{h_A^2} \right)} \right], \quad \text{em que} \quad h_A^2 = \frac{\sum_{i=1}^{100} \sigma_a^2 + \sum_{i=1}^{100} \sigma_d^2}{\sum_{i=1}^{100} \sigma_a^2 + \sum_{i=1}^{100} \sigma_d^2 + \sigma_e^2}$$

$$\text{e } h_{Ak}^2 = \frac{\sum_{i=1}^{99} \sigma_{d \neq k}^2 + \sum_{i=1}^{100} \sigma_a^2}{\sum_{i=1}^{100} \sigma_a^2 + \sum_{i=1}^{100} \sigma_d^2 + \sigma_e^2}$$

em que d_{rk} é o efeito de dominância da marca k reparametrizado; h_{Ak}^2 é a herdabilidade no sentido amplo considerando a variância de dominância individual das 99 marcas de maior efeito, desconsiderando a variância de dominância individual da marca k ; d_k é o efeito de dominância da marca k entre os 100 maiores efeitos de dominância das marcas sem a reparametrização; $\sum_{i=1}^{99} \sigma_{d \neq k}^2$ é o somatório da variância de dominância individual das 99 marcas de maior efeito, desconsiderando a variância de dominância individual da marca k .

TABELAS SUPLEMENTARES

Tabela S1 Correlações aditiva entre caracteres estimadas pelo modelo GBLUP-MV-A.

Caracteres	AIE (0,3)	CE (0,3)	NFE (0,3)	PG (0,3)
AP	0,6509	-0,1717	0,0724	0,4664
AIE	-	-0,1570	0,6473	0,3153
CE	-	-	-0,1639	0,3221
NFE	-	-	-	-0,0555
Caracteres	AIE (0,5)	CE (0,5)	NFE (0,5)	PG (0,5)
AP	0,6960	-0,0135	0,1217	0,6115
AIE	-	0,1260	0,3885	0,5712
CE	-	-	-0,1458	0,3277
NFE	-	-	-	0,1334
Caracteres	AIE (0,7)	CE (0,7)	NFE (0,7)	PG (0,7)
AP	0,9697	0,8405	-0,5736	0,9245
AIE	-	0,8433	-0,4662	0,9009
CE	-	-	-0,5699	0,8984
NFE	-	-	-	-0,5550
Caracteres	AIE (0,5)	CE (0,7)	NFE (0,7)	PG (0,3)
AP	0,6597	-0,0403	0,0169	0,4772
AIE	-	0,1147	0,5313	0,4913
CE	-	-	-0,0590	0,3575
NFE	-	-	-	0,1020

Correlações aditivas entre os caracteres altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG), ajustados os valores fenotípicos dos híbridos para faixas de herdabilidade 0,3; 0,5; 0,7, conforme representado dentro do parêntesis.

Tabela S2 Correlações aditiva, dominante e genética total entre caracteres estimada pelo GBLUP-MV-AD

Caracteres	Efeitos	AIE (0,3)	CE (0,3)	NFE (0,3)	PG (0,3)
AP	α	0,5928	-0,1637	0,1143	0,4738
	δ	0,5638	0,6535	-0,3579	0,7463
	g	0,5793	0,0768	-0,0928	0,5336
AIE	α	-	-0,1634	0,6026	0,3065
	δ	-	0,4079	-0,4000	0,1723
	g	-	0,0177	0,1379	0,2527
CE	α	-	-	-0,1222	0,3480
	δ	-	-	0,0697	0,8407
	g	-	-	-0,0591	0,4424
NFE	α	-	-	-	-0,0945
	δ	-	-	-	0,1492
	g	-	-	-	-0,0192
Caracteres	Efeitos	AIE (0,5)	CE (0,5)	NFE (0,5)	PG (0,5)
AP	α	0,6447	-0,0315	0,2026	0,5806
	δ	0,4945	0,2024	-0,2330	-0,2486
	g	0,5830	0,0227	0,0412	0,3770
AIE	α	-	0,1418	0,4830	0,5512
	δ	-	0,3236	0,0865	-0,2716
	g	-	0,1794	0,3204	0,3186
CE	α	-	-	-0,1088	0,2999
	δ	-	-	0,0208	0,6263
	g	-	-	-0,0744	0,3452
NFE	α	-	-	-	0,0950
	δ	-	-	-	0,4852
	g	-	-	-	0,1795
Caracteres	Efeitos	AIE (0,7)	CE (0,7)	NFE (0,7)	PG (0,7)
AP	α	0,7374	-0,1421	0,1467	0,3688
	δ	0,6359	0,4804	0,1313	0,2385
	g	0,6981	0,0051	0,1409	0,3265
AIE	α	-	0,1246	0,3747	0,3355
	δ	-	0,2371	0,1739	-0,0574
	g	-	0,1470	0,2954	0,2329
CE	α	-	-	-0,1128	0,3831
	δ	-	-	0,0963	0,6638
	g	-	-	-0,0584	0,4177
NFE	α	-	-	-	-0,0036
	δ	-	-	-	0,3197
	g	-	-	-	0,0621

Continuação...

