



HELICIO DUARTE PEREIRA

**PREDIÇÃO GENÔMICA AMPLA DE
HÍBRIDOS SIMPLES DE MILHO
CONSIDERANDO MODELO NÃO ADITIVO**

LAVRAS-MG

2015

HELICIO DUARTE PEREIRA

**PREDIÇÃO GENÔMICA AMPLA DE HÍBRIDOS SIMPLES DE
MILHO CONSIDERANDO MODELO NÃO ADITIVO**

Dissertação 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 Mestre.

Orientador

Dr. Renzo Garcia Von Pinho

LAVRAS-MG

2015

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

Pereira, Helcio Duarte.

Predição genômica ampla de híbridos simples de milho considerando modelo não aditivo / Helcio Duarte Pereira. – Lavras : UFLA, 2015.

78 p. : il.

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

Orientador(a): Renzo Garcia Von Pinho.

Bibliografia.

1. melhoramento genético. 2. epistasia. 3. heterose. 4. modelos bayesianos. I. Universidade Federal de Lavras. II. Título.

HELICIO DUARTE PEREIRA

**PREDIÇÃO GENÔMICA AMPLA DE HÍBRIDOS SIMPLES DE
MILHO CONSIDERANDO MODELO NÃO ADITIVO**

Dissertação 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 Mestre.

APROVADA em 13 de fevereiro de 2015.

Dr. Rodrigo Oliveira de Lima UFV

Dr. Samuel Pereira de Carvalho UFLA

Dr. Renzo Garcia Von Pinho

Orientador

LAVRAS-MG

2015

AGRADECIMENTOS

Primeiramente à Universidade Federal de Lavras (UFLA) por meio do Departamento de Agricultura (DAG), pela oportunidade concedida para a realização do mestrado em Fitotecnia. Também à Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela concessão da bolsa de estudos. Duas figuras que forneceram a base para todo o enredo desenrolar-se.

A todos os funcionários do DAG, cada um cumprindo sua função e sendo essencial para a concretização deste trabalho. Certamente meu orientador Renzo Garcia Von Pinho, secretárias, funcionários de campo do setor de grandes culturas e demais professores, com os quais mantive contato mais próximo, serão sempre lembrados com especial cuidado.

Agradecimento especial aos integrantes do Grupo do Milho da UFLA, pessoas que me acolheram no decorrer da caminhada e com as quais tenho hoje as mais estreitas relações de amizade nessa nova universidade. Além é claro de todo o aprendizado que obtive trabalhando com esse grupo. No meu caso, sinceramente, tão importante quanto o conhecimento que adquiri com as disciplinas da pós-graduação.

Não poderia deixar de mencionar a crucial importância da ajuda que recebi de dois grandes amigos e colegas de pós-graduação da UFLA: Jhonathan Pedroso Rigal dos Santos e Luiz Paulo Miranda Pires. Dois elementos sem os quais este trabalho não teria sequer começado. Minha dívida com vocês é impagável, em qualquer sentido que se possa imaginar!

RESUMO

Avaliou-se por simulação o comportamento de modelos de predição genômica frente à inclusão de efeitos não aditivos de dominância e/ou epistasias na predição de híbridos simples de milho para o caráter peso de espiga. Os objetivos foram avaliar modelos G-BLUP de predição de híbridos e compará-los com os modelos Bayesian Ridge Regression (BRR), Bayes A, Bayes LASSO, Bayes C, Bayes B e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS), implementados no pacote BGLR do software R, com inclusão ou ausência de efeitos não aditivos em três cenários de herdabilidade. Utilizaram-se dados do banco de germoplasma de milho do Departamento de Agricultura dos EUA para encontrar valores paramétricos de componentes de variância, a partir de dados fenotípicos disponíveis, e com isso simulou-se a obtenção de 400 híbridos simples entre dois grupos de linhagens retirados do total de dados. Os modelos foram comparados via soma de quadrados dos erros de predição (PRESS), acurácia preditiva, estimação de componentes de variância, correlação dos valores estimados e simulados e por fim, herdabilidade dos vários tipos de efeitos genéticos presentes em cada situação. Os modelos respondem positivamente à inclusão de efeitos não aditivos de dominância. As estimativas também tornam-se mais precisas quanto maior a herdabilidade considerada. As performances dos modelos variam de acordo com a faixa de herdabilidade considerada. As piores estimativas de efeitos genéticos estão relacionadas aos efeitos epistáticos. Assim, conclui-se que as estimativas obtidas com os modelos G-BLUP, principalmente, e com os demais modelos testados tornam-se mais precisas quando se consideram os efeitos não aditivos para um caráter quantitativo como o peso de espigas de milho.

Palavras-chave: Melhoramento genético. Epistasia. Heterose. Modelos bayesianos.

ABSTRACT

It was evaluated by simulation the behavior of genomic prediction models against the inclusion of non-additive effects of dominance and/or epistasis at the prediction of single cross hybrids of maize for the ear weight character. The objectives were to assess G-BLUP models of hybrid prediction with inclusion of non-additive effects and compare with the Bayesian Ridge Regression (BRR), Bayes A, Bayes LASSO, Bayes C, Bayes B and Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) models, implemented in the BGLR package of R software, with inclusion or absence of non-additive effect in three scenarios of heritability. It was used data from the maize germplasm bank of the United States Department of Agriculture to find parametric values of variance components, from phenotypic data available, and with it simulated the achievement of 400 single cross hybrids between two groups of lines taken from the whole data available. The models were compared by sum of squares of prediction errors (PRESS), accuracy of prediction, estimation of variance components, correlation between simulated and estimated values and finally heritability of several kinds of genetics effects present in every situation. Altogether, all the models responded positively to the inclusion of non-additive effects, especially dominance. For some comparative criteria, the G-BLUP models performed notably higher to the others, but the opposite was not found for any of the other models. Slight differences among the performance of the models occurred according to the heritability range considered. The worst estimates of genetic effects were always related to the epistatic effects. So, it was concluded that the resulting estimates with the G-BLUP models, mainly, and the other models tested became more accurate when taking into account non-additive effects for a quantitative character as ear weight in maize.

Keywords: Plant breeding. Epistasis. Heterosis. Bayesian models

LISTA DE TABELAS

Tabela 1 Grupos heteróticos de linhagens definidos pela análise de componentes principais	44
Tabela 2 Componentes de variância paramétricos usados na simulação	45
Tabela 3 Acurácias preditivas e PRESS (entre parênteses) dos modelos G-BLUP A, Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS), para as três faixas de herdabilidades paramétricas simuladas (h^2) e modelo englobando apenas efeitos aditivos	46
Tabela 4 Acurácias preditivas e PRESS (entre parênteses) dos modelos G-BLUP dominante (G-BLUP AD), G-BLUP epistático (G-BLUP EP), Bayes A, Bayes B, Bayes C, Bayes LASSO e Bayesian Ridge Regression (BRR), para as 3 faixas de herdabilidades paramétricas simuladas (h^2) e modelo englobando efeitos não aditivos de dominância (todos) e ainda epistáticos (apenas G-BLUP EP)	46
Tabela 5 Componentes de variância paramétricos (S) e estimados pelos modelos, nas três faixas de herdabilidade (h^2)	53
Tabela 6 Correlações entre os efeitos genéticos simulados e estimados pelos modelos, nas três faixas de herdabilidades (h^2)	55
Tabela 7 Acurácias preditivas dos modelos G-BLUP aditivo (G-BLUP A), G-BLUP dominante (G-BLUP AD), G-BLUP epistático (G-BLUP EP), Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) considerando somente efeitos aditivos ou incluindo também os não aditivos, na faixa de herdabilidade de 0,3	57

Tabela 8 Acurácias preditivas dos modelos G-BLUP aditivo (G-BLUP A), G-BLUP dominante (G-BLUP AD), G-BLUP epistático (G-BLUP EP), Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) considerando somente efeitos aditivos ou incluindo também os não aditivos, na faixa de herdabilidade de 0,5	58
Tabela 9 Acurácias preditivas dos modelos G-BLUP aditivo (G-BLUP A), G-BLUP dominante (G-BLUP AD), G-BLUP epistático (G-BLUP EP), Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) considerando somente efeitos aditivos ou incluindo também os não aditivos, na faixa de herdabilidade de 0,7	59
Tabela 10 Estimativas da herdabilidade dos efeitos da variância aditiva/variância genética total para os modelos Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR), Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) e G-BLUP A, na condição de modelos incluindo apenas efeitos aditivos e para as três faixas de herdabilidade paramétricas simuladas (h^2)	65
Tabela 11 Estimativas de herdabilidade para os modelos Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR), G-BLUP dominante (G-BLUP AD) e G-BLUP epistático (G-BLUP EP) na condição de modelos incluindo efeitos não aditivos para a herdabilidade paramétrica de 0,3	67
Tabela 12 Estimativas de herdabilidade para os modelos Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR), G-BLUP dominante (G-BLUP AD) e G-BLUP epistático (G-BLUP EP) na condição de modelos incluindo efeitos não aditivos para a herdabilidade paramétrica de 0,5	68

Tabela 13 Estimativas de herdabilidade para os modelos Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR), G-BLUP dominante (G-BLUP AD) e G-BLUP epistático (G-BLUP EP) na condição de modelos incluindo efeitos não aditivos para a herdabilidade paramétrica de 0,7	69
--	----

LISTA DE GRÁFICOS

Gráfico 1 Acurácias preditivas dos modelos G-BLUP incluindo apenas efeitos aditivos (G-BLUP A), efeitos aditivos e dominantes (G-BLUP AD) e por fim efeitos aditivos, dominantes e epistáticos (G-BLUP EP), para as três faixas de herdabilidade simuladas...	49
---	----

LISTA DE FIGURAS

Figura 1 Análise de componentes principais das linhagens endogâmicas do banco de germoplasma do Departamento de Agricultura dos Estados Unidos (USDA). O eixo X representa o primeiro componente principal (pc1) e o eixo Y o segundo componente principal (pc2)	36
--	----

SUMÁRIO

1	INTRODUÇÃO	12
2	REFERENCIAL TEÓRICO	15
2.1	Heterose no melhoramento de milho	15
2.2	Predição do desempenho de híbridos na cultura do milho	18
2.3	Uso de marcadores moleculares no melhoramento genético e predição de híbridos de milho	21
2.4	Seleção genômica ampla (GWS) no melhoramento.....	23
2.5	Modelos mistos/BLUP no melhoramento de plantas.....	25
2.6	Técnicas bayesianas no melhoramento genético.....	28
2.7	Efeitos não aditivos nos modelos de predição	29
3	MATERIAL E MÉTODOS	32
3.1	Descrição do banco de dados utilizado	32
3.2	Obtenção dos valores paramétricos das marcas	33
3.3	Simulação dos híbridos simples.....	35
3.4	Estimativas do peso de espigas com inclusão ou ausência dos efeitos não aditivos.....	39
3.5	Comparação dos modelos	42
4	RESULTADOS E DISCUSSÃO.....	44
4.1	Grupos de linhagens e valores obtidos dos efeitos genéticos paramétricos	44
4.2	Acurácias preditivas e soma de quadrados dos erros de predição para todos os modelos.....	45

4.3	Modelo G-BLUP considerando efeitos aditivos e não aditivos	49
4.4	Modelos G-BLUP e bayesianos	56
4.5	Herdabilidades dos componentes de variância	65
5	CONCLUSÕES.....	70
	REFERÊNCIAS	71

1 INTRODUÇÃO

No melhoramento genético da cultura do milho é comum trabalhar com grande número de cruzamentos, o que gera uma significativa demanda de trabalho e recursos (BERNARDO, 1994). Desse modo, têm sido investigadas algumas técnicas com o fim de auxiliarem na avaliação dos cruzamentos a serem feitos em um programa de melhoramento.

Propostas iniciais foram apresentadas no sentido de direcionar os cruzamentos mais promissores, baseando-se na divergência genética entre os pais. Isso se fundamenta no fato da heterose relacionar-se com a diferença de frequência alélica entre os genitores e a divergência genética entre esses ser um indício de sua ocorrência. Entretanto, tais premissas não garantiram o sucesso dessas metodologias.

Em seguida, um novo recurso apresentava-se disponível aos melhoristas, que era a informação extraída diretamente do DNA dos indivíduos, graças aos avanços na Genética Molecular. Assim, os primeiros usos desse recurso foram no sentido de tentar identificar genes (locos marcados) relacionados com os caracteres de interesse, processo conhecido como seleção assistida por marcadores. Logo, seria possível selecionar os candidatos em função das marcas que esses produzissem pela genotipagem, usando técnicas de marcadores moleculares. Novamente, tal premissa não concordou com os resultados alcançados quando os materiais eram avaliados a campo para características quantitativas. Isso decorre do fato de que tais características são controladas por muitos genes de pequeno efeito e muito influenciadas pelas condições de ambiente. Assim, a seleção de apenas alguns locos para tal classe de caráter é insuficiente para abrigar toda a complexidade genética envolvida.

