



WAGNER MATEUS COSTA MELO

**MODELOS MISTOS NO ESTUDO DA
HETEROSE E PREDIÇÃO DO DESEMPENHO
DE HÍBRIDOS DE MILHO**

LAVRAS - MG

2013

WAGNER MATEUS COSTA MELO

**MODELOS MISTOS NO ESTUDO DA HETEROSE E PREDIÇÃO DO
DESEMPENHO DE HÍBRIDOS DE MILHO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia/Fitotecnia, área de concentração em Produção Vegetal, para a obtenção do título de Doutor.

Orientador

Dr. Renzo Garcia Von Pinho

Coorientador

Dr. Márcio Balestre

LAVRAS – MG

2013

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Melo, Wagner Mateus Costa.

Modelos mistos no estudo da heterose e predição do
desempenho de híbridos de milho / Wagner Mateus Costa Melo. –
Lavras : UFLA, 2013.

120 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2013.

Orientador: Renzo Garcia Von Pinho.

Bibliografia.

1. *Zea mays*. 2. Pleiotropia. 3. Caracteres complexos. 4. Análise bayesiana. 5. Arquitetura genética. 6. Melhoramento genético. I. Universidade Federal de Lavras. II. Título.

CDD – 633.1523

WAGNER MATEUS COSTA MELO

**MODELOS MISTOS NO ESTUDO DA HETEROSE E PREDIÇÃO DO
DESEMPENHO DE HÍBRIDOS DE MILHO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia/Fitotecnia, área de concentração em Produção Vegetal, para a obtenção do título de Doutor.

APROVADA em 28 de Fevereiro de 2013.

Dr. Márcio Balestre	UFLA
Dr. Adriano Teodoro Bruzi	UFLA
Dr. João Cândido de Souza	UFLA
Dr. Glauco Vieira de Miranda	UFV

Dr. Renzo Garcia Von Pinho
Orientador

LAVRAS - MG

2013

DEDICO

À memória de meus avôs Antônio Francisco da Costa e Geraldo Sumbém;

As minhas avós Celuta de Melo e Luzia Sumbém;

Ao meu pai, José Mateus de Melo, aos meus irmãos; Emílio, Fernanda e

Patrícia, pelo estímulo.

EM ESPECIAL

À minha esposa, Camila Fonseca Melo, a minha filha, Helena Fonseca Melo, e a

minha mãe, Ieda Maria Costa Melo.

AGRADECIMENTOS

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

À Universidade Federal de Lavras e a Universidade Federal de Uberlândia, pela oportunidade de realização do curso de Doutorado;

As empresas Monsanto do Brasil e Advanta Comércio de Sementes, por terem aceitado a realização deste curso durante a os dias de trabalho;

Ao orientador Prof. Dr Renzo Garcia Von Pinho, pela orientação, amizade, dedicação, disponibilidade, paciência e pelos ensinamentos transmitidos durante o curso;

Ao coorientador Márcio Balestre, o qual me auxiliou muito no decorrer destes anos e por todos os ensinamentos prestados;

A minha esposa, por ter me incentivado para a realização deste curso;

Ao Dr. Gilberto Pozzar, por todo incentivo na realização deste curso durante o tempo que trabalhei na Monsanto do Brasil;

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

RESUMO GERAL

O objetivo com este trabalho foi determinar os efeitos epistáticos na heterose do milho por meio de um único modelo, com base em análises de di, tri e tetralelos e com a informação de marcadores de microssatélites. Além disto, objetivou-se prever o desempenho de híbridos de milho e validar o método de efeitos totais das marcas associadas utilizando validação cruzada (prever os híbridos e validar com os resultados já obtidos dos mesmos em ensaios) e resultados de ensaios conduzidos com os híbridos preditos. Dois conjuntos de dados foram utilizados. Para determinar a importância dos efeitos epistáticos utilizou-se 51 linhagens de diferentes origens obtendo-se 6 híbridos duplos, 14 híbridos triplos e 58 híbridos simples, totalizando 78 híbridos. Para o experimento de predição utilizou-se os mesmos 58 híbridos simples. Setenta e nove marcadores de microssatélites foram utilizados na genotipagem das 51 linhagens e produziram 79 marcas. Esses marcadores estão distribuídos ao longo dos 10 grupos de ligação do milho. Essa informação foi utilizada na construção da matriz de informação de parentesco e na construção das matrizes dos efeitos totais das marcas associadas (ETMA) em que os valores das marcas foram utilizados na matriz de efeitos aleatórios no modelo linear misto. De acordo com este trabalho foi possível desmembrar a heterose e entender melhor a participação dos efeitos epistáticos na mesma quando analisa conjuntamente híbridos simples, duplos e triplos sob um mesmo modelo. Sob esta abordagem, foi possível compreender a heterose nos seus diversos componentes. Também foi possível verificar que a utilização de marcadores moleculares teve uma tendência de melhora da acurácia dos efeitos genéticos epistáticos e de dominância. Com diferentes níveis de desbalanceamento do modelo, a acurácia preditiva do método ETMA variou de 0,10 a 0,96. Além disto, quando os híbridos não testados na população de treino tiveram seus valores preditos e testados em ensaios, a acurácia foi de 0,30, valor este excelente quando valida-se a predição em locais totalmente diferentes aos locais utilizados para gerar os valores de predição. No que diz respeito ao descarte de bons híbridos, o método de efeitos totais das marcas associada acertou em 50% dos híbridos selecionados nos ensaios regionais e houve perda seletiva em apenas 33% dos casos. De acordo com esses resultados pode-se inferir que, com os recursos computacionais disponíveis atualmente e técnicas estatísticas avançadas, pode-se expandir o conceito de cruzamentos para fronteiras muito além das tradicionais capacidades gerais e específicas de combinação. Também se verificou que a validação preditiva em diferentes condições de cultivo é viável e que os resultados de validação cruzada representam de forma contundente a situação real de desempenho no campo.

Palavras-chave: Milho Híbrido. Melhoramento Genético. Zea Mays. Pleiotropia. Caracteres complexos. Análise Bayesiana. Arquitetura Genética.

GENERAL ABSTRACT

The purpose of this study was to determine the epistatic effects in heterosis of maize by means of a single model based on analyses of di-, tri- and tetrallelic, and analyses with the information from microsatellite markers. In addition, the aim was to predict maize hybrid performance and validate the total effect of associated mark (TEAM) method using cross validation (predict the hybrids and validate with the results already obtained from them in trials) and results of trials conducted with the predicted hybrids. Two sets of data were used. To determine the importance of the epistatic effects, 51 lines from different origins were used, obtaining 6 double hybrids, 14 three-way hybrids and 58 single hybrids, for a total of 78 hybrids. For the prediction experiment, the same 58 single hybrids were used. Seventy-nine microsatellite markers were used in genotyping of the 51 lines and they produced 79 marks. These markers are distributed along the 10 maize linkage groups. This information was used in constructing the relationship information matrix and in construction of the matrices of the total effects of associated mark (TEAM) where the values of the marks were used in the random effects matrix in the mixed linear model. According to the study, it was possible to break down heterosis and better understand the participation of epistatic effects in it when jointly analyzing single, double and three-way hybrids under the same model. Under this approach, it was possible to understand heterosis in its diverse components. It was also possible to verify that with the use of molecular markers there was a trend toward improvement of the accuracy of the epistatic and dominance genetic effects. With different levels of imbalance of the model, the predictive accuracy of the TEAM method ranged from 0.10 to 0.96. In addition, when the hybrids not tested in the training population had their values predicted and tested in trials, the accuracy was 0.30. This is an excellent value when the prediction in locations totally different from the locations used for generating the prediction values is validated. In regard to discard of good hybrids, the total of the method total effects of associated mark was right in 50% of the hybrids selected in the regional trials and there was selective loss in only 33% of the cases. According to these results, it may be inferred that with the computational resources currently available and advanced statistical techniques, the concept of crosses may be expanded to frontiers far beyond the traditional general and specific combination capacities. It was also verified that predictive validation under different growing conditions is viable and that the results of cross validation represent the real field performance situation in a striking manner.

Keywords: Hybrid Maize. Genetic Breeding. Zea Mays. Pleiotropy. Complex traits. Bayesian Analysis. Genetic Architecture.

LISTA DE FIGURAS

PRIMEIRA PARTE

- Figura 1 Mudança no valor genotípico de um indivíduo por ocasião da mudança do alelo a pelo alelo A 20
- Figura 2 Resposta aditiva no valor genotípico devido aos efeitos dos genes A e B26
- Figura 3 Resposta não aditiva no valor genotípico devido à interação dos alelos dos genes A e B 27
- Figura 4 Esquema de produção dos híbridos simples, triplos e duplos de milho.....31

SEGUNDA PARTE - ARTIGOS

ARTIGO 1

- Figura 1 - Sistema de cruzamento em 51 linhagens de milho. Os círculos superiores se referem ao background de cada linhagem.79
- Figura 2 - Decomposição da variância genética em seus componentes descritos pela variância aditiva, dominante, aditiva x aditiva, aditiva x dominante e dominante x dominante.82
- Figura 3 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando um parâmetro no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.....83
- Figura 4 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando dois parâmetros no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.....84

Figura 5 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando três parâmetros no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.....	85
Figura 6 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando quatro parâmetros no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.....	86
Figura 7 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando cinco parâmetros no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.....	87
Figura 8 - Decomposição do valor genético do híbrido em efeitos aditivos + epistasia dominante-dominante. A linha azul se refere ao valor genético total, a linha pontilhada preta se refere à soma dos efeitos aditivos e epistático dominante-dominante.....	88
Figura 9 - Decomposição do valor genético do híbrido em efeitos aditivos e epistasia dominante-dominante. A linha vermelha é a participação dos efeitos aditivos (pais) e a linha verde se refere ao efeito dominante-dominante proveniente do cruzamento em questão.....	89
Figura 10 - Cluster das linhagens com base na epistasia do tipo dominante-dominante. A graduação nas linhas do vermelho ao azul mostra os cruzamentos com epistasia variando do negativo ao positivo. Foi utilizado um método simples de agrupamento baseado no algoritmo K-means que visa minimizar as variâncias de pontos dentro de cada cluster.	90

ARTIGO 2

- Figura 1 Relação dos cruzamentos realizados no experimento de treino (Quadrante preto) com os híbridos testados em ensaios regionais (vermelho), Os híbridos selecionados na população de treino estão em verde, Os híbridos selecionados no ensaio regional estão representados pela letra B, Os híbridos avaliados no ensaio regional que foram selecionados na população de treino apresentam seus quadrantes delimitados pela cor verde,..... 114
- Figura 2 Ajuste do método dos efeitos totais das marcas associadas (ETMA) considerando apenas o efeito aditivo das marcas (ETMA1) e efeitos aditivos + efeitos de dominância (ETMA2), ... 115
- Figura 3 Credibilidade dos valores preditos dos híbridos não testados pelo método ETMA2 na população de treino e valores observados nos ensaios nacionais, 116
- Figura 4 Ajuste do método dos efeitos totais das marcas associadas (ETMA2) por genitor com mais de quatro observações (genitores de 1 a 6), 117
- Figura 5 Ajuste do método dos efeitos totais das marcas associadas (ETMA2) por genitor com mais de quatro observações (genitores de 7, 8 e 9), 118
- Figura 6 Ajuste do método dos efeitos totais das marcas associadas (ETMA2) por genitor com mais de quatro observações (genitores de 10, 11, 13, 14, 18 e 26), 119
- Figura 7 Credibilidade seletiva entre os híbridos testados na população de treino e híbridos avaliados no ensaio regional considerando o método tradicional e totais das marcas associadas (ETMA2), 120

LISTA DE TABELAS

PRIMEIRA PARTE

Tabela 1	Probabilidade ρ dos híbridos herdarem os alelos dos genitores p_1 $p_2 p_3$ e p_4 , considerando quatro fundadores L1 L2, L3 e L4	32
----------	---	----

SEGUNDA PARTE - ARTIGOS

ARTIGO 1

Tabela 1	Distribuição dos 79 marcadores de microssatélites nos 10 grupos de ligação (GL) do milho	78
Tabela 2	Seleção de modelo com base no critério bayesiano de informação (BIC) para os efeitos aditivos, dominantes e epistáticos considerando o modelo sem informação de marcadores moleculares.....	80
Tabela 3	Seleção de modelo com base no critério bayesiano de informação (BIC) para os efeitos aditivos, dominantes e epistáticos considerando o modelo com informação de marcadores moleculares.....	81

ARTIGO 2

Tabela 1.	Distribuição dos 79 marcadores de microssatélites nos 10 grupos de ligação (GL) do milho	112
Tabela 2 e 3.	Média, mediana e moda das correlações entre a predição de híbridos não testados e os resultados da população treino, utilizando diferentes tamanhos de desbalanceamento no modelo aditivo (A – Tabela 2) e no modelo aditivo + dominante (A+D – Tabela 3).....	113

SUMÁRIO

PRIMEIRA PARTE	
1	INTRODUÇÃO 13
2	REFERENCIAL TEÓRICO 16
2.1	Heterose em milho 16
2.2	Epistasia 18
2.3	Parentesco entre indivíduos, similaridade por estado, similaridade por descendência e suas aplicações no melhoramento genético da cultura do milho 33
2.4	Predição do desempenho de híbridos de milho 36
3	CONCLUSÕES 41
	REFERÊNCIAS 42
	SEGUNDA PARTE - ARTIGOS 52
	ARTIGO 1 Controle genético do desempenho de híbridos de milho utilizando pedigrees complexos e marcadores de microssatélites..... 52
	ARTIGO 2 Predição de híbridos simples de milho utilizando a abordagem de efeitos totais das marcas associadas com validação cruzada e em ensaios regionais..... 91

PRIMEIRA PARTE

1 INTRODUÇÃO

Ao iniciar um programa de melhoramento de milho, uma das principais dificuldades enfrentadas pelo melhorista é a escolha das linhagens a serem utilizadas para síntese de híbridos. Essa escolha baseia-se normalmente no desempenho *per se* e na capacidade de combinação das linhagens parentais e, considerando esse último critério, busca-se separar esses genitores em grupos heteróticos de modo a maximizar o vigor híbrido (BALESTRE et al., 2008a).

O estabelecimento de grupos heteróticos pode ser obtido de diversas maneiras. O mais comum baseia-se na informação da capacidade específica de combinação de linhagens ou mesmo populações provenientes de cruzamentos biparentais como é o caso das populações de melhoramento, ou populações derivadas dos cruzamentos de linhagens elites. Pode estabelecer grupos heteróticos também por meio das distâncias genéticas das linhagens ou populações por meio de marcadores moleculares.

O uso de marcadores para estabelecimento de grupos heteróticos advém da relação teórica entre distância genética e heterose (FALCONER; MACKAY, 1990). Baseados nessa relação, diversos estudos foram realizados com o objetivo de prever cruzamentos utilizando as distâncias genéticas (DG) (LANZA et al., 1997; BETRÁN et al., 2003; AMORIM et al., 2006). Embora esses autores observassem relação positiva entre heterose e DG, resultados conflitantes foram constatados por Balestre et al. (2008b), Legesse et al. (2008), Dhliwayo et al. (2009), Devi e Singh (2011).

Esses resultados conflitantes sugerem que a utilização de marcadores moleculares para predição de híbridos de milho e estabelecimento de grupos heteróticos ou até mesmo o desmembramento da heterose é restritiva; seja pela

base teórica frouxa, onde a heterose seria função apenas das diferenças nas frequências alélicas e o efeito de dominância, ou pela fraca associação dessas marcas com QTLs de grande efeito controlando o caráter.

Uma alternativa à predição de cruzamentos via distância genética seria a utilização de modelos mistos – modelo que engloba os efeitos fixos e os aleatórios - associados à informação de pedigree (BERNARDO 1994; 1995). Neste caso, os marcadores moleculares seriam de grande utilidade na predição de híbridos visto que apresentam estimativas acuradas do parentesco quando a informação de pedigree é escassa ou inexistente. Mais recentemente, Schrag et al. (2009; 2010) sugeriram substituir a matriz de parentesco pela matriz de marcas observadas ou efeito total das marcas associadas (ETMA), onde uma população de treino é utilizada para aferir valores genéticos a cada marca e, posteriormente, realiza-se a validação através de híbridos não testados. Essa metodologia mostrou-se superior aos métodos tradicionais de predição (SCHRAG et al., 2009).

Embora esses modelos tenham apresentado resultados satisfatórios, é notório que, para a predição de híbridos ou da heterose, é preciso considerar que esse fenômeno pode surgir de vários outros eventos genético tais como a epistasia, colinearidade genética e até mesmo pleiotropia (BALESTRE et al. 2012).

No que concerne à epistasia e sua participação na heterose em milho, estudos de mapeamento de QTLs não têm verificado grande importância desse fenômeno (LU; ROMERO-SEVERSON; BERNARDO, 2003; FRASCAROLI et al., 2007, GARCIA et al., 2008; SCHÖN et al., 2010). Esses resultados podem surgir pela utilização de modelos restritos onde apenas as epistasias em QTLs de grande efeito são levadas em consideração. Essa especulação se justifica pela recente demonstração de Melchinger et al. (2007), que derivou o controle da heterose de forma analítica e sugeriu que a epistasia pode ter papel fundamental

no controle genético da heterose em milho, principalmente no que diz respeito à interferência da epistasia do tipo aditiva x aditiva que compõe esse efeito. Contudo, é sabido que em dialelos tradicionais a ausência de contrastes ortogonais para obtenção dos efeitos epistáticos e alto efeito de multicolinearidade impede que os efeitos epistáticos sejam estimados e, dessa forma, a capacidade geral de combinação confunde efeitos aditivos e interação do tipo aditiva x aditiva e impede que esses efeitos sejam estudados de forma isolada. Da mesma forma, a capacidade específica de combinação carrega todos os efeitos epistáticos e mais os desvios de dominância. Baseado nisto, é fácil imaginar que a CEC representa um conjunto de efeitos de interação inter e intra-alélicas.

Uma forma de superar o problema do confundimento dos efeitos de dominância com efeitos epistáticos é tomar o dialelo tradicional como uma forma de pedigree complexo que pode envolver acasalamentos entre linhagens, entre híbridos e entre linhagens com híbridos gerando um sistema intrincado de cruzamentos (COCKERHAM, 1961, RAWLINGS; COCKERHAM, 1962a;b). Esses autores propuseram a análise de tri e tetralelos para estimar alguns efeitos epistáticos através de contrastes de efeitos. Porém, uma vez que se considera esse sistema dentro de um único pedigree, pode-se utilizar uma única análise para estimar todos os efeitos epistáticos decorrentes destes acasalamentos.

O objetivo com este trabalho foi determinar os efeitos epistáticos na heterose do milho utilizando pedigrees complexos reunindo sob um único modelo, os di, tri e tetralelos em análises com e sem a informação de marcadores de microssatélites. Além disto, objetivou-se prever o desempenho de híbridos de milho e legitimar o método de efeitos totais das marcas associadas utilizando validação cruzada e resultados de ensaios.

2 REFERENCIAL TEÓRICO

2.1 Heterose em milho

De forma genérica, o termo heterose descreve a superioridade do híbrido F_1 em relação a médias dos parentais (SHULL, 1909). A exploração da heterose em milho é considerada um exemplo de sucesso de melhoramento via hibridação; no entanto, o controle genético desse fenômeno ainda é desconhecido e motivo de grande especulação.

Três principais hipóteses têm sido levantadas acerca do controle genético da heterose. A hipótese de dominância atribui à heterose um efeito cumulativo de alelos que apresentam dominância ou dominância parcial (JONES, 1918). Esses alelos mascaram os alelos deletérios recessivos no genótipo heterozigoto de modo que o híbrido torna-se superior às médias dos parentais se houver dominância completa ou dominância parcial. Além disso, a fase de repulsão de tais genes pode gerar pseudosobredominância.

Por outro lado, a hipótese de sobredominância atribui ao genótipo heterozigoto a superioridade em relação a ambos os parentais homozigotos e não apenas à média dos homozigotos (SHULL, 1909; EAST, 1936). Por fim, a heterose pode surgir devido à interação alélica nos locos, diferentes controladores do caráter – hipótese epistática (POWERS, 1944).

Além dessas hipóteses, outras teorias acerca da heterose têm sido levantadas tais como: colinearidade genética - onde a heterose é função da complementação de locos e não alélica (HOCHHOLDINGER; HOECKER, 2007) -, e heterose multiplicativa - sugere que a heterose pode surgir devido à ação multiplicativa de QTLs presentes em subcaracteres ou componentes da produção (SCHNELL; COCKERHAM, 1992). Nessa última hipótese e, dado que os componentes da produção podem apresentar QTL pleiotrópicos em

relação à produção *per se*, é esperado que o grau de pleiotropia exercida por esses QTLs influenciem fortemente a heterose.

Os estudos da heterose por meio de análise puramente fenotípica normalmente apresentam resultados conflitantes e interpretações equivocadas acerca do controle genético (TROYER, 2006). Sob esse cenário, a análise fenotípica e genotípica - via marcadores moleculares - pode fornecer uma nova percepção acerca do controle genético desse fenômeno tão importante no melhoramento de milho.

Um dos primeiros trabalhos relacionados ao controle genético da heterose em milho via marcadores moleculares foi realizado por Stuber et al. (1992) que utilizaram o delineamento III – que é o cruzamento da F2 com os dois pais -, sugerido por Comstock e Robinson (1952), (MELCHINGER et al., 2007). Os resultados de Stuber et al. (1992) sugeriram que a heterose é controlada principalmente por genes com sobredominância e que existe evidência de epistasia atuando sobre esse fenômeno. Posteriormente, Cockerham e Zeng (1996) desenvolveram métodos estatísticos para análise de QTLs envolvidos no controle genético da heterose com a incorporação de contrastes epistáticos. Nesse trabalho, os autores observaram que a heterose é controlada por QTLs com efeito de dominância e com pequena participação de efeitos epistáticos.

Embora diversos trabalhos subsequentes a estes fossem realizados com objetivo de desmembrar a heterose, resultados inconclusivos foram obtidos, principalmente no que se refere à participação da epistasia no controle genético da heterose (LU; ROMERO-SEVERSON; BERNARDO, 2003; FRASCAROLI et al., 2007, GARCIA et al., 2008; SCHÖN et al., 2010). A partir desses resultados duas questões ficaram em aberto: (i) o ajuste dos efeitos epistáticos apenas em QTLs de grande efeito pode não desenhar a real contribuição desse fenômeno no controle genético da heterose, requerendo dessa maneira a

aplicação de modelos complexos para o estudo da distribuição da epistasia ao longo do genoma (MOORE, 2005; HUGHES, 2005; LEHNER, 2011); (ii) a epistasia pode não ser de caráter-específica e poderia ser mais bem estudada dentro do conceito da multiplicidade de QTLs pleiotrópicos uma vez que esses dois fenômenos (pleiotropia e epistasia) podem ser indissociáveis (SCHNELL; COCKERHAM, 1992; CHEVERUD et al., 2004; WOLF et al., 2005; WOLF et al., 2006; PAVLICEV et al., 2008; MACKAY; STONE; AYROLES, 2009).

2.2 Epistasia

O termo “epistasia” foi definido por Bateson (1909) para descrever um desvio da segregação mendeliana no qual um alelo - de um determinado loco - inibe a ação de outro alelo localizado em loco distinto; ao invés de ocorrerem interações alélicas apenas intra-locos, também foi sugerido por Bateson (1909) que os alelos dos diferentes locos também interagem entre si, sub ou superexpressando determinados genes.

Atualmente existe grande conflito a respeito da definição da epistasia, convergindo em certa confusão de interpretação (CORDELL, 2002). Os geneticistas têm utilizado o termo epistasia para descrever três eventos distintos: relação funcional entre os genes, ordenamento genético em rotas regulatórias e diferenças quantitativas em alelos de efeito específico. Phillips (2008) classificou a epistasia em três tipos: (i) Epistasia funcional – que descreve a ação molecular onde proteínas ou outros produtos gênicos interagindo entre si em rotas metabólicas -, (ii) epistasia composicional - se refere à ideia clássica de epistasia onde um alelo de um determinado loco interfere na expressão de outro loco, sendo então restrita ao indivíduo - e (iii) epistasia estatística – que foi adotada por Sir Ronald Fisher, que considera a epistasia como um desvio médio resultante da combinação de alelos nos diferentes locos em uma população de

genótipos. Esse último tipo é o mais estudado na genética quantitativa, pois, devido à quantidade de locos envolvidos na adaptabilidade de um indivíduo torna-se quase impossível estimar a epistasia composicional para todos os locos (PHILLIPS, 2008).

Para entender a base da epistasia estatística precisa-se entender o modelo que a descreve e para isto necessita-se distinguir interação intragênica e intergênica.

Considerando um loco com dois alelos - A e a - cujas frequências são p e q . Sob o modelo Fisheriano, tem que o efeito de uma substituição alélica é dado pelo coeficiente angular de uma regressão linear do número de alelos que o indivíduo carrega sobre seu valor genotípico ou pela derivada da média populacional. Na figura abaixo, dois modelos distintos podem ser utilizados para obtenção dos efeitos alélicos:

- a) uma regressão linear (linha pontilhada);
- b) ou um modelo fatorial $3k$, onde o i -ésimo nível possui efeito linear (azul) ou quadrático ou de curvatura (vermelho).

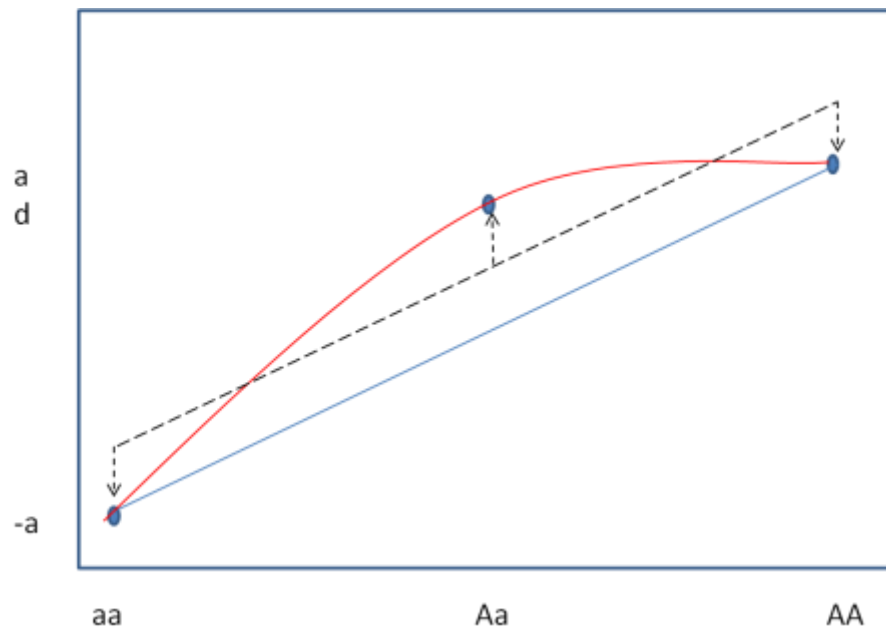


Figura 1 Mudança no valor genotípico de um indivíduo por ocasião da mudança do alelo *a* pelo alelo *A*

O modelo linear de Fisher utiliza uma reta que minimiza as distâncias entre o genótipo e valor genotípico; observa-se, porém, um desvio (seta) decorrente do ajuste de um modelo linear em um modelo quadrático. Assim, tomando a média geral desses indivíduos como: $m = a(p - q) + 2pqd$, a mudança no valor genotípico pela troca do alelo *a* pelo *A* pode ser dada por: $a + (q - p)d = \alpha$ = coeficiente angular da regressão linear. Por consequência, o efeito do alelo *A* é dado por $q\alpha$ e o efeito do alelo *a* é dado por $-p\alpha$. Então, se o indivíduo possui 2 alelos *A* (*AA*) o preditor genotípico torna-se $2q\alpha$; se tiver apenas um alelo *A* tem-se $(q-p)\alpha$ e nenhum alelo favorável (*aa*), o preditor linear é $-2p\alpha$. Assim, o valor genético aditivo ou valor reprodutivo do indivíduo é dado pela soma dos alelos que esse indivíduo possui. Faz-se necessário distinguir esse efeito do efeito *b*, obtido pelo contraste dos indivíduos homozigotos.