Caracteres	Efeitos	AIE (0,5)	CE (0,7)	NFE (0,7)	PG (0,3)
AP	α	0,6162	-0,0749	0,0958	0,4865
	δ	0,5653	0,2262	-0,2504	0,7575
	g	0,5929	0,0072	-0,0488	0,5444
AIE	α	-	0,0757	0,5651	0,4894
	δ	-	0,2178	0,0505	0,0969
	g	-	0,1087	0,3440	0,3514
CE	α	-	-	-0,1107	0,2918
	δ	-	-	0,1175	0,4929
	g	-	-	-0,0511	0,3250
NFE	α	-	-	-	0,0837
	δ	-	-	-	-0,0480
	g	-	-	-	0,0461

Correlações aditivas entre os caracteres altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG), ajustados os valores fenotípicos dos híbridos para faixas de herdabilidade 0,3; 0,5; 0,7.

Tabela S3 Componentes de variância paramétricos dos valores genéticos aditivos, dominância e genético total dos híbridos simples

	AP	AIE	CE	NFE	PG
σ_a^2	23,36	74,65	112,10	0,80	0,35
σ_d^2	19,07	60,55	33,00	1,00	0,06
σ_g^2	42,44	135,20	145,10	1,80	0,41

Caracteres: altura de planta (AP), altura de inserção de espiga (EIA), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG).

Tabela S4 Componentes de variância genética e soma do quadrado dos erros de predição dos efeitos genéticos estimados pelo modelo GBLUP-UV-A, GBLUP-UV-AD em faixas de herdabilidade (em parêntesis) fixas (em parêntesis) em 0,3; 0,5 e 0,7 para os cinco caracteres, respectivamente

	Modelos	AP (0,3)	AIE (0,3)	CE (0,3)	NFE (0,3)	PG (0,3)
σ_a^2		30,99	66,09	100,38	1,14	0,35
$SQRP_\alpha$	UV-A	1634,48	6605,54	8148,16	20,09	17,07
$SQRP_g$		8105,92	29535,98	20820,88	407,74	39,23
<hr/>						
σ_a^2		28,82	68,31	104,71	1,11	0,34
σ_δ^2		23,88	61,97	39,77	0,82	0,09
σ_g^2		52,71	130,29	144,48	1,93	0,43
$SQRP_\alpha$	UV-AD	1675,30	5924,20	7825,70	25,20	17,00
$SQRP_\delta$		4621,70	11272,20	6102,10	96,90	11,90
$SQRP_g$		5060,10	16645,50	13269,10	113,10	27,40
<hr/>						
		AP (0,5)	AIE (0,5)	CE (0,5)	NFE (0,5)	PG (0,5)
σ_a^2		43,13	90,51	138,00	1,26	0,39
$SQRP_\alpha$	UV-A	872,30	2523,20	2392,50	17,40	7,30
$SQRP_g$		7237,20	25100,10	14706,10	396,70	27,50
σ_a^2		41,16	84,15	137,45	1,16	0,35
σ_δ^2		25,08	69,09	25,93	0,74	0,07
σ_g^2		66,24	153,24	163,38	1,90	0,42
$SQRP_\alpha$	UV-AD	593,50	1767,70	2358,10	18,70	5,90
$SQRP_\delta$		2236,70	7439,50	4728,30	78,00	7,80
$SQRP_g$		2493,00	8758,20	6558,20	89,60	12,30
<hr/>						
		AP (0,7)	AIE (0,7)	CE (0,7)	NFE (0,7)	PG (0,7)
σ_a^2		51,45	100,51	174,91	1,07	0,49
$SQRP_\alpha$	UV-A	992,80	863,60	863,60	9,20	3,60
$SQRP_g$		6894,80	23992,80	12802,50	385,60	24,40
<hr/>						
σ_a^2		33,11	94,27	147,76	0,91	0,36
σ_δ^2		20,88	59,58	27,89	0,56	0,06
σ_g^2		54,00	153,86	175,64	1,47	0,41
$SQRP_\alpha$	UV-AD	709,80	894,80	865,40	8,60	3,60
$SQRP_\delta$		1683,70	4458,10	2906,30	43,90	5,00
$SQRP_g$		1636,00	4817,40	3279,10	44,90	7,10