O próximo passo surgiu como um aprimoramento da seleção assistida por marcadores e diferenciava-se dessa por não se preocupar em identificar locos individuais significativamente relacionados com o caráter sob interesse, desconsiderando se são de grande ou pequeno efeito. Assim,

essa técnica proposta por Meuwissen, Hayes e Goddard (2001), conhecida como seleção genômica ampla, do inglês “Genome Wide Selection - GWS”, faz uso de toda a informação obtida pela genotipagem na inferência do valor genético do indivíduo. No entanto, permaneceu sem ser explorada por alguns anos, no melhoramento vegetal, e mais recentemente ganhou impulso com o concomitante desenvolvimento de novas técnicas moleculares de menores custos e também de modelos estatísticos capazes de lidar com tal volume de informação.

A teoria dos modelos mistos, desenvolvida por Henderson (1975) para lidar com as dificuldades da manipulação de dados no melhoramento animal, passou a ser aplicada também no melhoramento de plantas. Para tal, faz-se necessário o uso de informações de parentesco, atualmente baseada em marcadores moleculares e não mais apenas em informações de *pedigree*, como originalmente proposto. Os estimadores BLUP (*Best Linear Unbiased Prediction*), presentes nos modelos mistos, são os mais utilizados para a predição de valores genéticos nos modelos desenvolvidos até os dias atuais e contam com boas propriedades estatísticas e base teórica já validada. Entretanto, a comunidade científica está trabalhando ainda, predominantemente, com modelos que somente incluem efeitos aditivos, portanto, não levando em conta efeitos de dominância e epistasias. Isso não está em concordância com o que se aceita como possíveis explicações para o fenômeno da heterose. Assim, um fator crucial explorado no melhoramento do milho não está sendo considerado nos modelos de predição de desempenho de híbridos.

O novo desafio passa a ser desenvolver modelos estatísticos com boa capacidade preditiva e fundamentados na teoria genética. Então, surge a necessidade de considerar na modelagem os efeitos heteróticos tão explorados no melhoramento do milho. Um melhor modelo de predição resultaria em maior eficiência dos programas de melhoramento e economia de recursos pela decisão acertada dos indivíduos de fato superiores.

Os objetivos com esse trabalho são: (i) avaliar modelos G-BLUP de predição de híbridos com inclusão de efeitos não aditivos e (ii) comparar modelos G-BLUP com outros reportados na literatura, quais sejam: Bayesian Ridge Regression (BRR), Bayes A, Bayes LASSO, Bayes C, Bayes B e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS), com inclusão ou ausência de efeitos não aditivos.

2 REFERENCIAL TEÓRICO

2.1 Heterose no melhoramento de milho

Uma das maiores contribuições econômicas da genética, se não a maior, foi o desenvolvimento do milho híbrido. Tipicamente, o híbrido entre duas linhas puras ultrapassa o desempenho das populações de cruzamento ao acaso a partir da qual as linhagens foram derivadas (CROW, 2008). Esse fato recebe a denominação de heterose e tem sido o alicerce do melhoramento do milho até os dias atuais. Heterose refere-se ao fenômeno no qual o híbrido F1 exibe características fenotípicas superiores à média dos pais (EAST, 1909; SHULL, 1908).

Com tais benefícios, não é surpreendente que o melhoramento de culturas alimentares e biocombustíveis seja baseado em princípios que controlam a heterose, porém esses princípios ainda não estão bem entendidos (LIPPMAN; ZAMIR, 2007).

O milho é o melhor exemplo de cultura em que mais se explora o efeito da heterose no melhoramento genético. Isso decorre de seu sistema de acasalamento, sendo uma planta alógama e sofrendo severa redução de desempenho com a endogamia. Por conseguinte, é rotineiro em programas de melhoramento dessa cultura a contínua realização de cruzamentos e avaliação da descendência.

O emprego de linhagens em programas de melhoramento de milho está associado justamente à busca do vigor híbrido, ou heterose, fenômeno que proporciona grande produtividade em híbridos provenientes do cruzamento de parentais que exibem alta divergência entre si (PATERNIANI et al., 2008).

Qualquer discussão sobre a melhoria de rendimento do milho deve reconhecer o papel que a heterose desempenha no rendimento de grãos e os processos fisiológicos que subsidiam a produção. O aumento no rendimento do milho comercial durante os últimos 150 anos nos EUA começou com a

larga adoção de híbridos pelos agricultores no final de 1930 e, conseqüentemente, melhoria de rendimento e híbridos, no caso do milho, são altamente associados (TOLLENAAR; LEE, 2006).

A heterose tem sido amplamente estudada em milho por causa de (i) sua grande expressão para a produção de grãos (100-200%), (ii) sua exploração intensiva na criação de híbridos de milho, e (iii) os pré-requisitos biológicos favoráveis da cultura, como grande coeficiente de multiplicação e facilidade de cruzamentos controlados. Embora muitas hipóteses tenham sido sugeridas para explicar a heterose, suas bases genética, fisiológica e bioquímica ainda permanecem em grande parte inexplicáveis (REIF; HALLAUER; MELCHINGER, 2005).

De uma forma simples, o oposto da heterose é a depressão por endogamia, na qual sucessivas autofecundações reduzem os locos em heterozigose e a adaptabilidade geral do organismo. Depressão por endogamia é provavelmente causada pela fixação de alelos deletérios dentro da linhagem. Uma questão fundamental é se a reversão dos efeitos da endogamia no híbrido F1 (via complementação dos alelos deletérios com “bons” alelos do outro parental) é suficiente para explicar a heterose, ou se outros mecanismos atuam (SPRINGER; STUPAR, 2007). Embora melhoristas continuem a melhorar linhagens e definir as melhores combinações parentais, há muitas questões relacionadas às bases genética e molecular da heterose, a principal sendo quais genes estão envolvidos (LIPPMAN; ZAMIR, 2007).

Desde os primórdios, na década de 1910 e 1920, há duas hipóteses para explicar a heterose: a hipótese da sobredominância, em que o aumento do desempenho é principalmente devido aos locos nos quais o heterozigoto desempenha-se melhor do que ambos homozigotos (SHULL, 1908), e a hipótese da dominância, em que no híbrido os alelos recessivos deletérios trazidos por um dos pais são ocultados por alelos dominantes do outro pai (BRUCE, 1910).

Heterose em nível populacional é inferida a partir da correlação entre adaptabilidade da população e nível geral de heterozigosidade, e tem sido debatido se locos específicos mostram efeitos de sobredominância ou se “heterozigosidade *per se*” confere heterose de forma geral. A distinção é importante, porque no segundo caso sugere-se que não existem genes específicos ou alelos que são responsáveis pelo vigor híbrido, ao passo que no primeiro genes individuais com efeitos fenotípicos específicos e diferenciados estão envolvidos (LIPPMAN; ZAMIR, 2007).

O debate acerca das hipóteses da dominância/sobredominância torna-se ainda com mais nuances quando contribuições de epistasias são consideradas. Epistasia é definido como a interação entre dois ou mais locos afetando o fenótipo de um caráter. Com relação à heterose, um loco de ação dominante pode interagir com um loco sobredominante para maiores ganhos heteróticos. Portanto, o *background* genético e as interações alélicas podem ter efeito na contribuição heterótica de um loco individual (SPRINGER; STUPAR, 2007). Três grandes mecanismos para explicar a heterose foram propostos: dominância, sobredominância, e epistasia, mas suas importâncias relativas não estão claramente elucidadas (LARIEPE et al., 2012).

O desenvolvimento de marcadores moleculares para análises genéticas foi um grande passo na direção de analisar o tipo de ação gênica sustentando a heterose. Dois procedimentos principais têm sido usados: (i) a investigação da relação entre heterose e divergência genética entre as linhagens parentais e (ii) mapeamento de QTL (*Quantitative Trait Loci*) para identificação de segmentos cromossômicos envolvidos com a heterose (GARCIA et al., 2008; LARIEPE et al., 2012).

Mesmo entre híbridos de populações F2 de milho, em que não se explora o máximo de heterose entre as possibilidades de cruzamentos, Doná et al. (2011) encontraram significância dos efeitos de dominância e epistasia na manifestação da produtividade de grãos, altura de espiga e florescimento masculino.

2.2 Predição do desempenho de híbridos na cultura do milho

Híbridos comerciais são obtidos do cruzamento de linhagens de grupos heteróticos distintos. Infelizmente, os melhoristas são incapazes de avaliar todas as combinações de híbridos simples entre tais linhagens por ser uma quantidade proibitiva. Para ilustrar, existem 100 híbridos possíveis entre 10 linhagens do grupo heterótico X e 10 do grupo Y. Porém se o melhorista tem 50 linhagens do grupo X e 50 do grupo Y o número de possíveis híbridos aumenta para 2500 (BERNARDO, 1994, 1996).

Genótipos que manifestam uma mesma capacidade de combinação ou resposta heterótica quando cruzados com genótipos distintos de outro germoplasma pertencem a um mesmo grupo heterótico. Padrão heterótico refere-se à resposta específica do cruzamento entre dois grupos heteróticos, resultando em híbridos que expressam alta heterose e performance (MELCHINGER; GUMBER, 1998).

Em vista da grande quantidade de cruzamentos necessários para a produção de híbridos nos programas de melhoramento é de se esperar que o interesse em reduzi-los, baseando-se em fundamentos genéticos, esteve e ainda mantém-se sob intensa investigação. De maneira geral, os métodos fazem uso de informações de parentesco/similaridade entre indivíduos de uma população, de forma que apenas alguns precisam ser testados. Esses fornecerão estimativas de valores genéticos a serem usadas na predição da performance dos demais, que guardam certa relação de parentesco com os anteriores.

Uma proposta pioneira de predição de desempenho e, por conseguinte, economia de recursos, surgiu com Jenkins (1934) para a predição de híbridos duplos, que eram os mais cultivados nos EUA na época. Jenkins sugeriu 4 modelos de predição chamados de A, B, C e D. O método B provou ser o mais adequado, baseando-se na performance dos híbridos simples não parentais envolvendo as 4 linhagens em questão, e foi

largamente utilizado na predição não só de híbridos duplos como também triplos (MELCHINGER et al., 1987).

Embora os métodos de predição propostos por Jenkins tenham sido largamente usados eles apresentam uma importante limitação: a necessidade de avaliar todas as possibilidades de híbridos simples. O que ainda continua demandando bastante trabalho e despesas. Uma alternativa poderia ser prever os híbridos duplos baseando-se em informação de *pedigree*. Sendo possível determinar a covariância ou similaridade genética entre híbridos duplos, a informação das linhagens genitoras ou dos híbridos simples parentais não é mais necessária (BALESTRE; VON PINHO; SOUZA, 2011).

Sabendo-se do fenômeno da heterose entre linhagens contrastantes, os melhoristas de milho propuseram como cruzamentos mais promissores aqueles entre linhagens/populações mais distantes do que entre as mais relacionadas. Uma maior divergência entre duas populações aumenta a probabilidade de seleção complementar para alelos favoráveis em diferentes locos (REIF; HALLAUER; MELCHINGER, 2005).

A avaliação da divergência genética é muito utilizada pelos melhoristas de milho para a seleção de parentais. Esse método visa selecionar aqueles materiais mais promissores para cruzamento, diminuindo os gastos e o tempo necessário para a realização de várias combinações híbridas, muitas vezes desnecessárias. Os esforços são, portanto, concentrados naquelas combinações provenientes de materiais mais divergentes (RINALDI et al., 2007).

Entretanto, de maneira geral, altas estimativas de divergência genética entre as linhagens não se correlacionam com a produtividade do híbrido. Logo, não é possível prever sobre a CEC (capacidade específica de combinação) e produtividade de grãos nos cruzamentos baseando-se apenas na divergência genética entre as linhagens envolvidas (PATERNIANI et al., 2008; SCHRAG et al., 2009).