Deve-se frisar que o “efeito de um alelo” apresenta um desvio em relação ao valor genotípico do indivíduo por ajustar um modelo linear em uma resposta claramente quadrática quando a dominância está presente. O erro do modelo de Fisher pode ser quantificado como:

$$a = m + 2q\alpha + \delta_1;$$

$$d = m + (q - p)\alpha + \delta_2 \quad e$$

$$e - a = m - 2p\alpha + \delta_3,$$

onde δ é o erro do ajuste do modelo. Para o indivíduo AA o valor genotípico considerando um modelo linear agora se torna $a = m + 2q\alpha + \delta_1$; então $a - m - 2q\alpha = \delta_1$. Se substituindo os valores tem-se que:

$$a - (p - q)a - 2pqd - 2q[a + (q - p)d] = \delta_1 \rightarrow 2q[a - pd - a - qd + pd] = \delta_1$$

pode-se afirmar que esse desvio pode ser dado por:

$$-2q^2d = \delta_1 \quad (1)$$

Esse desvio é classificado de desvio de dominância, cuja interpretação não deve ser confundida com o valor genotípico do indivíduo heterozigoto ou o efeito de dominância obtido da diferença do indivíduo heterozigoto em relação à média. Para os demais indivíduos os desvios de regressão são dados por:

$$Aa = d = m + (q - p)\alpha + \delta_2 \text{ então:}$$

$$2pqd = \delta_2, \quad (2)$$

e $-a = m - 2p\alpha + \delta_3$, então

$$-2p^2d = \delta_3. \quad (3)$$

Para uma população com frequência alélica $p=0.5$, o valor genotípico esperado na forma linear seria, $AA = m + a - 0,5d$ $Aa = m + 0,5d$ e $aa = m - a - 0,5d$.

Outra forma de abordar esse problema é considerar que cada gene é um fator de um fatorial 3^k , e no caso específico de um gene tem-se a seguinte descrição.

$$\begin{bmatrix} AA \\ Aa \\ aa \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} m \\ \alpha \\ \delta \end{bmatrix} \rightarrow \text{Tornando a matrix ortogonal teremos:}$$

$$\begin{bmatrix} AA \\ Aa \\ aa \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1/3 \\ 1 & 0 & -2/3 \\ 1 & -1 & 1/3 \end{bmatrix} \begin{bmatrix} m \\ \alpha \\ \delta \end{bmatrix} \quad (4)$$

A terceira coluna é o efeito quadrático do gene A referente ao número de alelos favoráveis nos indivíduos (número de alelos elevado ao quadrado), e representa a curvatura da Figura 1. O efeito dos desvios de dominância pode ser repadronizado por 3 e tem-se que:

$$\begin{bmatrix} AA \\ Aa \\ aa \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} m \\ \alpha \\ \delta \end{bmatrix} \quad (5)$$

Assim é fácil verificar que a última coluna $(1 \ -2 \ 1)^T$ representa o contraste quadrático ou a curvatura dos desvios de dominância da Figura 1. Por exemplo, se o efeito de dominância é ausente, é esperado que $AA - Aa = Aa - aa$. Então o contraste de curvatura pode ser dado por $AA - 2Aa + aa = 0$ $(1 - 2 + 1)$, ou $0 = -AA + 2Aa - aa = (-1 + 2 - 1)$, dependendo da concavidade da parábola. Uma rápida inspeção desse contraste revela um teste de hipótese para curvatura ou dominância a saber: $H_0 : Aa = (AA + aa)/2$. Ou seja, essa é a hipótese de não curvatura ou sem a presença do efeito de dominância. Assim, $H_0 : 2Aa = AA + aa \rightarrow AA - 2Aa + aa = 0$.

Esse contraste representa uma concavidade positiva. No entanto, pode-se expressar o mesmo contraste utilizando a concavidade da Figura 1. Neste caso: $H_0 : 2Aa = AA + aa \rightarrow -AA + 2Aa - aa = 0$. Substituindo na matriz:

$$\begin{bmatrix} AA \\ Aa \\ aa \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} m \\ \alpha \\ \delta \end{bmatrix} \quad (6)$$

Cocherham (1954) demonstrou que, em populações com frequências alélicas p e q , o k -ésimo fator terá a sua matriz de delineamento ponderada pelas frequências alélicas. Nesse caso, pode-se substituir os valores acima por:

$$\begin{bmatrix} AA \\ Aa \\ aa \end{bmatrix} = \begin{bmatrix} 1 & 2q & -2q^2 \\ 1 & q-p & 2pq \\ 1 & -2p & -2p^2 \end{bmatrix} \begin{bmatrix} m \\ \alpha \\ d \end{bmatrix} \quad (7)$$

Nessa matriz pode-se notar que cada regressor representa a parte conhecida das frequências alélicas no modelo linear de Fisher, ou a parte conhecida do valor genotípico. Porém os parâmetros desconhecidos são α e d , mas o geneticista quer saber na verdade a e d . Se $p=q=0,5$ tem-se que:

$$\begin{bmatrix} AA \\ Aa \\ aa \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1/2 \\ 1 & 0 & 1/2 \\ 1 & -1 & -1/2 \end{bmatrix} \begin{bmatrix} m \\ a \\ d \end{bmatrix} \quad (8)$$

Esse modelo foi denominado de métrica F_2 , pois considera as frequências alélicas obtidas em populações derivadas do cruzamento de duas linhagens endogâmicas. Com essa matriz, fica claro que $AA = m + a - 0,5d$, $Aa = m + 0,5d$ e $aa = m - a - 0,5d$, ou seja, o mesmo valor genotípico obtido na regressão linear de Fisher. Com esses resultados pode-se interpretar o efeito de dominância de duas formas: i) interpretação residual de Fisher, ou seja, que o efeito de dominância existe apenas como um resíduo de um ajuste linear incorreto, mas útil, ii) interpretação de Cocherham, ou seja, dominância existe na sua natureza e o modelo correto para descrevê-la é um modelo polinomial.

A solução para os efeitos aditivos e de dominância pode ser dada por:

$$\begin{bmatrix} m \\ a \\ d \end{bmatrix} = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \\ -1/2 & 1 & -1/2 \end{bmatrix} \begin{bmatrix} AA \\ Aa \\ aa \end{bmatrix} \quad (9)$$

Neste caso em particular o contraste corresponde exatamente aos contrastes da estimativa dos efeitos aditivos e dominantes na escala simplificada. No entanto, deve-se entender que isto só vale para frequência alélica igual a 0,5. Para outros valores de frequência alélica os contrastes estimam o efeito da substituição alélica e efeito de dominância.

Pode-se concluir até o momento que o valor genotípico de um indivíduo expresso na forma linear possui como parâmetros desconhecidos o efeito de substituição alélica e o desvio. Utilizando a matriz ponderada de Cockerham, pode-se verificar que o desvio pode ser expresso em termos de efeito de dominância (daí o termo desvio de dominância), porém, o efeito da substituição alélica não pode ser expresso em função dos efeitos aditivos, a menos na frequência alélica 0,5. No entanto, o valor predito do indivíduo devido à matriz de delineamento ponderada pelas frequências alélicas representa o valor genético aditivo ou valor reprodutivo e sua variância depende a transformação do efeito da substituição alélica. Essa distinção faz-se necessária, pois comumente o geneticista se atrapalha em definir que parâmetros ele está estimando no seu modelo linear.

Definida a matriz de contrastes ortogonais para a solução dos efeitos aditivos e dominantes, pode-se definir um modelo para dois genes; ou um fatorial 3^2 . Sob essa ótica pode-se considerar os genes A e B, cujo modelo fatorial simplificado é dado por:

$$\begin{aligned} y &= A_i + A_o + B_l + B_q \\ y &= \alpha_1 + d_1 + \alpha_2 + d_2 \end{aligned} \tag{10}$$

Onde y é o valor genotípico dos indivíduos e l e Q são os efeitos lineares e quadráticos dos fatores A e B (Figura 1). Assim, utilizando a denominação de Cockerham tem-se que para as combinações dos genes A e B:

$$\begin{bmatrix} AABB \\ AABb \\ AAbb \\ AaBB \\ AaBb \\ Aabb \\ aaBB \\ aaBb \\ aabb \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1/2 & 1 & -1/2 \\ 1 & 1 & -1/2 & 0 & 1/2 \\ 1 & 1 & -1/2 & -1 & -1/2 \\ 1 & 0 & 1/2 & 1 & -1/2 \\ 1 & 0 & 1/2 & 0 & 1/2 \\ 1 & 0 & 1/2 & -1 & -1/2 \\ 1 & -1 & -1/2 & 1 & -1/2 \\ 1 & -1 & -1/2 & 0 & 1/2 \\ 1 & -1 & -1/2 & -1 & -1/2 \end{bmatrix} \begin{bmatrix} m \\ a \\ d_1 \\ a_2 \\ d_2 \end{bmatrix}$$

Em um modelo sem epistasia a resposta esperada desses genes pode ser observada na figura abaixo:

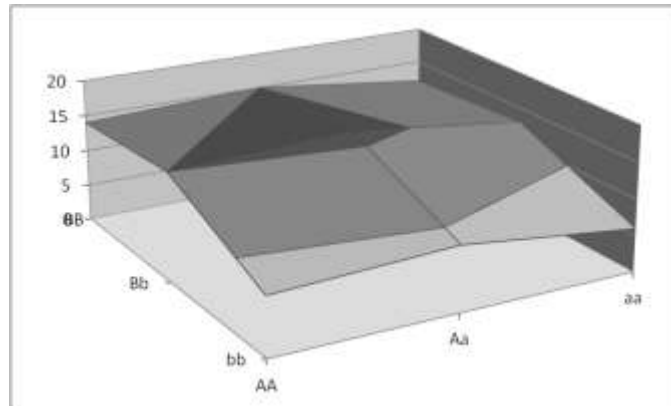


Figura 2 Resposta aditiva no valor genotípico devido aos efeitos dos genes A e B

Nesta figura fica evidente que há resposta à adição do gene B é linear e a superfície de resposta, em cada nível, é plana.

No entanto, se existe uma interação entre os alelos desses genes, ou de forma genérica, uma interação entre genes, uma curvatura é esperada e tem-se a seguinte distorção na figura:

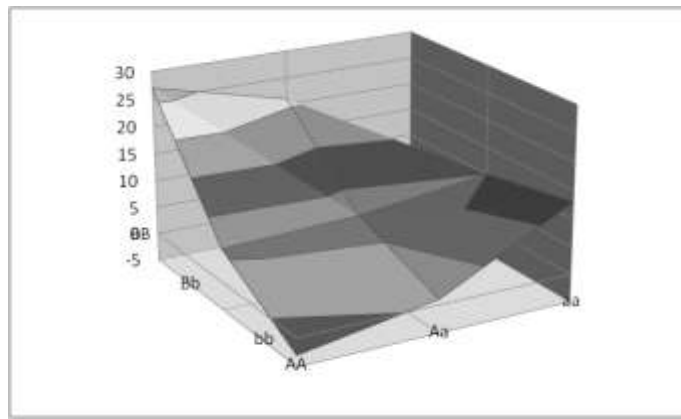


Figura 3 Resposta não aditiva no valor genotípico devido à interação dos alelos dos genes A e B

Sob o modelo de interação ou epistasia pode-se verificar que o modelo aditivo (A+B) é ineficaz para estimar os efeitos da interação ou, como denominado por Cordell, (2002), a epistasia estatística. Assim, pode-se utilizar um modelo fatorial 3^2 completo que é dado por:

$$\begin{aligned}
 y &= A_l + A_q + B_l + B_q + A_l B_l + A_l B_q + A_q B_l + A_q B_q \\
 y &= \alpha_1 + d_1 + \alpha_2 + d_2 + \alpha_1 \alpha_2 + \alpha_1 d_2 + \alpha_2 d_1 + d_1 d_2 \\
 y &= (\alpha_1 + \alpha_2) + (d_1 + d_2) + \alpha_1 \alpha_2 + \alpha_1 d_2 + \alpha_2 d_1 + d_1 d_2 \quad (11)
 \end{aligned}$$

A nova matriz será dada por:

$$\begin{array}{l}
 \left[\begin{array}{l}
 AABB \\
 AABb \\
 AAbb \\
 AaBB \\
 AaBb \\
 Aabb \\
 aaBB \\
 aaBb \\
 aabb
 \end{array} \right] = \left[\begin{array}{cccccccc}
 1 & 1 & -1/2 & 1 & -1/2 & 1 & -1/2 & -1/2 & 1/4 \\
 1 & 1 & -1/2 & 0 & 1/2 & 0 & 1/2 & 0 & -1/4 \\
 1 & 1 & -1/2 & -1 & -1/2 & -1 & -1/2 & 1/2 & 1/4 \\
 1 & 0 & 1/2 & 1 & -1/2 & 0 & 0 & 1/2 & -1/4 \\
 1 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 1/4 \\
 1 & 0 & 1/2 & -1 & -1/2 & 0 & 0 & -1/2 & -1/4 \\
 1 & -1 & -1/2 & 1 & -1/2 & -1 & 1/2 & -1/2 & 1/4 \\
 1 & -1 & -1/2 & 0 & 1/2 & 0 & -1/2 & 0 & -1/4 \\
 1 & -1 & -1/2 & -1 & -1/2 & 1 & 1/2 & 1/2 & 1/4
 \end{array} \right] \left[\begin{array}{l}
 m \\
 \alpha_1 \\
 d_1 \\
 \alpha_2 \\
 d_2 \\
 \alpha_1 \alpha_2 \\
 \alpha_1 d_2 \\
 \alpha_2 d_1 \\
 d_1 d_2
 \end{array} \right]
 \end{array}$$

Esse modelo epistático pode ser aplicado em qualquer população estruturada. Porém, em populações desestruturadas ou pedigrees complexos, a frequência alélica normalmente é desconhecida e, neste caso, Cockerham (1980) sugere um modelo fatorial alternativo quando os fundadores são conhecidos.

Considerando uma população que será gerada por k fundadores (por exemplo, linhagens). Considerando que a linhagem I possui o genótipo AAbb. Essa linhagem será cruzada com outra linhagem, aqui chamada de linhagem II, com genótipo aaBB. Na definição de Cockerham (1980), o modelo ideal para descrever o genótipo do descendente é um modelo fatorial com quatro fatores, cada fator descrevendo o efeito do alelo que esse indivíduo recebeu do parental I e II. Como em espécies diploides um indivíduo recebe um alelo do parental masculino (m) e outro alelo do parental feminino (f), pode-se descrever o valor genotípico de um indivíduo proveniente de um cruzamento qualquer sem efeito de epistasia como:

$$G_{ijkl} = m + \alpha_i^f + \alpha_j^m + \delta_{ij}^{mf} + \alpha_k^f + \alpha_l^m + \delta_{kl}^{mf} \quad (12)$$

Onde $i j$ representam os alelos A e a respectivamente e $k l$ representam B e b . Uma vez que o efeito de dominância - como já foi demonstrado - representa a interação intra-alélica, a epistasia - embora tratada como interação gênica - nada mais é do que a interação interalélica. Sendo assim, pode-se expandir esse modelo 4^k para todas as interações de segunda ordem (quadráticas) e tem-se que:

$$\begin{aligned}
G_{ijkl} = & m + \alpha_{A_i}^f + \alpha_{A_j}^m + \delta_{A_i A_j}^{mf} + \alpha_{B_k}^f + \alpha_{B_l}^m + \delta_{B_k B_l}^{mf} \\
& + (\alpha\alpha)_{A_i B_k}^{ff} + (\alpha\alpha)_{A_i B_l}^{fm} + (\alpha\alpha)_{A_j B_k}^{mf} + (\alpha\alpha)_{A_j B_l}^{mm} \\
& + (\alpha\delta)_{A_i (B_k B_l)}^{m(mf)} + (\alpha\delta)_{A_j (B_k B_l)}^{f(mf)} + (\delta\alpha)_{(A_i A_j) B_k}^{(mf)f} + (\delta\alpha)_{(A_i A_j) B_l}^{(mf)m} + \\
& (\delta\delta)_{(A_i A_j)(B_k B_l)}^{(mf)(mf)}
\end{aligned} \tag{13}$$

Esse modelo fatorial é tratado de forma diferente do modelo (11), pois o parâmetro é o efeito do alelo e não da substituição alélica. Dessa forma, esse modelo apresenta uma maior quantidade de parâmetros a ser estimada. Em linhagens, como os alelos são iguais, ou seja, $\alpha_{A_i}^f = \alpha_{A_j}^m \rightarrow \alpha_{A_i}^f + \alpha_{A_j}^m = 2\alpha_{A_i}^f = 2A_a$, pode-se simplificar o modelo acima por:

$$G = m + 2(A_a + A_b) + D_a + D_b + 4(AA)_{ab} + 2(AD)_{ab} + 2(DA)_{ab} + (DD)_{ab} \tag{14}$$

Para caracteres quantitativos, assumindo um modelo infinitesimal, tem-se que $(A_a + A_b)$ pode ser confundido em A , assim como:

$$D_a + D_b = D; (AA)_{ab} = AA;$$

$$2(AD)_{ab} + 2(DA)_{ab} = 2[(AD)_{ab} + (DA)_{ab}] = 2AD \text{ e } (DD)_{ab} = DD.$$

Assim, sob o modelo infinitesimal:

$$G = m + 2A + D + 4AA + 2AD + DD \quad (15)$$

Não há neste modelo o genótipo para o k-ésimo gene. Em cruzamentos controlados, porém, pode-se delinear a matriz dos efeitos aditivos e dominantes com base no número de alelos fundadores, ou como no exemplo dado, o número de linhagens endogâmicas envolvidas nos cruzamentos. Deve-se substituir a matriz desconhecida pela esperança dessa matriz, baseada no pedigree dos cruzamentos. Com base no modelo 15, tem-se que para o i-ésima linhagem parental envolvida no cruzamento: $\sum_1 \rho_i A_i = 2A_i \rightarrow \sum_1 \rho_i / 2 = 1$, onde $\rho_i / 2$ representa a probabilidade do indivíduo receber o alelo ρ_i da i-ésima linhagem. Da mesma forma $\sum_{ij} \varpi_{ij} D_{ij} = D_{ij} \rightarrow \sum_{ij} \varpi_{ij} = 1$, onde ϖ_{ij} é a probabilidade que um alelo venha do parental i e o outro alelo do parental j .

Então o modelo 15 pode ser novamente descrito como:

$$G = m + \sum_i \rho_i A_i + \sum_{ij} \varpi_{ij} D_{ij} + \left(\sum_i \rho_i A_i \right)^2 + \left(\sum_i \rho_i A_i \right) \left(\sum_{ij} \varpi_{ij} D_{ij} \right) + \left(\sum_{ij} \varpi_{ij} D_{ij} \right)^2$$

$$G = m + \sum_i \rho_i A_i + \sum_{ij} \varpi_{ij} D_{ij} + \sum_{ij} \rho_i \rho_j (AA)_{ij} + \sum_i \sum_{jk} \rho_i \varpi_{jk} (AD)_{i(jk)} + \sum_{ij} \sum_{kl} \varpi_{ij} \varpi_{kl} (DD)_{ij} \quad (16)$$

O modelo epistático 16 é simplesmente um modelo multiplicativo infinitesimal envolvendo inúmeros efeitos alélicos e de dominância. Também mostra que apenas com a matriz de pedigree pode-se prever todos os efeitos epistáticos e de dominância, conhecendo os fundadores. Por exemplo, considerando o seguinte pedigree com quatro linhagens:

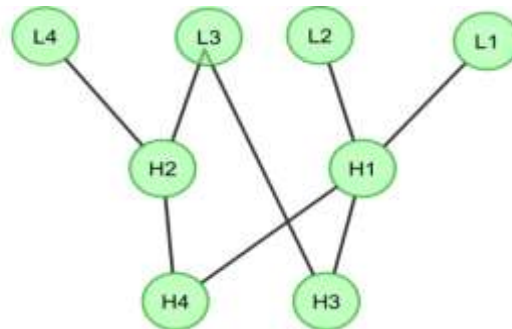


Figura 4 Esquema de produção dos híbridos simples, triplos e duplos de milho

No modelo acima, L representa as linhagens genitoras e H os híbridos resultantes. Nesse pequeno pedigree três tipos comuns de híbridos são observados: H2 e H1 são híbridos simples, H3 é um híbrido triplo e H4 um híbrido duplo. O modelo completo epistático pode ser obtido sabendo a probabilidade ρ dos alelos fundadores serem transmitidos para cada um dos híbridos. Por exemplo, supondo que a linhagem L1 tenha dois alelos semelhantes AA, e a linhagem L2 tenha dois alelos aa. Então a probabilidade do híbrido H1 herdar A de L1 é 1, e a probabilidade de herdar a de L2 é também 1. Utilizando essa lógica pode-se calcular a probabilidade de que cada indivíduo herde os alelos dos seus genitores. Isso é calculado na Tabela 1.

Tabela 1 Probabilidade ρ dos híbridos herdarem os alelos dos genitores $p_1 p_2 p_3$ e p_4 , considerando quatro fundadores L1 L2, L3 e L4

	P1	P2		p_1	p_2	p_3	p_4	ρ_1	ρ_2	ρ_3	ρ_4	ϖ_1	ϖ_2	ϖ_3	ϖ_4	ϖ_{12}	ϖ_{13}	ϖ_{14}	ϖ_{23}	ϖ_{24}	ϖ_{34}
L1	-	-	<i>m</i>	1	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0
			<i>f</i>	1	0	0	0														
L2	-	-	<i>m</i>	0	1	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0
			<i>f</i>	0	1	0	0														
L3	-	-	<i>m</i>	0	0	1	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0
			<i>f</i>	0	0	1	0														
L4	-	-	<i>m</i>	0	0	0	1	0	0	0	2	0	0	0	1	0	0	0	0	0	0
			<i>f</i>	0	0	0	1														
H1	L1	L2	<i>m</i>	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0
			<i>f</i>	0	1	0	0														
H2	L3	L4	<i>m</i>	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1
			<i>f</i>	0	0	0	1														
H3	H1	L3	<i>m</i>	0,5	0,5	0	0	0,5	0,5	1	0	0	0	0	0	0	0,5	0	0,5	0	0
			<i>f</i>	0	0	1	0														
H4	H1	H2	<i>m</i>	0,5	0,5	0	0	0,5	0,5	0,5	0,5	0	0	0	0	0	0,25	0,25	0,25	0,25	0
			<i>f</i>	0	0	0,5	0,5														

Nessa tabela, pode-se observar que, quando são analisados apenas híbridos simples, a matriz de delineamento dos efeitos aditivos (ρ) e dominantes (ω) se assemelha às matrizes das capacidades gerais e específicas de combinação, respectivamente. Como esses efeitos normalmente estão confundidos com os efeitos epistáticos se pode separar esses efeitos e verificar realmente o quanto os efeitos aditivos contribuem para o desempenho dos híbridos de milho.

2.3 Parentesco entre indivíduos, similaridade por estado, similaridade por descendência e suas aplicações no melhoramento genético da cultura do milho

O melhoramento genético normalmente utiliza uma estrutura populacional conhecida para estimar componentes da variância. No entanto, em alguns delineamentos de cruzamentos, ao se assumir um modelo fixo – ou mesmo misto na análise de variância –, ignora-se por completo a covariância entre parentes e assume-se, por exemplo, a média fenotípica como objeto de seleção. Essa abordagem foi, e ainda é aplicada em programas de melhoramento comerciais e em instituições públicas e tem como desvantagem a não incorporação de toda informação genotípica entre parentes; ou seja, os métodos convencionais não sabem lidar com a seguinte questão: O que o genótipo de um indivíduo pode fornecer de informação a respeito do comportamento de seu parente? A resposta a esta pergunta foi dada por Piepho et al. (2008). Esses autores apresentaram diversos exemplos nos quais a informação de parentesco pode contribuir na seleção genética e estimativa dos componentes da variância.

No melhoramento animal e vegetal, o cálculo do parentesco entre indivíduos pode ser obtido a partir do coeficiente de coancestria, que é a probabilidade do indivíduo herdar alelos idênticos dos pais ou ter alelos

idênticos por descendência (FALCONER; MACKAY, 1990). No contexto dos modelos mistos, os valores genotípicos são considerados aleatórios e a introdução da matriz de informação de parentesco faz com que os eBLUPs (melhor preditor linear não viesado) sejam correlacionados. Assim, quanto maior for o grau de parentesco entre indivíduos, maior será a recuperação intergenotípica.

Para que ocorra um ganho com a utilização da informação de parentesco é necessário que as estimativas desse parâmetro sejam acuradas. O problema é que esses valores são obtidos com várias restrições que nem sempre podem ser alcançadas na prática (NYQUIST, 1991; HOLLAND; NYQUIST; CERVANTES-MARTINEZ, 2003). Dentre essas restrições, Piepho et al. (2008) destaca: (i) os genótipos são oriundos de uma população idealizada, (ii) na população base os indivíduos não são relacionados, e (iii) a população base está em equilíbrio de fase ou mesmo em equilíbrio de Hardy-Weinberg. Resumindo, assume-se que essa população não sofreu seleção. Como se pode observar, essa suposição é difícil de ser atendida, principalmente por estas populações estarem em constante seleção.

Assim, se a estrutura populacional é ignorada, a variância genética pode ser viciada e os termos “Melhor” e “Não Viesado” utilizados nos BLUPs dos valores genotípicos não são mais garantidos.

A utilização da informação de pedigree na matriz das equações de modelos mistos tem melhorado de forma significativa a predição dos valores genotípicos dos candidatos à seleção por considerar seu histórico reprodutivo. Conforme argumenta Lynch e Wash (1998), mesmo com as limitações dos métodos de “verossimilhança” – no contexto subentende-se “modelos mistos” como um caso particular da inferência de verossimilhança, essa abordagem permite a introdução de matrizes de parentesco que servem como ferramentas para correção de vieses devido à seleção. No caso do melhoramento vegetal fica

evidente que a utilização da informação de parentesco pode contribuir grandemente para a acurácia de predição, pois normalmente faz-se uso de populações provenientes de cruzamentos bi-parentais e, no caso do melhoramento genético do milho, cruzamentos dialélicos com linhagens endogâmicas.

Henderson (1952; 1977) propôs a utilização de modelos mistos em dialelos. Posteriormente, Bernardo (1993) utilizou a matriz de parentesco obtida pelo coeficiente de coancestria e marcadores moleculares para predição do desempenho de híbridos de milho. Os resultados obtidos por esse autor incentivaram diversos outros trabalhos que buscaram predizer o desempenho de híbridos de milho utilizando marcadores moleculares para a construção das matrizes de parentesco (SCHRAG et al., 2009; 2010 ; FERREIRA et al., 2010).

Esses exemplos demonstram a importância da utilização de marcadores moleculares no melhoramento genético do milho. No entanto, alguns autores destacam que a aplicação dos dados de similaridade genética devem ser aplicados com ressalvas (JORDAN et al., 2005; FERREIRA et al., 2010).

A similaridade genética é função apenas do comprimento da molécula de DNA observada durante o processo de eletroforese, ou seja, observa-se apenas o “estado” alélico do gene. Portanto, a similaridade genética é uma medida de identidade entre dois indivíduos de acordo com o “estado” observado do gene sem considerar de quais genitores esses dois indivíduos herdaram tal gene. Por outro lado, a identidade por descendência (IPD) considera não apenas o estado do gene, mas também a probabilidade de que determinados alelos sejam herdados de genitores comuns. Neste caso, pode-se construir estruturas de matrizes de (co) variâncias que estejam de acordo com as pressuposições da genética quantitativa e do modelo genético. Dentro desse contexto, sugere-se utilizar as similaridades genéticas corrigidas pela IPD; embora em alguns casos

essa correção possa trazer prejuízos à análise por adotar matrizes aditivas e dominantes como identidades (FERREIRA et al., 2010).