Variância aditiva (σ_a^2), de dominância (σ_δ^2) e genética total (σ_g^2) e a soma do quadrado dos erros de predição aditivos ($SQRP_\alpha$), de dominância ($SQRP_\delta$) e genético total ($SQRP_g$) para os caracteres altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE), peso de grãos (PG).

Tabela S5 Componentes de variância genética e soma do quadrado dos erros de predição dos efeitos genéticos estimados pelo modelo GBLUP-MV-A, GBLUP-MV-AD em faixas de herdabilidade (em parêntesis) fixas em 0,3; 0,5 e 0,7 para os cinco caracteres, respectivamente, e por ordem determinada por sorteio.

	Modelos	AP (0,3)	AIE (0,3)	CE (0,3)	NFE (0,3)	PG (0,3)
σ_a^2		32,66	65,74	100,77	1,01	0,38
$SQRP_\alpha$	MV-A	1552,80	6880,80	8538,30	32,00	13,50
$SQRP_g$		7944,50	29703,10	21293,90	415,60	36,10

σ_a^2		28,68	61,63	106,92	1,03	0,36
σ_δ^2		20,54	53,86	26,68	0,88	0,08
σ_s^2		49,22	115,49	133,59	1,92	0,44
$SQRP_\alpha$	MV-AD	1627,20	5617,29	8605,41	34,93	14,14
$SQRP_\delta$		4769,87	12022,08	6336,99	119,64	16,12
$SQRP_g$		5105,89	17095,04	14013,16	132,14	27,45

		AP (0,5)	AIE (0,5)	CE (0,5)	NFE (0,5)	PG (0,5)
σ_a^2		48,22	101,76	138,12	1,24	0,44
$SQRP_\alpha$	MV-A	1013,60	2481,10	2495,70	13,80	6,90
$SQRP_g$		7189,60	25108,60	14801,90	393,40	26,90

σ_a^2		40,57	86,30	138,32	1,16	0,36
σ_δ^2		22,89	67,06	23,09	0,71	0,06
σ_s^2		63,46	153,36	161,40	1,87	0,42
$SQRP_\alpha$	MV-AD	872,18	2001,84	2516,43	16,66	5,60
$SQRP_\delta$		2272,55	7112,43	4867,94	82,66	11,56
$SQRP_g$		2387,58	8386,19	6712,12	89,22	16,19

		AP (0,7)	AIE (0,7)	CE (0,7)	NFE (0,7)	PG (0,7)
σ_a^2		569,03	737,22	766,66	1,70	2,59
$SQRP_\alpha$	MV-A	2158,40	2160,30	1908,60	11,70	6,60
$SQRP_g$		5229,30	22562,40	11294,90	386,10	21,90

σ_a^2		33,52	86,84	145,80	0,89	0,36
σ_δ^2	MV-AD	18,38	58,87	25,56	0,56	0,04

Continuação...

σ_g^2		51,90	145,70	171,36	1,46	0,40
$SQRP_\alpha$		738,25	881,22	821,72	10,10	3,24
$SQRP_\delta$		1792,12	5071,81	3428,82	47,17	5,63
$SQRP_g$		1758,59	5356,71	3713,70	47,33	7,75
		AP (0,3)	AIE (0,5)	CE (0,7)	NFE (0,7)	PG (0,3)
σ_a^2		32,54	93,66	172,15	1,05	0,39
$SQRP_\alpha$	MV-A	1811,50	2752,40	841,40	12,40	13,80
$SQRP_g$		8150,80	25369,80	12799,70	388,50	35,60
σ_a^2		27,70	78,38	148,17	0,91	0,35
σ_δ^2		22,27	68,27	26,27	0,58	0,08
σ_g^2		49,97	146,65	174,45	1,49	0,43
$SQRP_\alpha$	MV-AD	1739,31	2360,33	857,16	10,07	13,43
$SQRP_\delta$		4559,40	7967,44	3106,47	45,96	15,25
$SQRP_g$		5096,80	9275,91	3411,84	47,33	26,52

Variância aditiva (σ_a^2), de dominância (σ_δ^2) e genética total (σ_g^2) e a soma do quadrado dos erros de predição aditivos ($SQRP_\alpha$), de dominância ($SQRP_\delta$) e genético total ($SQRP_g$) para os caracteres altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE), peso de grãos (PG).