A correlação entre divergência genética e estimativas de parâmetros quantitativos não é consistente na maioria das pesquisas (CAIXETA et al., 2013). O trabalho de Guimarães et al. (2007) evidencia que estimativas de correlações entre variáveis moleculares (distância genética) com estimativas de parâmetros quantitativos são complexas, de pouca consistência e difícil interpretação. Nesse trabalho, com a cultura do milho, os autores encontraram correlação entre divergência genética e heterose, porém o mesmo não foi observado para divergência e capacidade específica de combinação, o que é de certo modo contraditório. A baixa correlação entre divergência genética e heterose deve-se à baixa acurácia em predizer os desvios de dominância por marcadores moleculares (CAIXETA et al., 2013). No entanto, levando-se em conta a complexidade de fenótipos heteróticos no contexto de compreender-se as análises quantitativas, seria possível obter vantagens dos recursos genômicos para descobrir os fatores mendelianos que definem a heterose. Somente então será possível visualizar, a longo prazo, a questão do papel da heterose na adaptabilidade de populações e mais efetivamente explorar os princípios fundamentais da heterose no melhoramento vegetal (LIPPMAN; ZAMIR, 2007).

No cenário recente do uso cada vez maior da tecnologia de duplo haploide (DH) em milho a predição de desempenho das melhores combinações híbridas com uso de marcadores torna-se especialmente atrativa. Isso decorre do fato de que no desenvolvimento de linhagens pela tecnologia DH essas não são submetidas à avaliação *per se* ou em cruzamentos (*testcross*), diferentemente do processo tradicional de autofecundações por gerações sucessivas, que dá condições de seleção. Portanto, o grande número de linhagens obtidas com a tecnologia DH, representando uma amostra de linhagens do cruzamento parental, embora com a vantagem de acelerar o processo, impõe uma alta demanda de recursos para os testes de campo (SCHRAG et al., 2009).

Como em um programa de melhoramento milhares de linhagens precisam ser avaliadas para sua performance em *testcross*, modelos de

predição de potencial genético baseados em dados genotípicos de linhagens não avaliadas são altamente desejáveis (ALBRECHT et al., 2011).

2.3 Uso de marcadores moleculares no melhoramento genético e predição de híbridos de milho

Um novo recurso para uso no melhoramento foi a informação diretamente no material genético dos indivíduos sob avaliação/seleção. Isso foi possível graças à evolução da Genética Molecular e suas áreas correlatas como a Biotecnologia. Com isso, o emprego de marcadores moleculares no melhoramento foi um processo imediato.

Predição de valores genéticos baseados em *pedigree* tem sido usada com sucesso no melhoramento de plantas e animais por muito tempo. No entanto, há um limite na taxa anual de progresso genético que pode ser alcançada com predições baseadas em *pedigree*. Marcadores moleculares permitem descrever o genoma de indivíduos em um grande número de locos, revelando a arquitetura genética do caráter, e isso abre possibilidades para acuradas predições de valores genéticos em estádios precoces do desenvolvimento (DE LOS CAMPOS et al., 2013; GUO et al., 2013; HEFFNER; SORRELLS; JANNINK, 2009). O desenvolvimento das tecnologias de identificação de genes usando as ferramentas da Biotecnologia fornecem amplas oportunidades para os pesquisadores melhorarem ainda mais as modernas cultivares (JENA; MACKILL, 2008).

Ao longo dos últimos 30 anos o uso da genotipagem permitiu a caracterização e mapeamento de genes e vias metabólicas em plantas, assim como o estudo da diversidade de espécies, evolução, caracterização de germoplasma e pureza genética de sementes. *Single Nucleotide Polymorphism* (SNP) surgiu como o marcador mais usado devido à sua abundância no genoma e a relativa facilidade na determinação de sua frequência com custo compatível (DESCHAMPS; LLACA; MAY, 2012).

Programas de melhoramento estritamente convencionais dependem largamente da seleção fenotípica, da experiência do melhorista e conhecimento sobre o controle genético de caracteres de importância econômica. Grande quantidade de dados moleculares está disponível a partir de estudos para importantes caracteres de interesse agrônomo, o que por sua vez, se melhor explorada no melhoramento, pode auxiliar a seleção genotípica (XU; CROUCH, 2008). Obviamente a validação da seleção inicial baseada apenas em dados genotípicos precisa ser feita pela avaliação final de campo nas condições-alvo que o novo material genético encontrará (HEFFNER; SORRELLS; JANNINK, 2009).

Com o uso de marcadores de DNA no melhoramento a eficiência e precisão podem ser melhoradas. Essa nova modalidade de melhoramento é conhecida como seleção assistida por marcadores (SAM) (COLLARD; MACKILL, 2008). Isso não significa a substituição da tradicional avaliação fenotípica pela genotípica apenas, mas sim que essa última pode contribuir para economia de recursos e trabalho desempenhando um papel de “*screening*” inicial. Conforme ressaltado por Guo et al. (2013), cruzamentos e delineamentos experimentais de campo são cruciais para análises genéticas e predição de híbridos.

Uma tecnologia eficiente e de custo competitivo para a SAM deve ser desenvolvida para permitir aos melhoristas avaliarem os genótipos por meio de genomas completos e recombinar os genes de importância agrônoma a partir de diferentes fontes (JENA; MACKILL, 2008; XU; CROUCH, 2008).

Com o advento de marcadores moleculares no final dos anos 1980, o mapeamento de locos de caracteres quantitativos (QTL), com a ajuda de marcadores ligados, forneceu uma nova e poderosa ferramenta para estudar a arquitetura genética de caracteres quantitativos complexos em genética e melhoramento de plantas. Estudos de QTL's têm sido publicados para uma série de caracteres agrônomicos, fisiológicos e outros. Alguns deles foram

especificamente concebidos para estudar a base genética da heterose e a depressão endogâmica (REIF; HALLAUER; MELCHINGER, 2005).

O desafio, no entanto, é a utilização de tal informação no melhoramento genético. Tais caracteres são controlados por numerosos genes e exibem baixa herdabilidade e forte interação com o ambiente. Assim, seu uso limitado para esses caracteres é atribuído a várias razões, incluindo QTL's pouco informativos ou específicos para a população e ambiente estudados, QTL's não fortemente ligados ao gene, polimorfismo e problemas associados com arraste de ligação (ASHRAF; FOOLAD, 2013).

Xu e Crouch (2008) relatam que informação gênica não tem sido efetivamente usada na maioria dos programas de melhoramento pelo fato de que para muitos caracteres complexos a relação entre informação gênica e variação fenotípica ainda não é bem definida e também pela falta de ferramentas apropriadas para uso de tal informação, ainda que a relação seja conhecida. A evolução de tais ferramentas passa por sua incorporação em softwares de simulação e decisão, permitindo aos melhoristas tomarem decisões rápidas e precisas.

2.4 Seleção genômica ampla (GWS) no melhoramento

Uma modalidade de SAM que surgiu recentemente é a seleção genômica ampla, do inglês "*Genome Wide Selection*" - GWS. Ela veio para melhor explorar a informação genotípica ao lidar com caracteres quantitativos, primeiramente no melhoramento animal e posteriormente em culturas perenes. Seu uso em culturas anuais é mais recente. Tal técnica trabalha com maior quantidade de informações e por isso necessita de genotipagem mais intensa. Isso esbarra no problema do custo de marcadores moleculares, pois no melhoramento vegetal a quantidade de indivíduos a serem avaliados é muito grande. No futuro, os custos de genotipagem diminuirão, mas é improvável que os custos da fenotipagem também reduzam, portanto direcionando os objetivos no sentido de redução da

fenotipagem e aumento da genotipagem (HEFFNER; SORRELLS; JANNINK, 2009). Assim, novas tecnologias de marcadores podem reduzir significativamente o custo da SAM, incluindo aí a GWS. Se a efetividade dos novos métodos for validada e os equipamentos puderem ser obtidos com facilidade, isso deve proporcionar a SAM tornar-se mais largamente aplicada em programas de melhoramento genético de plantas (COLLARD; MACKILL, 2008).

No geral, estudos empíricos têm confirmado a superioridade da seleção genômica sobre a predição baseada em parentesco, conforme foi antecipado por estudos de simulação (DE LOS CAMPOS et al., 2013). As acurácias observadas com a seleção genômica oferecem a possibilidade de que futuras linhagens elite e parentais sejam selecionadas pelo seu valor genômico em vez de seus dados fenotípicos (HEFFNER; SORRELLS; JANNINK, 2009).

Os marcadores moleculares podem ser testados para sua associação com o caráter de interesse no melhoramento e suas estimativas de efeitos alélicos podem ser empregadas na predição de desempenho dos futuros híbridos. A soma dos efeitos alélicos fornece uma estimativa do valor genético do indivíduo (SCHRAG et al., 2009).

No entanto, na GWS, utiliza-se um conjunto numeroso de marcadores, sem necessidade de estabelecer sua associação com os caracteres a serem melhorados e, então, faz-se a predição de fenótipos (BERNARDO, 2014). Predições baseadas em *pedigree* usam informações de parentesco para predizer o valor genético. No entanto, tal procedimento não explora a similaridade genética entre indivíduos não aparentados. Portanto, uma vantagem da seleção genômica reside em sua habilidade para utilizar informações de indivíduos mais distantemente relacionados (DAETWYLER et al., 2013). Isso é possível, pois a matriz de variâncias e covariâncias genéticas é construída baseando-se nas marcas em comum que os indivíduos compartilham e não em suas relações por ascendência.

Apenas dados genotípicos necessitam ser usados na população sob melhoramento para a seleção dos indivíduos com os maiores valores genéticos. Na GWS uma população inicial de treinamento é usada para obtenção de dados genotípicos e fenotípicos a partir de uma numerosa quantidade de marcadores para capturar toda a variância aditiva do caráter em consideração. Valores genéticos são então preditos na população sob melhoramento somente baseando-se no genótipo e efeitos estimados das marcas (DESCHAMPS; LLACA; MAY, 2012; GODDARD; HAYES, 2007).

Considere o exemplo apresentado por Bernardo e Yu (2007): se os candidatos são genotipados para 256 SNP's, distribuídos ao longo do genoma, na GWS os efeitos nos locos quantitativos (QTL) de todas as 256 marcas são ajustados como efeitos aleatórios em um modelo linear. O valor do caráter é então predito como a soma do valor genético do indivíduo por meio de todas as 256 marcas e a seleção é baseada nessa predição.

O conceito de seleção genômica é baseado na hipótese de que com suficiente densidade de marcadores todos os polimorfismos que contribuem para a variação no caráter estarão em desequilíbrio de ligação com marcadores, segregando na população sob estudo (GODDARD; HAYES, 2007). Modelos de predição genômica são desenvolvidos baseados em numerosas populações de treinamento para a qual dados genotípicos (marcas) e fenotípicos estão disponíveis. O valor genético dos candidatos à seleção pode então ser predito baseando-se em informação genômica (ALBRECHT et al., 2011).

Pelo fato da seleção genômica empregar todos os dados de marcadores para o caráter, espera-se maior acurácia na estimativa/predição do valor genético dos indivíduos e, portanto, maior eficiência no melhoramento de plantas (ASHRAF; FOOLAD, 2013).

2.5 Modelos mistos/BLUP no melhoramento de plantas

A base fundamental do melhoramento de plantas é a seleção de indivíduos específicos com caracteres desejáveis. A seleção tipicamente envolve avaliação de uma população sob melhoramento para um ou mais caracteres em ensaios de campo (resistência a doenças, tolerância a estresses, e outros) ou testes químicos (qualidade de grãos, por exemplo). O objetivo do melhoramento é reunir mais combinações de alelos favoráveis em novos materiais (COLLARD; MACKILL, 2008).

No melhoramento animal essa seleção é uma tarefa difícil e forçou o desenvolvimento de novas metodologias na estimativa de valores genéticos. A mais usada e reconhecida são os modelos mistos de Henderson (1975). Nesse procedimento, existem efeitos fixos e aleatórios no modelo e que são estimados simultaneamente, encontrando-se o BLUP (*Best Linear Unbiased Prediction*) dos candidatos à seleção. Essa é uma estimativa do efeito aleatório, presente no valor genotípico dos indivíduos, já que esses são formados a partir de uma amostra aleatória de alelos de uma população de referência.

Para a avaliação de cruzamentos, quando esses representam uma amostra aleatória proveniente de alguma população, portanto um efeito aleatório no modelo, o método dos quadrados mínimos não resulta nas melhores estimativas (HENDERSON, 1977).

Uma grande mudança nos procedimentos de avaliação genética ocorreu nos anos de 1970. Índices de seleção foram substituídos pela metodologia de modelos mistos de Henderson. Uma vantagem desses novos procedimentos é que para um dado modelo os resultados têm as propriedades de melhores preditores lineares não viesados (BLUP) (POWELL; NORMAN, 2006).