2.4 Predição do desempenho de híbridos de milho

Em um programa de melhoramento de milho a busca por genótipos com alta heterose envolve a aplicação de sistemas de cruzamentos cujo produto final são os híbridos; esta é, sem dúvida alguma, a principal atividade desempenhada há mais de um século no melhoramento genético de milho. Com efeito, pode-se dizer que os “dialelos” são os principais métodos aplicados na cultura e que todos os demais métodos de melhoramento são suplementares a essa finalidade última que é a obtenção de híbridos altamente produtivos.

A quantidade de cruzamentos a ser realizada a partir de um conjunto de linhagens pode ser tornar um problema quando o número desses parentais é alto. Mesmo quando existem grupos heteróticos bem definidos, a quantidade de cruzamentos pode ser tornar tão elevada que a realização de todos os cruzamentos torna-se inviável. Por exemplo, considerando 50 linhagens pertencentes a dois grupos e suponha que essas linhagens serão inter cruzadas em um esquema fatorial 50x50. Nestas circunstâncias são possíveis 2500 híbridos, o que não é problema em um programa de médio e grande porte. Mas, se o melhorista possui 100 linhagens por grupo, a síntese de híbridos já se torna onerosa (10.000). Por outro lado, se um melhorista não possui grupos heteróticos bem definidos, o cruzamento biparental de 100 linhagens, em um sistema dialélico, resultaria em 4450 híbridos e tornaria a avaliação desse material também dispendiosa. Dessa forma, pode-se observar que o direcionamento de cruzamentos é um fator crucial em um programa de melhoramento moderno.

Essa otimização normalmente é feita de forma empírica, ou seja, o melhorista utiliza-se da experiência e direciona vários cruzamentos e, por consequência, descarta vários outros. Genericamente, quanto mais cruzamentos são descartados, mais híbridos com alto potencial são rejeitados e mesmo um melhorista experiente não pode prever de forma acurada a estrutura genética do híbrido e suas características apenas com base nos atributos das linhagens pertencentes aos grupos heteróticos. A solução para este entrave seria a utilização de métodos que permitam prever o desempenho de híbridos de forma acurada e assim aperfeiçoar o programa pelo descarte de cruzamentos indesejáveis.

Esse problema foi estudado primeiramente por Jenkins (1934). Esse autor utilizou o desempenho de híbridos simples para a predição de híbridos duplos. Jenkins sugeriu quatro métodos de predição denominados métodos A, B, C e D. Dentre esses quatro métodos propostos, o método B foi o que apresentou melhores resultados. Esse método representa a predição com base nas linhagens não parentais e foi durante vários anos utilizados para a predição de híbridos duplos (HALLAUER; CARENA; MIRANDA-FILHO, 2010).

Posteriormente, Eberhart (1964) e Cockerham (1967) aprimoraram as ideias de Jenkins, lapidando seus modelos genéticos muito frouxos. Todos esses modelos visavam à predição de híbridos duplos e triplos, mas eram de pouca utilidade para predição de híbridos simples.

Com o advento do uso de marcadores moleculares vislumbrou-se a possibilidade de prever híbridos simples utilizando a distância genética entre as linhagens. Uma vez que a heterose pode ser descrita em função da distância genética e do efeito de dominância quando desconsidera a epistasia (MELCHINGER, 1999), pode-se conceber que quanto mais divergente os genitores, maior a heterose esperada.

Utilizando essa teoria simplificada da genética quantitativa, vários estudos foram conduzidos a partir da década de 90 visando prever o comportamento de híbridos com base na distância genética (SMITH et al., 1994; LANZA et al., 1997; BARBOSA et al., 2003; BALESTRE et al., 2008a; DEVI; SINGH, 2011; DI et al., 2012) . No entanto, os resultados obtidos por essa abordagem são inconclusivos e, de maneira geral, não são confiáveis para ser utilizados em larga escala em programas de melhoramento.

Uma nova abordagem foi desenvolvida paralelamente baseada nos trabalhos de Henderson (1952) utilizando modelos mistos e pedigrees complexos (BERNARDO, 1994). Esse autor unificou as ideias de modelos mistos e marcadores moleculares na predição de híbridos simples.

Segundo Bernardo (1994; 1995), uma das utilizações do melhor preditor linear não viesado (BLUP) no melhoramento genético vegetal consiste em prever o desempenho genético de híbridos que não estejam presentes ou que tenham sido perdidos. Esse autor considera relevante o uso de técnicas preditivas como uma alternativa viável para diminuir os custos e também o tempo de avaliação de híbridos pela diminuição substancial do número de híbridos a serem avaliados. A partir de covariâncias genéticas entre as linhagens a serem cruzadas, pode-se prever o desempenho genotípico de alguns dos seus cruzamentos, sem que eles tenham sido obtidos e avaliados. Essa técnica tem grande potencial para auxiliar o melhoramento genético, dado ao elevado número de híbridos possíveis de serem obtidos com poucas linhagens. Apesar da correlação entre o desempenho genotípico do cruzamento predito e o seu valor fenotípico serem baixa em muitos casos, essa abordagem apresenta vantagens (BERNARDO, 1994). Dessa forma, a utilização do BLUP tem mostrado eficiência considerável para tal finalidade (BERNARDO, 1994; 1995; 1996a; b).

Nesse contexto, Nejati-Javeremi, Smith e Gibson (1997) sugere o emprego da similaridade no estado, pois, segundo esses autores, esta possui a

vantagem de se utilizar toda a semelhança alélica existente entre os indivíduos avaliados. No entanto, pelo fato de levar em consideração a co-ocorrência de bandas (0 e 1) e não a probabilidade de genitores possuírem alelos herdados de um genitor em comum, a similaridade no estado pode superestimar o parentesco.

Diante disso, Bernardo (1993) propõe o uso de um coeficiente de coancestria que utiliza informação de similaridade genética, no intuito de corrigir possíveis vieses da similaridade no estado. Entretanto, esse método somente é válido para dialelos parciais nos quais não ocorrem cruzamentos endogâmicos, podendo ser utilizada somente quando se tem informação prévia a respeito de grupos heteróticos. Como alternativa a essa restrição, alguns autores (LYNCH E RITLAND, 1999; QUELLER; GOODNIGHT, 1989; RITLAND, 1996 e WANG, 2002) sugerem procedimentos não viesados para estimação conjunta de parentesco por meio de marcadores moleculares que podem ser utilizados em dialelos completos e parciais. Segundo Balestre (2009), o coeficiente de parentesco proposto por Lynch e Ritland (1999) pode ser utilizado em dialelos completos, pois as medidas de parentesco obtidas por esse estimador levam em consideração todos os indivíduos em avaliação. Além disso, mesmo utilizando dados moleculares, esse estimador de parentesco considera a probabilidade condicional de que dois indivíduos possuam alelos herdados de genitores em comum. De modo que somente indivíduos que atendam a essa premissa participam da predição de híbridos não avaliados. Dessa forma, aqueles indivíduos que possuem parentesco igual a zero não contribuirão na predição, concorrendo para resultados mais coerentes.

Mais recentemente, Schrag et al. (2009, 2010) propuseram a substituição da matriz de parentesco obtida por marcadores pelo próprio valor das marcas dos genótipos. Essa metodologia foi denominada TEAM ou “*total effects of associated markers*” ou ETMA (efeito total das marcas associadas). Nessa abordagem, a matriz do delineamento da CGC e CEC são substituídas pela

matriz de marcadores moleculares e utiliza-se uma regressão aleatória (RR-BLUP) onde valores genéticos são atribuídos a cada marca. Nesse contexto, o valor genotípico do híbrido é recuperado por meio dos efeitos totais dos marcadores como se fosse uma seleção genômica. Esse método consiste basicamente de se obter os valores genéticos das marcas a partir de uma população de treino e, posteriormente, se valida a predição no conjunto genético restante que pode ser tanto uma geração avançada ou a própria população de treino via validação cruzada.

Esse método tem se mostrado eficiente na predição de híbridos e recentemente Massman et al. (2012) estendeu essa abordagem para a seleção genômica propriamente dita e alcançou resultados encorajadores. Assim, o atual estado da arte de predição de híbridos em programas de melhoramento é a aplicação direta do ETMA ou seleção genômica em cruzamentos dialélicos e predição dos híbridos não testados.

3 CONCLUSÕES

Pelo o que foi discutido na revisão de literatura, pode-se concluir que o fenômeno da epistasia pode ser estudado em delineamentos úteis ao melhorista, tais como dialelos, trialelos e tetraalelos.

Também se verifica que existe uma grande incerteza no que diz respeito à predição de híbridos de milho utilizando marcadores moleculares e que novas abordagens devem ser sugeridas, principalmente aquelas em que a quantidade de informação seja maior do que a simples utilização de distâncias genéticas.

REFERÊNCIAS

AMORIM, E. P. et al. Genetic distance based on SSR and grain yield of inter an intrapopulational maize single cross hybrids. **Maydica**, Bergamo, v. 51, p. 507-513, 2006. Disponível em: <http://www.maydica.org/articles/51_507.pdf>. Acesso em: 20 set. 2012.

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

BALESTRE, M. et al. Genetics distances estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids. **Genetics and Molecular Research**, Ribeirão Preto, v. 7, n. 1, p. 65-73, jan. 2008a. Disponível em: <<http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=LILACS&lang=p&nextAction=lnk&exprSearch=553772&indexSearch=ID>>. Acesso em: 20 set. 2012.

BALESTRE, M. et al. Potential of maize single-cross hybrids for extraction of inbred lines using the mean components and mixed models with microsatellite marker information. **Genetics and Molecular Research**, Ribeirão Preto, v. 7, n. 4, p. 1106-1118, Oct. 2008b.

Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19048489>>. Acesso em: 20 set. 2012.

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**: international journal of plant breeding research, Berlin, v. 125, n. 3, p. 479-493, Mar. 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22437491>>. Acesso em: 20 set. 2012.

BARBOSA, A. M. M. et al. Relationship of intra- and interpopulation tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. **Euphytica**, Netherlands, v. 130, n. 1, p. 87-99, 2003. Disponível em: <<http://link.springer.com/article/10.1023%2FA%3A1022381008721?LI=true>>. Acesso em: 20 set. 2012.

BATESON, W. **Mendel's principles of heredity**. Cambridge: University Cambridge, 1909.

BERNARDO, R. Genetic models for predicting maize performance in unbalanced yield trial data. **Crop Science**, Madison, v. 35, n. 1, p. 141-147, Jan./Feb. 1995.

BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science**, Madison, v. 34, n. 1, p. 20-25, Jan./Feb. 1994. Disponível em: <<https://www.agronomy.org/publications/cs/abstracts/34/1/CS0340010020?access=0&view=pdf>>. Acesso em: 20 set. 2012.

BERNARDO, R. Best linear unbiased prediction of maize single cross performance. **Crop Science**, Madison, v. 36, n. 1, p. 50-56, Jan./Feb. 1996a. Disponível em: <<https://www.crops.org/publications/cs/abstracts/36/1/CS0360010050?access=0&view=pdf>>. Acesso em: 20 set. 2012.

BERNARDO, R. Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. **Crop Science**, Madison, v. 36, n. 4, p. 862-866, July/Aug. 1996b. Disponível em: <<http://agris.fao.org/agris-search/search/display.do?f=1997%2FUS%2FUS97016.xml%3BUS9730326>>. Acesso em: 20 set. 2012.

BERNARDO, R. Estimation of coefficient of coancestry using molecular markers in maize. **Theoretical and Applied Genetics**: international journal of plant breeding research, Berlin, v. 85, n. 8, p. 1055-1062, Feb. 1993. Disponível em: <<http://link.springer.com/article/10.1007%2FBF00215047?LI=true>>. Acesso em: 20 set. 2012.

BETRÁN, F. J. et al. Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and nonstress environments. **Crop Science**, Madison, v. 43, n. 3, p. 797-806, May/June 2003. Disponível em: <<https://www.soils.org/publications/cs/abstracts/43/3/797?access=0&view=pdf>>. Acesso em: 20 set. 2012.

CHEVERUD, J. et al. Pleiotropic effects on mandibular morphology II: differential epistasis and genetic variation in morphological integration. **Journal of Experimental Zoology**, Hoboken, v. 302, n. 5, p. 424–435, Sept. 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15384169>>. Acesso em: 20 set. 2012.

COCKERHAM, C. C. An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariances among Relatives When Epistasis Is Present. **Genetics**. Baltimore, v. 39 n.6, p. 859–882, Nov 1954.

COCKERHAM, C. C. Implications of genetic variances in a hybrid breeding program. **Crop Science**, Madison, v. 1, p. 47–52, 1961.

COCKERHAM, C. C. Prediction of double crosses from single crosses. L. **Theoretical and Applied Genetics**: international journal of plant breeding research, Berlin, v. 37, n. 4, p. 160-169, 1967. Disponível em: <<http://link.springer.com/article/10.1007%2FBF00329523>>. Acesso em: 20 set. 2012.

COCKERHAM, C. C. Random and fixed effects in plant genetics. **Theoretical and Applied Genetics**: international journal of plant breeding research, Berlin, v. 56, n. 3, p. 119-131, May 1980. Disponível em: <<http://link.springer.com/article/10.1007%2FBF00265082?LI=true>>. Acesso em: 20 set. 2012.

COCKERHAM, C. C.; ZENG, Z. B. Design III with marker loci. **Genetics**, Baltimore, v. 143, n. 3, p. 1437-1456, July 1996. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1207411/>>. Acesso em: 20 set. 2012.

COMSTOCK, R. E.; ROBINSON, H. F. Estimation of average dominance genes. In: GOWEN, J. W. (Ed.). **Heterosis**. Ames: Iowa State College, 1952. Cap. 30, p. 494-516.

CORDELL, H. J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. **Humans Molecular Genetics**, v. 11, n. 20, p. 2463–2468, 2002. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/12351582>>. Acesso em: 20 set. 2012.

DEVI, P.; SINGH, N. K. Heterosis, molecular diversity, combining ability and their interrelationships in short duration maize (*Zea mays* L.) across the environments. **Euphytica**: netherlands journal of plant breeding, Wageningen, v. 178, n. 1, p. 71–81, 2011. Disponível em: <http://www.researchgate.net/publication/225159428_Heterosis_molecular_diversity_combining_ability_and_their_interrelationships_in_short_duration_maize_%28Zea_mays_L.%29_across_the_environments>. Acesso em: 20 set. 2012.

DHLIWAYO, T. et al. Combining ability, genetic distances, and heterosis among elite cimmyt and ita tropical maize inbred lines. **Crop Science**, Madison, v. 49, n. 4, p. 1201-1210, July 2009. Disponível em: <<https://www.crops.org/publications/cs/abstracts/49/4/1201?access=0&view=article>>. Acesso em: 20 set. 2012.

DI, R. et al. Predictive potential of microsatellite markers on heterosis of fecundity in crossbred sheep. **Molecular Biology Reports**, v. 39, n. 3, p. 2761-2766, 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21674186>>. Acesso em: 20 set. 2012.

EAST, E. M. Heterosis. **Genetics**, Baltimore, v. 21, n. 4, p. 375-397, 1936.

EBERHART, S. A. Theoretical relations among single, three-way, and double cross hybrids.

Biometrics, v. 20, n. 3, p. 522–539, 1964. Disponível em: <<http://www.jstor.org/discover/10.2307/2528493?uid=2129&uid=2&uid=70&uid=4&sid=21101596760741>>. Acesso em: 20 set. 2012.

FALCONER, D. S.; MACKAY, T. F. **Introduction to quantitative genetics**. 4. ed. Essex: Longman, 1990. 463 p.

FERREIRA, D. V. et al. Prediction of maize hybrid performance using similarity in state and similarity by descent information. **Genetics and Molecular Research**, v. 9, n. 4, p. 2381-2394, Dec. 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21161887>>. Acesso em: 20 set. 2012.

FRASCAROLI, E. et al. Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. **Genetics**, Baltimore, v. 176, n. 1, p. 625–644, May 2007. Disponível em: <<http://www.genetics.org/content/176/1/625.abstract>>. Acesso em: 20 set. 2012.

GARCIA, A. A. et al. Quantitative trait mapping and the genetic basis of heterosis in maize and rice. **Genetics**, Baltimore, v. 180, n. 3, p. 1707-1724, Nov. 2008. Disponível em: <<http://www.genetics.org/content/180/3/1707.abstract>>. Acesso em: 20 set. 2012.

HALLAUER, A. R.; CARENA, M. J.; MIRANDA-FILHO, J. B. **Quantitative genetics in maize breeding**. New York: Springer, 2010. 680 p.

HENDERSON, C. R. Specific and general combining ability. In: GOWEN, J. W. (Ed.). **Heterosis**. Ames: Iowa State College, 1952. p. 352–370.

HENDERSON, C. R. Prediction of the merits of single crosses. **Theoretical and Applied Genetics**, Berlin, v. 49, n. 6, p. 273–282, 1977. Disponível em: <http://link.springer.com/article/10.1007%2F978-3-642-75133-7_11>. Acesso em: 20 set. 2012.

HOCHHOLDINGER, F.; HOECKER, N. Towards the molecular basis of Heterosis. **Trends in Plant Science**, v. 12, n. 9, p. 427-432, Aug./Sept. 2007. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17720610>>. Acesso em: 20 set. 2012.

HOLLAND, J. B.; NYQUIST, W. E.; CERVANTES-MARTINEZ, C. T. Estimating and interpreting heritability for plant breeding: an update. **Plant Breeding Reviews**, v. 22, p. 9–112, 2003. Disponível em: <http://www.ars.usda.gov/research/publications/publications.htm?seq_no_115=134062>. Acesso em: 20 set. 2012.

HUGHES, T. R. Universal epistasis analysis. **Nature Genetics**, London, v. 37, n. 5, p. 457-458, May 2005. Disponível em: <<http://www.nature.com/ng/journal/v37/n5/full/ng0505-457.html>>. Acesso em: 20 set. 2012.

JENKINS, M. T. Methods of estimating the performance of double crosses in corn. **Journal of American Society of Agronomy**, Washington, v. 26, n. 3, p. 199-204, Mar. 1934. Disponível em: <<https://www.crops.org/publications/aj/abstracts/26/3/AJ0260030199?access=0&view=pdf>>. Acesso em: 20 set. 2012.

JONES, D. F. The effects of inbreeding and crossbreeding upon development. **Connecticut Agricultural Experimental Station Bulletin**, Washington, v. 297, p. 5-100, 1918. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1091458/>>. Acesso em: 20 set. 2012.

JORDAN, D. R. et al. Comparison of identity by descent and identity by state for detecting genetic regions under selection in a sorghum pedigree breeding program. **Molecular Breeding**, Dordrecht, v. 14, n. 4, p. 441-454, Jan. 2005. Disponível em: <<http://www.ingentaconnect.com/content/klu/molb/2005/00000014/00000004/0000901?crawler=true>>. Acesso em: 20 set. 2012.

LANZA, L. L. et al. Genetic distance of inbred lines and prediction of maize singlecross performance using RAPD markers. **Theoretical and Applied Genetics**, Berlin, v. 94, n. 8, p. 1023-1030, June 1997. Disponível em: <<http://link.springer.com/article/10.1007%2Fs001220050510?LI=true>>. Acesso em: 20 set. 2012.

LEGESSE, B. W. et al. Relationship between hybrid performance and AFLP based genetic distance in highland maize inbred lines. **Euphytica**, Netherlands, v. 162, p. 313-323, 2008. Disponível em: <<http://repository.up.ac.za/handle/2263/9804>>. Acesso em: 20 set. 2012.

LEHNER, B. Molecular mechanisms of epistasis within and between genes. **Trends in Genetics**, London, v. 27, n. 8, p. 323-331, Aug. 2011. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21684621>>. Acesso em: 20 set. 2012.

LU, H.; ROMERO-SEVERSON, J.; BERNARDO, R. Genetics basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. **Theoretical and Applied Genetics**, Berlin, v. 107, n. 3, p. 494-502, Aug. 2003. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/12759730>>. Acesso em: 20 set. 2012.

LYNCH, M.; RITLAND, K. Estimation of pairwise relatedness with molecular markers. **Genetics**, Baltimore, v. 152, n. 4, p. 1753-1766. Aug. 1999. Disponível em: <<http://www.genetics.org/content/152/4/1753.abstract>>. Acesso em: 20 set. 2012.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980 p.

MACKAY, T. F.; STONE, E. A.; AYROLES, J. F. The genetics of quantitative traits: challenges and prospects. **Nature Reviews Genetics**, London, v. 10, n. 8, p. 565-577, Aug. 2009. Disponível em: <<http://www.nature.com/nrg/journal/v10/n8/abs/nrg2612.html>>. Acesso em: 20 set. 2012.

MASSMAN, J. M. et al. Genomewide predictions from maize single-cross data. **Theoretical and Applied Genetics**, v. 126, n. 1, p. 13-22, ago. 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22886355>>. Acesso em: 20 set. 2012.

MELCHINGER, A. E. et al. The role of epistasis in the manifestation of heterosis: a systems-oriented approach. **Genetics**, Baltimore, v. 177, n. 3, p. 1815-1825, Nov. 2007. Disponível em: <<http://www.genetics.org/content/177/3/1815>>. Acesso em: 20 set. 2012.

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

MOORE, J. H. A global view of epistasis. **Nature Genetics**, London, v. 37, n. 1, p. 13-14, Jan. 2005. Disponível em: <<http://www.nature.com/ng/journal/v37/n1/full/ng0105-13.html>>. Acesso em: 20 set. 2012.

NEJATI-JAVEREMI, A.; SMITH, C.; GIBSON, J. P. Effect of total allelic relationship on accuracy of evaluation and response to selection. **Journal of Animal Science**, Champaign, v. 75, n. 7, p. 1738-1745, July 1997. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9222829>>. Acesso em: 20 set. 2012.

NYQUIST, W. E. Estimation of heritability and prediction of selection response in plant populations. **CRC Critical Reviews in Plant Science**, Boca Raton, v. 10, n. 3, p. 235–322, 1991. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/07352689109382313#preview>>. Acesso em: 20 set. 2012.

PAVLICEV, M. et al. Genetic variation in pleiotropy: differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. **Evolution**, Malden, v. 62, n. 1, p. 199-213, Jan. 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18005158>>. Acesso em: 20 set. 2012.

PHILLIPS, P. C. Epistasis-the essential role of gene interactions in the structure and evolution of genetic systems. **Nature Review Genetics**, v. 9, n. 11, p. 855–867, Nov. 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2689140/>>. Acesso em: 20 set. 2012.

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, Netherlands, v. 161, n. 1-2, p. 209-228, May 2008. Disponível em: <<http://link.springer.com/article/10.1007%2Fs10681-007-9449-8?LI=true>>. Acesso em: 20 set. 2012.

POWERS, L. An expansion of Jones' theory for the explanation of heterosis. **The American Naturalist**, Chicago, v. 78, n. 776, p. 275-280, May/June 1944. Disponível em: <<http://www.jstor.org/discover/10.2307/2458261?uid=2129&uid=2&uid=70&uid=4&sid=21101587197233>>. Acesso em: 20 set. 2012.

QUELLER, D. C.; GOODNIGHT, K. F. Estimating relatedness using genetic markers. **Evolution**, Houston, v. 43, n. 2, p. 258-275, Mar. 1989. Disponível em: <<http://www.jstor.org/discover/10.2307/2409206?uid=2129&uid=2&uid=70&uid=4&sid=21101587197233>>. Acesso em: 20 set. 2012.

RAWLINGS, J.; COCKERHAM, C. C. Analysis of double cross hybrid populations. **Biometrics**, Washington, v. 18, n. 2, p. 229-244, June 1962a. Disponível em: <<http://www.jstor.org/discover/10.2307/2527461?uid=2129&uid=2&uid=70&uid=4&sid=21101587197233>>. Acesso em: 20 set. 2012.

RAWLINGS, J.; COCKERHAM, C. C. Triallel analysis. **Crop Science**, Madison, v. 2, p. 228-231, 1962b.

RITLAND, K. Estimators for pairwise relatedness and individual inbreeding coefficients. **Genetical Research**, Cambridge, v. 67, n. 2, p. 175-185, Apr. 1996. Disponível em: <<http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=1794000>>. Acesso em: 20 set. 2012.

SCHNELL, F. W.; COCKERHAM, C. C. Multiplicative vs. arbitrary gene action in heterosis. **Genetics**, Baltimore, v. 131, n. 2, p. 461-469, June 1992. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/1644280>>. Acesso em: 20 set. 2012.

SCHÖN, C. C. et al. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 321-332, Jan. 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19911156>>. Acesso em: 20 set. 2012.

SCHRAG, T. A. et al. Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. **Theoretical Applied Genetics**, Berlin, n. 118, n. 4, p. 741-751, Feb. 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19048224>>. Acesso em: 20 set. 2012.

SCHRAG, T. A. et al. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. **Theoretical Applied Genetics**, Berlin, v. 120, n. 2, p. 451-461, Jan. 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19916002>>. Acesso em: 20 set. 2012.

SHULL, G. H. A pure line method of corn breeding. **American Breeders Association Report**, Washington, v. 4, p. 296-301, 1909.

SMITH, S. et al. Association among inbred lines of maize using RFLP and amplification technologies (AFLP and AP-PCR), and correlations with pedigree, F1 yield and heterosis. **Maize Genetics Newsletter**, Oxford, v. 68, p. 45, 1994.

STUBER, C. W. et al. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. **Genetics**, Baltimore, v. 132, n. 3, p. 823-839, Nov. 1992. Disponível em: <<http://www.genetics.org/content/132/3/823>>. Acesso em: 20 set. 2012.

TROYER, A. F. Adaptedness and heterosis in corn and mule hybrids. **Crop Science**, Madison, v. 46, n. 2, p. 528-543, Mar. 2006

WANG, J. An estimator for pairwise relatedness using molecular markers. **Genetics**, Baltimore, p. 160, n. 3, p. 1203-1215, Mar. 2002. Disponível em: <<http://www.genetics.org/content/160/3/1203>>. Acesso em: 20 set. 2012.

WOLF, J. B. et al. Epistatic leiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. **Genetics**, Baltimore, v. 171, n. 2, p. 1–12, Oct. 2005.

WOLF, J. B. et al. The contribution of epistatic pleiotropy to the genetic architecture of covariation among polygenic traits in mice. **Evolution Development**, Malden, v. 8, n. 5, p. 468–476, Sept./Oct. 2006. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/16925682>>. Acesso em: 20 set. 2012.

SEGUNDA PARTE - ARTIGOS

**ARTIGO 1 Controle genético do desempenho de híbridos de milho
 utilizando pedigrees complexos e marcadores de
 microssatélites**

Artigo redigido nas normas da Revista Científica Theoretical and Applied
Genetics

Controle genético do desempenho de híbridos de milho utilizando pedigrees complexos e marcadores de microssatélites

Wagner Mateus Costa Melo¹, Marcio Balestre², Renzo Garcia Von Pinho¹

¹ Departamento de Agricultura, Universidade Federal de Lavras. C.P 3037. Lavras. Minas Gerais, Brasil.

²Departamento de Ciências Exatas, Universidade Federal de Lavras, C.P 3037. Lavras, Minas Gerais, Brasil

Resumo

O objetivo do autor com este trabalho foi determinar a importância dos efeitos epistáticos na heterose do milho por meio de um único modelo, com base em análises de di, tri e tetra-alelos e com a informação de marcadores de microssatélites. Para isto utilizou-se 51 linhagens de diferentes origens obtendo-se dessa forma 6 híbridos duplos, 14 híbridos triplos e 58 híbridos simples, totalizando 78 híbridos. Setenta e nove marcadores de microssatélites foram utilizados na genotipagem das 51 linhagens. Esses marcadores estão distribuídos ao longo dos 10 grupos de ligação do milho. Essa informação foi utilizada na construção da matriz de informação de parentesco. De acordo com este trabalho foi possível desmembrar a heterose e entender melhor a participação dos efeitos epistáticos na mesma quando analisa conjuntamente híbridos simples, duplos e triplos em um mesmo modelo. Sob esta abordagem, é possível compreender a heterose nos seus diversos componentes. Também foi possível verificar que com a utilização de marcadores moleculares houve uma tendência na melhora da acurácia e dos efeitos genéticos epistáticos e de dominância. Assim, com os recursos computacionais disponíveis atualmente, técnicas estatísticas avançadas e uso de marcadores moleculares é possível entender melhor a heterose e os fatores que a controlam.