FIGURAS SUPLEMENTARES

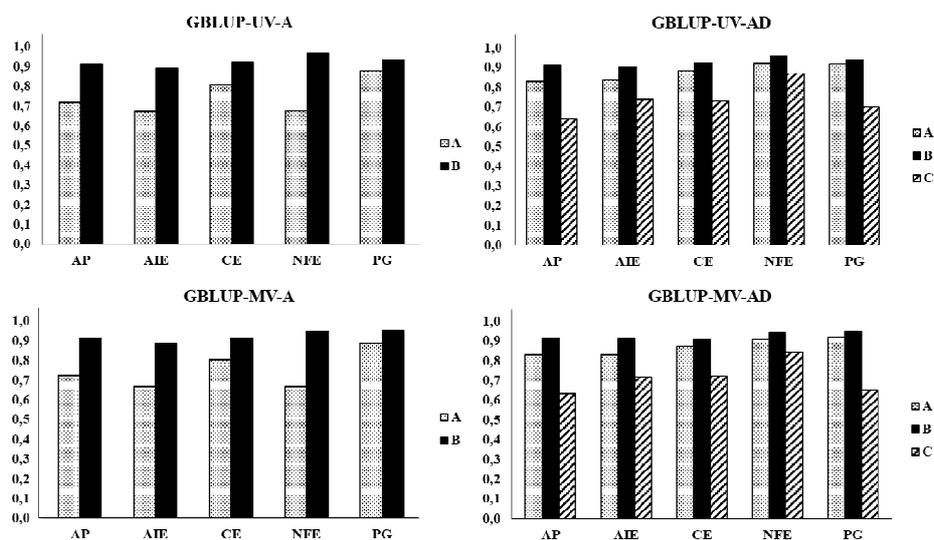


Figura S1 Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,3. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

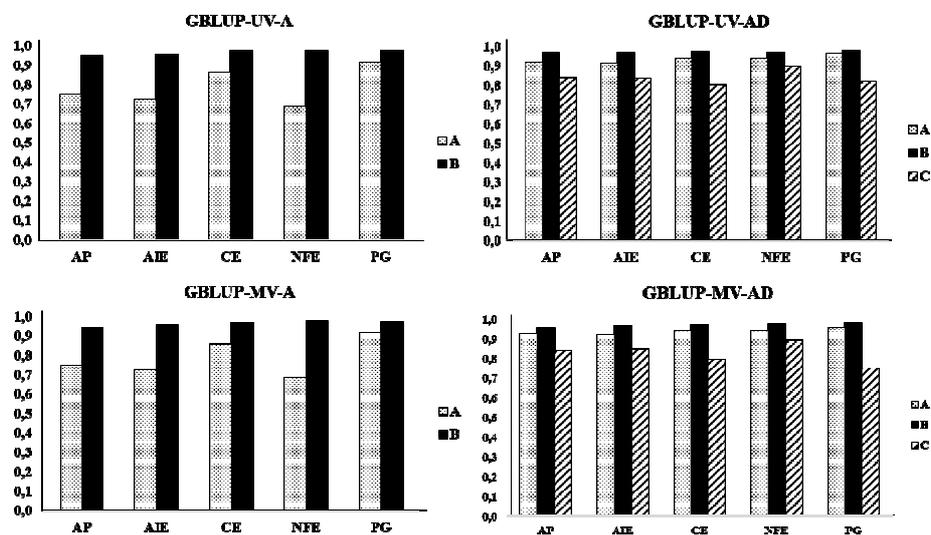


Figura S2 Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,5. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