Para uso dos modelos mistos na predição de híbridos não testados é necessário fazer uso da informação de parentesco entre os indivíduos avaliados e os que ainda serão. Isso pode ser feito por meio de informação de *pedigree*, porém nem sempre essa está disponível. Assim, no contexto do BLUP, o uso da informação de marcadores moleculares é uma alternativa

quando os *pedigrees* dos pais não estão disponíveis, contribuindo para a obtenção de estimativas mais acuradas do valor genotípico dos híbridos não testados (FERREIRA et al., 2010).

Os híbridos simples comumente são avaliados em diferentes grupos de ensaios, locais ou anos. O método BLUP permite predições com tal análise desbalanceada. Para tal seleciona-se um grupo de híbridos como preditores e estima-se sua variância genética. Essa população de preditores pode então ser usada para prever a performance dos futuros cruzamentos derivados da mesma combinação de linhagens (BERNARDO, 1994). No BLUP a performance dos híbridos não testados é feita a partir de (i) a performance dos híbridos avaliados e (ii) o parentesco genético entre as linhagens parentais (BERNARDO, 1996).

Uma grande quantidade de dados será gerada com o emprego de programas de SAM em larga escala e esses dados precisam ser analisados e integrados com outros tipos de informações para se fazer decisões de seleção em curto período de tempo. Portanto, ferramentas de suporte à decisão são essenciais para acelerar esse processo, mantendo a acurácia e precisão (XU; CROUCH, 2008).

O trabalho de Fritsche Neto et al. (2010) evidencia a robustez do procedimento de predição REML/BLUP (*Restricted Maximum Likelihood / Best Linear Unbiased Prediction*). No caso, os pesquisadores simularam vários graus de desbalanceamento de híbridos a terem seus valores genotípicos preditos, em um conjunto de 256 híbridos simples, e constataram que a acurácia entre os valores observados e as predições manteve-se acima de 90% com até 50% de desbalanceamento.

Desde sua proposta no melhoramento animal, o BLUP tem sido estimado com severos desbalanceamentos (ARNHOLD et al., 2012). Ferreira et al. (2010) também encontraram boa eficiência ao avaliarem o BLUP na predição de híbridos simples a partir de linhagens com pouca informação de similaridade/parentesco.

2.6 Técnicas bayesianas no melhoramento genético

Na seleção genômica é preciso obter a estimativa do efeito dos marcadores, a partir da população de treinamento, para a posterior avaliação do valor genético dos indivíduos candidatos à seleção. O procedimento de como encontrar essas estimativas de efeitos pode ter um impacto na eficiência da seleção como um todo.

Com as modernas tecnologias de genotipagem o número de marcas, e, portanto, o número de parâmetros a serem estimados, podem facilmente exceder o número de observações. Para contornar os problemas que surgem com esse cenário procedimentos de estimação realizando seleção variável, “*shrinkage*” das estimativas, ou uma combinação de ambas são comumente usadas (DE LOS CAMPOS et al., 2013).

Os primeiros estudos genômicos concluíram que métodos bayesianos com capacidade de modelar locos com variâncias específicas eram superiores àqueles que atribuíam equivalência de variâncias para todos os locos (p.ex. G-BLUP). Essa conclusão foi confirmada verdadeira apenas quando poucos QTL's tinham uma grande contribuição para a variância genética, indicando a importância de se examinar a arquitetura genética com muitos QTL's (DAETWYLER et al., 2013). Assim, os métodos bayesianos desempenham-se melhor quando o caráter é controlado por pequeno número de genes. Porém, o G-BLUP é mais robusto, ou seja, tem boa performance em várias situações de arquitetura genética.

Devido à alta demanda computacional requerida pelos métodos bayesianos seu emprego no melhoramento de plantas tem sido restrito e procedimentos frequentistas têm predominado na solução dos problemas regulares (BALESTRE et al., 2013).

Teoricamente é esperado que um método que é flexível o bastante para modelar qualquer tipo de arquitetura genética (ex.: Bayes-C, Bayes-B, quando todos os parâmetros são estimados a partir dos dados) desempenhará bem sobre uma grande variedade de caracteres com diferentes

características, entretanto, evidências empíricas nem sempre confirmam isso (DE LOS CAMPOS et al., 2013).

A estimativa BLUP do efeito de um marcador pode ser interpretada como uma estimativa de mínimos quadrados que foi regredida em direção a zero. Esse é também o caso para estimativas de efeito de marcas sob os métodos Bayes-A e Bayes-B, mas esses fazem um *shrinkage* de forma não linear, de modo que uma estimativa de mínimos quadrados que é pequena, relativa ao erro padrão, é aproximada para zero e estimativas maiores sofrem *shrinkage* menos severo (GODDARD; HAYES; MEUWISSEN, 2010).

Na estatística bayesiana, parâmetros como a variância explicada por um loco são assumidos, a priori, provir de uma determinada distribuição de probabilidade. Consequentemente a variância pode variar entre os locos, e combinando a informação a priori com a dos dados obtidos resulta em uma estimativa do parâmetro. Esse procedimento parece mais realista do que assumir que a variância entre os locos é fixa, como no caso do método BLUP (MEUWISSEN; HAYES; GODDARD, 2001). Bayes-B apresenta uma *priori* mais realista porque espera-se que algumas regiões do genoma não carreguem nenhum QTL, assim algumas marcas deveriam ter estimativas de efeito iguais a zero (HEFFNER; SORRELLS; JANNINK, 2009).

Não se conhece a verdadeira distribuição dos efeitos dos marcadores, mas para alguns caracteres existem poucos QTL's com efeitos que são maiores do que seria amostrado de uma distribuição normal. Também parece provável que quando o número de marcadores usados aumenta a pressuposição de que muitos têm efeito igual a zero é esperada. Portanto, Bayes-B parece ser amplamente útil – raramente ele desempenha-se pior do que o BLUP e em alguns casos é significativamente melhor (GODDARD; HAYES; MEUWISSEN, 2010).

2.7 Efeitos não aditivos nos modelos de predição

Os efeitos genéticos podem ser decompostos em três classes de efeitos: (i) aditivos; (ii) dominantes; e (iii) epistáticos. Efeitos aditivos são considerados como efeitos lineares de cada alelo que são integralmente herdáveis ao longo de gerações. Efeitos de dominância são efeitos quadráticos, oriundos da interação de alelos de um loco, por exemplo, o valor resultante de dois alelos juntos, especificamente, na constituição de um dado genótipo. Por fim, efeitos epistáticos são desvios adicionais devido à combinação de dois ou mais genes, ou seja, a falta de aditividade dos valores genotípicos de dois genes diferentes (FALCONER; MACKAY, 1996; ZENG; WANG; ZOU, 2005). Justamente nesse ponto fundamenta-se a heterose, tão reconhecida e trabalhada no melhoramento do milho. Logo, uma boa predição deve considerar seu efeito, o que não tem sido verificado na prática.

Grandes melhorias em certos caracteres têm sido efetuadas pela seleção na variância genética aditiva e utilizando modelos estritamente aditivos. Entretanto, há alguns caracteres que devem ter importantes componentes genéticos não aditivos. Se os melhoristas soubessem a relativa magnitude destes componentes, uma seleção mais acurada poderia ser praticada (HENDERSON, 1985b).

Índices de seleção e BLUP têm se restringido largamente para modelos genéticos aditivos. Isso tem sido uma consequência da falta de conhecimento de componentes de variância genéticos não aditivos e também devido a falta de entendimento de como incorporar estas variâncias, ainda que conhecidas, nos procedimentos de avaliação (HENDERSON, 1985a). Viana et al. (2009) estimaram os efeitos genéticos aditivos e heteróticos trabalhando com 10 híbridos simples comerciais e suas gerações F2. Destaca-se a contribuição média encontrada devido à heterose para o caráter produtividade de grãos, que foi de 69,36%. Mostrando, como esperado, a importância dos efeitos não aditivos no desempenho do milho.

Programas de melhoramento de híbridos seriam beneficiados a partir de uma maneira confiável de predição de fenótipos por meio de um melhor

entendimento da base genética da heterose (LARIEPE et al., 2012). O nível de heterose para caracteres específicos varia e não é correlacionado em diferentes híbridos da mesma espécie. Isso indica que heterose não é o resultado da ação de um único loco, nem que a heterose simplesmente reflete a extensão da heterozigosidade entre os pais (SPRINGER; STUPAR, 2007).

Para abranger todas as questões considerando as bases genéticas de caracteres complexos e otimizar os processos de melhoramento, simulação computacional e modelagem estão se tornando ferramentas essenciais (XU; CROUCH, 2008).

Entre os modelos aditivos, as regressões bayesianas e o G-BLUP (Genomic-BLUP) são os mais comumente usados. G-BLUP é atrativo porque sua implementação é simples usando softwares já existentes, é um método pragmático e robusto para predição de caracteres poligênicos (DE LOS CAMPOS et al., 2013; TECHNOW et al., 2014).

Em estudo comparando modelos com efeitos apenas aditivos e incluindo efeito não aditivo de dominância na predição do rendimento de híbridos, portanto, um caráter para o qual a variância devida à dominância representa parte substancial da variância genética total (BERNARDO, 1996) encontrou superioridade preditiva para o modelo com inclusão de efeitos não aditivos. O autor ainda menciona que na prática o único custo adicional de tal consideração é o maior tempo de computação dos dados. No entanto, atualmente, esse fator não é mais considerado restritivo devido ao desenvolvimento de processadores e algoritmos mais rápidos.

3 MATERIAL E MÉTODOS

3.1 Descrição do banco de dados utilizado

No presente trabalho, os dados de genotipagem e fenotipagem dos materiais genéticos foram obtidos a partir do banco de dados do Departamento de Agricultura dos Estados Unidos (USDA-ARS). Os dados genotípicos estão disponibilizados publicamente a partir da publicação do trabalho de Romay et al. (2013), em que foram caracterizados 2.815 acessos de linhagens de milho com 681.257 SNPs/GBS distribuídos ao longo dos 10 cromossomos da espécie.

Os dados fenotípicos estão disponíveis publicamente no banco de dados Panzea (2014). Esses dados são relacionados às linhagens endogâmicas derivadas da população NAM (*Nested Association Mapping population*) (MCMULLEN et al., 2009), que também foram genotipadas no trabalho realizado por Romay et al. (2013).

No banco de dados Panzea (2014), há disponibilidade de dados fenotípicos para 57 caracteres de 5.000 linhagens endogâmicas recombinantes, derivadas dos cruzamentos entre a linhagem B73 com outras 25 linhagens endogâmicas de alta diversidade genética – desses cruzamentos derivou-se à população NAM. Assim, esses cruzamentos foram planejados visando representar o *pool* gênico mundial da cultura do milho. Os dados fenotípicos das linhagens foram obtidos por avaliações em ensaios realizados no período de 2006 a 2009, pelos grupos de pesquisa envolvidos no projeto.

Para utilização das informações de tal banco de dados, inicialmente foi necessária uma etapa de pré-processamento das informações. Para isso, foram removidas todas as marcas redundantes, e separadas 27.000 marcas informativas do total das 681.257 marcas disponibilizadas no banco. Esse número de marcas SNP's/GBS é uma quantidade comum em genotipagens executadas em programas de melhoramento. A seleção foi baseada em uma amostragem sistemática, com o sorteio da primeira marca entre a 1^o marca e

a 25^o marca, e as demais selecionadas linearmente a cada 25 posições de marcas até a cobertura completa dos 10 grupos de ligação do milho. Logo, verifica-se que o procedimento cobriu densamente e uniformemente todos os 10 grupos de ligação do milho.

Em relação ao conjunto de dados fenotípicos foi escolhido o caráter peso de espiga, que mais se aproxima ao caráter produtividade de grãos, dentre os disponíveis no banco de dados. Esses dados fenotípicos foram utilizados para estimar os valores genéticos paramétricos dos 27.000 marcadores processados do conjunto total de marcas descrito anteriormente. Os dados fenotípicos separados foram obtidos em ensaios conduzidos no ano de 2006, na safra de verão, na cidade de Aurora estado de Nova York, na latitude 42,75° N e longitude 76,70° W.

O método estatístico utilizado será discutido na seção subsequente. O processo de manipulação de dados do projeto Panzea foi realizado em uma primeira etapa no software TASSEL (BRADBURY et al., 2007), devido ao volume de informações (número de marcas genotipadas – 681.257), para permitir a subdivisão desse volume em subconjuntos menores. Posteriormente fez-se a amostragem sistemática descrita anteriormente.