Palavras-Chave: Milho; Melhoramento genético; Heterose; modelos mistos; SSR; identidade no estado.

Genetic control of maize hybrid performance using complex pedigrees and microsatellite markers

Wagner Mateus Costa Melo¹, Marcio Balestre², Renzo Garcia Von Pinho¹

¹Departamento de Agricultura, Universidade Federal de Lavras. Cp 3037. Lavras. Minas Gerais, Brasil.

²Departamento de Ciências Exatas, Universidade Federal de Lavras, C.P 3037. Lavras, Minas Gerais, Brazil

Abstract

The purpose of this study was to determine the importance of epistatic effects on maize heterosis by means of a single model, based on analyses of di-, tri- and tetrallelic and also with the information from microsatellite markers. For that reason, 51 lines from different backgrounds were used, thus obtaining 6 double hybrids, 14 three-way hybrids and 58 single hybrids, for a total of 78 hybrids. Seventy-nine microsatellite markers were used in the genotyping of the 51 lines. These markers are distributed along the 10 maize linkage groups. This information was used in construction of the relationship information matrix. According to this study, it was possible to break down heterosis and better understand the participation of epistatic effects in it when jointly analyzing single, double and three-way hybrids under the same model. Under this approach, it is possible to understand heterosis in its diverse components. It was also possible to verify that with the use of molecular markers there was a trend toward improvement of the accuracy of the epistatic and dominance genetic effects. Thus, with the computational resources currently available, advanced statistical techniques and use of molecular markers, it is possible to better understand heterosis and the factors that control it.

Keywords: Maize; Genetic breeding; Heterosis; Mixed models; SSR; Identity in the state.

Introdução

O desmembramento de caracteres complexos tem sido um dos principais desafios dos geneticistas (Mackay et al., 2009). A utilização de ferramentas moleculares, técnicas estatísticas avançadas e uso de redes genéticas tem contribuído enormemente para melhor compreensão dos efeitos de locos relacionados a caracteres quantitativos e suas inter-relações (Balestre et al., 2012).

A compreensão do funcionamento de caracteres complexos pode ser útil em várias áreas da genética. No melhoramento de plantas alguns fenômenos tais como a epistasia e pleiotropia são de extrema importância, principalmente no que se refere à heterose (Balestre et al., 2012).

Recentemente, alguns trabalhos têm buscado desmembrar a heterose em caracteres quantitativos em milho (Melchinger et al., 2007; Larièpe, et al., 2012; Riedelsheimer et al., 2012). No entanto, resultados conflitantes são obtidos e as hipóteses que tentam descrever esse fenômeno vão desde a simples complementação de locos de efeito aditivo - hipótese dominante (Troyer, 2006) - a explicações mais elaboradas envolvendo dominância completa, sobredominância, colinearidade genética, epistasia, pleiotropia e ação multiplicativa de genes relacionado a componentes de produção (Schnell e Cockerham, 1992; Hochholdinger e Hoecker, 2007).

No que concerne à epistasia e sua participação na heterose em milho, estudos de mapeamento de QTLs não têm verificado grande importância desse fenômeno (Lu, et al., 2003, Frascaroli et al., 2007; Garcia et al. 2008; Shon et al. 2010). Esses resultados podem surgir pela utilização de modelos restritos os quais apenas as epistasias em QTLs de grande efeito são levadas em consideração. A especulação de que a epistasia pode contribuir para o controle genético da heterose remonta o trabalho de Powers (1944) e Van Der Veen (1959), passando por Schnell e Cockerham (1992) e Zeng et al. (2005) e se

justifica pela recente demonstração de Melchinger et al. (2007), que derivou o controle da heterose de forma analítica e sugeriu que a epistasia pode ter papel fundamental no controle genético da heterose em milho, principalmente no que diz respeito à interferência da epistasia do tipo aditiva x aditiva que compõe esse efeito. Conforme destacado por Melchinger et al. (2007), se os dois parentais são totalmente contrastantes é possível que a heterose possa ocorrer mesmo que o efeito de dominância seja nulo nos dois locos que controlam o caráter. Além disso, linhagens com epistasia aditiva-aditiva positiva tendem a subestimar o papel da dominância na heterose e o contrário também é verdadeiro.

No melhoramento genético de milho, o conhecimento da epistasia pode ajudar no processo de hibridação - onde esse efeito é passível de fixação - uma vez que em populações esse efeito é altamente influenciado pelas frequências alélicas e normalmente confundido na variância aditiva (Hill et al., 2008). Contudo, é sabido que em dialelos tradicionais a ausência de contrastes ortogonais para obtenção dos efeitos epistáticos e alto efeito de multicolinearidade impede que os efeitos epistáticos sejam estimados de forma independente (Roso et al., 2005). Dessa forma, a capacidade geral de combinação termina por confundir os efeitos aditivos e interação do tipo aditiva x aditiva e impede que esses efeitos sejam estudados de forma isolada.

Da mesma forma, a capacidade específica de combinação carrega todos os efeitos epistáticos e mais os desvios de dominância e, baseado nisto, a decomposição da heterose apenas em efeitos CEC e CGC pode ser pouco informativa. Se a CGC absorve os efeitos aditivos e epistasia aditiva-aditiva, é fácil imaginar que a CEC representa um conjunto de efeitos de interações inter e intra alélicas que, embora complexas e altamente confundidas, podem ser fixadas nos híbridos tornando a sua decomposição de extrema importância no direcionamento de cruzamentos.

Uma forma de superar o problema do confundimento dos efeitos de dominância com efeitos epistáticos é tomar o dialelo tradicional como uma forma de pedigree complexo que pode envolver acasalamentos entre linhagens, entre híbridos e entre linhagens com híbridos gerando um sistema intrincado de cruzamentos (Cockerham 1961, Rawlings e Cockerham 1962a, 1962b). Esses autores propuseram a análise de tri e tetra-alelos para estimar alguns efeitos epistáticos por meio de contrastes de efeitos. Porém, uma vez que se considera esse sistema dentro de um único pedigree, pode-se utilizar uma única análise para estimar todos os efeitos epistáticos decorrentes destes acasalamentos diminuindo enormemente o efeito de confundimento por assumir um modelo de encolhimento (Roso et al., 2005).

Neste trabalho buscou-se determinar a importância dos efeitos epistáticos na heterose do milho por meio de um único modelo, com base em análises de di, tri e tetra-alelos e ainda com a informação de marcadores de microssatélites.

Material e Métodos

Informações experimentais

No inverno de 2010, no campo experimental de Uberlândia foram produzidos os híbridos, descritos no item materiais utilizados, para a utilização nos ensaios. O experimento foi conduzido no modelo experimental de blocos incompletos DBI com duas repetições e cada parcela possuía quatro linhas de 5m com espaçamento entre linhas de 70 cm. Foi avaliada a produção de grãos neste ensaio em 7 locais distribuídos na região sul do Brasil (Vacaria RS plantio 10/10/2010, Abelardo Luz SC plantio 14/10/2010, Arapoti PR plantio 16/10/2010, Candoi PR plantio 24/10/2010, Canoinhas PR plantio 26/10/2010, Castro PR plantio 28/10/2010 e Ponta Grossa PR plantio 02/11/2010). O peso de

grãos/parcela foi corrigido para umidade de 13% e convertido em t.ha⁻¹. As adubações de plantio e de cobertura foram feitas de acordo com a recomendação de adubação para cada área, os tratos culturais foram feitos para controle de lagarta do cartucho (*Spodoptera Frugiperda*) e da espiga (*Helicoverpa Zea*) e também para o controle do mato.

Materiais utilizados

Cinquenta e uma linhagens de diferentes origens foram utilizadas nos cruzamentos. Dos cruzamentos dessas linhagens foram obtidos 6 híbridos duplos, 14 híbridos triplos e 58 híbridos simples, totalizando 78 híbridos. O sistema de cruzamentos pode ser observado na Figura 1. As testemunhas utilizadas foram P30F53Herculex, DKB390 YieldGard, DKB240 e AG8025.

Marcadores de microssatélites

Setenta e nove marcadores de microssatélites foram utilizados na genotipagem das 51 linhagens (Tabela 1). Esses marcadores estão distribuídos ao longo dos 10 grupos de ligação do milho. Esses marcadores foram utilizados na construção das matrizes de similaridade genética aplicando o coeficiente de Jaccard conforme citado por (Baestres et al. 2008). As matrizes aditivas foram obtidas pela similaridade no estado sem qualquer correção. A matriz dos efeitos de dominância e matrizes epistáticas foram calculadas respectivamente por:

$$G_d = \frac{1}{4} \left[p(x_1^m = x_2^m) p(x_1^f = x_2^f) + p(x_1^m = x_2^f) p(x_1^f = x_2^m) \right] \quad (1)$$

$$G_{aa_{kl}} = g_{a_{ii'}} g_{a_{jj'}}; \quad k = 1, 2, \dots, n_a; \quad l = 1, 2, \dots, n_a \quad (2)$$

$$G_{ad_{kl}} = g_{a_{ii'}} g_{d_{jj'}}; \quad k = 1, 2, \dots, n_{ad}; \quad l = 1, 2, \dots, n_{ad}$$

$$G_{dd_{kl}} = g_{d_{ii'}} g_{d_{jj'}}; \quad k = 1, 2, \dots, n_{dd}; \quad l = 1, 2, \dots, n_{dd} \quad (3)$$

onde i, i', j e j' se referem ao i -ésimo efeito comum entre as epistasias e g_a, g_d à probabilidade de que $i=i'$ e $j=j'$ dada pela similaridade genética aditiva e de dominância entre os genitores. Essas matrizes epistáticas são semelhantes às obtidas pelo produto de Hadamard das matrizes G_a e G_d no modelo animal individual.

Modelo gamético em dialelos

O modelo gamético, também conhecido como modelo fatorial de Cockerham, pode ser aplicado em qualquer pedigree onde a frequência alélica dos genitores é conhecida (Cockerham, 1980). Essa modelagem foi aplicada por Cockerham (1961) e Rawlings e Cockerham (1961a, 1961b) em cruzamentos de linhagens onde as frequências alélicas das progênies são conhecidas e assumem valores iguais a $(p = q = 1/2)$; $(p = 1/2, q = s = 1/4)$ e $(p = q = s = u = 1/4)$ que são casos especiais dos delineamentos de dialelos, trialelos e tetraalelos, respectivamente.

Na definição de Cockerham (1980), o modelo ideal para descrever o genótipo do híbrido descendente do cruzamento de linhagens endogâmicas seria através de um fatorial com quatro fatores; cada fator descrevendo o efeito do alelo que esse indivíduo recebeu do parental n . Em espécies diploides, se um indivíduo recebe um alelo do parental masculino (m) e outro alelo do parental feminino (f), o valor genotípico desse indivíduo proveniente de um cruzamento qualquer, sem efeito de epistasia, é dado por:

$$G_{ijkl} = m + \alpha_i^f + \alpha_j^m + \delta_{ij}^{mf} + \alpha_k^f + \alpha_l^m + \delta_{kl}^{mf}$$

onde i, j representam os alelos A e a respectivamente e k, l representam B e b . Uma vez que o efeito de dominância representa a interação intra-alélica, a

epistasia nada mais é do que a interação interalélica (embora usualmente tratada como interação gênica). Sendo assim, pode-se expandir esse modelo 4^k para todas as interações de segunda ordem (quadráticas) e tem-se que:

$$\begin{aligned} G_{ijkl} = & m + \alpha_{A_i}^f + \alpha_{A_j}^m + \delta_{A_i A_j}^{mf} + \alpha_{B_k}^f + \alpha_{B_l}^m + \delta_{B_k B_l}^{mf} \\ & + (\alpha\alpha)_{A_i B_k}^{ff} + (\alpha\alpha)_{A_i B_l}^{fm} + (\alpha\alpha)_{A_j B_k}^{mf} + (\alpha\alpha)_{A_j B_l}^{mm} \\ & + (\alpha\delta)_{A_i (B_k B_l)}^{m(mf)} + (\alpha\delta)_{A_j (B_k B_l)}^{f(mf)} + (\delta\alpha)_{(A_i A_j) B_k}^{(mf) f} + (\delta\alpha)_{(A_i A_j) B_l}^{(mf) m} + \\ & (\delta\delta)_{(A_i A_j) (B_k B_l)}^{(mf) (mf)} \end{aligned}$$

Dessa forma, esse modelo apresenta uma maior quantidade de parâmetros a ser estimado do que o modelo animal, quando se considera mais de 2 genitores. Em linhagens completamente endogâmicas, se a probabilidade de os alelos serem idênticos no estado é igual a 1, ou seja, $\alpha_{A_i}^f = \alpha_{A_j}^m \rightarrow \alpha_{A_i}^f + \alpha_{A_i}^m = 2\alpha_{A_i}^f = 2a_a$, pode-se simplificar o modelo acima por:

$$G = m + 2(a_a + a_b) + d_a + d_b + 4(aa)_{ab} + 2(ad)_{ab} + 2(da)_{ab} + (dd)_{ab}$$

Para caracteres quantitativos, assumindo um modelo infinitesimal, tem-se que $(a_a + a_b)$ pode ser confundido em a^* , assim como: $d_a + d_b = d^*$, $(aa)_{ab} = aa^*$; $2(ad)_{ab} + 2(da)_{ab} = 2[(ad)_{ab} + (da)_{ab}] = 2a^* d^*$ e $(dd)_{ab} = d^* d^*$. Assim, sob o modelo infinitesimal:

$$G = m + 2a^* + d^* + 4a^* a^* + 2a^* d^* + d^* d^* \quad (4)$$

Nesse modelo o genótipo para o k-ésimo gene não é observado. Porém em cruzamentos controlados pode-se delinear a matriz dos efeitos aditivos e dominantes com base no número de linhagens genitoras envolvidas nos

cruzamentos. Então, deve-se substituir a matriz desconhecida pela esperança dessa matriz, baseada no pedigree dos cruzamentos. Com base no modelo 4, tem-se para a i -ésima linhagem parental envolvida no cruzamento:

$$\sum_1 \rho_i a_i = 2a_i \rightarrow \sum_1 \rho_i / 2 = 1, \text{ onde } \rho_i / 2 \text{ representa a probabilidade do}$$

indivíduo receber o alelo ρ_i da i -ésima linhagem. Da mesma forma

$$\sum_{ij} \varpi_{ij} d_{ij} = d_{ij} \rightarrow \sum_{ij} \varpi_{ij} = 1, \text{ onde } \varpi_{ij} \text{ é a probabilidade que um alelo venha do}$$

parental i e o outro alelo do parental j .

O modelo 4 pode ser novamente descrito como:

$$G = m + \sum_i \rho_i a_i + \sum_{ij} \varpi_{ij} d_{ij} + \left(\sum_i \rho_i a_i \right)^2 + \left(\sum_i \rho_i a_i \right) \left(\sum_{ij} \varpi_{ij} d_{ij} \right) + \left(\sum_{ij} \varpi_{ij} d_{ij} \right)^2$$

$$G = m + \sum_i \rho_i a_i + \sum_{ij} \varpi_{ij} d_{ij} + \sum_{ij} \rho_i \rho_j (aa)_{ij} + \sum_i \sum_{jk} \rho_i \varpi_{jk} (ad)_{i(jk)} + \sum_{ij} \sum_{kl} \varpi_{ij} \varpi_{kl} (dd)_{ij}$$
(5)

Em forma matricial, para todo o pedigree, tem-se:

$$G = \mu j + Z_1 a + Z_2 d + \Omega aa + \Psi ad + \Phi dd$$
(6)

onde Z_1, Z_2, Ω, Ψ e Φ são matrizes obtidas pela probabilidade de determinado híbrido herdar os alelos das respectivas linhagens.

Neste estudo, um modelo linear misto foi adotado para a produção de grãos. Esse modelo é dado por:

$$y = X\beta + Z_1 a + Z_2 d + \Omega aa + \Psi ad + \Phi dd + \varepsilon$$
(7)

Onde y é o vetor das médias fenotípicas, X é a matriz dos efeitos fixos (média geral, ambientes e blocos), Z_1, Z_2, Ω, Ψ e Φ são as matrizes dos efeitos aleatórios, onde Z_1 – Efeitos aditivos, Z_2 – Desvios de dominância, e os demais são respectivamente as epistasias aditiva/aditiva, aditiva/dominante e dominante/dominante e \mathcal{E} é o erro experimental.

A solução para os efeitos genéticos pode ser obtida utilizando as equações dos modelos mistos já descritas por Henderson (1984). Para estimação das variâncias referentes aos efeitos aditivos, dominantes e epistáticos utilizou-se o algoritmo EM (*expectation-maximization*), com dois passos de Newton-Raphson.

Modelo com e sem a informação de marcadores

O modelo 7 é um modelo geral de análise no qual se assume que os genitores são não aparentados, ou seja, que $P(a_a = a_b) = 0$. Neste caso, as matrizes de (co)variâncias dos efeitos genéticos introduzidas nas equações de modelos mistos são dadas por:

$$G_a = I\sigma_a^2; G_d = I\sigma_d^2; G_{aa} = I\sigma_{aa}^2; G_{ad} = I\sigma_{ad}^2; G_{dd} = I\sigma_{dd}^2.$$

No entanto, se algumas linhagens são provenientes de origens semelhantes, é possível que $P(a_a = a_b) \neq 0$ e exista covariância nos efeitos alélicos (aditivos, dominantes e epistáticos). Assim, uma medida dessa covariância pode ser dada pelas matrizes G_a, G_d, G_{aa}, G_{ad} e G_{dd} , já descritas anteriormente no item com informação de marcadores.

Para seleção dos modelos, utilizou-se o critério bayesiano de informação (BIC), testando todas as combinações possíveis dos efeitos aditivos, dominantes e epistáticos. (Carlborg e Haley, 2004).

Os modelos com e sem a informação de parentesco foram comparados em relação à acurácia preditiva dada por: $r_i = \sqrt{1 - d_{ii} \frac{\sigma_e^2}{\sigma_j^2}}$ onde i é o i -ésimo elemento da diagonal da matriz das equações de modelos mistos e j é o j -ésimo componente da variância genética.

Para verificar se os valores genotípicos e componentes da variância genética estão realmente descrevendo o desempenho dos híbridos, uma análise foi realizada via SAS proc mixed para esses mesmos dados, porém considerando o efeito de híbridos como aleatório, ou seja:

$$y = X\beta + Z_1g + \varepsilon \quad (8)$$

onde g é o valor genotípico do híbrido enquanto os demais efeitos já foram descritos no modelo 7. Neste caso, se o modelo 7 é correto, espera-se que a combinação linear dos seus efeitos reconstrua g , assim como a combinação linear dos seus componentes de variância reconstrua σ_g^2 .

Resultados

Sessenta e dois modelos diferentes foram utilizados para descrever os efeitos genéticos, sendo 31 sem a informação de parentesco e os outros 31 com a recuperação intergenotípica (Tabela 2 e 3). O modelo com menor critério de informação em ambas as abordagens foi do tipo epistático dominante x dominante.

Nessas duas tabelas pode-se observar que o ajuste do modelo aditivo ou dominante - quando avaliados de forma isolada ou conjuntamente - foi menos eficiente que os modelos epistáticos na descrição da heterose, pois o BIC para estes modelos foi maior que o BIC para os modelos epistáticos (Tabela 2 e 3). Além disso, quando foi considerado um modelo com dois parâmetros, como

normalmente o melhorista utiliza em delineamento de dialelos, um melhor ajuste foi obtido no modelo que considera o efeito médio dos parentais (aditivo) mais a epistasia do tipo dominante x dominante. Na Tabela 3 essa afirmação torna-se evidente pela Deviance dos efeitos $a+dd$, pelo menor BIC e também pelo resultado do modelo completo, onde apenas esses dois efeitos convergiram em variâncias não nulas, mostrando desta forma, que o modelo completo convergiu para o modelo $a+dd$.

Os resultados descritos nessas duas tabelas sugerem que existe um forte confundimento das variâncias aditiva, dominante e epistáticas em pedigrees complexos, o que pode ocorrer no caso de desbalanceamento ou ajuste de efeitos epistáticos que definem, por natureza, um modelo fatorial onde o confundimento de efeitos já é bem descrito na literatura (Roso et al., 2005). Esse efeito torna-se evidente pela grande alteração nas magnitudes dos parâmetros genéticos por ocasião da sua adição ou deleção do modelo (Tabela 2 e 3). Por exemplo, quando se tentou descrever o desempenho dos híbridos com base apenas no efeito de dominância, a variância desse efeito foi $\sigma_d^2 = 0,29$. Quando ajustado em um modelo aditivo/dominante, o valor de 0,29 foi decomposto em $\sigma_a^2 = 0,11$ e $\sigma_d^2 = 0,13$. Posteriormente, ajustou-se esse efeito junto com todos os demais epistáticos ($d+aa+ad+dd$) o que normalmente é observado na estimativa das capacidades específica de combinação (Cockerham, 1961). Neste caso, esse valor caiu a 10% da estimativa original e no modelo completo foi praticamente nulo (Tabela 2). O mesmo ocorreu com os demais efeitos do modelo, que embora fossem eficientes em decompor a variância genética (variância de híbridos - Figura 2), apresentaram também grande influência dos parâmetros adicionados ou retirados do modelo genético.

De maneira geral, a decomposição da variância genética em efeitos aditivos, dominantes e epistáticos foi satisfatória (Figura 2). Para comprovar se o

efeito genotípico pode ser decomposto em efeitos aditivos, dominantes e epistáticos, o modelo 7 foi analisado considerando apenas os efeitos de genótipos (híbridos). Nesse sentido, espera-se que a variância de híbridos seja um produto das combinações lineares das variâncias aditiva, dominante e epistáticas. Se o modelo genético é aceitável, espera-se que a variância de híbridos seja separada em partes correspondentes aos efeitos genéticos que controlam o caráter. Na figura 2, esse resultado é demonstrado e valida a decomposição genética pelo modelo genético adotado.

A utilização da informação de marcadores incorporada à matriz de parentesco ocasionou uma tendência em aumentar a variância genética e a acurácia preditiva dos efeitos dominantes e epistáticos (Figura 3 a 7). Com exceção da acurácia aditiva, à medida que se aumentou o número de parâmetros no modelo, a utilização de marcadores na predição dos efeitos genéticos tendeu a melhorar a acurácia. Esse resultado sugere que, mesmo utilizando uma medida de similaridade no estado, é possível melhorar a acurácia preditiva principalmente em modelos com maior número de parâmetros. No entanto, deve-se atentar para o fato de a similaridade no estado causar perda de acurácia dos efeitos aditivos do modelo (Figura 3 a 7). Mesmo não apresentando uma melhora unânime em todos os efeitos do modelo, de forma geral, a incorporação da informação de marcadores proporcionou um pequeno ganho na estimativa dos efeitos epistáticos e dominante.

Para verificar a importância dos efeitos epistáticos de origem dominante x dominante na composição da média dos híbridos e por consequência na heterose, realizou-se a decomposição do valor genotípico dos híbridos (obtidos via modelo 8) em efeitos aditivos e epistático do tipo dominante x dominante; este foi o modelo com menor deviance depois do modelo completo quando se considera a abordagem com maior acurácia. Na figura 8 podemos verificar o ajuste de $r^2=0,99$ entre o valor obtido no modelo de híbridos (valor genotípico) e

modelo aditivo e epistático dominante x dominante. Também foi possível verificar que o efeito epistático dominante x dominante foi mais importante que a média dos efeitos dos alelos parentais, efeito aditivo, figura 9.

Na figura 10, podemos observar dois grupos distintos de cruzamentos. O agrupamento dos parentais de acordo com a epistasia dominante x dominante gerou dois grupos bem caracterizados de cruzamentos, obtidos de modo a maximizar a heterose. As epistasias negativas ou indesejáveis são dadas em vermelho e, de forma gradual, as linhas azuis mostram os cruzamentos com epistasias favoráveis. Apenas um cruzamento favorável foi inserido no cluster da direita, os demais foram alocados dentro do cluster esquerdo. Assim, com base nesse grafo de nós e arestas é possível o melhorista estabelecer grupos heteróticos e facilitar a orientação de cruzamentos.

Discussões

O controle genético da heterose ainda é obscuro é recheado de hipóteses que permeiam situações das mais simples, como a mera complementação alélica das linhagens sob efeito de dominância, a situações mais complexas que envolvem epistasia, pleiotropia e colinearidade genética (Powers, 1944; Schnell e Cockerham, 1992; Hochholdinger e Hoecker, 2007; Balestre et al., 2012). Nesse estudo, observou-se que a superioridade dos híbridos em relação ao efeito dos parentais foi função da epistasia do tipo dominante x dominante, quando consideramos um parâmetro e do tipo aditiva epistática dominante x dominante quando consideramos dois parâmetros. Este resultado foi obtido considerando o modelo com menor BIC e o modelo completo; sendo que este último convergiu para variâncias quase nulas dos efeitos de dominância e epistasias aditiva x aditiva e aditiva x dominante. Com efeito, podemos inferir que o modelo completo convergiu no modelo aditivo e dominante x dominante.

Embora pareça paradoxal sugerir que o controle genético da heterose seja devido à epistasia dominante x dominante quando os efeitos de dominância são nulos, esse fenômeno é extremamente aceitável quando imaginamos o exemplo apresentado por Carlborg e Haley (2004) (Box 1) para epistasia dominante x dominante e aditiva x aditiva. Segundo esses autores, é possível que os contrastes dos efeitos genotípicos médios de cada QTL possam ser nulos, mas quando avaliados conjuntamente e incorporando efeitos epistáticos no modelo, padrões de efeitos genéticos como aqueles apresentados nas figuras d e e descritas em Carlborg e Haley (2004) podem ser detectados. Exemplos semelhantes foram apresentados por Zeng et al (2005) para esse fenômeno, mas sob a ótica de escolha de modelos.

Devido a esse fato, Carlborg e Haley (2004) sugerem que a epistasia seja, sempre que possível incorporada ao modelo genético de modo a encontrar padrões que são negligenciados em modelos mais simples. Essa modelagem conjunta permitiu, por exemplo, que Hua et al. (2003) desmembrasse a heterose em diferentes níveis de dominância entre os QTLs e mais o efeito epistático dominante x dominante. Segundo esses autores, a epistasia do tipo dominante x dominante seria a principal responsável pelo controle da heterose em arroz sugerindo então que o modelo heterótico surge dos diferentes graus de dominância mais epistasia dominante x dominante, ou seja,

$$h = \sum_{i \in Q} d_i + \sum_{i \neq j \in Q} d_i d_j$$

Embora o modelo $d+dd$ - sugerido por Hua et al. (2003) como unificador da heterose - descrito na tabela 3 apresentasse um bom ajuste em relação aos modelos com 2 parâmetros (exceto em relação a $a+dd$), sua justificativa baseada nos ajustes das métricas F_∞ ou F_2 apresenta alguns

problemas quanto à interpretação da heterose. Em linhagens contrastantes, e sob a métrica F_{∞} , espera-se que, para dois locos, a heterose seja função de: $h = d_1 + d_2 + d_1d_2 - a_1a_2$. Na tabela 3, pode-se observar que esse modelo apresentou um bom ajuste, mas muito mais pelo componente $d+dd$ do que devido à participação da epistasia aditiva x aditiva.

Por outro lado, caso se utilize a métrica F_2 de Cocherham (1996), se tem que a heterose pode ser descrita em função apenas de: $h = d_1 + d_2 - a_1a_2$. Segundo Melchinger et al (2007), umas das vantagens dessa abordagem é que os efeitos são ortogonais e dessa forma não há confundimento entre os efeitos de dominância e epistasia dominante x dominante como na métrica F_{∞} . No entanto esse modelo apresentou um ajuste muito inferior em relação ao modelo esperado pela métrica F_{∞} (Tabela 3).