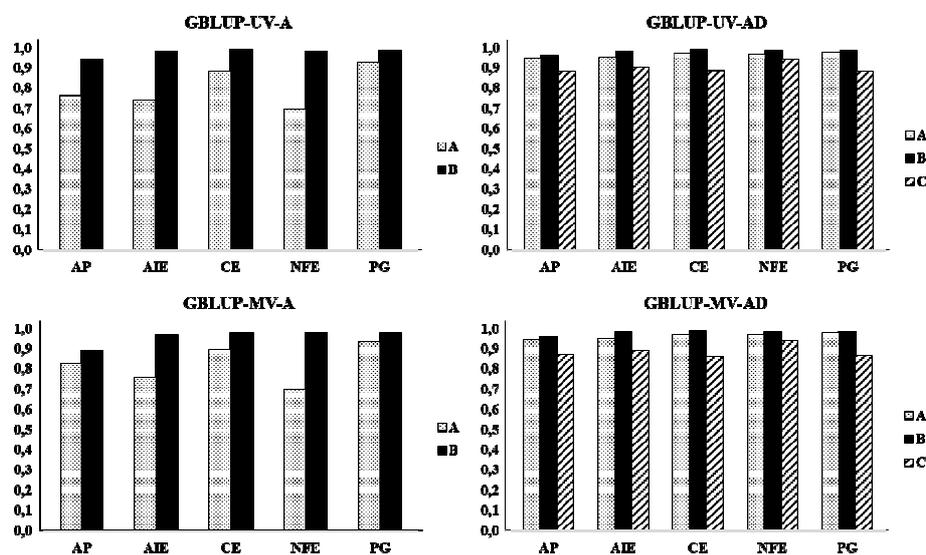


Figura S3 Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade fixa de 0,7. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

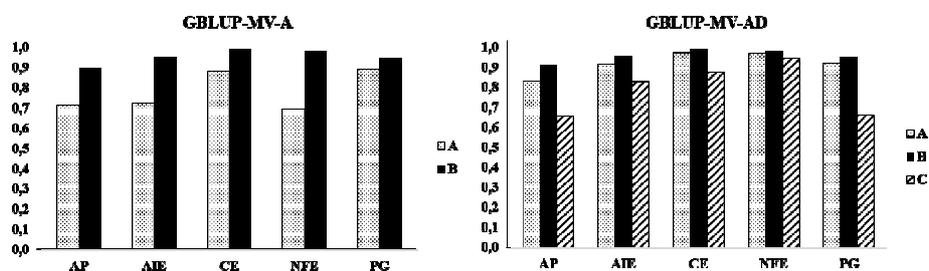


Figura S4 Correlações entre efeitos genéticos paramétricos e estimados dos caracteres ajustados em herdabilidade sorteada. Modelos utilizados: GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) correlações entre os efeitos genéticos totais paramétricos e estimados; (B) correlações entre os efeitos aditivos paramétricos e estimados; (C) correlações entre os efeitos de dominância paramétricos e estimados.

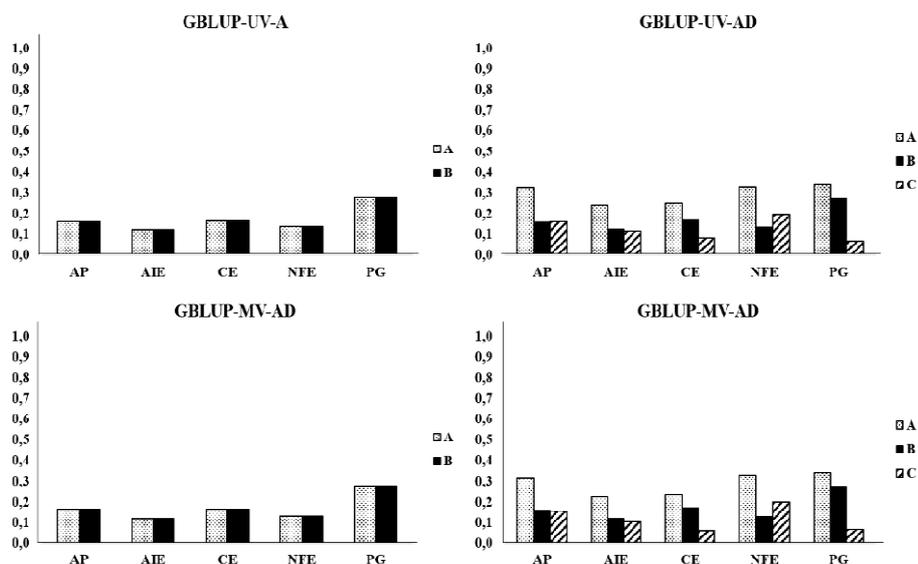


Figura S5 Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixada 0,3. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.