3.2 Obtenção dos valores paramétricos das marcas

O método Bayes B foi utilizado para estimação dos efeitos genéticos dos 27.000 marcadores moleculares descritos anteriormente. Esse método apresenta em geral um critério mais sensível para detecção de QTL's (MEUWISSEN; HAYES; GODDARD, 2001). Na etapa seguinte os híbridos foram simulados tomando essa estimação como valores paramétricos. O modelo genético estatístico utilizado nesse procedimento foi:

$$\mathbf{Y} = \mathbf{X}\mathbf{u} + \mathbf{W}_A\mathbf{a} + \mathbf{W}_D\mathbf{d} + \boldsymbol{\varepsilon} \quad (1)$$

em que \mathbf{Y} : vetor de observações $n \times 1$; \mathbf{X} : matriz de incidência dos efeitos fixos $n \times p$ conhecida; \mathbf{u} : vetor de efeitos fixos não conhecidos; \mathbf{W}_A e \mathbf{W}_D : matrizes de incidência dos efeitos aleatórios $n \times q$ conhecidas; \mathbf{a} e \mathbf{d} : vetores de efeitos aleatórios aditivos e dominantes das marcas $q \times 1$ não conhecidos, respectivamente e $\boldsymbol{\varepsilon}$: vetor dos efeitos residuais $n \times 1$. As matrizes de incidência \mathbf{W}_A e \mathbf{W}_D foram construídas seguindo a métrica de Cockerham (1954):

$$W_A \begin{pmatrix} 2 - 2p_k \\ 1 - 2p_k \\ -2p_k \end{pmatrix} \text{ para os genótipos } W_A \begin{pmatrix} AA \\ Aa \\ aa \end{pmatrix} \quad (2)$$

$$W_D \begin{pmatrix} -2(1 - p_k)^2 \\ 2p_k(1 - p_k) \\ -2p_k^2 \end{pmatrix} \text{ para os genótipos } W_D \begin{pmatrix} AA \\ Aa \\ aa \end{pmatrix} \quad (3)$$

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

Uma descrição completa sobre as especificações das distribuições de probabilidade dos efeitos aleatórios do modelo geral acima, para o uso do método Bayes B, pode ser encontrada em los Campos et al. (2013) e Pérez e De los Campos (2014).

Todas as análises bayesianas foram realizadas no pacote BGLR (PÉREZ; DE LOS CAMPOS, 2014) do software R (R CORE TEAM, 2014), com a função *BGLR*, ajustada para 10.000 iterações com os primeiros 1.000 ciclos descartados como *burn in* para inspeções gráficas. O parâmetro de mistura π (100/27.000), do modelo Bayes B foi ajustado visando o aproveitamento de informações genéticas para o ajuste de 100 efeitos

aditivos e de dominância das marcas. O Bayes B foi opção por encolher (*shrinkage*) com maior intensidade os efeitos das marcas.

3.3 Simulação dos híbridos simples

Foram construídos 400 híbridos simples via cruzamento artificial no software R (R CORE TEAM, 2014) em um esquema de cruzamento dialélico parcial sem recíprocos 20 (**Grupo 1** de linhagens) x 20 (**Grupo 2** de linhagens).

Os dois grupos de linhagens mencionados foram definidos graficamente por análise de componentes principais (Figura 1). Foi realizado o procedimento de decomposição espectral da matriz de parentesco aditivo (**A**) entre as 5.000 linhagens, e em seguida plotado os dois primeiros componentes principais (detalhes sobre a matriz de parentesco serão dados adiante). O **Grupo 1** foi definido pelas linhagens mais próximas da linhagem B73 no gráfico, e o **Grupo 2** pelas linhagens próximas da linhagem Mo17. Essas duas linhagens também foram integradas em cada um dos grupos. As linhagens B73 e Mo17 são linhagens elite para produção de grãos e produzem híbridos superiores quando cruzadas (GARCIA et al., 2008). Logo, espera-se grande efeito devido à heterose nos híbridos resultantes desse cruzamento.

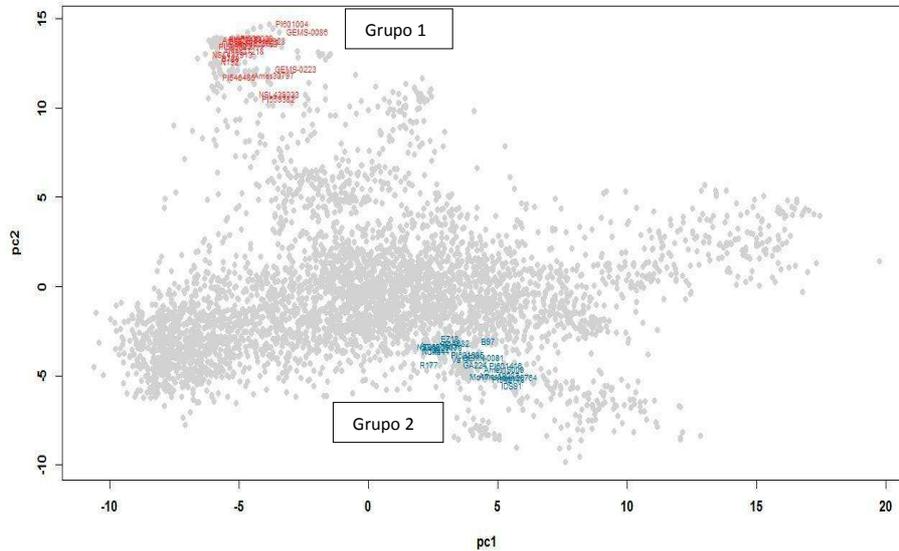


Figura 1 Análise de componentes principais das linhagens endogâmicas do banco de germoplasma do Departamento de Agricultura dos Estados Unidos (USDA). O eixo X representa o primeiro componente principal (pc1) e o eixo Y o segundo componente principal (pc2)

O estado genotípico de cada loco dos híbridos no cruzamento artificial foi definido pela esperança da contribuição alélica de cada linhagem parental no híbrido em questão. A esperança matemática foi computada por:

$$E(Z_{ij}) = p_{\alpha j} p_{\Omega j} 2 + p_{\alpha j} (1 - p_{\Omega j}) 1 + p_{\Omega j} (1 - p_{\alpha j}) 1 \quad (4)$$

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

Por exemplo, se for cruzada uma linhagem α com dois alelos favoráveis em um dado loco, com uma linhagem Ω com zero alelos favoráveis no mesmo loco, a esperança será dada por:

$$E(\mathbf{Z}_{ij}) = \mathbf{1} \times \mathbf{0} \times \mathbf{2} + \mathbf{1} \times (\mathbf{1} - \mathbf{0}) \times \mathbf{1} + \mathbf{0} \times (\mathbf{1} - \mathbf{1}) \times \mathbf{1} = \mathbf{1} \quad (5)$$

Na matriz de marcas das linhagens (\mathbf{Z}) houve dados perdidos, que são comuns em procedimentos de genotipagem com a técnica SNPs/GBS. Para eliminar essa restrição, realizou-se a imputação dos locos perdidos por meio da função *mean* do pacote rrBLUP do software R. Após esse processo os estados genotípicos dos locos variaram entre os intervalos de 0, 1 e 2. Para finalidade de cálculo de esperança, realizou-se interpolação numérica durante a estimação da frequência alélica favorável para o loco perdido, quando isso ocorreu para os locos de uma determinada linhagem parental. Quando as duas linhagens apresentaram o mesmo loco como perdido, durante o processo de cruzamento artificial, o estado genotípico do loco foi definido pelo mesmo procedimento de imputação descrito anteriormente, porém de forma direta na matriz \mathbf{Z} dos híbridos, após o procedimento de esperança nos locos em que houve pelo menos a presença de um dos locos como não perdido nas linhagens parentais.

O valor fenotípico dos híbridos foi obtido por:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{W}_{\alpha} \mathbf{a} + \mathbf{W}_{d} \mathbf{d} + \mathbf{W}_{\alpha\alpha} \mathbf{a}\mathbf{a} + \mathbf{W}_{\alpha d} \mathbf{a}\mathbf{d} + \mathbf{W}_{d\alpha} \mathbf{d}\mathbf{a} + \mathbf{W}_{dd} \mathbf{d}\mathbf{d} + \boldsymbol{\epsilon} \quad (6)$$

em que $\boldsymbol{\mu}$: média amostral obtida pelo Bayes B usando os dados fenotípicos reais das linhagens; \mathbf{a} : vetor de efeitos aditivos das marcas também estimados pelo Bayes B; \mathbf{d} : vetor de dominância das marcas estimado; $\mathbf{a}\mathbf{a}$ (efeito epistático do tipo aditivo-aditivo), $\mathbf{a}\mathbf{d}$ (efeito epistático do tipo aditivo-dominante), $\mathbf{d}\mathbf{a}$ (efeito epistático do tipo dominante-aditivo) e $\mathbf{d}\mathbf{d}$ (efeito epistático do tipo dominante-dominante) são vetores dos efeitos

epistáticos simulados que serão descritos na sequência; \mathbf{e} : vetor de resíduos amostrado de uma distribuição normal univariada ajustando as observações para cenários de herdabilidade 0,3, 0,5 e 0,7; detalhes sobre o ajuste da herdabilidade também serão dados na sequência; W_{aa} , W_{ad} , W_{da} e W_{dd} são matrizes de desvios epistáticos construídas conforme estabelecido por Cockerham (1954). Para construção dessas matrizes foram separadas as colunas relacionadas aos maiores 233 efeitos aditivos estimados na análise das linhagens de milho (construção das matrizes de desvio de epistasia incluindo de efeitos aditivos entre locos, por exemplo, W_{aa} , W_{ad} e W_{da}), e as colunas que envolvem os 233 maiores efeitos de dominância (construções que envolvem efeitos de dominância, por exemplo, W_{ad} , W_{da} e W_{dd}). Detalhes sobre a métrica de Cockerham (1954), do ponto de vista matricial e usando marcadores moleculares, podem ser encontrados em Kao e Zeng (2002) e Zeng, Wang e Zou (2005).

Os efeitos aditivos e de dominância, conforme já mencionado, foram estimados no conjunto de dados reais pelo método Bayes B. Os efeitos epistáticos foram simulados pela elaboração de todas as combinações duas-a-duas entre os valores genéticos dos marcadores (233 maiores valores) para cada classe de efeito epistático, sem recíprocos, com o objetivo de obter vetores de efeitos epistáticos das classes **aa**, **ad**, **da** e **dd**. Por exemplo, na construção de um efeito epistático do tipo **aa** que surge da interação de um loco com valor **a** de 0,5, com um outro loco com valor **a** de 0,5, o efeito epistático que surge da interação entre esses locos do tipo **aa** será 0,25 (produto dos dois efeitos).

As faixas de herdabilidade 0,3, 0,5 e 0,7 foram ajustadas nas observações pela magnitude dos resíduos (\mathbf{e}). Conforme também enfatizado, os valores do vetor \mathbf{e} foram amostrados de uma distribuição normal residual. As variâncias residuais foram definidas por:

$$V\mathbf{e} = \frac{(1 - h^2)V\mathbf{g}}{h^2} \quad (7)$$

em que h^2 : herdabilidade no sentido amplo desejada para ajuste e $V\mathbf{g}$: variância genética total do caráter.

3.4 Estimativas do peso de espigas com inclusão ou ausência dos efeitos não aditivos

Após o procedimento de simulação foi executado o procedimento analítico dos híbridos simples simulados com o uso de modelos G-BLUP. O modelo G-BLUP geral utilizado para avaliar a capacidade de predição de valores genéticos paramétricos dos híbridos foi:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\epsilon} \quad (8)$$

em que \mathbf{y} ($n \times 1$): vetor das observações fenotípicas dos híbridos simples; \mathbf{X} ($n \times p$): à matriz de incidência de efeitos fixos, que no caso desse trabalho foi apenas a média amostral; $\boldsymbol{\beta}$ ($1 \times p$): vetor de efeitos fixos (média amostral); \mathbf{Z} ($n \times q$): matriz de incidência de efeitos aleatórios, no caso do G-BLUP uma matriz identidade ($n \times n$); \mathbf{g} : valor genético (aleatório) dos híbridos, que dependendo do modelo apresentou um padrão de \mathbf{a} , \mathbf{d} , \mathbf{aa} , \mathbf{ad} e \mathbf{dd} que será descrito na sequência e $\boldsymbol{\epsilon}$ ($n \times 1$): vetor de efeitos residuais.