Um bom exemplo das diferenças entre as abordagens F_{∞} e F_2 foi descrito por Zeng et al (2005) - Tabelas 3 e 4. Esses autores demonstraram que o confundimento entre os efeitos aditivos, dominantes e epistáticos pode ser drástico dependendo do modelo adotado. Isto não quer dizer que o modelo adotado neste trabalho possa ser duvidoso quanto à interpretação genética, pelo contrário, o modelo foi capaz de desmembrar de forma eficiente os efeitos dos híbridos e realizar a decomposição da variância genética de maneira bem precisa (Figura 2). No entanto, como a abordagem do modelo gamético difere dos modelos de QTLs F_{∞} e F_2 , (ambos assumem frequências alélicas populacionais nas gerações F_2 ou F_{∞} e não se ajustam a casos de cruzamentos complexos onde as frequências alélicas diferem de 1/2) a nossa interpretação da heterose não pode se restringir apenas aos resultados esperados por estas duas abordagens. Assim, os resultados deste trabalho sugerem que a diferença entre as médias dos parentais e o desempenho dos híbridos pode ser explicada apenas pela epistasia dominante x dominante.

Como o modelo adotado neste trabalho foi o uso do valor genotípico, é esperado que o confundimento seja amenizado (Roso et al. 2005). Porém, devido a esse efeito de multicolinearidade, teste de componente da variância via modelos completos e reduzidos pela análise de deviance ficam comprometidos pelo confundimento das variâncias e sendo assim foi optado por utilizar um critério de informação ao invés da abordagem tradicional, pois é uma abordagem livre de testes.

De maneira geral a utilização da informação de marcadores moleculares ocasionou uma tendência à melhorar a acurácia preditiva dos efeitos genéticos. Esse resultado tem sido observado em diversos outros trabalhos (Balestre et al., 2008, Ferreira et al., 2010). Embora a utilização da similaridade no estado não seja ideal, pois superestima a covariância genética entre indivíduos, neste trabalho esse efeito foi contrabalanceado pela melhora nas estimativas dos efeitos epistáticos.

No melhoramento de plantas, o melhorista busca sempre aperfeiçoar cruzamentos de modo a obter o máximo vigor híbrido utilizando um número mínimo de cruzamentos. Sendo assim, o agrupamento de cruzamentos com base na heterose é de grande utilidade. Nesse trabalho foi utilizado um método simples de agrupamento baseado no algoritmo *K-means* que visa minimizar as variâncias de pontos dentro de cada cluster. Ainda que essa abordagem fosse utilizada nesse estudo apenas como facilitador na separação de grupos heteróticos ou para aperfeiçoar cruzamentos, abordagens utilizando de redes bayesianas ou algoritmos genéticos podem ser de grande valia nesses casos e trabalhos estão sendo desenvolvidos nesse sentido.

Este trabalho demonstrou que é possível explorar grande quantidade de informação quando são analisados conjuntamente híbridos simples, duplos e triplos sob um mesmo modelo. Nesta abordagem, é possível desmembrar a heterose em diversos componentes e, com base nisto, definir qual a melhor

estratégia de cruzamentos a ser utilizada. A utilização de marcadores moleculares ocasionou uma tendência em melhorar a acurácia dos efeitos genéticos epistáticos e de dominância. Assim, com os recursos computacionais disponíveis atualmente e técnicas estatísticas avançadas pode-se expandir o conceito de cruzamentos para fronteiras muito além das tradicionais capacidades gerais e específicas de combinação, muito comumente utilizadas pelos melhoristas para escolha de cruzamentos e parentais.

Agradecimentos

A Universidade Federal de Lavras (UFLA) e a empresa Advanta Comércio de Sementes pelo apoio.

Referências

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

BALESTRE M, VON PINHO R.G; DE SOUZA JUNIOR C.L; DE SOUSA BUENO FILHO J.S. Bayesian mapping of multiple traits in maize: the importance of pleiotropic effects in studying the inheritance of quantitative traits. **Theoretical and Applied Genetics**. Mar. 2012.

CARLBORG, O.; HALEY, C.S;. Epistasis: too often neglected in complex trait studies? **Nature Review Genetics**,v.5, n. 8, p. 618-25, Aug 2004.

COCKERHAM, C. C. Implications of genetic variances in a hybrid breeding program. **Crop Science**. v.1. p.47-52. 1961.

COCKERHAM, C. C. Random and fixed effects in plant genetics. **Theoretical Applied. Genetics**.v.56, n.3, p.119-131. May 1980.

COCKERHAM, C. C.; ZENG, Z.B. Design III with marker loci. **Genetics**, Baltimore, v.143, n. 3, p.1437-1456, jul. 1996.

FERREIRA, D.V.; VON PINHO R.G.; BALESTRE, M.; OLIVEIRA, R.L. Prediction of maize hybrid performance using similarity in state and similarity by descent information. **Genetics Molecular Research**.v.9 n. 4, p. 2381-2394. Dec 2010

FRASCAROLI, E.; CANÈ, M.A.; LANDI, P.; PEA, G.; GIANFRANCESCHI, L.; VILLA, M.;MORGANTE, M.; PÈ M.E. Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. **Genetics**, Baltimore, v. 176, n. 1, p. 625–644. May. 2007.

GARCIA, A.A; WANG, S.; MELCHINGER, A.E.; ZENG, Z.B.; Quantitative trait mapping and the genetic basis of heterosis in maize and rice. **Genetics**, Baltimore, v.180, n.3, p. 1707 -1724, Nov. 2008.

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

HILL, W.G.; GODDARD, M.E.; VISSCHER, P.M. Data and theory point to mainly additive genetic variance for complex traits. **PLoS Genetics**. V.4, n. 2, p. 1-10, Feb . 2008

HOCHHOLDINGER, F.; HOECKER, N. Towards the molecular basis of Heterosis. **TRENDS in Plant Science**, v.12, n.9. 2007

HUA, J; XING, Y; WU, W. et. al. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. **PNAS** v. 100, n.5 p. 2574–2579. March. 2003.

LARIÈPE, A.; MANGIN, B.; JASSON, S. et al. The Genetic Basis of Heterosis: Multiparental Quantitative Trait Loci Mapping Reveals Contrasted Levels of Apparent Overdominance Among Traits of Agronomical Interest in Maize (*Zea mays* L.). **Genetics**.v. 190, n. 2, p. 795-811 . Feb. 2012

LU, H.; ROMERO-SEVERSON, J.; BERNARDO, R. Genetics basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. **Theoretical and Applied Genetics**, Berlin, v. 107, n.3, p. 494-502, Ago. 2003.

MACKAY, T.F.; STONE, E.A.; AYROLES, J.F. The genetics of quantitative traits: challenges and prospects. **Nature Reviews. Genetics**, London, v. 10, n. 8, p. 565-577, Aug. 2009.

MELCHINGER, A. E.; UTZ, H. F.; PIEPHO, H.P.; ZENG, Z.B.; SCHÖN, C. C. The role of epistasis in the manifestation of heterosis: A systems-oriented approach. **Genetics**, Baltimore, v.177, n. 3 p. 1815–1825, Nov. 2007.

POWERS, L. An expansion of Jones' theory for the explanation of heterosis. **American Nat.** n.78, p. 275-280. 1944

RAWLINGS, J.; COCKERHAM C. C. Analysis of double cross hybrid populations. **Biometrics**. v. 18. p. 229. 244. 1962a.

RAWLINGS, J.; COCKERHAM C. C. Triallel analysis. **Crop Science**. v.2, p.228-231. 1962b

RIEDELSEIMER, C.; CZEDIK-EYSENBERG, A.; GRIEDER, C. et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. **Nature Genetics** v.44, n. 2, p. 217-220 Jan. 2012

ROSO, V.M.; SCHENKEL, F.S.; MILLER, S.P.; SCHAEFFER, L.R.; Estimation of genetic effects in the presence of multicollinearity in multibreed beef cattle evaluation. *Journal Animal Science*. v. 83, n. 8, p. 1788-1800 Aug. 2005

SCHNELL, F.W.; COCKERHAM, C.C. Multiplicative vs. arbitrary gene action in heterosis. **Genetics**, Baltimore, v. 131, n. 2, p. 461-469, Jun.1992.

SCHÖN, C.C.; DHILLON, B.S.; UTZ, H.F.; MELCHINGER, A.E. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 321-332, Jan. 2010.

TROYER, A. F. Adaptedness and heterosis in corn and mule hybrids. **Crop Science**, Madison, v. 46, n. 2, p. 528-543, Mar. 2006

VAN DER VEEN, J. H. Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. **Genetica**, Dordrecht, v. 30, n.1, p. 201–232, 1959.

ZENG, Z.B.; WANG, T.; ZOU W. Modeling quantitative trait loci and interpretation of models. **Genetics**, Baltimore, v.169, n.3, p.1711-1725, Mar. 2005.

BUCHANAN, A.V.; SHOLTIS, S.; RICHTSMEIER, J; WEISS, K.M. What are genes “for” or where are traits “from”? What is the question? **Bioessays**, Hoboken, v. 31, n. 2, p. 198–208, Feb. (2009).

CARBONE, M.A.; JORDAN, K.W.; LYMAN, R.F.; HARBISON, S.T.; LEIPS, J.; MORGAN, T.J.; DeLUCA, M.; AWADALLA, P.; MACKAY, T.F. Phenotypic variation and natural selection at *Catsup*, a pleiotropic quantitative trait gene in *Drosophila*. **Current Biology**. Cambridge v. 16, n. 9, p. 912-919, May 2006.

CHEVERUD, J.; EHRICH, T.H.; VAUGHN, T.T.; KOREISHI, S.F.; LINSEY, R.B.; PLETSCHER, L.S. Pleiotropic effects on mandibular morphology II: differential epistasis and genetic variation in morphological integration. **Journal of Experimental Zoology**, Hoboken, v.302, n. 5, p.424–435. Sep. 2004

FLINT, J.; MACKAY, T.F.C. Genetic architecture of quantitative traits in mice, flies, and humans. **Genome Research**, Woodbury, v.19, n. 5, p.723– 73, May. 2009

FRASCAROLI, E.; CANÈ, M.A.; LANDI, P.; PEA, G.; GIANFRANCESCHI, L.; VILLA, M.;MORGANTE, M.; PÈ M.E. Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. **Genetics**, Baltimore, v. 176, n. 1, p. 625–644. May. 2007.

GARCIA, A.A; WANG, S.; MELCHINGER, A.E.; ZENG, Z.B.; **Quantitative trait mapping and the genetic basis of heterosis in maize and rice. Genetics**, Baltimore, v.180, n.3, p. 1707 -1724, Nov. 2008.

GILBERT, H.; LE ROY, P. Comparison of three multitrait methods for QTL detection. **Genetics, Selection, Evolution**, London, v.35, n. 3, p. 281–304, May. 2003

GOMMANS, W.M.; MULLEN, S.P.; MAAS, S. RNA editing: a driving force for adaptive evolution? **Bioessays**. Hoboken, v. 31, n. 10, p. 1137-1145, Oct. 2009.

GRUNEBERG, H. An analysis of the "pleiotropic" effects of a new lethal mutation in the rat (*mus norvegicus*). **Proceeding of the Royal Society**. London, v.125, n.3, p. 123-144, Mar.1938.

HASTINGS, W.K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, London, v. 57, n. 1, p. 97-109, Feb. 1970.

HUBERTS, D.H.; VAN DER KLEI, I.J. Moonlighting proteins: an intriguing mode of multitasking. **Biochimica et Biophysica Acta**, Amsterdam, v.1803, n.4, p.520-525, Apr. 2010.

JIANG, C.; ZENG, Z. Multiple trait analysis of genetic mapping for quantitative trait loci. **Genetics**, Baltimore, v. 140, n. 3, p. 1111-1127, jul. 1995.

LINCOLN, S.E.; DALY, M.J.; LANDER, E.S. **Constructing genetic maps with Mapmaker Exp 3.0**. 3rd ed. Cambridge: Whitehead Institute for Biometrical Research, 1992. 230p.

LIU, J.; LIU, Y.; LIU, X; DENG, H.W. Bayesian mapping of quantitative trait loci for multiple complex traits with the use of variance components. **American Journal of Human Genetics**, Cambridge, v. 81, n. 2, p. 304–320, Aug. 2007.

LU, H.; ROMERO-SEVERSON, J.; BERNARDO, R. Genetics basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. **Theoretical and Applied Genetics**, Berlin, v. 107, n.3,p. 494-502, Ago. 2003.

MACKAY, T.F.C. The genetic architecture of quantitative traits. **Annual Review of Genetics**. Palo Alto, v.35, p. 336-339, 2001.

MACKAY, T.F.; STONE, E.A.; AYROLES, J.F. The genetics of quantitative traits: challenges and prospects. **Nature Reviews. Genetics**, London, v. 10, n. 8, p. 565-577, Aug. 2009.

MALOSETTI, M.; RIBAUT, J.M.; VARGAS, M.; CROSSA, J.; VAN EEUWIJK, F.A. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). **Euphytica**, Dordrecht, v. 161, n.1, p.241–257, 2008.

MAYDANOVYCH, O.; BEAL, P.A. Breaking the central dogma by RNA editing. **Chemical Reviews**. Washington, v. 106, n. 8, p. 3397-3411, Aug. 2006.

MELCHINGER, A. E.; UTZ, H. F.; PIEPHO, H.P.; ZENG, Z.B.; SCHÖN, C. C. The role of epistasis in the manifestation of heterosis: A systems-oriented approach. **Genetics**, Baltimore, v.177, n. 3 p. 1815–1825, Nov. 2007.

METROPOLIS, N.; ROSEMBLUT, A.W.; ROSEMBLUT, M.N.; TELLER, A.H.; TELLER, E. Equations of state calculations by fast computing machines. **Journal of Chemical Physics, New York**, v. 21, 1087-1092, 1953.

PAVLICEV, M.; KENNEY-HUNT, J.P.; NORGARD, E.A.; ROSEMAN, C.C.; WOLF, J.B.; CHEVERUD, J.M. Genetic variation in pleiotropy: differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. **Evolution**. Malden, v. 62, n. 1, p. 199-213, Jan. 2008.

SCHÖN, C.C.; DHILLON, B.S.; UTZ, H.F.; MELCHINGER, A.E. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 321-332, Jan. 2010.

SCHNELL, F.W.; COCKERHAM, C.C. Multiplicative vs. arbitrary gene action in heterosis. **Genetics**, Baltimore, v. 131, n. 2, p. 461-469, Jun.1992.

SIBOV, S.T.; SOUZA JÚNIOR, C.L.; GARCIA, A.A.F.; GARCIA, A.F.; SILVA, A.R.; MANGOLIN, C.A.; BENCHIMOL, L.L.; SOUZA, A.P. Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 1. Map construction and localization of loci showing distorted segregation. **Hereditas**, Oxford, v. 139, p. 96-106, 2003.

SILVA, A.R. Análise genética de caracteres quantitativos em milho com o delineamento III e marcadores moleculares. Piracicaba, 2002. 143p. Tese (Doutorado) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

SHANNON, P.; MARKIEL, A.; OZIER, O.; BALIGA, N.S.; WANG, J.T.; RAMAGE, D.; AMIN, N.; SCHWIKOWSKI, B.; IDEKER, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome Research**. New York, v.13, n.11, p. 2498-2504, Nov. 2003.

STEARNS, F.W. One hundred years of pleiotropy: a retrospective. **Genetics**, Baltimore, v. 186, n.3, p.767-773, Nov. 2010.

SU, Z.; ZENG, Y.; GU, X. A preliminary analysis of gene pleiotropy estimated from protein sequences. **Journal of Experimental Zoology**, Hoboken, v. 314, n. 2, p. 115-122, Mar. 2010

VAN EEUWIJK, F.A.; BOER, M.; TOTIR, R. L.; BINK, M.; WRIGHT, D.; WINKLER, C. R.; PODLICH, D.; BOLDMAN, K.; BAUMGARTEN A.; SMALLEY M.; ARBELBIDE, M.; BRAAK, C.J.F.T.; COOPER, MARK. Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 429-440, Jan. 2010.

XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, Baltimore, v. 163, n.2, p. 789–801, Feb. 2003.

XU, C.; WANG, X.; LI, Z.; XU, S.; Mapping QTL for multiple traits using Bayesian statistics. **Genetics Research**, New York, v. 91, n. 1, p. 23-37, Feb. 2009.

WAGNER, G.P.; KENNEY-HUNT, J.P.; PAVLICEV, M.; PECK, J.R.; WAXMAN, D.; CHEVERUD JM. Pleiotropic scaling of gene effects and the 'cost of complexity'. **Nature**, London, v.27, n. 7186, p. 470-472. Mar. 2008.

WAGNER, G.P.; ZHANG, J. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. **Nature Reviews, Genetics**, London, v.12, p. 204-213, Mar. 2011.

WANG, H.; ZHANG, Y.M.; LI, X.; MASINDE, G.L.; MOHAN, S.; BAYLINK, D.J.; XU, S. Bayesian shrinkage estimation of quantitative trait loci parameters. **Genetics**, Baltimore, v. 170, n. 1, p. 465-480, May 2005.

WELLER, J.I.; WIGGANS, G.R.; VANRADEN, P.M.; RON, M. Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multitrait experiment. **Theoretical Applied Genetics**, Berlin, v. 92, n.4, p. 998–1002, Oct.1996

WOLF, J. B.; LEAMY, L.L.; ROUTMAN, E. J.; CHEVERUD, J. M. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. **Genetics**, Baltimore, v. 171, n.2, p.1–12, Oct. 2005.

WOLF, J. B.; POMP, D.; EISEN, E. J.; CHEVERUD, J. M.; LEAMY, L. L. The contribution of epistatic pleiotropy to the genetic architecture of covariation among polygenic traits in mice. **Evolution Development**, Malden, v.8, n.5, p. 468–476, Oct. 2006.

WRIGHT, S. **Evolution and the genetics of populations**. Chicago, University of Chicago Press. 1968. 480 p.

YAN, W.; HUNT, L. A. Interpretation of Genotype x Environment Interaction for Winter Wheat Yield in Ontario. **Crop Science**.Madison, v. 41, n.1 p. 19-25, Jan-Feb. 200.

YANG, R.; XU, S. Bayesian Shrinkage Analysis of Quantitative Trait Loci for Dynamic Traits. **Genetics**, Baltimore, v.2, n.176, p. 1169–1175. Jul. 2007.

YANG, R.C. Epistasis of quantitative trait loci under different gene action models. **Genetics**, Baltimore, v.167, n.3, p. 1493–1505. Jul. 2004.

ZENG, Z.B.; WANG, T.; ZOU W. Modeling quantitative trait loci and interpretation of models. **Genetics**, Baltimore, v.169, n.3, p.1711-1725, Mar. 2005.

Tabela 1 Distribuição dos 79 marcadores de microssatélites nos 10 grupos de ligação (GL) do milho.

Marcador	GL	Bin	Marcador	GL	Bin	Marcador	GL	Bin	Marcador	GL	Bin
bnlg1179	1	1,01	dupssr08	3	3,09	bnlg1200	7	7,01	umc1139	8	8,01
bnlg1014	1	1,01	bnlg1496	3	3,09	bnlg1808	7	7,02	bnlg1056	8	8,01
bnlg1007	1	1,02	umc1136	3	3,09	bnlg1305	7	7,03	bngl2082	8	8,03
bnlg1614	1	1,02	phi072	4	4,01	umc1342	7	7,04	bngl1067	8	8,03
bnlg1866	1	1,03	umc1101	4	4,09	bnlg2259	7	7,04	umc1858	8	8,04
umc1128	1	1,07	umc1109	4	4,10	umc1154	7	7,05	phi015	8	8,08
phi037	1	1,08	umc1197	4	4,11	umc1075	8	8,01	bnlg1131	8	8,09
bnlg1643	1	1,08	umc1058	4	4,11	umc1414	8	8,01	bnlg2122	9	9,01
umc1725	1	1,11	phi019	4	4,11	bnlg1194	8	8,01	umc1040	9	9,01
umc1797	1	1,12	umc1591	5	5,04	phi119	8	8,02	bnlg1724	9	9,01
umc1079	2	2,06	umc1482	5	5,04	umc1034	8	8,03	umc1078	9	9,05
bnlg1036	2	2,06	bnlg1237	5	5,05	phi115	8	8,03	umc1310	9	9,06
dupssr24	2	2,08	bnlg1118	5	5,07	mmc412	8	8,03	umc1319	10	10,01
bnlg1520	2	2,09	bnlg1371	6	6,01	umc2146	8	8,03	bnlg1079	10	10,03
umc1970	3	3,01	umc1006	6	6,02	phi121	8	8,03	umc2043	10	10,05
bnlg1601	3	3,05	umc1887	6	6,03	umc2147	8	8,03	bnlg1074	10	10,05
bnlg1160	3	3,06	umc1918	6	6,04	umc1157	8	8,03	umc1061	10	10,06
umc1148	3	3,07	bnlg1740	6	6,07	umc1202	8	8,05	bnlg1360	10	10,07
umc1167	3	3,08	phi089	6	6,08	bngl240	8	8,06	umc1084	10	10,07
bnlg1108	3	3,08	umc1066	7	7,01	umc1933	8	8,08			

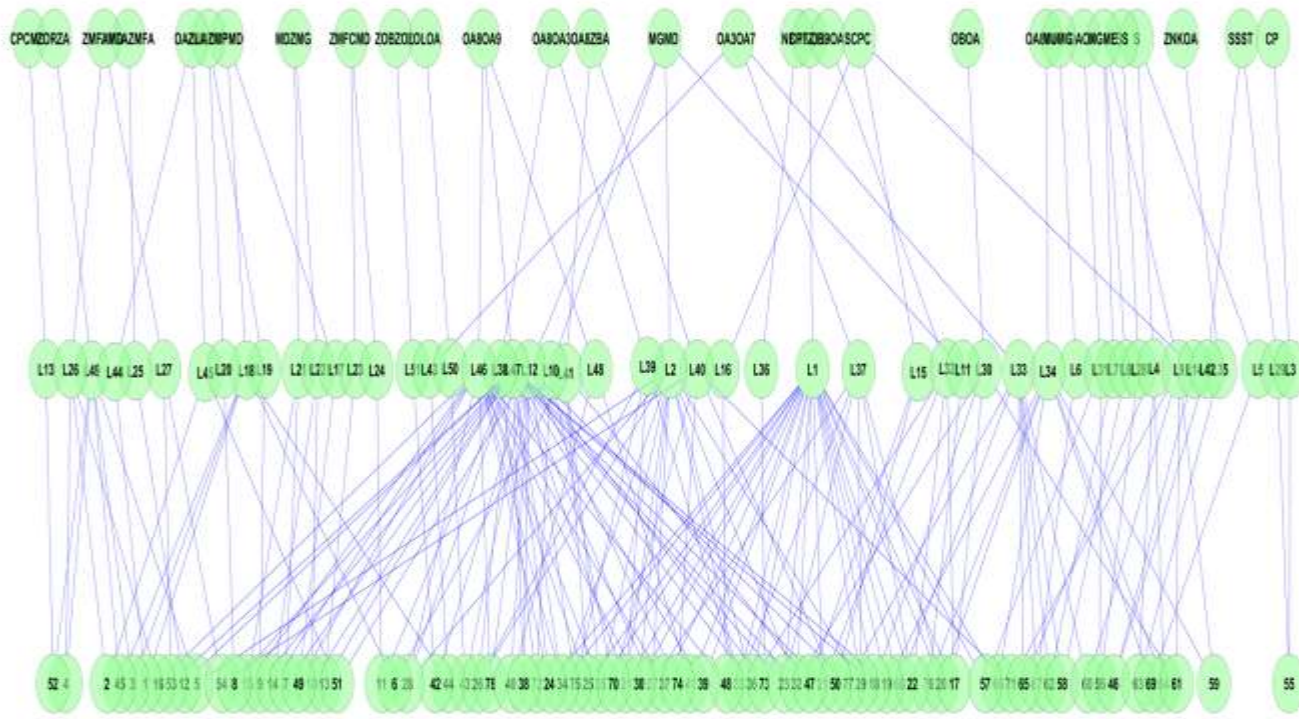


Figura 1 - Sistema de cruzamento em 51 linhagens de milho. Os círculos superiores se referem ao background de cada linhagem.

Tabela 2 Seleção de modelo com base no critério bayesiano de informação (BIC) para os efeitos aditivos, dominantes e epistáticos considerando o modelo sem informação de marcadores moleculares.

Modelo	BIC	DEV	σ_a^2	σ_d^2	σ_{aa}^2	σ_{ad}^2	σ_{dd}^2
<i>dd</i>	2703,04	2689,05	-	-	-	-	0,30
<i>ad</i>	2706,12	2692,13	-	-	-	0,15	-
<i>aa</i>	2707,25	2693,26	-	-	0,05	-	-
<i>aa+dd</i>	2711,77	2690,79	-	-	0,03	-	0,07
<i>ad+dd</i>	2712,93	2691,94	-	-	-	0,08	0,07
<i>aa+ad</i>	2712,95	2691,96	-	-	0,03	0,07	-
<i>d+ad</i>	2713,98	2692,99	-	0,07	-	0,12	-
<i>d</i>	2714,05	2700,06	-	0,29	-	-	-
<i>a+dd</i>	2714,32	2693,33	0,10	-	-	-	0,08
<i>a+ad</i>	2714,40	2693,41	0,09	-	-	0,08	-
<i>d+aa</i>	2714,60	2693,62	-	0,07	0,04	-	-
<i>a+ad+dd</i>	2714,81	2693,82	0,09	-	-	0,05	0,07
<i>a+aa</i>	2715,37	2694,38	0,07	-	0,03	-	-
<i>d+dd</i>	2716,27	2695,28	-	0,05	-	-	0,12
<i>a+aa+dd</i>	2717,91	2689,93	0,06	-	0,02	-	0,09
<i>aa+ad+dd</i>	2717,97	2689,99	-	-	0,02	0,06	0,07
<i>d+aa+dd</i>	2718,05	2690,07	-	0,05	0,03	-	0,10
<i>a+d+dd</i>	2718,10	2690,11	0,09	0,04	-	-	0,11
<i>d+ad+dd</i>	2718,46	2690,47	-	0,05	-	0,08	0,10
<i>a+d</i>	2720,06	2699,07	0,11	0,13	-	-	-
<i>d+aa+ad</i>	2720,24	2692,26	-	0,04	0,02	0,06	-
<i>a+aa+ad</i>	2721,27	2693,28	0,08	-	0,01	0,05	-
<i>a+d+ad</i>	2721,81	2693,82	0,09	0,03	-	0,06	-
<i>a+d+aa</i>	2722,68	2694,70	0,07	0,05	0,03	-	-
<i>a+aa+ad+dd</i>	2724,15	2689,17	0,05	-	0,01	0,04	0,06
<i>a+d+aa+dd</i>	2724,91	2689,93	0,06	0,04	0,02	-	0,08
<i>d+aa+ad+dd</i>	2725,20	2690,22	-	0,03	0,02	0,05	0,07
<i>a+d+ad+dd</i>	2725,61	2690,63	0,09	0,02	-	0,04	0,06
<i>a+d+aa+ad</i>	2726,93	2691,95	0,05	0,03	0,01	0,05	-
<i>Completo</i>	2731,83	2689,86	0,10	0,00	0,00	0,01	0,13
<i>a</i>	2750,75	2736,76	0,21	-	-	-	-

Tabela 3 Seleção de modelo com base no critério bayesiano de informação (BIC) para os efeitos aditivos, dominantes e epistáticos considerando o modelo com informação de marcadores moleculares.