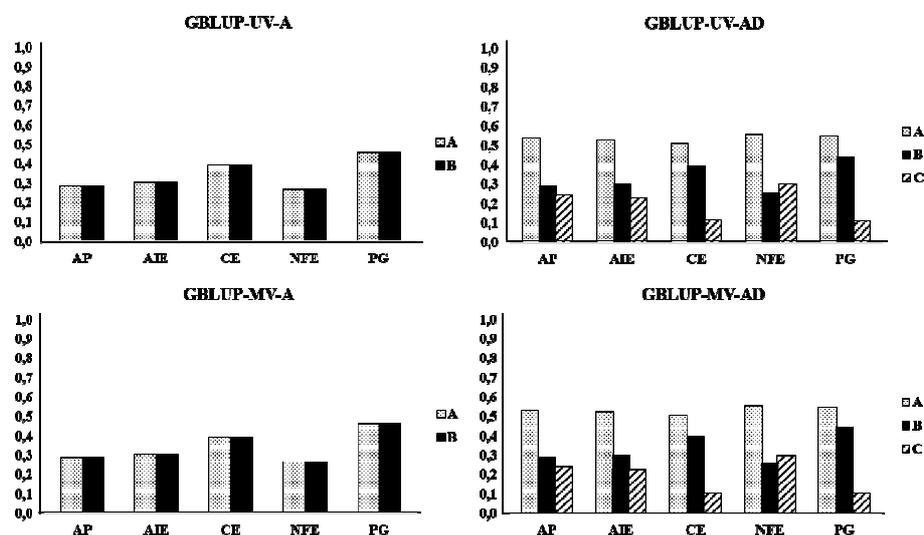


Figura S6 Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixada 0,5. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.

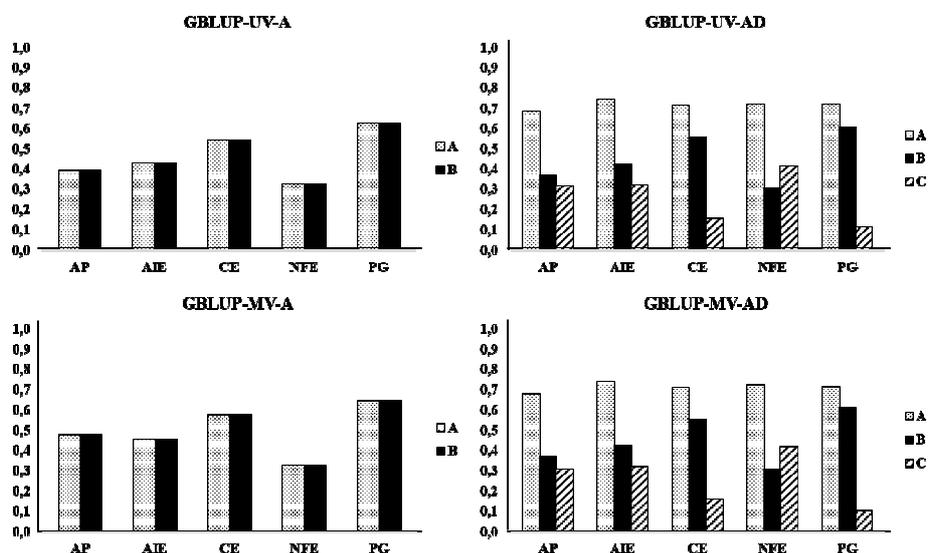


Figura S7 Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade fixada 0,7. Modelos utilizados: GBLUP-UV-A, GBLUP-UV-AD, GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.

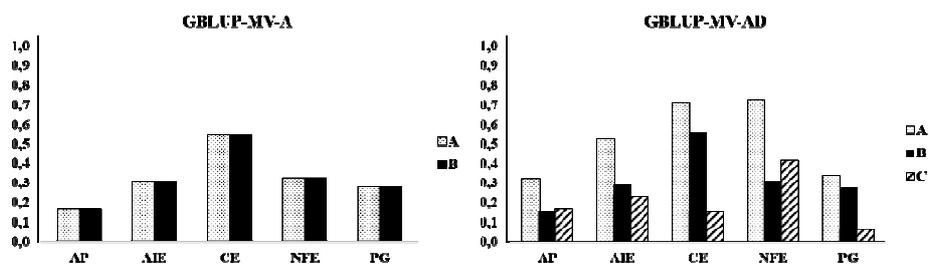


Figura S8 Coeficientes de herdabilidade dos caracteres ajustados em herdabilidade sorteada. Modelos utilizados: GBLUP-MV-A e GBLUP-MV-AD. Caracteres analisados: Altura de planta (AP), altura de inserção de espiga (AIE), comprimento de espiga (CE), número de fileiras por espiga (NFE) e peso de grãos (PG). (A) Coeficientes de herdabilidade no sentido amplo; (B) Coeficientes de herdabilidade aditiva; (C) Coeficientes de herdabilidade de dominância.