Foram utilizadas três configurações de modelos G-BLUP: (i) G-BLUP A com a estimação de apenas efeitos \mathbf{a} no modelo; (ii) G-BLUP AD com a inclusão apenas de efeitos \mathbf{a} e \mathbf{d} ; (iii) G-BLUP EP incluindo efeitos \mathbf{a} , \mathbf{d} , \mathbf{aa} , \mathbf{ad} e \mathbf{dd} .

Os pressupostos do modelo descrito acima são: $\mathbf{y} \sim MVN(\boldsymbol{\mu}, \mathbf{V})$; $\mathbf{g} \sim MVN(\mathbf{0}, \mathbf{G})$; $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \mathbf{R})$; em que \mathbf{V} : matriz de variância-covariância

fenotípica; **G**: matriz de variância-covariância genética e **R**: matriz de variância-covariância residual.

Todos os híbridos simulados nas configurações de herdabilidade 0,3; 0,5; e 0,7 foram analisados pelas três classes de modelos.

As matrizes de parentesco aditiva (**A**) e de dominância (**D**) utilizadas nesse trabalho foram as propostas por Vitezica, Varona e Legarra (2013):

$$\mathbf{A} = \frac{\mathbf{W}_A \mathbf{W}_A'}{2\sum pq} \quad (9)$$

$$\mathbf{D} = \frac{\mathbf{W}_D \mathbf{W}_D'}{4\sum (pq)^2} \quad (10)$$

em que **p**: frequência do alelo favorável por loco e **q** frequência do alelo desfavorável; **W_A**: matriz de desvio dos marcadores codificados por 2 (dois alelos favoráveis), 1 (um alelo favorável), 0 (nenhum alelo favorável) centralizados em **2p** (média do alelo favorável em um dado loco); **2∑pq**: soma da variâncias aditivas dos locos; **W_D**: matriz de desvios de dominância para cada loco **i**; **4∑(pq)²**: soma da variância dos desvios de dominância e

$$\mathbf{w}_i = \begin{cases} -2q^2 \\ 2pq \\ -2p^2 \end{cases} \text{ para os genótipos } \begin{cases} A^1 A^1 \\ A^1 A^2 \\ A^2 A^2 \end{cases}$$

Como ocorreram problemas de autovalor negativo nas matrizes aditiva e dominante, foi feita a decomposição espectral da matriz de acordo com o exposto nos trabalhos de Calinski et al. (2005) e Colvin e Dykstra (1991). Primeiramente, anularam-se os elementos negativos na diagonal da matriz de autovalores e adicionou-se a constante de valor 0.0001 de maneira decrescente ao longo da diagonal. Posteriormente, reconstruiu-se a matriz de parentesco decomposta e obteve-se uma matriz positiva definida,

possibilitando a aplicação do processo gaussiano necessário para uso do G-BLUP descrito nesse trabalho.

As matrizes de parentesco epistáticas foram obtidas conforme Henderson (1985b) por produto de Hadamard entre as matrizes de parentesco \mathbf{A} e \mathbf{D} , assim, $\mathbf{AA}=\mathbf{A}\#\mathbf{A}$, $\mathbf{AD}=\mathbf{A}\#\mathbf{D}$, $\mathbf{DD}=\mathbf{D}\#\mathbf{D}$, em que # refere-se ao produto de Hadamard.

Os componentes de variância durante a análise foram otimizados pela maximização da função REML de Patterson e Thompson (1971) com uso do algoritmo Expectation Maximization (EM) de Dempster (1977).

Com o objetivo de comparar a capacidade preditiva do modelo G-BLUP incluindo efeitos não aditivos com outros modelos preditivos disponíveis na literatura, foi empregado o pacote BGLR do software R (PÉREZ; DE LOS CAMPOS, 2014). Assim, foi executada a análise dos híbridos simulados com os seguintes modelos: Bayesian Ridge Regression (BRR); Bayes A; Bayes Lasso; Bayes C; Bayes B e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS). As análises foram realizadas conforme as configurações padrões das funções do pacote. Como de entrada foi utilizada a matriz de desvios dos marcadores com a construção definida pela equação (2) já reportada e o vetor de observações dependendo da configuração dos dados simulados. Especialmente no caso do RKHS, foi utilizado como *kernel* a matriz de distância euclidiana entre os marcadores e parâmetro “*bandwish*” configurado como 0,25, conforme mencionado na descrição de um exemplo do uso do pacote em Pérez e De los Campos (2014).

Foram estimados efeitos aditivos e de dominância com esses modelos, exceto para o RKHS. Essa metodologia preditiva carece de fundamentação genética por não se basear em matrizes de parentesco aditivo ou dominante e sim em distâncias euclidianas. Desse modo, não ocorre a decomposição ortogonal dos efeitos genéticos caso se tente incluir efeitos não aditivos em um modelo RKHS, ou seja, estão todos confundidos nas estimativas em geral superestimadas desses modelos. Maiores detalhes sobre

as propriedades estatísticas desse modelo podem ser encontradas em Morota e Gianola (2014).

Os efeitos epistáticos não foram considerados pela inviabilidade computacional de tal inclusão. O que motivou essa comparação é o fato de não haver na literatura, por enquanto, trabalhos comparando o modelo G-BLUP incluindo efeitos não aditivos com os outros modelos descritos.

De forma geral, desde Meuwissen, Hayes e Goddard (2001) vários trabalhos compararam o desempenho preditivo de modelos aditivos e concluiu-se que não há grande discrepância entre modelos, e que quando existe geralmente varia em função da espécie e da arquitetura genética do caráter (DE LOS CAMPOS et al., 2013; GIANOLA; MOROTA; CROSSA, 2014; HESLOT, et al., 2012; HOWARD; CARRIQUIRY; BEAVIS, 2014). Um dos objetivos neste trabalho foi verificar se isso também é verdadeiro com a inclusão de efeitos não aditivos no modelo G-BLUP durante predições de híbridos simples de milho, para o caráter escolhido (peso de espigas).

3.5 Comparação dos modelos

A capacidade preditiva dos modelos G-BLUP com efeitos aditivos e não aditivos foi avaliada por meio da correlação de Pearson entre os valores estimados e os paramétricos. O mesmo procedimento foi feito para os resultados obtidos pelos outros modelos preditivos disponíveis no pacote BGLR (PÉREZ; DE LOS CAMPOS, 2014). Também, para todos os modelos foi computada a herdabilidade dos efeitos genéticos por:

$$h^2_{g^*y} = \frac{\sigma_{g^*}^2}{\sigma_{g^*}^2 + \sigma_e^2} \quad (11)$$

em que g^* poderá ser: **a**, **d**, **aa**, **ad**, **dd** ou combinações dependendo do modelo.

Para os modelos G-BLUP calculou-se as estimativas dos vários componentes de variância dos efeitos genéticos aditivos e não aditivos, dependendo do modelo, assim como as correlações entre os valores estimados e simulados.

Por fim, estimou-se a Soma de Quadrados dos Erros de Predição (PRESS) por:

$$PRESS = \sum_{\mathbf{1}}^n (\hat{\theta} - \theta)^2 \quad (12)$$

em que $\hat{\theta}$: corresponde ao valor genético estimado; θ : valor genético paramétrico e n : número total de efeitos genéticos (400 híbridos).

Todas as análises foram programadas e executadas no software R (R CORE TEAM, 2014).

4 RESULTADOS E DISCUSSÃO

4.1 Grupos de linhagens e valores obtidos dos efeitos genéticos paramétricos

Na Tabela 1 são representados os dois grupos heteróticos usados na construção dos híbridos, que foram definidos em função do resultado da análise de componentes principais entre as 5.000 linhagens disponíveis no banco de dados.

Tabela 1 Grupos heteróticos de linhagens definidos pela análise de componentes principais

Grupo 1		Grupo 2	
B73	GEMS-0086	Mo17	EZ18
NC328	PI01004	Mo44	NC44
Ames22753	PI559382	R177	Va17
PI539927	PI539923	B97	GEMN-0081
Ames27151	NSL438033	GA224	IDS91
Ames27218	GEMS-0223	PI601685	PI601416
PI538009	N192	A682	Ames19008
NSL437913	PI46485	Ames10261	Ames19287
A679	Ames30797	Ames27178	Ames26764
B109	PI550473	NSL437903	PI542778

Fonte: Panzea (2014)

Os valores dos componentes de variância paramétricos obtidos a partir dos dados fenotípicos do conjunto de 5.000 linhagens e usados na construção dos híbridos encontram-se na Tabela 2.

Tabela 2 Componentes de variância paramétricos usados na simulação

Parâmetros*	Va	Vd	Vaa	Vad	Vdd	Vg
Valores	20,7	14,3	2,17	0,0792	0,00249	37,3

* Va: variância dos efeitos aditivos; Vd: variância dos efeitos de dominância; Vaa: variância dos efeitos epistáticos aditivos-aditivos; Vad: variância dos efeitos epistáticos aditivos-dominantes; Vdd: variância dos efeitos epistáticos dominantes-dominantes; Vg: variância genética total para o caráter peso de espiga

4.2 Acurácias preditivas e soma de quadrados dos erros de predição para todos os modelos

Nas Tabelas 3 e 4 são representadas as acurácias preditivas e a soma de quadrados dos erros de predição (PRESS) das estimativas do peso de espiga dos híbridos simples para os modelos estudados no trabalho, nas três faixas de herdabilidade simuladas e na condição de considerar ou não efeitos não aditivos na modelagem.

A magnitude das acurácias, levando em conta os modelos, as faixas de herdabilidade e a inclusão de efeitos não aditivos, pode ser considerada alta em comparação a outros estudos de simulação e também com dados reais reportados na literatura. Conforme concluiu Albrecht et al. (2011), Crossa et al. (2014), Desta e Ortiz (2014) e Technow et al. (2014) o tamanho da população de estudo é um dos fatores que influencia em tal estimativa. Os primeiros autores mencionados atribuíram a elevada magnitude da acurácia encontrada em seu trabalho (em torno de 0,7) ao fato de trabalharem com uma população de 1.377 linhagens, o que é bem menor do que o utilizado no presente trabalho (5.000), assim como todos os outros relatos encontrados na literatura até o momento.

Tabela 3 Acurácias preditivas e PRESS (entre parênteses) dos modelos G-BLUP A, Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS), para as três faixas de herdabilidades paramétricas simuladas (h^2) e modelo englobando apenas efeitos aditivos

Modelo	$h^2 = 0,3$	$h^2 = 0,5$	$h^2 = 0,7$
G-BLUP A	0,6966 (7.066,08)	0,7599 (5.839,64)	0,7725 (5.483,76)
Bayes A	0,7073 (7.048,04)	0,7764 (5.416,12)	0,7729 (5.478,56)
Bayes B	0,7164 (7.001,80)	0,7771 (5.400,80)	0,7909 (5.095,12)
Bayes C	0,7132 (7.001,08)	0,7735 (5.482,28)	0,7853 (5.213,28)
Bayes LASSO	0,7359 (6.774,84)	0,7898 (5.128,76)	0,7769 (5.392,44)
BRR	0,7056 (7.057,36)	0,7616 (5.773,44)	0,7702 (5.537,40)
RKHS	0,6967 (63.223,56)	0,7956 (302.923,44)	0,9155 (15.276,16)

Tabela 4 Acurácias preditivas e PRESS (entre parênteses) dos modelos G-BLUP dominante (G-BLUP AD), G-BLUP epistático (G-BLUP EP), Bayes A, Bayes B, Bayes C, Bayes LASSO e Bayesian Ridge Regression (BRR), para as 3 faixas de herdabilidades paramétricas simuladas (h^2) e modelo englobando efeitos não aditivos de dominância (todos) e ainda epistáticos (apenas G-BLUP EP)

Modelo	$h^2 = 0,3$	$h^2 = 0,5$	$h^2 = 0,7$
G-BLUP AD	0,8331 (4.255,36)	0,9039 (2.527,92)	0,9439 (1.711,92)
G-BLUP EP	0,8333 (4.228,40)	0,9019 (2.550,40)	0,9436 (1.666,32)
Bayes A	0,8406 (4.676,64)	0,8406 (4.529,28)	0,9424 (1.801,12)
Bayes B	0,8400 (4.727,32)	0,8334 (4.989,56)	0,9421 (1.814,40)
Bayes C	0,8399 (4.533,60)	0,8420 (4.383,36)	0,9436 (1.728,16)
Bayes LASSO	0,8389 (4.758,52)	0,8391 (4.757,20)	0,9438 (1.703,52)
BRR	0,8410 (4.184,16)	0,8416 (4.248,16)	0,9444 (1.697,20)

Outro fato que contribuiu para as altas acurácias preditivas reportadas nas Tabelas 3 e 4 foi a elevada densidade de marcadores utilizada (27.000). Isso dá sustentação ao princípio de aplicação da predição genômica, qual seja, o fato de que para cada loco polimórfico existe(m) marcador(es) associado(s) em desequilíbrio de ligação (ALBRECHT et al., 2011; CHARMET; STORLIE, 2012; CROSSA et al., 2014; DESTA; ORTIZ 2014).