Modelo	BIC	DEV	σ_a^2	σ_d^2	σ_{aa}^2	σ_{ad}^2	σ_{dd}^2
<i>dd</i>	2703,10	2689,11	-	-	-	-	0,30
<i>Ad</i>	2708,38	2694,39	-	-	-	0,14	-
<i>AA</i>	2708,70	2694,71	-	-	0,06	-	-
<i>a+dd</i>	2710,87	2689,88	0,10	-	-	-	0,20
<i>d+dd</i>	2711,18	2690,20	-	0,11	-	-	0,19
<i>aa+dd</i>	2711,50	2690,51	-	-	0,03	-	0,13
<i>ad+dd</i>	2711,57	2690,59	-	-	-	0,08	0,12
<i>D</i>	2714,04	2700,05	-	0,33	-	-	-
<i>a+d+dd</i>	2714,50	2693,52	0,07	0,08	-	-	0,15
<i>aa+ad</i>	2715,26	2694,27	-	-	0,03	0,07	-
<i>d+ad</i>	2716,20	2695,21	-	0,07	-	0,11	-
<i>a+ad</i>	2716,20	2695,21	0,07	-	-	0,10	-
<i>a+AA</i>	2716,30	2695,32	0,07	-	0,05	-	-
<i>d+AA</i>	2716,35	2695,36	-	0,08	0,04	-	-
<i>d+aa+dd</i>	2718,91	2690,92	-	0,06	0,02	-	0,11
<i>a+aa+dd</i>	2718,91	2690,92	0,05	-	0,02	-	0,11
<i>d+ad+dd</i>	2718,97	2690,99	-	0,05	-	0,07	0,11
<i>a+ad+dd</i>	2719,03	2691,05	0,05	-	-	0,07	0,10
<i>aa+ad+dd</i>	2719,10	2691,12	-	-	0,02	0,05	0,09
<i>a+d</i>	2720,83	2699,84	0,12	0,18	-	-	-
<i>d+aa+ad</i>	2722,66	2694,68	-	0,04	0,02	0,06	-
<i>a+aa+ad</i>	2722,87	2694,89	0,05	-	0,02	0,06	-
<i>a+d+ad</i>	2723,68	2695,70	0,06	0,05	-	0,08	-
<i>a+d+AA</i>	2724,06	2696,08	0,06	0,07	0,03	-	-
<i>a+d+aa+dd</i>	2726,27	2691,29	0,05	0,05	0,02	-	0,10
<i>a+d+ad+dd</i>	2726,29	2691,31	0,04	0,04	-	0,06	0,09
<i>d+aa+ad+dd</i>	2726,34	2691,36	-	0,04	0,02	0,05	0,08
<i>a+aa+ad+dd</i>	2726,45	2691,47	0,04	-	0,01	0,05	0,08
<i>a+d+aa+ad</i>	2730,18	2695,21	0,04	0,04	0,02	0,06	-
<i>Completo</i>	2731,68	2689,70	0,09	0,00	0,00	0,01	0,20
<i>A</i>	2752,26	2738,27	0,36	-	-	-	-

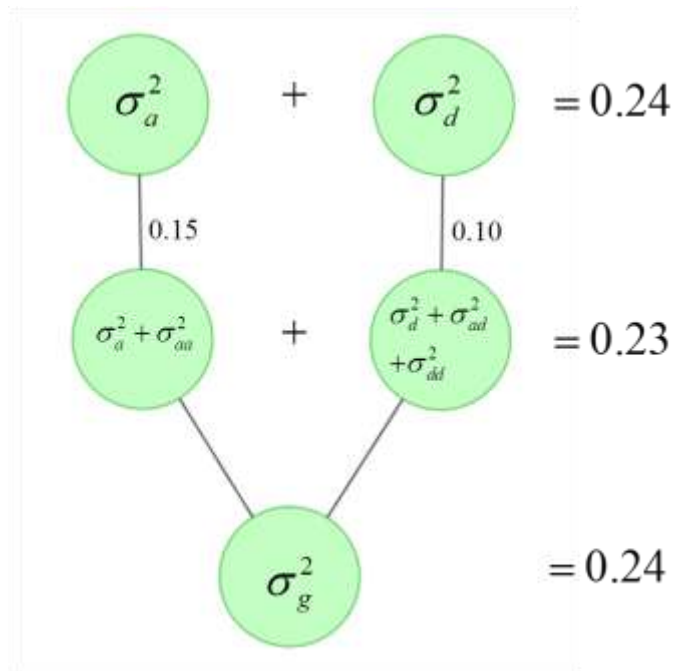


Figura 2 - Decomposição da variância genética em seus componentes descritos pela variância aditiva, dominante, aditiva x aditiva, aditiva x dominante e dominante x dominante.

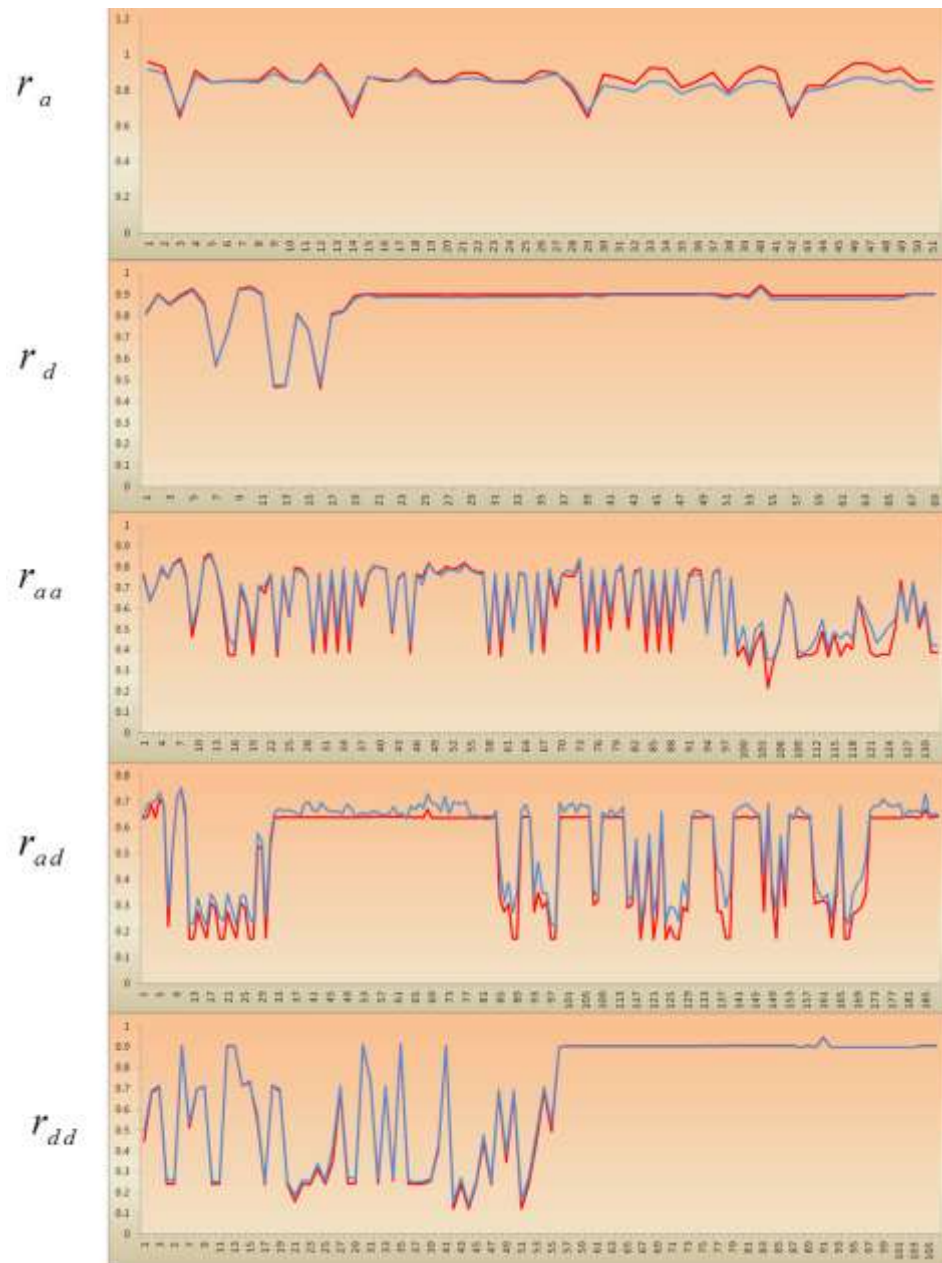


Figura 3 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando um parâmetro no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.

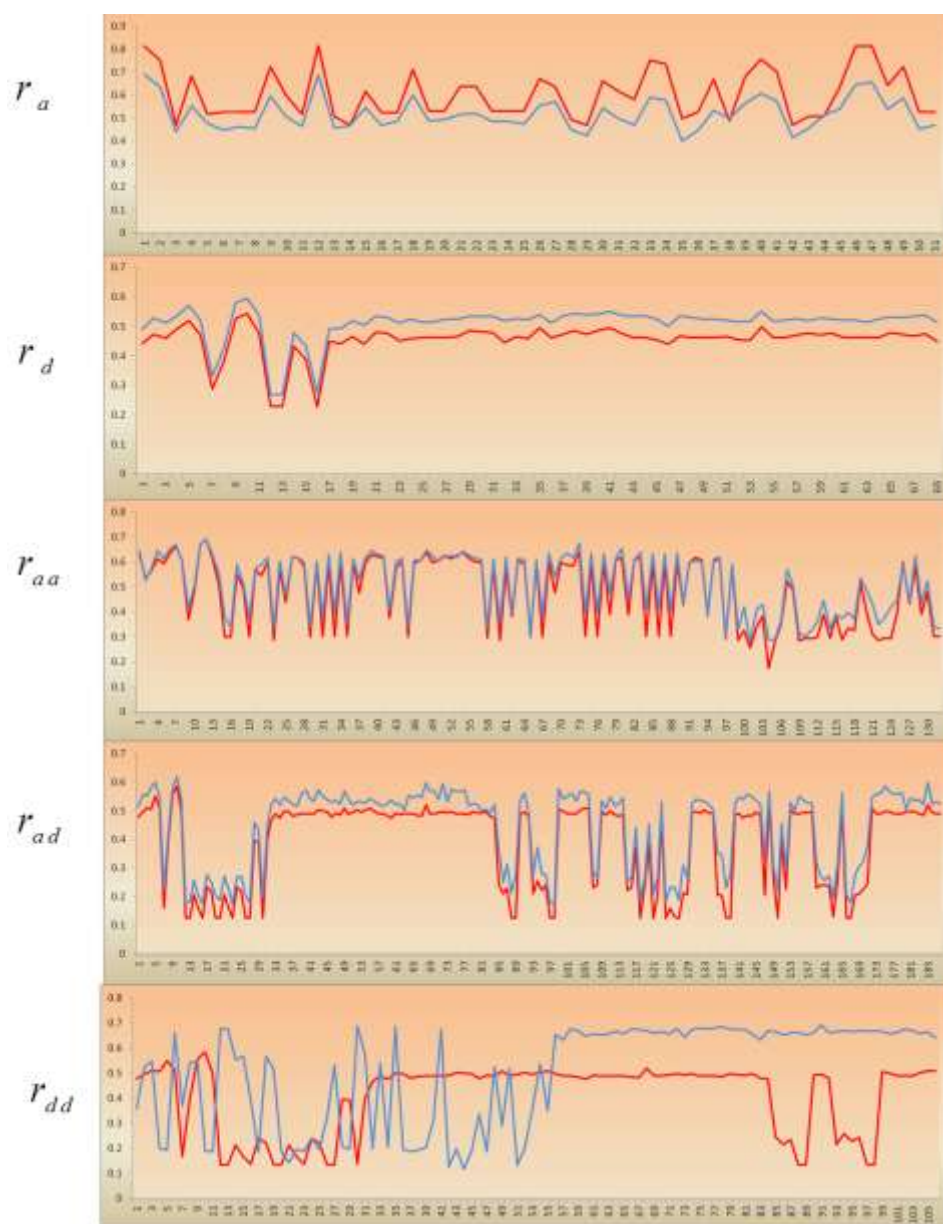


Figura 4 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando dois parâmetros no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.

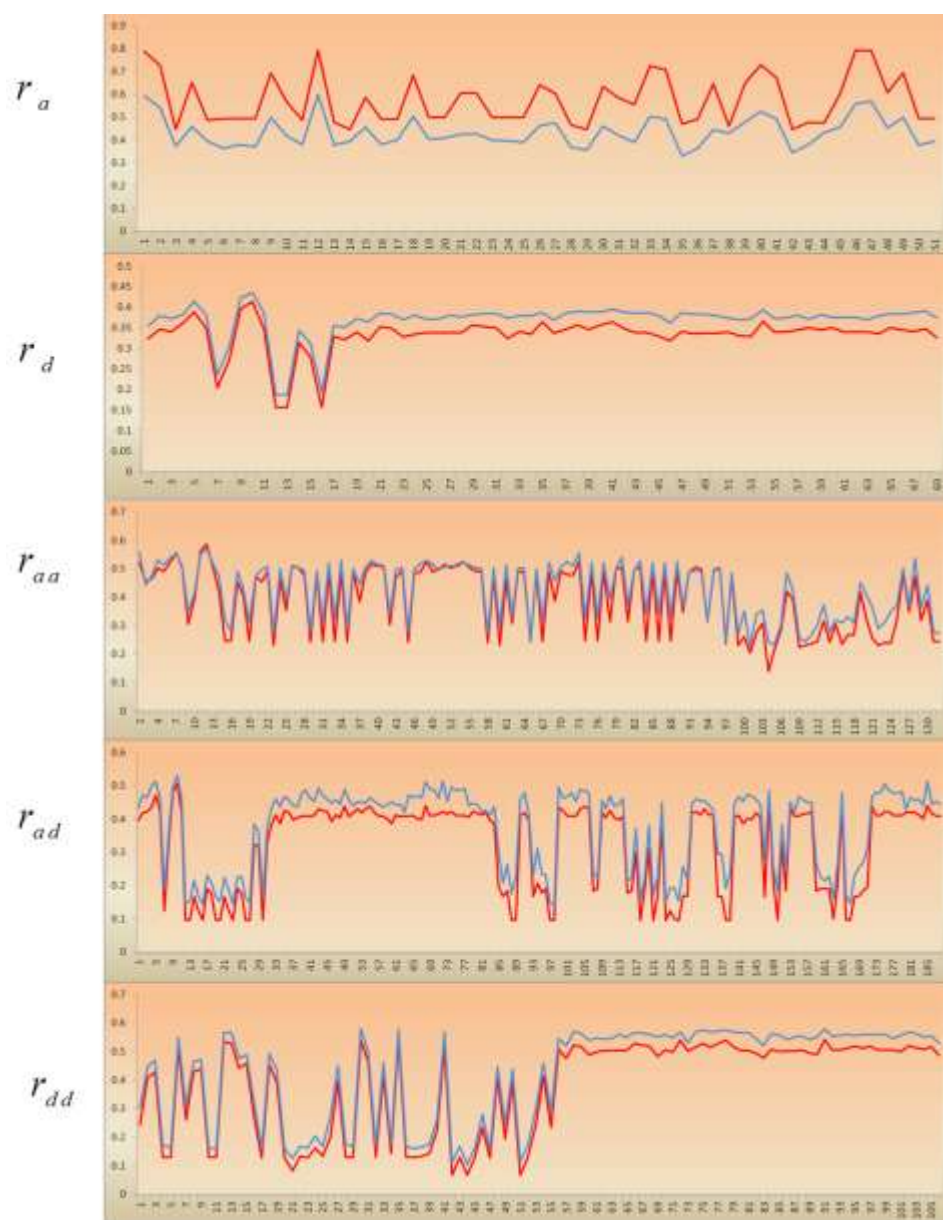


Figura 5 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando três parâmetros no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.

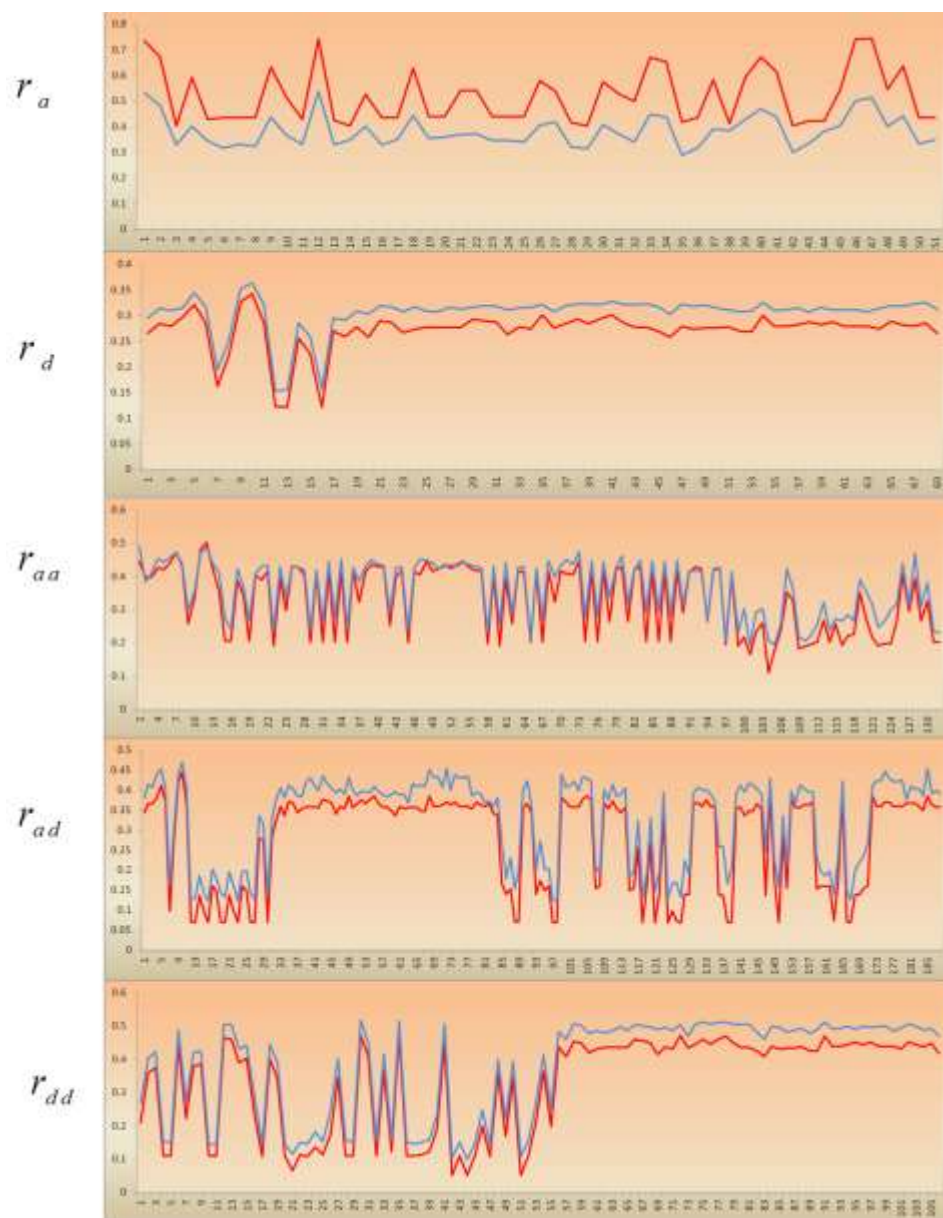


Figura 6 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando quatro parâmetros no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.

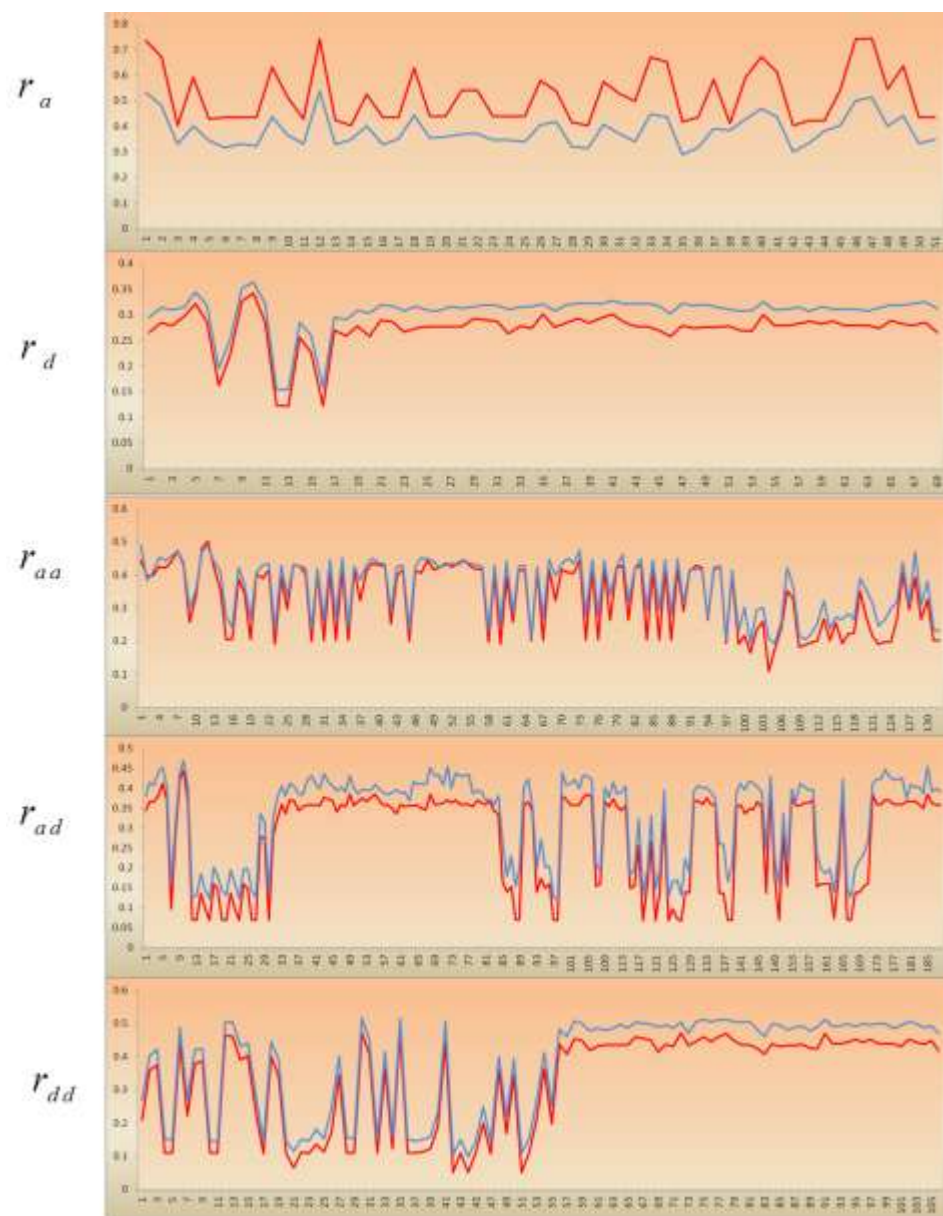


Figura 7 - Acurácia preditiva média dos efeitos genéticos aditivo, dominante e epistáticos considerando cinco parâmetros no modelo. A linha vermelha se refere ao modelo sem informação e a linha azul ao modelo com informação de marcadores.



Figura 8 - Decomposição do valor genético do híbrido em efeitos aditivos + epistasia dominante-dominante. A linha azul se refere ao valor genético total, a linha pontilhada preta se refere à soma dos efeitos aditivos e epistático dominante-dominante.

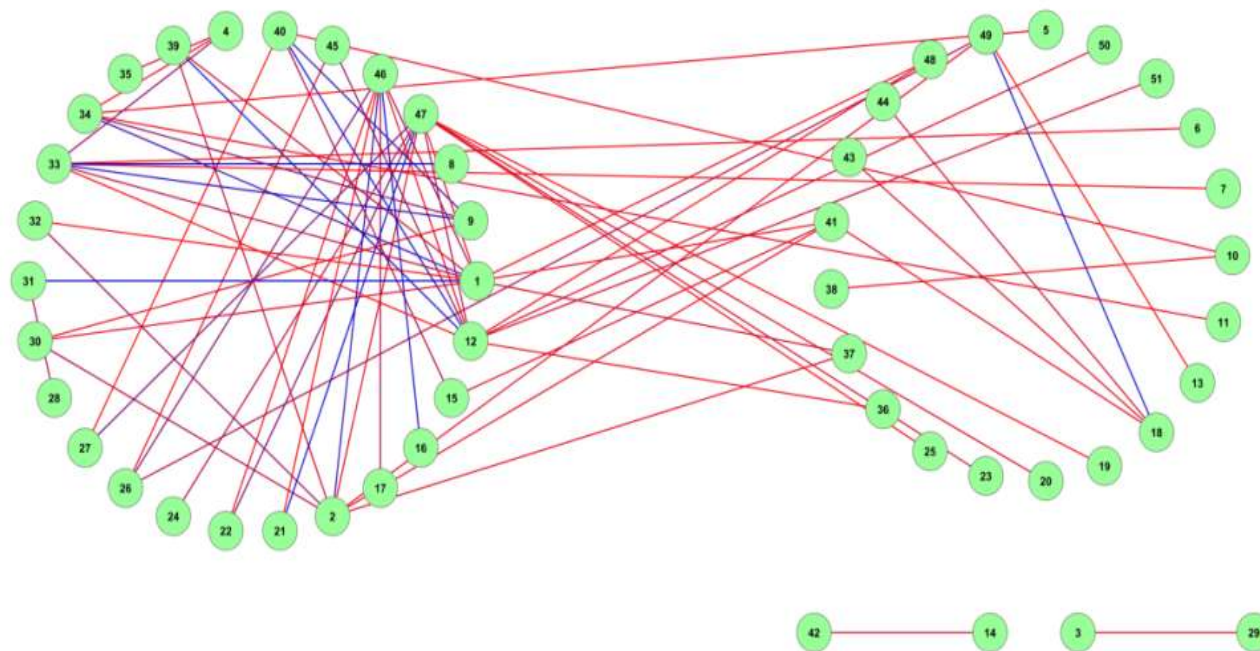


Figura 10 - Cluster das linhagens com base na epistasia do tipo dominante-dominante. A graduação nas linhas do vermelho ao azul mostra os cruzamentos com epistasia variando do negativo ao positivo. Foi utilizado um método simples de agrupamento baseado no algoritmo K-means que visa minimizar as variâncias de pontos dentro de cada cluster.

ARTIGO 2 **Predição de híbridos simples de milho utilizando a abordagem de efeitos totais das marcas associadas com validação cruzada e em ensaios regionais**

Artigo redigido nas normas da Revista Científica Genetics and Molecular
Research

Predição de híbridos simples de milho utilizando a abordagem de efeitos totais das marcas associadas com validação cruzada e em ensaios regionais

Wagner Mateus Costa Melo¹, Marcio Balestre², Renzo Garcia Von Pinho¹

¹ Departamento de Agricultura, Universidade Federal de Lavras. C.P 3037. Lavras. Minas Gerais, Brasil.

² Departamento de Ciências Exatas, Universidade Federal de Lavras, C.P 3037. Lavras, Minas Gerais, Brasil

Resumo

O objetivo com este trabalho foi prever o desempenho de híbridos de milho e verificar se o método de efeitos totais das marcas associadas (ETMA) prediz corretamente os híbridos. Foram feitas dos tipos de verificação, a validação cruzada (prever os híbridos e validar com os resultados já obtidos dos mesmos em ensaios) e resultados de ensaios conduzidos com os híbridos preditos. Os ensaios de verificação dos híbridos preditos foram conduzidos em seis locais do Sul do Brasil (Guarapuava PR, Vacaria RS, Ipiranga PR, Sananduva RS, Itapeva SP e Faxinal dos Guedes SC) na safra 2010/11. Os ensaios regionais foram conduzidos em outras seis cidades no sul do Brasil (Sananduva RS, Vacaria RS, Guarapuava PR, Ipiranga PR, Itapeva SP e Faxinal dos Guedes SC) na safra 2011/12. Para isto, utilizou-se cinquenta e uma linhagens de diferentes origens e, a partir de cruzamentos dessas linhagens, foram obtidos 58 híbridos simples em um sistema de diallelo parcial entre 27 linhagens do grupo heterótico Swan e 24 linhagens do grupo heterótico Amarelo Dentado. Setenta e nove marcadores de microssatélites foram utilizados na genotipagem das 51 linhagens. Com a informação gerada por esses marcadores foram construídas as matrizes dos efeitos aditivos e dominantes. Observou-se que, sob diferentes níveis de tamanho de amostra, a acurácia preditiva variou de 0,10 a 0,96. Além disto, quando os híbridos não testados na população de treino

tiveram seus valores preditos para os ensaios regionais, a acurácia foi de 0,30. No que diz respeito à perda seletiva, o método de efeitos totais das marcas associada acertou em 50% dos híbridos selecionados nos ensaios regionais, valor considerado extremamente bom, pois a predição e checagem desta foram realizadas em ambientes totalmente diferentes. Também houve perda em apenas 33% dos casos, ou seja, dos materiais preditos como bons somente 33% foram considerados como ruins de acordo com os resultados dos ensaios regionais. Os resultados desse trabalho mostram que a validação preditiva em diferentes condições de cultivo é possível e que os resultados de validação cruzada representam de forma contundente a situação real de desempenho no campo.