Bernardo e Yu (2007) reforçaram a afirmação anterior com um estudo de simulação em milho no qual nas maiores densidades de marcadores resultaram as melhores respostas preditivas. Ainda segundo os autores, a densidade de marcadores é especialmente importante quando o caráter é controlado por maior número de QTL's e nos cenários de maior herdabilidade.

Por sua vez, esperava-se que tal densidade de marcadores acentuasse as diferenças entre os modelos, conforme apontado por Crossa et al. (2013), ressaltando que em baixa densidade de marcadores o pobre desequilíbrio de ligação entre esses e os QTL's age como um fator limitando a acurácia de predição e a capacidade de cada modelo expressar seu potencial. Todavia, como visto nas Tabelas 3 e 4, de modo geral as diferenças em acurácia entre os modelos foram muito discretas.

Como já reportado de maneira generalizada na literatura (ALBRECHT et al., 2011; CROSSA et al., 2014; DE LOS CAMPOS et al., 2010; JIA; JANNINK, 2012), tanto no melhoramento animal quanto no vegetal, a acurácia de predição baseando-se em parentesco a partir de marcadores supera o uso de parentesco baseando-se em *pedigree*. No trabalho em questão foi utilizado apenas o parentesco definido pelos marcadores, tanto para os métodos bayesianos quanto para os métodos derivados dos modelos mistos, nesse último caso caracterizando o “*Genomic BLUP*” ou G-BLUP.

De forma geral, para todos os modelos e nos dois tipos de arquitetura genética simuladas, somente aditiva e também não aditiva, nas maiores

herdabilidades as acurácias de predição também foram mais elevadas, assim como esperado e relatado pela maioria dos trabalhos no assunto.

Para a soma de quadrados dos erros de predição (PRESS), conforme já alertado anteriormente no trabalho, o modelo RKHS apresentou os maiores valores (Tabela 3) para as três faixas de herdabilidade, embora isso não tenha afetado as acurácias preditivas desse modelo. Para o cenário considerado na Tabela 3 com o aumento da herdabilidade ocorreu uma diminuição das estimativas de PRESS, mas somente até a herdabilidade de 0,5. A mudança de herdabilidade de 0,5 para 0,7 pouco afetou a PRESS dos modelos, assim como as diferenças entre os modelos, com exceção do RKHS, que foram de magnitudes bem pequenas.

Quando se incluiu efeitos não aditivos aos modelos para predição (Tabela 4) o comportamento da PRESS frente ao aumento da herdabilidade diferiu do ocorrido anteriormente, pois agora as reduções ocorreram até a maior faixa de herdabilidade simulada (0,7). Esse é um comportamento coerente, já que quanto maior a herdabilidade de um caráter menos resíduos existem entre os dados, o que proporciona estimativas mais próximas dos valores paramétricos, e logo, menores estimativas de erros.

Apenas na faixa de herdabilidade de 0,5, para a situação de efeitos não aditivos inclusos, os modelos G-BLUP se destacaram com as menores PRESS. Nos demais casos, novamente as diferenças entre os modelos foram bastante reduzidas, acompanhando o que ocorreu para a acurácia de predição.

Comparando as arquiteturas genéticas de apenas efeitos aditivos ou com inclusão de efeitos não aditivos, para os mesmos modelos dentro de cada faixa de herdabilidade, nota-se que a inclusão de efeitos não aditivos de dominância e/ou epistasia contribuiu para melhorar as estimativas visto as menores PRESS. Isso indica uma melhoria na captura de efeitos genéticos pelos modelos, resultando em estimativas mais fiéis dos valores paramétricos.

Com base nos resultados aqui encontrados percebe-se que é possível conseguir boas predições de performance de híbridos simples de milho, validando o potencial da predição genômica como uma ferramenta no melhoramento. Como os custos com a genotipagem estão diminuindo, programas de melhoramento poderiam fazer uso de modelos de predição genômica a fim de predizer o valor genotípico dos novos cruzamentos antes desses alcançarem o estágio de testes de campo (PÉREZ-RODRÍGUEZ et al., 2012).

4.3 Modelo G-BLUP considerando efeitos aditivos e não aditivos

De modo geral, verifica-se que a inclusão de efeitos não aditivos no modelo G-BLUP aumentou a acurácia de predição do desempenho dos híbridos, para as três faixas de herdabilidade estudadas no trabalho (Gráfico 1). O acréscimo em acurácia somente ocorreu em comparação ao modelo G-BLUP apenas aditivo, sem diferença ao se comparar os modelos considerando os efeitos dominantes ou incluindo ainda os epistáticos.

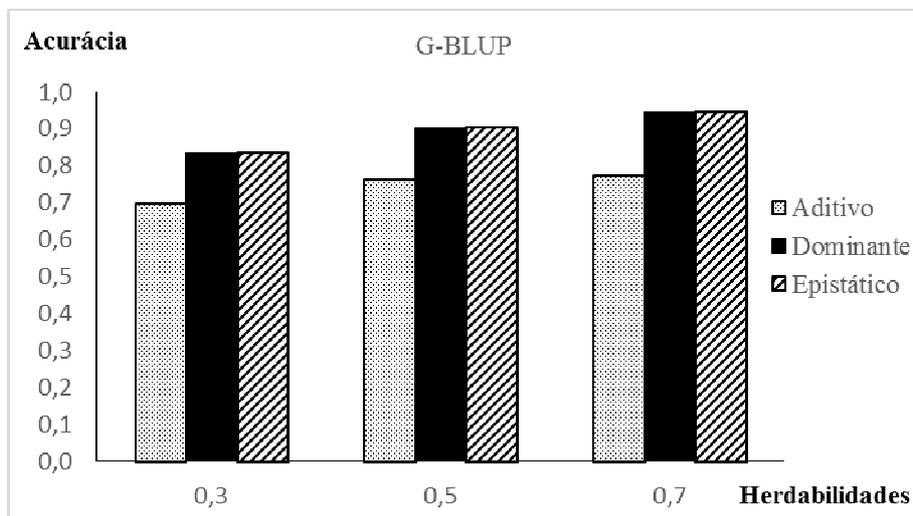


Gráfico 1 Acurácias preditivas dos modelos G-BLUP incluindo apenas efeitos aditivos (G-BLUP A), efeitos aditivos e dominantes (G-

BLUP AD) e por fim efeitos aditivos, dominantes e epistáticos (G-BLUP EP), para as três faixas de herdabilidade simuladas

Igualmente trabalhando com híbridos advindos do cruzamento das linhagens B73 e Mo17, Garcia et al. (2008) concluíram que a dominância é a principal causa da heterose em milho. Os resultados aqui encontrados também evidenciam esse fato, visto as pequenas magnitudes das variâncias devido aos efeitos epistáticos e a falta de resposta do modelo G-BLUP à sua inclusão.

Em trabalho sobre a utilidade dos modelos mistos (BLUP) para a predição do desempenho de híbridos de milho, Bernardo (1994) concluiu sobre o potencial de tal metodologia para essa finalidade. No caso o autor utilizou outro marcador molecular, o RFLP (*Restriction Fragment Length Polymorphisms*), e um conjunto de dados (54 híbridos) bem menor do que o usado no presente trabalho. Ainda assim, as correlações entre os valores preditos e os observados (acurácia) aproximaram-se bastante das encontradas neste trabalho, variando de 0,688 a 0,80; o que segundo o autor é de moderada a alta magnitude. Outra constatação por ele relatada foi a superioridade do modelo considerando a variância devido à dominância, do mesmo modo que o resultado obtido no presente estudo. Ressalta ainda a importância que a variância atribuída à dominância tem para o caráter produtividade em milho.

Também em concordância com os resultados aqui encontrados, Su et al. (2012) concluíram à respeito da vantagem de incluir efeitos não aditivos em um modelo de predição G-BLUP. Uma significância encontrada pelos autores e não verificada nesse trabalho foi o efeito da inclusão de efeitos epistáticos em um modelo já considerando efeitos aditivos e dominantes. O trabalho desses autores guarda semelhanças com o aqui realizado, como trabalhar com um caráter quantitativo, população numerosa e usar dados reais para estimar os efeitos das marcas. Entretanto, a pesquisa referia-se a uma espécie animal e provavelmente com outra arquitetura genética, visto a

significância da inclusão dos efeitos epistáticos. Porém, serve para ilustrar a importância de incluir efeitos não aditivos na predição de caracteres quantitativos. Vale lembrar que modelos de predição genômica foram desenvolvidos pelo melhoramento animal e estendidos com sucesso para o melhoramento vegetal.

Em outro trabalho avaliando a acurácia de modelos mistos (BLUP) na predição de caracteres na cultura do milho, Bernardo (1996) estudou dois tipos de modelos, aditivo e dominante, para quatro caracteres. Novamente, os resultados evidenciaram a superioridade do G-BLUP considerando a variância de dominância na modelagem, porém tal vantagem não foi consistente para os diferentes caracteres estudados. O autor encontrou a tendência de que para caracteres em que a variância de dominância contabiliza importante parcela da variância genética total o acréscimo em acurácia, sobre o modelo estritamente aditivo, foi maior do que quando essa variância teve pequena participação. Sugere ainda que o uso de modelos aditivos deve ocorrer quando se tem informação *a priori* sobre a pequena contribuição da variância de dominância para a variância genética total, o que ocorreu para o caráter umidade de grãos entre os caracteres por ele estudados. No entanto, para o componente de produção peso de espiga estudado no presente trabalho essa recomendação não se aplica, visto ser aceito que esse caráter está sob controle de vários genes que mostram expressiva resposta à heterose e endogamia, indicando haver importante contribuição dos efeitos de dominância. O desempenho do modelo estritamente aditivo reforça essa afirmação (Gráfico 1).

Corroborando a ideia apresentada no parágrafo anterior Nishio e Satoh (2014) não encontraram diferença na acurácia de predição entre modelos G-BLUP apenas aditivo e G-BLUP considerando efeitos de dominância, para dois caracteres estudados no contexto do melhoramento animal. No trabalho desses autores a variância explicada pela dominância era de apenas 9,6% e 6,3% para cada caráter, portanto, de fato uma pequena fração.

Na opinião de Crossa et al. (2014) quando a predição genômica visa o valor genotípico dos indivíduos para fins de liberação como cultivares comerciais, ambos os efeitos, aditivos e não aditivos, são importantes para determinar o valor comercial de híbridos de milho.

Dudley e Johnson (2013) propuseram uma metodologia diferente da utilizada aqui para incorporar efeitos epistáticos em um modelo de predição genômica. Os autores concluíram que somente ocorreram ganhos em acurácia quando os marcadores e as interações passaram por uma pré-seleção, incluindo na modelagem somente os significativos, segundo determinado nível de probabilidade. Vale destacar que nesse trabalho os autores consideraram apenas caracteres qualitativos dos grãos de milho, diferentemente do considerado no trabalho em questão.

Guo et al. (2012) apontam como causa da falta generalizada de resposta à inclusão de efeitos epistáticos o grande viés que é integrado no modelo ao combinar efeitos principais e epistáticos dos QTL's.

Ao comparar-se a qualidade dos modelos G-BLUP em estimar os componentes de variância (Tabela 5) verifica-se que para a variância dos efeitos aditivos (V_a) ocorreu uma diminuição na sua magnitude e ao mesmo tempo uma aproximação ao valor paramétrico verdadeiro com a inclusão de efeitos não aditivos de dominância e epistáticos, nas três faixas de herdabilidade. A redução nas estimativas significa que gradativamente essa variância tornou-se melhor isolada das demais, reduzindo o confundimento. A maior aproximação ao valor paramétrico indica a capacidade de resposta do modelo G-BLUP à inclusão dos efeitos dominantes e epistáticos para o caráter peso de espigas em milho.