Palavras-chave: Milho híbrido, Melhoramento Genético, BLUP, marcadores moleculares, ETMA, seleção genômica

Prediction of single maize hybrids using the total associated mark effects approach with cross validation and in regional trials

Wagner Mateus Costa Melo¹, Marcio Balestre², Renzo Garcia Von Pinho¹

¹ Departamento de Agricultura, Universidade Federal de Lavras. Cp 3037. Lavras. Minas Gerais, Brazil.

² Departamento de Ciências Exatas, Universidade Federal de Lavras, C.P 3037. Lavras, Minas Gerais, Brazil.

Abstract

The purpose of this study was to predict the performance of maize hybrids and verify if the method of the total effects of associated mark (TEAM) correctly predicts the hybrids. Verification was performed in two ways, cross validation (predicting the hybrids and validating them with the results already obtained from them in trials) and results of trials conducted with the predicted hybrids. Verification trials of the predicted hybrids were carried out in six locations in the South of Brazil (Guarapuava/PR, Vacaria/RS, Ipiranga/PR, Sananduva/RS, Itapeva/SP and Faxinal dos Guedes/SC) in the season 2010/11. The regional trials were carried out in another six location in the South of Brazil (Sananduva RS, Vacaria RS, Guarapuava PR, Ipiranga PR, Itapeva SP e Faxinal dos Guedes SC) in the season 2011/12. For this purpose, fifty-one lines from different backgrounds were used, and from the crosses of these lines, 58 single hybrids were obtained in a partial diallel system among 27 lines of the Swan heterotic group and 24 lines of the Amerelo Dentado heterotic group. Seventy-nine microsatellite markers were used in the genotyping of the 51 lines. With the information generated by these markers, the matrices of the additive and dominant effects were constructed. It was observed that under different size levels of the sample, the predictive accuracy varied from 0,10 to 0,96. In addition, when the hybrids not tested in the training population had their values

predicted for the regional trials, the accuracy was 0,30. In regard to selective loss, the total effects of associated mark method was right in 50% of the hybrids selected in the regional trials, a value considered to be extremely good because the prediction and checking of it were performed in totally different environments. There was also a loss in only 33% of the cases, i.e., of the materials predicted as good, only 33% were considered as bad according to the results of the regional trials. The results of this study show that predictive validation under different growing conditions is possible and that the results of cross validation represent the real field performance situation in a striking manner.

Keywords: Hybrid maize, Genetic breeding, BLUP, molecular markers, ETMA, genomic selection

Introdução

Um programa de melhoramento de milho visa à obtenção de híbridos cada vez mais produtivos e para isto é necessário à realização de vários cruzamentos e geração de diversos híbridos a serem testados nas redes de ensaios. Embora existam várias estratégias de melhoramento, que podem ir desde o melhoramento populacional à obtenção e seleção de haploides para síntese de linhagens, pode-se dizer que a escolha de parentais e delineamentos de cruzamentos (fatoriais, dialelos, trialelos, tetraalelos) são pontos importantes desse processo.

Mesmo existindo diversas técnicas de cruzamentos o planejamento dos cruzamentos é realizado considerando ou a “experiência” do melhorista, ou por meio de características desejáveis das linhagens parentais e grupos heteróticos

preestabelecidos (HALLAUER, CARENA, MIRANDA-FILHO 2010). Embora a “experiência” do melhorista seja de grande utilidade, o emprego de métodos de predição de híbridos serve de ferramenta para o direcionamento dos cruzamentos (BUENO-FILHO, GILMOR 2003).

Outro fator a ser levado em consideração na escolha de cruzamentos é o número de linhagens a ser utilizada na síntese de híbridos. Isto se torna problemático à medida que o número de linhagens aumenta e o leque de cruzamentos se torna tão grande que inviabiliza economicamente a obtenção de todos os cruzamentos (BERNARDO 1994). Dessa forma, é inevitável que o melhorista realize vários cruzamentos e descarte outros. O problema dessa prática é que vários cruzamentos promissores não são realizados e outros tantos cruzamentos inferiores são avaliados. Assim, torna-se fundamental que, mesmo avaliando apenas uma parcela dos cruzamentos, o melhorista seja capaz de prever todos os cruzamentos não avaliados de modo a recuperar genótipos superiores que seriam descartados.

Os primeiros trabalhos de predição de híbridos simples utilizando marcadores moleculares foram feitos com a distância genética entre as linhagens. Uma vez que a heterose pode ser descrita em função da distância genética e do efeito de dominância quando desconsideramos a epistasia (MELCHINGER, 1999), pode-se conceber que quanto mais divergente os genitores, maior a heterose esperada.

Utilizando essa teoria simplificada da genética quantitativa, vários estudos foram conduzidos a partir da década de 90 visando prever o comportamento de híbridos com base na distância genética (SMITH et al., 1994; LANZA et al., 1997; BARBOSA et al. 2003; BALESTRE et al. 2008; DEVI e SINGH, 2011; DI et al. 2012) . No entanto, os resultados obtidos por essa abordagem são inconclusivos e, de maneira geral, não refletem confiabilidade para utilização em larga escala em programas de melhoramento.

Uma nova abordagem foi desenvolvida paralelamente à predição por marcadores, mas baseada nos trabalhos de Henderson (1952), onde se utilizou modelos mistos, pedigrees complexos e marcadores moleculares (BERNARDO 1994). Esse autor unificou as análises de modelos mistos e marcadores moleculares na predição de híbridos simples. A técnica consiste basicamente de se utilizar a matriz de parentesco dos candidatos à seleção nas equações dos modelos mistos para recuperar híbridos não avaliados com base na sua covariância genética com os híbridos testados. Essa técnica tem sido eficiente na predição de híbridos nas mais diversas situações (CHARCOSSET et al. 1998; SCHRAG et al. 2007; FERREIRA et al. 2010; GOWDA et al. 2012).

Recentemente, Schrag et al. (2007, 2009, 2010) propuseram a substituição da matriz de similaridade genética utilizada na técnica de Bernardo pelo genótipo do próprio marcador na matriz das equações de modelos mistos. Essa metodologia foi denominada TEAM ou “*total effects of associated markers*” ou ETMA efeito total das marcas associadas. Nessa abordagem, a matriz do delineamento da CGC e CEC é substituída pela matriz aditiva e dominante dos marcadores moleculares e utiliza-se uma regressão aleatória (RR-BLUP) onde valores genéticos são atribuídos a cada marca. Nesse contexto, o valor genotípico do híbrido é recuperado por meio dos efeitos totais dos marcadores, como se fosse uma seleção genômica. Esse método busca a obtenção dos valores genéticos das marcas a partir de uma população de treino e, posteriormente, se valida a predição no conjunto genético restante que pode ser tanto os híbridos não testados ou a própria população de treino via validação cruzada. A validação cruzada é a validação da predição utilizando os resultados dos ensaios dos materiais utilizados na população treino.

Essa metodologia tem se mostrado mais eficiente que o método proposto por Bernardo (1994) na predição de híbridos (SCHRAG et al. 2009). De forma

mais contundente Massman et al. (2012) estendeu essa abordagem para a seleção genômica propriamente dita.

Baseado nisto, o objetivo com este trabalho foi validar o método de efeitos totais das marcas associadas utilizando validação cruzada e resultados de ensaios.

Material e Métodos

Informações experimentais

No inverno de 2010, no campo experimental de Uberlândia cinquenta e uma linhagens de diferentes origens foram cruzadas para a formação da população de treino. Dos cruzamentos dessas linhagens foram obtidos 58 híbridos simples em um sistema de dialelo parcial (Figura 1). Estes materiais foram avaliados em sete locais distribuídos na região sul do Brasil (Vacaria RS plantio 10/10/2010, Abelardo Luz SC plantio 14/10/2010, Arapoti PR plantio 16/10/2010, Candoi PR plantio 24/10/2010, Canoinhas PR plantio 26/10/2010, Castro PR plantio 28/10/2010 e Ponta Grossa PR plantio 02/11/2010). Posteriormente, no verão 2010/11 foram produzidos mais 119 novos híbridos resultantes do inter cruzamento entre estas 51 linhagens totalizando 175 genótipos a serem avaliados no segundo ano de avaliação, que estamos considerando como ensaios regionais. Assim, esses 175 e mais 1419 híbridos de outras origens que foram avaliados conjuntamente em ensaios regionais conduzidos em seis cidades no sul do Brasil (Sananduva RS plantio 18/10/2011, Vacaria RS plantio 16/10/2011, Guarapuava PR plantio 14/10/2011, Ipiranga PR plantio 20/10/2011, Itapeva SP plantio 24/10/2011 e Faxinal dos Guedes SC 28/10/2011). As condições dos ensaios regionais foram às mesmas descritas para os ensaios da população treino, ou seja, ensaios de primeiro ano, embora as cidades não fossem necessariamente as mesmas.

Os dois experimentos, dos dois anos, foram conduzidos no modelo experimental de blocos incompletos com duas repetições e cada parcela possui quatro linhas de 5 m, com espaçamento entre linhas de 70 cm. Foi avaliada a produção de grãos, sendo que o peso da parcela foi corrigido para umidade de 13% e convertido em t.ha⁻¹. As adubações de plantio e de cobertura foram feitas de acordo com a recomendação de adubação para cada área, os tratos culturais foram feitos para controle de lagarta do cartucho (*Spodoptera Frugiperda*) e da espiga (*Helicoverpa Zea*) e também para o controle do mato.

Informações Moleculares

Setenta e nove marcadores de microssatélites foram utilizados na genotipagem das 51 linhagens (Tabela 1). Esses marcadores estão distribuídos ao longo dos 10 grupos de ligação do milho.

A matriz de informação de marcadores foi construída utilizando o código 1 para presença do alelo t na marca m na linhagem i e 0 caso contrário. Essa codificação facilita a construção das matrizes aditivas e dominantes dos híbridos.

Utilizando essa codificação e considerando que a recombinação é irrelevante em linhagens homozigotas, a matriz aditiva dos híbridos foi construída da seguinte forma:

$$A_{ik} = \begin{cases} 2 & \text{se } a_{1i} = a_{1j} = 1 \\ 1 & \text{se } a_{1i} \neq a_{1j} \\ 0 & \text{se } a_{1i} = a_{1j} = 0 \end{cases} \quad \forall a_{1i} \vee a_{1j} = 1, 0$$

onde a é fenótipo do t -ésimo alelo da marca m nas linhagem i e j .

De maneira semelhante, a matriz dos efeitos de dominância foi construída utilizando a seguinte relação:

$$\Delta_{lk} = \begin{cases} a_{1i} \times a_{1j} \vee a_{2i} \times a_{2l} & \text{se } t_i = t_j \\ a_{1i} \times a_{2j} + a_{1j} \times a_{2i} & \text{caso contrário} \end{cases}$$

sendo 1 e 2 os t-ésimos alelos da marca m . No primeiro caso, tem o desvio de dominância para os homozigotos e, no segundo caso, o desvio para a complementação heterozigota.

Análise dialélica e efeito total das marcas associadas

Os dados fenotípicos dos híbridos foram analisados utilizando o modelo de dialelos parciais e modelo de marcadores.

No primeiro caso assumiu-se o seguinte modelo linear:

$$y = X\beta + Z_1a_1 + Z_2a_2 + Z_3d + Z_4i_1 + Z_5i_2 + Z_6w + \xi \quad (1)$$

Onde y é a média fenotípica na parcela, X é a matriz de incidência dos efeitos fixos (blocos, repetições e ambientes), Z_1 a Z_6 correspondem às matrizes de incidências dos efeitos aleatórios. Os parâmetros a_1 , a_2 , d , i_1 , i_2 e w correspondem aos efeitos da CGC do grupo1 (CGC1), CGC do grupo2 (CGC2), CEC, interação CGC1xAmbientes, CGC2 x Ambientes e CEC x Ambientes respectivamente, ξ é o resíduo do modelo. As estimativas dos efeitos fixos, dos componentes da variância fenotípica e as predições dos efeitos aleatórios foram obtidas via máxima verossimilhança restrita (REML), utilizando o algoritmo EM (*expectation- maximization*).

No modelo de marcadores, as matrizes de incidência dos efeitos dos parentais e combinação específica foram substituídas pelas matrizes dos efeitos aditivos e dominantes das marcas. Assim, o modelo de efeitos totais das marcas associadas foi dado por:

$$y = X\beta + Aa + \Delta d + \xi \quad (2)$$

onde A corresponde aos efeitos aditivos e Δ aos desvios de dominância. Ao contrário do modelo RR-BLUP - normalmente utilizado na seleção genômica, o modelo 2 se caracteriza como um modelo misto já que os efeitos ambientais são levados em consideração na matriz dos efeitos fixos; além disto, a interação do valor dado a cada marca por ambientes foi colocada no resíduo devido ao alto custo computacional envolvido na estimação do efeito de interação de cada alelo e interação alélica com o ambiente.

O valor genético aditivo e de dominância de cada híbrido foi recuperado pelo somatório dos efeitos aditivos de cada alelo que o indivíduo possui; em

outras palavras, $\alpha_i = \sum_{j=1}^n \lambda a_j : \lambda = \{0, 1, 2\}$ e $\delta_i = \sum_{j=1}^k \phi d_j : \phi = \{0, 1\}$ onde k é o

número total de interações alélicas l dentro de cada marca m , e n é o número de alelos observados nas m marcas.

As estimativas dos componentes das variâncias fenotípicas e efeitos fixos foram obtidos junto com a predição dos efeitos aditivos e dominantes contidos em cada marca via máxima verossimilhança restrita. A variância genética total recuperada foi tomada como comum para cada marca e dada por:

$$\sigma_g^2 = n\sigma_a^2 + k\sigma_d^2 \text{ onde } \sigma_a^2 = \left(\sum_{j=1}^n a_j^2 + \text{traço}[I_n W_n^{-1} \sigma^2] \right) / n \quad \text{e}$$

$$\sigma_d^2 = \left(\sum_{j=1}^k d_j^2 + \text{traço}[I_n W_2^{-1} \sigma^2] \right) / k, \text{ sendo } W_n^{-1} \text{ e } W_k^{-1} \text{ submatrizes da}$$

inversa da matriz de equações de modelos mistos.

Validação cruzada e validação com dados reais

A validação cruzada foi realizada no conjunto de híbridos da população treino. Para isso, diferentes níveis de desbalanceamento foram aplicados aos dados e os valores genotípicos dos híbridos tirados para o desbalanceamento foram recuperados.

Os níveis de desbalanceamento variaram de 5 a 51%. Utilizou-se o modelo aditivo e aditivo + dominante para predição do valor genotípico de cada híbrido. A correlação entre valores preditos e observados foi utilizada como parâmetro na validação cruzada. Também se utilizou os dados brutos das distâncias genéticas como preditor das combinações específicas dos híbridos. A distância genética foi calculada como $1 - s_{ij}$, onde s_{ij} é a similaridade genética de Jaccard.

Na segunda fase do trabalho, as combinações híbridas não testadas na fase de treino foram preditas e comparadas com os resultados dos ensaios regionais; ou seja, todas as combinações possíveis das 51 linhagens foram preditas e aquelas presentes nos ensaios regionais foram utilizadas para a validação da técnica. A credibilidade da técnica foi obtida via ajuste quadrático de regressão entre as CEC preditas no conjunto de treino e os valores observados nos ensaios regionais.

Resultados

A variância da capacidade geral do grupo 1 foi de $0,055 \text{ t.ha}^{-1}$ (linhagens 1-27; Figura 1) enquanto a variância genética do grupo II foi de $0,077 \text{ t.ha}^{-1}$ (linhagens 28-51; Figura 1). Considerando a soma das duas variâncias genéticas temos $0,132 \text{ t.ha}^{-1}$, valor este superior ao obtido para a variância da capacidade específica de combinação $0,063 \text{ t.ha}^{-1}$. As três variâncias praticamente decompõem a variância genética total obtida na análise de híbridos

(0,22 t.ha⁻¹). A herdabilidade foi de 0,65, sendo que a variância fenotípica foi fortemente influenciada pela interação híbridos por ambientes. A interação híbridos por ambientes foi três vezes maior que a variância genética e 5 vezes maior que a variância residual ($\sigma_{CGCi}^2 = 0,45 t^2.ha^{-2}$ $\sigma_{CECi}^2 = 0,23 t^2.ha^{-2}$), sendo assim pode-se inferir que o desempenho dos híbridos foi altamente influenciado pela sua interação com os ambientes.

A análise por marcadores moleculares nas linhagens revelou 636 alelos diferentes, com uma média de 8,05 alelos por loco. Os cruzamentos realizados, figura 1, geraram 1168 efeitos de dominância e que somados aos efeitos aditivos totalizam 1804 efeitos genéticos que praticamente descreveram toda a variância genética de híbridos ($\sigma_g^2 = n\sigma_a^2 + k\sigma_d^2 = 0,25 t.ha^{-1}$). Esse resultado pode ser confirmado na figura 2, onde o ajuste do modelo ETMA-2, que inclui os efeitos aditivos e dominantes, foi mais efetivo em descrever o desempenho dos híbridos do que o modelo ETMA-1 que descreve apenas os efeitos aditivos das marcas. Na figura 2 pode-se verificar que os efeitos aditivos foram mais importantes que os efeitos de dominância, pois o ajuste entre os valores preditos com o modelo estritamente aditivo e o valor genotípico foi de 0,84 e com a adição no modelo dos efeitos dominantes o ajuste passou para 0,95, ou seja, acrescentou 0,11 no ajuste com a adição dos efeitos de dominância. Estes resultados confirmam a relação de magnitude das variâncias de CGC e CEC obtidas na análise dialélica.

O processo de validação cruzada demonstrou que a credibilidade de predição do método ETMA 1 e 2 foi de razoável a baixa, sendo que o modelo que inclui os efeitos de dominância foi levemente superior ao modelo estritamente aditivo, como pode ser observado na figura 2. Os resultados da tabela 2 mostram à média, a mediana e a moda das correlações entre os valores preditos e observados quando consideramos o modelo estritamente aditivo (ETMA1) e aditivo + dominante (ETMA2). De acordo com estes resultados

pode-se observar que à medida que o número de híbridos tirados aumenta (aumento do desbalanceamento), a moda das correlações entre os valores observados e preditos do modelo ETMA-1 tende assintoticamente ao valor de correlação da distância genética e produção de grãos 0,10, tabela 2 tamanho amostra 0*. Vale destacar que esse valor foi muito baixo e não significativo pelo teste de Mantel. O modelo ETMA-2, que estamos considerando o modelo aditivo e com os efeitos de dominância, este foi pouco afetado pela quantidade de híbridos desbalanceados e manteve a correlação mais estável, com a moda da distribuição das correlações na faixa dos 0,34 mesmo quando mais de 50% dos híbridos foram descartados.

No modelo ETMA-2 o ajuste entre os híbridos preditos e observados nos ensaios regionais foi de $R^2 = 0,09$ (Figura 3). Nessa figura podemos observar que boa parte dos híbridos de alto desempenho nos ensaios regionais foi também àqueles preditos com alto desempenho (circulados de verde) e que os híbridos de baixo desempenho nos ensaios regionais também foram preditos com baixa produtividade (circulados de vermelho). Esse resultado demonstra que, mesmo com um baixo ajuste, de maneira geral, o método descartou os híbridos de baixo desempenho e selecionou os híbridos de alto desempenho.

Na figura de 4 a 6 temos o ajuste preditivo do desempenho de híbridos por parental, considerando os parentais que participaram, como genitor, de pelo menos quatro híbridos. Os resultados demonstram que, para a maioria dos parentais, a predição conduziu a seleção dos melhores e descarte dos piores híbridos, sendo que, apenas os parentais 5, 10, 11 e 14, a predição conduziu à seleção dos piores híbridos e descarte dos melhores. Nestas figuras observa-se que a capacidade de predição foi alta para os cruzamentos envolvendo os parentais 6 e 9 e baixa para as linhagens 10 e 11, sugerindo que a credibilidade do método pode variar de linhagem para linhagem.

O ajuste entre o desempenho dos híbridos testados no conjunto de treinamento (ensaios do primeiro ano) e seu desempenho nos ensaios regionais (ensaio de segundo ano) é dado na Figura 7. Essa figura mede a credibilidade da seleção nos ambientes de treino (ensaios do primeiro ano) e a seleção nas condições dos ensaios do segundo ano. A segunda parte desta figura mostra o ajuste da predição, considerando o modelo aditivo e dominante (ETMA2), com os valores observados nos ensaios regionais. Dado que a herdabilidade dos híbridos na população de treino foi de 0,65 seria esperado que esse valor se repetisse ao longo dos outros ambientes já que a CEC e CGC são livres do efeito de interação. No entanto a figura 6 mostra que esse ajuste foi de 0,23, e quando se utilizou o desempenho predito pelo método ETMA-2 esse ajuste caiu quase pela metade (0,13), bem próximo ao ajuste obtido entre os híbridos preditos e os valores obtidos nos ensaios regionais (0,09) (Figura 3). Pode-se inferir que quando a interação genótipos por ambientes é alta - como observado nesse trabalho - os BLUP's dos híbridos não podem ser extrapolados para qualquer ambiente, mesmo estando livres dos efeitos de interação e os ambientes estiverem dentro de uma mesma macrorregião. O mesmo podemos extrapolar em relação a recomendação de cultivares.

A figura 1 resume toda dinâmica desse trabalho. Nessa figura pode-se observar os híbridos avaliados na população de treino, ensaios de primeiro ano (células pretas), os híbridos avaliados no ensaio regional, ensaios de segundo ano (células vermelhas) e os híbridos preditos e selecionados a 20% (células verdes). As células vermelhas e pretas se referem aos híbridos que foram avaliados nos ensaios do primeiro e do segundo ano. As células com a letra B se referem aos híbridos selecionados nos ensaios de segundo ano. Assim, aquelas células com a letra B e que estão marcadas em verde referem-se aos híbridos preditos como bons e confirmados nos ensaios de segundo ano, ou seja, aqueles híbridos onde a seleção foi correta. Já as células vermelhas marcadas em verde e

sem a letra B referem-se aos híbridos preditos como bons, mas não selecionados nos ensaios do segundo ano, ou seja, o erro de seleção. Fica claro que essa figura representa uma função de perda do tipo 0 e 1, ou seja, 0 é quando selecionou-se corretamente (perda nula) e 1 quando a seleção foi incorreta (perda total). Considerando esse parâmetro e contando as células da figura verifica-se que houve acerto em 50% e erro em 33%. Em outras palavras, dos híbridos selecionados nos ensaios regionais, o método foi capaz de acertar em 50%, sendo este, um valor extremamente bom, pois a predição foi feita com base em resultados de ambientes extremamente diferentes (locais dos ensaios do primeiro ano de avaliação) dos ambientes utilizados para a validação da técnica (locais dos ensaios do segundo ano de avaliação). Além disto, foram selecionados apenas 33% dos híbridos descartados nos ensaios regionais, sendo um valor extremamente bom considerando o número de marcas utilizadas e o número de híbridos testados. Esse resultado mostra que o método é um melhor para o descarte de híbridos de baixo desempenho do que na seleção dos melhores híbridos.

Discussões

Houve uma tendência na melhora da predição dos valores genotípicos utilizando os marcadores moleculares, dado que os BLUP's da CEC e CGC puderam ser decompostos em 1804 efeitos aditivos e dominantes e o somatório desses efeitos praticamente recuperaram a variância genética de híbridos. A melhora na predição utilizando marcadores só não foi maior, pois os modelos ETMA não incluíram a interação por ambientes, sendo esse efeito confundido com a variância residual. Foi possível verificar que a interação genótipos por ambientes teve grande participação na variância fenotípica dos híbridos e foi a principal causa da falta de correlação genética entre os híbridos nas sete

localidades avaliadas dos ensaios de segundo ano, portanto caso fosse incluída a interação por ambiente a predição do valor genotípico teria sido muito melhor.

Os resultados de predição observados nesse estudo são animadores, dado o tamanho da população de treino (58 híbridos), o número de marcas utilizadas (78 marcas) e os locais de avaliação do primeiro e do segundo ano serem ambientes totalmente diferentes. Por exemplo, Schrag et al (2009 e 2010) obtiveram altos valores de R^2 (0,16-0,65) utilizando um desbalanceamento de 50% em conjunto de 400 híbridos testados e mais de 1000 marcadores moleculares. Utilizando menos de 10% de marcas e 15% de híbridos na população de treino, nossos resultados alcançaram um ajuste de 0,15 para esse nível de desbalanceamento. Por outro lado, em comparação ao trabalho de Massman et al (2012) que, utilizando validação cruzada, obtiveram acurácias variando de 0,75 a 0,87 para um desbalanceamento de 10%, nossos resultados chegaram a 0,62 para esse mesmo nível de desbalanceamento e alcançou uma acurácia de 0,95 com 5% de híbridos desbalanceados (3 híbridos tirados do total de 58).

Outro resultado importante observado nesse trabalho foi à ineficácia do uso das distâncias genéticas para a predição do desempenho de híbridos de milho, correlação de 0,10 entre a distância genética e produção de grãos, valor este não significativo pelo teste de Mantel. Resultados semelhantes foram observados por Lessege et al. (2008), Devi e Singh (2011) e Gowda et al (2012). Assim pode-se inferir que é melhor utilizar a informação de marcadores como preditor em um modelo linear do que as distâncias genéticas como preditoras do desempenho de híbridos, dado que, mesmo utilizando um modelo desbalanceado, a quantidade de informação contida no ETMA é muito maior do que as distâncias per se.

No que concerne aos métodos de validação de predição, os trabalhos descritos na literatura utilizam apenas a validação cruzada como método para

acessar a acurácia e de forma alguma testam o valor predito dos híbridos em condições completamente diferentes da condição de treino, ou seja, sob diferentes ambientes e em competição com genótipos completamente distintos daqueles utilizados na validação cruzada. Nesse sentido, este trabalho distingue-se dos demais por levar a eficácia da técnica ao extremo das condições de predição e por considerar esse parâmetro para validação da técnica. Os resultados obtidos nesse estudo demonstraram que a acurácia de predição do desempenho de híbridos em diferentes locais e condições de cultivo distintas da população de treino é semelhante à obtida sob níveis de desbalanceamento acima de 30% (Tabela 1, Figura 3). Assim, pode-se utilizar a validação cruzada sob altos níveis de desbalanceamento para acessar a eficácia do método em diferentes condições de cultivo.

Além disso, observou-se que o valor de acurácia foi mais estável que a credibilidade dos valores genotípicos obtida pela herdabilidade, uma vez que a herdabilidade seletiva foi de 0,65 e a realizada foi de 0,22. Esse valor de 0,22 foi obtido pela regressão dos valores genotípicos dos híbridos testados na população de treino e os valores genotípicos desses mesmos híbridos nos ensaios regionais. Quando se utilizou o método ETMA para obtenção dos valores genotípicos desses mesmos híbridos observou-se que a credibilidade foi de 0,13, um pouco acima da credibilidade dos híbridos não testados na população de treino que foi de 0,09 (Figura 3).

Outro resultado interessante foi da capacidade do método em selecionar híbridos corretamente. Tomando a Figura 1 como um alvo e considerando o valor de 0 para cada vez que o híbrido certo foi selecionado e o valor de 1 para cada vez que o híbrido errado foi selecionado tem-se uma função de perda do tipo 0 e 1. Na figura 1 ficou evidente que 50% (30) dos melhores híbridos avaliados no ensaio regional foram aqueles selecionados com valores preditos na população de treino sob um índice de seleção proporcional a 20%. Além disso,

apenas 33% (20) dos híbridos descartados no ensaio regional foram selecionados pelos valores preditos na população de treino. Fica evidente que o método ETMA é mais eficiente em não cometer erros do tipo 1 (selecionar o errado) do que erros do tipo 0 (não selecionar o certo) e nesse caso a moda da distribuição de validação cruzada pode ser utilizada.

Neste trabalho foi possível verificar que a predição de híbridos, principalmente de milho, precisa ainda ser mais bem explorada. O aumento no número de marcas e de indivíduos na população de treino pode ser determinante para a validação da técnica. No entanto, tornou-se evidente que a validação em ensaios regionais, principalmente em locais e sob condições de cultivo diferentes das adotadas nas populações de treino, pode fornecer ao melhorista uma visão mais ampla sobre a eficácia da técnica. Nesse sentido, os resultados desse trabalho sugerem que a validação em diferentes condições de cultivo é possível e que os resultados de validação cruzada representam de forma contundente a situação real de desempenho no campo.

Agradecimentos

A Universidade Federal de Lavras e a empresa Advanta Comércio de Sementes.

Referências

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

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

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

BUENO FILHO, J. S. S. ; GILMOUR, S. G. . Planning Incomplete Block Experiments When Treatments Are Genetically Related. **Biometrics** (Washington), Washington, v. 59, n.2, p. 375-381, 2003.

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

DEVI, P.; SINGH N. K.; Heterosis, molecular diversity, combining ability and their interrelationships in short duration maize (*Zea mays* L.) across the environments. **Euphytica** v.178. p.71–81. 2011.

DI, R.; CHU, M.X.; LI, Y.L.; ZHANG, L.; FANG, L.; FENG, T.; CAO, G.L.; CHEN, H.Q.; LI, X.W. Predictive potential of microsatellite markers on heterosis of fecundity in crossbred sheep. **Molecular Biology Reports**. v.39, n. 3 p.2761-2766, 2012

FERREIRA, D.V.; VON PINHO, R.G.; BALESTRE, M.; OLIVEIRA, R.L. Prediction of maize hybrid performance using similarity in state and similarity by descent information. **Genetics and Molecular Research**. v. 9 n. 4, p.2381-2394, 2010

HALLAUER, A.R.; CARENA, M.J.; MIRANDA-FILHO, J. B. **Quantitative Genetics in Maize Breeding**. Springer , New York , 2010, 680 p.

HENDERSON, C.R. Specific and general combining ability. In: Gowen JW (ed) Heterosis. Iowa State College Press, Ames, p 352–370. 1952.

GOWDA, M.; ZHAO Y, MAURER, H. P.; WEISSMANN E. A.; WÜRSCHUM T, REIF J.C. Best linear unbiased prediction of triticales hybrid performance. **Euphytica**, versão on-line, 2012, DOI 10.1007/s10681-012-0784-z

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

LEGESSE, B.W.; MYBURG, A.A; PIXLEY, K.V; TWUMASI-AFRIYIE, S.; et al. Relationship between hybrid performance and AFLP based genetic distance in highland maize inbred lines. **Euphytica**, v. 162, p. 313-323, 2008.

MASSMAN, J.M.; GORDILLO, A.; LORENZANA, R.E.; BERNARDO, R.
Genomewide predictions from maize single-cross data. **Theoretical and Applied Genetics**. 2012. versão on-line.

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

SCHRAG T,A.; MAURER H,P.; MELCHINGER, A.E.; PIEPHO ,H.P.; PELEMAN, J.; FRISCH, M. Prediction of single-cross hybrid performance in maize using haplotype blocks associated with QTL for grain yield. . **Theoretical and Applied Genetics**, Berlin, v .114 n.2, p.1345–1355. 2007

SCHRAG, T.A.; MÖHRING, J.; MAURER, H.P.; DHILLON, B.S.; MELCHINGER, A.E.; PIEPHO, H.P.; SØRENSEN, A.P.; FRISCH, M. Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. **Theoretical Applied Genetics**, Berlin, n.118, n. 4, p. 741–751. 2009

SCHRAG, T.A.; MÖHRING J.; MELCHINGER, A.E.; KUSTERER, B.; DHILLON, B.S.;o • PIEPHO, H.P.; SØRENSEN, A.P.; FRISCH, M. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. **Theoretical Applied Genetics**, Berlin, n.120, n. 1, p. 451–461. 2010

SMITH, S.; LUK, S.; SOBRAL, B.; MUHAWISH, S.; PELEMAN, J.; ZABEAU, M. Association among inbred lines of maize using RFLP and amplification technologies (AFLP and AP-PCR), and correlations with pedigree, F1 yield and heterosis. **Maize Genetics Newsletter**, Oxford, v. 68, p. 45, 1994

Tabela 1. Distribuição dos 79 marcadores de microssatélites nos 10 grupos de ligação (GL) do milho.

Marcador	GL	Bin	Marcador	GL	Bin	Marcador	GL	Bin	Marcador	GL	Bin
bnlg1179	1	1,01	dupssr08	3	3,09	bnlg1200	7	7,01	umc1139	8	8,01
bnlg1014	1	1,01	bnlg1496	3	3,09	bnlg1808	7	7,02	bnlg1056	8	8,01
bnlg1007	1	1,02	umc1136	3	3,09	bnlg1305	7	7,03	bngl2082	8	8,03
bnlg1614	1	1,02	phi072	4	4,01	umc1342	7	7,04	bngl1067	8	8,03
bnlg1866	1	1,03	umc1101	4	4,09	bnlg2259	7	7,04	umc1858	8	8,04
umc1128	1	1,07	umc1109	4	4,10	umc1154	7	7,05	phi015	8	8,08
phi037	1	1,08	umc1197	4	4,11	umc1075	8	8,01	bnlg1131	8	8,09
bnlg1643	1	1,08	umc1058	4	4,11	umc1414	8	8,01	bnlg2122	9	9,01
umc1725	1	1,11	phi019	4	4,11	bnlg1194	8	8,01	umc1040	9	9,01
umc1797	1	1,12	umc1591	5	5,04	phi119	8	8,02	bnlg1724	9	9,01
umc1079	2	2,06	umc1482	5	5,04	umc1034	8	8,03	umc1078	9	9,05
bnlg1036	2	2,06	bnlg1237	5	5,05	phi115	8	8,03	umc1310	9	9,06
dupssr24	2	2,08	bnlg1118	5	5,07	mmc412	8	8,03	umc1319	10	10,01
bnlg1520	2	2,09	bnlg1371	6	6,01	umc2146	8	8,03	bnlg1079	10	10,03
umc1970	3	3,01	umc1006	6	6,02	phi121	8	8,03	umc2043	10	10,05
bnlg1601	3	3,05	umc1887	6	6,03	umc2147	8	8,03	bnlg1074	10	10,05
bnlg1160	3	3,06	umc1918	6	6,04	umc1157	8	8,03	umc1061	10	10,06
umc1148	3	3,07	bnlg1740	6	6,07	umc1202	8	8,05	bnlg1360	10	10,07
umc1167	3	3,08	phi089	6	6,08	bngl240	8	8,06	umc1084	10	10,07
bnlg1108	3	3,08	umc1066	7	7,01	umc1933	8	8,08			

Tabela 2 e 3. Média, mediana e moda das correlações entre a predição de híbridos não testados e os resultados da população treino, utilizando diferentes tamanhos de desbalanceamento no modelo aditivo (A – Tabela 2) e no modelo aditivo + dominante (A+D – Tabela 3).

Tamanho amostral	A			A+D		
	Média	Mediana	Moda	Média	Mediana	Moda
3 (5%)	0,28	0,51	0,95	0,39	0,68	0,94
6 (10%)	0,36	0,44	0,53	0,38	0,47	0,62
9 (15%)	0,21	0,30	0,44	0,36	0,35	0,32
12 (20%)	0,28	0,26	0,26	0,35	0,36	0,37
15 (25%)	0,26	0,29	0,39	0,32	0,32	0,34
17 (30%)	0,25	0,23	0,21	0,30	0,33	0,40
20 (35%)	0,22	0,24	0,27	0,29	0,30	0,39
25 (43%)	0,22	0,23	0,23	0,27	0,27	0,34
30 (51%)	0,15	0,14	0,09	0,27	0,28	0,34
*	0,10 ^{ns}			0,07		

* Correlação obtida via distância genética ^{ns} não significativo pelo teste de Mantel,

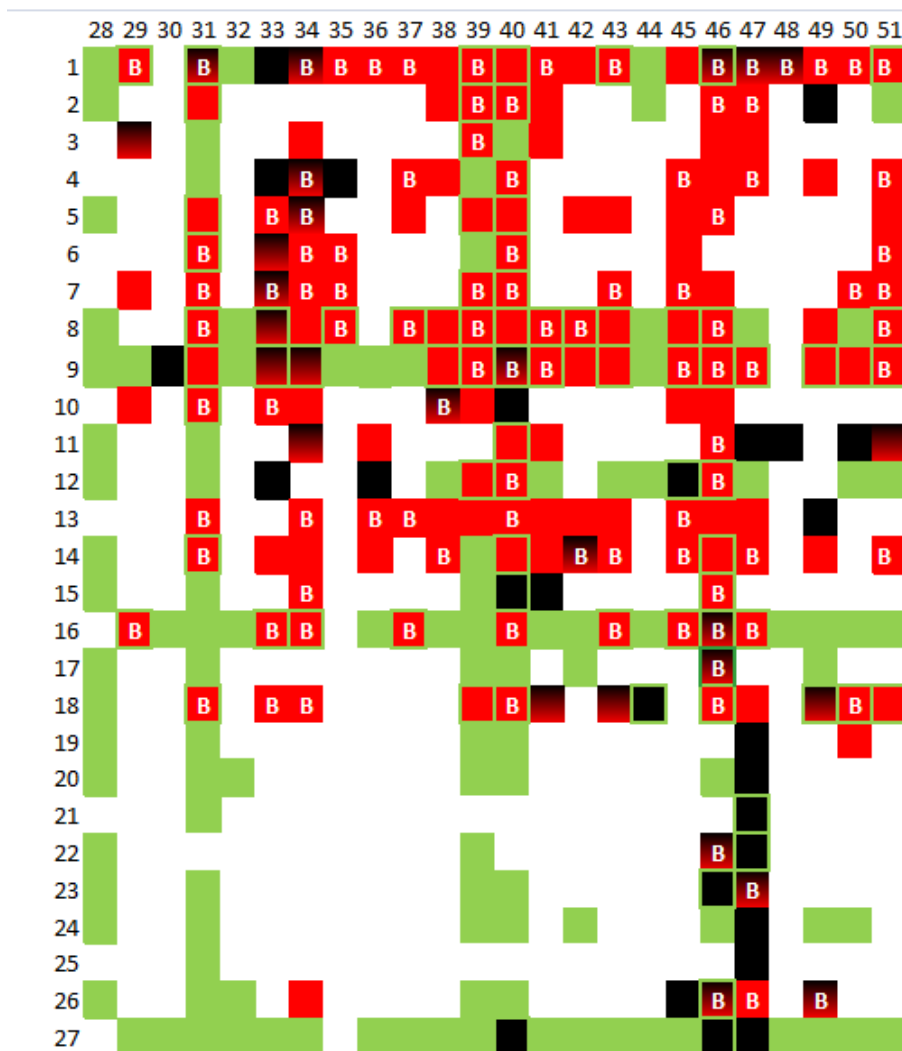


Figura 1 Relação dos cruzamentos realizados no experimento de treino (Quadrante preto) com os híbridos testados em ensaios regionais (vermelho). Os híbridos selecionados na população de treino estão em verde, Os híbridos selecionados no ensaio regional estão representados pela letra B, Os híbridos avaliados no ensaio regional que foram selecionados na população de treino apresentam seus quadrantes delimitados pela cor verde,

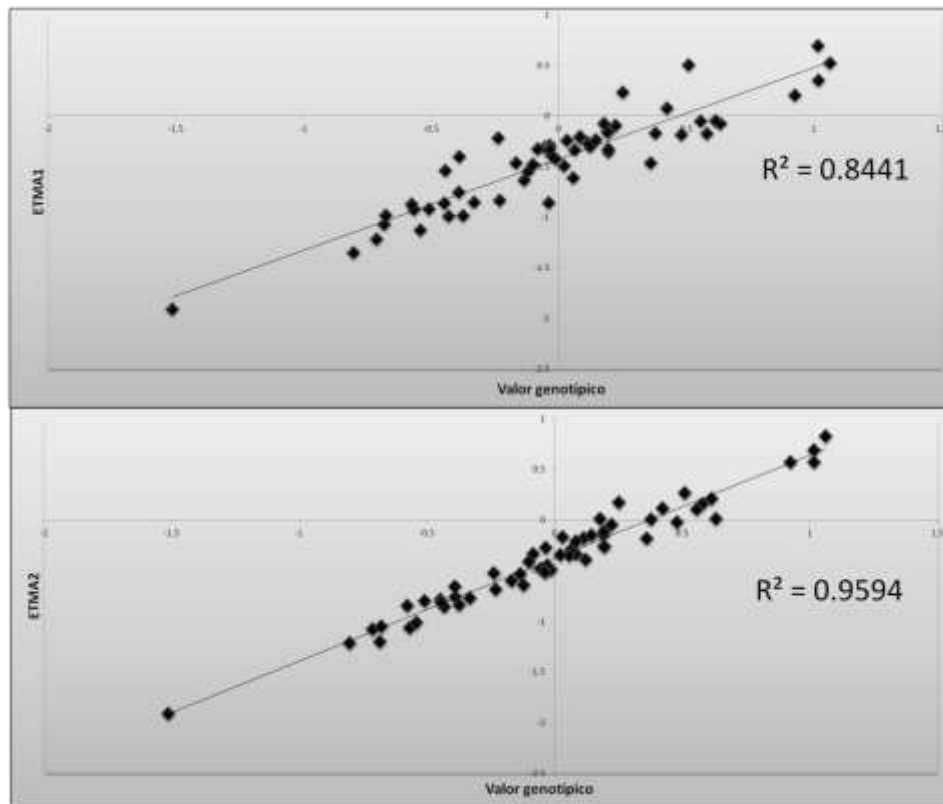


Figura 2 Ajuste do método dos efeitos totais das marcas associadas (ETMA) considerando apenas o efeito aditivo das marcas (ETMA1) e efeitos aditivos + efeitos de dominância (ETMA2),

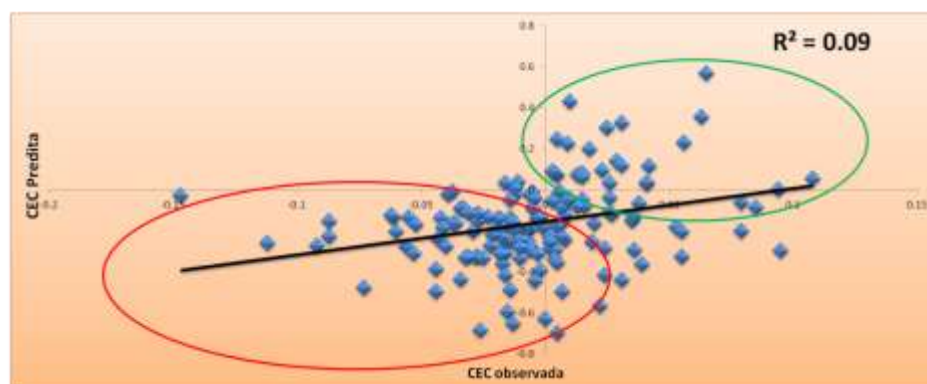


Figura 3 Credibilidade dos valores preditos dos híbridos não testados pelo método ETMA2 na população de treino e valores observados nos ensaios nacionais,

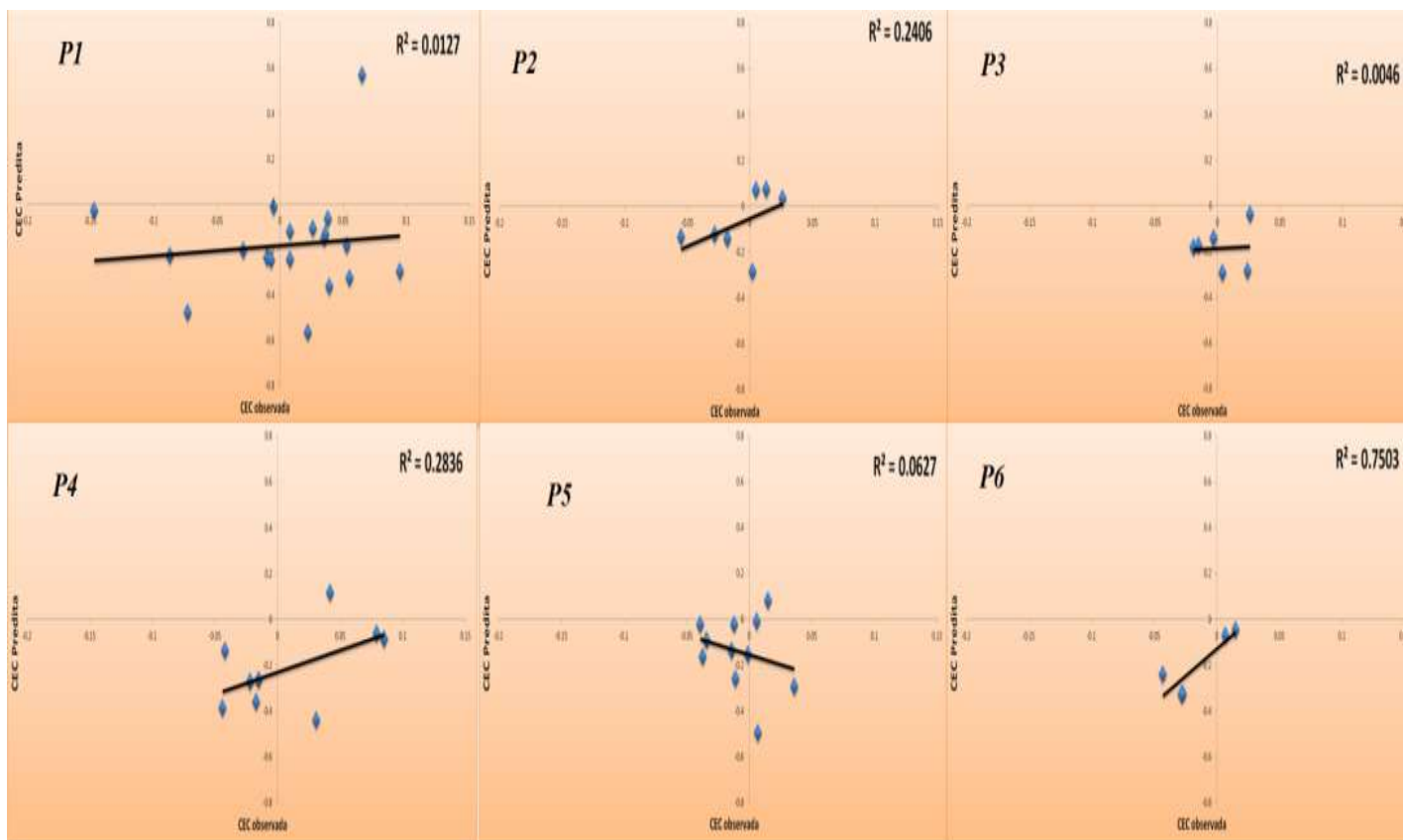


Figura 4 Ajuste do método dos efeitos totais das marcas associadas (ETMA2) por genitor com mais de quatro observações (genitores de 1 a 6),

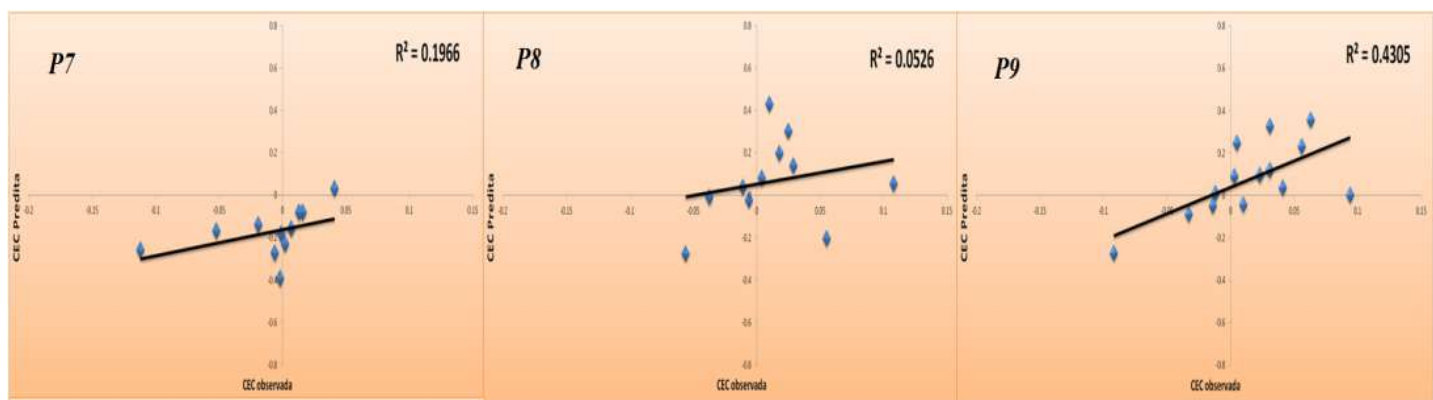


Figura 5 Ajuste do método dos efeitos totais das marcas associadas (ETMA2) por genitor com mais de quatro observações (genitores de 7, 8 e 9),

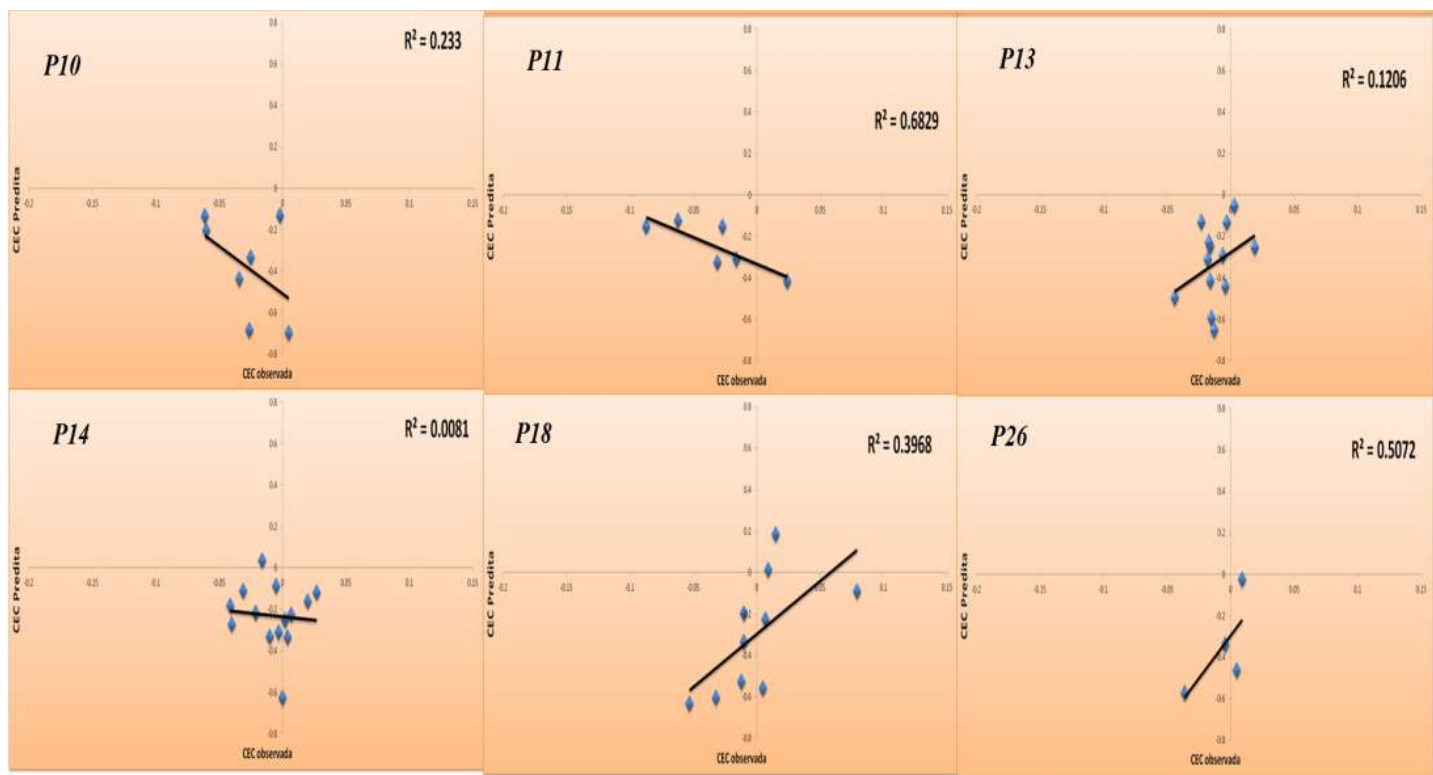


Figura 6 Ajuste do método dos efeitos totais das marcas associadas (ETMA2) por genitor com mais de quatro observações (genitores de 10, 11, 13, 14, 18 e 26),

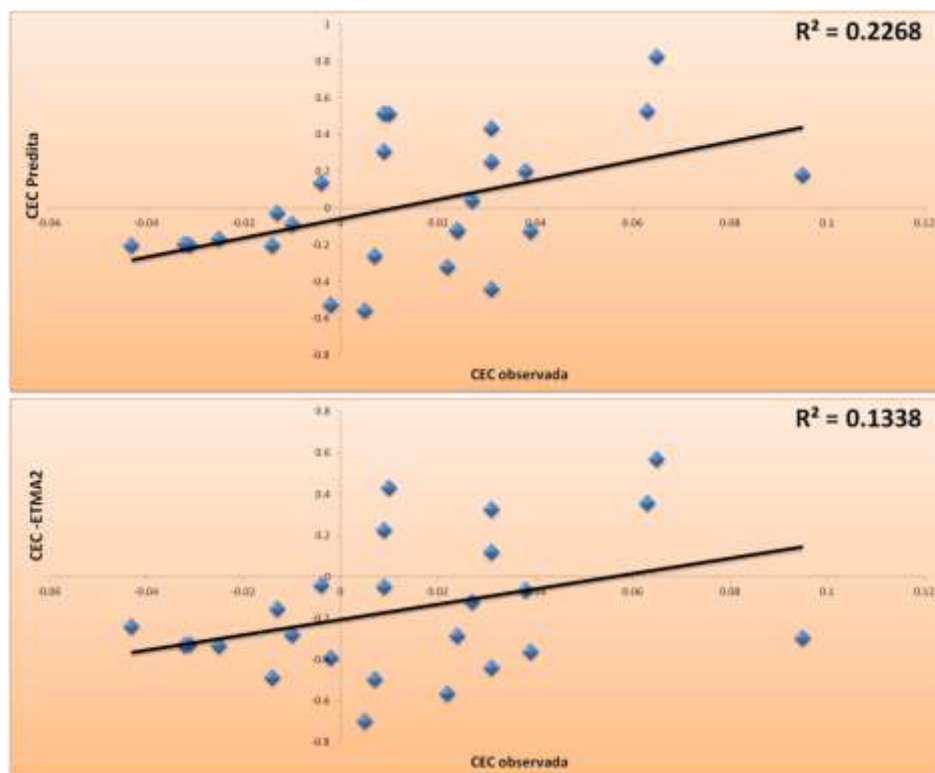


Figura 7 Credibilidade seletiva entre os híbridos testados na população de treino e híbridos avaliados no ensaio regional considerando o método tradicional e totais das marcas associadas (ETMA2),