Tabela 5 Componentes de variância paramétricos (S) e estimados pelos modelos, nas três faixas de herdabilidade (h^2)

		$h^2 = 0,3$	$h^2 = 0,5$	$h^2 = 0,7$
Parâmetros*	S	G-BLUP A		
Va	-	25,8227	26,6841	41,0145
Vg	-	25,8227	26,6841	41,0145
		G-BLUP AD		
Va	-	24,7678	21,2032	26,6961
Vd	-	12,5113	15,8379	14,8276
Vg	-	37,2791	37,0410	38,5237
		G-BLUP EP		
Va	20,7	23,8922	20,0839	21,1368
Vd	14,3	11,1155	13,9725	11,9878
Vaa	2,17	2,0353	2,2003	5,1167
Vad	0,0792	0,8026	1,0696	0,3495
Vdd	0,00249	0,6784	0,8224	0,2824
Vg	37,3	38,5240	38,1488	38,8731

* Va: variância dos efeitos aditivos; Vd: variância dos efeitos de dominância; Vaa: variância dos efeitos epistáticos aditivos-aditivos; Vad: variância dos efeitos epistáticos aditivos-dominantes; Vdd: variância dos efeitos epistáticos dominantes-dominantes; Vg: variância genética total para o caráter peso de espiga

Para a variância dos efeitos dominantes (Vd) o comportamento foi semelhante até certo ponto ao relatado para a Va, já que do modelo G-BLUP AD para o G-BLUP EP as estimativas reduziram-se, no entanto, não necessariamente ficaram mais precisas.

A inclusão dos efeitos de epistasia afetou muito pouco as estimativas, o que deveria ser esperado, já que a contribuição relativa da variância dos efeitos epistáticos para a variância genética é bem pequena para o caráter e população estudados.

O modelo G-BLUP EP estimou mais precisamente as variâncias de efeitos epistáticos aditivos-aditivos (V_{aa}) nas faixas de herdabilidade de 0,3 e 0,5. Para as variâncias dos demais efeitos epistáticos V_{ad} e V_{dd} na herdabilidade de 0,7 as estimativas foram mais fidedignas, ainda assim bem distantes do valor real simulado confirmando a dificuldade em se trabalhar com efeitos de epistasias na predição genômica.

De forma generalizada, a inclusão de efeitos não aditivos melhorou as estimativas da variância genética total (V_g). Basta comparar os desvios em relação ao valor paramétrico das estimativas fornecidas pelo G-BLUP A com as advindas do G-BLUP AD e G-BLUP EP. Nesse ponto chama a atenção que as estimativas mais precisas ocorreram quando se considerou apenas os efeitos de dominância na predição (G-BLUP AD), pois a inclusão dos efeitos epistáticos (G-BLUP EP) piorou-as ligeiramente.

Na Tabela 6 visualiza-se as correlações entre os efeitos genéticos estimados e simulados para as três configurações de modelos G-BLUP. Fica claro que os efeitos epistáticos tiveram as piores estimativas em vista das expressivas baixas correlações, principalmente as que envolviam efeitos de dominância ($r_{ad\hat{a}}$ e $r_{dd\hat{d}}$). Para os demais efeitos verificam-se as elevadas magnitudes de correlações com tendência de aumento com o acréscimo da herdabilidade (h^2).

Tabela 6 Correlações entre os efeitos genéticos simulados e estimados pelos modelos, nas três faixas de herdabilidades (h^2)

	$h^2 = 0,3$	$h^2 = 0,5$	$h^2 = 0,7$
Parâmetros*			
		G-BLUP A	
<i>raâ</i>	0,8849	0,9363	0,9244
<i>rgg</i>	0,6966	0,7599	0,7725
		G-BLUP AD	
<i>raâ</i>	0,8833	0,9612	0,9389
<i>rdâ</i>	0,7430	0,8005	0,8645
<i>rgg</i>	0,8331	0,9039	0,9439
		G-BLUP EP	
<i>raâ</i>	0,8833	0,9603	0,9385
<i>rdâ</i>	0,7442	0,8031	0,8584
<i>raaââ</i>	0,1816	0,1192	0,2955
<i>radââ</i>	0,0647	0,0249	0,0176
<i>rdâââ</i>	0,0482	-0,1180	-0,1312
<i>rgg</i>	0,8333	0,9019	0,9436

* *raâ*: correlação entre os valores paramétricos e estimados dos efeitos aditivos; *rdâ*: correlação entre os valores paramétricos e estimados dos efeitos dominantes; *raaââ*: correlação entre os valores paramétricos e estimados dos efeitos epistáticos aditivos-aditivos; *radââ*: correlação entre os valores paramétricos e estimados dos efeitos epistáticos aditivos-dominantes; *rdâââ*: correlação entre os valores paramétricos e estimados dos efeitos epistáticos dominantes-dominantes; *rgg*: correlação entre os valores paramétricos e estimados dos efeitos genéticos totais

Incluir efeitos epistáticos será útil se duas condições forem atendidas: em primeiro lugar a epistasia é importante para o caráter trabalhado e em segundo as interações epistáticas podem ser modeladas com precisão (LORENZANA; BERNARDO, 2009). No entanto, Heffner, Sorrells e Jannink (2009) alertam que antecipar os efeitos da epistasia no potencial da seleção genômica é difícil. Pérez-Rodríguez et al. (2012)

reforçam a observação anterior ao afirmarem que entender o impacto da epistasia em caracteres quantitativos continua um grande desafio.

Estudos de simulação permitem explorar maior variedade de situações com respeito à arquitetura genética do caráter e da população sob estudo, ao passo que dados reais oferecem a vantagem de refletir toda a complexidade envolvida. No entanto, a simulação de genomas e seus mecanismos genéticos são complexos. Há diferentes formas de variabilidade genômica e grande variedade de populações com históricos diferentes, assim como considerável incerteza sobre o modo e distribuição da ação gênica (DAETWYLER et al., 2013).

4.4 Modelos G-BLUP e bayesianos

De maneira análoga ao comportamento dos modelos G-BLUP na seção anterior, a inclusão de efeitos não aditivos (dominância) também melhorou a acurácia preditiva dos modelos bayesianos para a faixa de herdabilidade de 0,3 (Tabela 7).

Nota-se que na condição de modelos estritamente aditivos apenas o modelo Bayes LASSO mostrou uma ligeira diferença em relação aos demais. Para a situação considerando efeitos não aditivos na predição o comportamento dos modelos não diferiu. Importante destacar que nas duas arquiteturas genéticas consideradas os modelos G-BLUP desempenharam-se de maneira equivalente aos demais modelos.

Tabela 7 Acurácias preditivas dos modelos G-BLUP aditivo (G-BLUP A), G-BLUP dominante (G-BLUP AD), G-BLUP epistático (G-BLUP EP), Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) considerando somente efeitos aditivos ou incluindo também os não aditivos, na faixa de herdabilidade de 0,3

Modelo	Efeitos Aditivos	Efeitos Não aditivos
G-BLUP A	0,6966	-
G-BLUP AD	-	0,8331
G-BLUP EP	-	0,8333
Bayes A	0,7073	0,8406
Bayes B	0,7164	0,8400
Bayes C	0,7132	0,8399
Bayes LASSO	0,7359	0,8389
BRR	0,7056	0,8410
RKHS	0,6967	-

Já para a faixa de herdabilidade de 0,5 (Tabela 8) o ganho adicional com a inclusão de efeitos de dominância nos modelos foi menor do que o observado para a herdabilidade de 0,3. Novamente, na condição de modelos apenas aditivos diferenças expressivas não foram observadas. Por sua vez, na condição incluindo os efeitos não aditivos os modelos G-BLUP (G-BLUP AD e G-BLUP EP) destacaram-se.

Os modelos G-BLUP com efeitos dominantes e epistáticos não diferiram entre si, o que já era esperado pelo comportamento mostrado pelo Gráfico 1, no qual visualiza-se que a inclusão de efeitos epistáticos no modelo G-BLUP praticamente nada acrescenta na acurácia obtida com a inclusão apenas dos efeitos dominantes.

Por sua vez, também simulando híbridos simples de milho nessa faixa de herdabilidade e considerando apenas modelos aditivos, Meuwissen, Hayes e Goddard (2001) encontraram ligeira vantagem na acurácia preditiva dos modelos Bayes A e Bayes B sobre G-BLUP.

Tabela 8 Acurácias preditivas dos modelos G-BLUP aditivo (G-BLUP A), G-BLUP dominante (G-BLUP AD), G-BLUP epistático (G-BLUP EP), Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) considerando somente efeitos aditivos ou incluindo também os não aditivos, na faixa de herdabilidade de 0,5

Modelo	Efeitos Aditivos	Efeitos Não aditivos
G-BLUP A	0,7599	-
G-BLUP AD	-	0,9039
G-BLUP EP	-	0,9019
Bayes A	0,7764	0,8406
Bayes B	0,7771	0,8334
Bayes C	0,7735	0,842
Bayes LASSO	0,7898	0,8391
BRR	0,7616	0,8416
RKHS	0,7956	-

Na faixa de herdabilidade de 0,7 a situação inverteu-se e o comportamento dos modelos foi equivalente apenas quando da inclusão de efeitos de dominância. Na situação de modelos estritamente aditivos o modelo RKHS destacou-se dos demais na acurácia de predição do peso de espigas dos híbridos (Tabela 9). Para todos os modelos mais uma vez verifica-se a vantagem da inclusão de efeitos não aditivos na acurácia de predição.

Tabela 9 Acurácias preditivas dos modelos G-BLUP aditivo (G-BLUP A), G-BLUP dominante (G-BLUP AD), G-BLUP epistático (G-BLUP EP), Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR) e Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) considerando somente efeitos aditivos ou incluindo também os não aditivos, na faixa de herdabilidade de 0,7

Modelo	Efeitos Aditivos	Efeitos Não aditivos
G-BLUP A	0,7725	-
G-BLUP AD	-	0,9439
G-BLUP EP	-	0,9436
Bayes A	0,7729	0,9424
Bayes B	0,7909	0,9421
Bayes C	0,7853	0,9436
Bayes LASSO	0,7769	0,9438
BRR	0,7702	0,9444
RKHS	0,9155	-

Em um dos poucos trabalhos comparando predição de híbridos de milho entre modelos G-BLUP e bayesianos com efeitos de dominância, Technow et al. (2012) encontraram significância para a diferença entre G-BLUP e Bayes B somente quando considerou-se os efeitos de dominância nos modelos. Assim, esse resultado concorda com o encontrado no presente trabalho somente na situação de modelos aditivos, já que no modelo dominante a diferença encontrada pelos autores não foi aqui constatada.

Concordando com os resultados aqui encontrados, Jia e Jannink (2012) também não encontraram diferença em acurácia preditiva para modelos G-BLUP e Bayes A na condição de efeitos não aditivos presentes em baixa (0,1) ou média (0,5) herdabilidade e pequeno (20) ou grande (200) número de QTL's controlando um caráter.

Para caracteres controlados por grande número de QTL's, como o considerado nesse trabalho em que se simulou 100 QTL's com efeitos aditivos e dominantes, espera-se que a contribuição deles para a variância genética total varie. Essa pressuposição é considerada pelos modelos

bayesianos, já o contrário ocorre para os modelos mistos (estimadores BLUP) em que se assume equivalência de variâncias para os efeitos aleatórios do modelo (marcadores no caso). Além disso, espera-se que existam alguns marcadores de efeitos maiores (explicam maior parcela da variância genética do caráter) e o restante de pequenos efeitos. Como afirmado por De los Campos et al. (2010), quando o número de marcadores aumenta é esperado que alguns deles estejam localizados em regiões genômicas que não são associadas com a variabilidade genética do caráter quantitativo em avaliação.

Em tal cenário, Bernardo (2014) simulou um programa de melhoramento de milho e não encontrou diferença em acurácia para o 1º ciclo seletivo, tanto para igualdade de variâncias ou não entre os marcadores, utilizando a metodologia de modelos mistos para a predição. De forma semelhante, no presente trabalho nem os modelos G-BLUP nem os bayesianos diferiram na maior faixa de herdabilidade. A situação simulada pelo autor em maiores valores de herdabilidade foi apontada como favorável ao se considerar as desiguais variâncias dos marcadores.

De forma semelhante aos resultados do presente trabalho (exceto para herdabilidade de 0,7), Crossa et al. (2010) não verificaram superioridade preditiva dos modelos RKHS e Bayes LASSO sobre o G-BLUP trabalhando com modelos aditivos e com o caráter produtividade de grãos de milho em condições de disponibilidade hídrica contrastantes. Todavia, para caracteres relacionados aos florescimentos masculino e feminino o Bayes LASSO superou o RKHS e este por sua vez o G-BLUP. Assim, como já apontado por outros autores, a acurácia preditiva dos modelos depende da arquitetura genética do caráter em consideração. Nesse sentido os autores mencionados afirmaram que pode-se esperar melhor performance do Bayes LASSO para caracteres em que os efeitos aditivos são os principais. Entretanto, a incorreta, mas conveniente suposição de iguais variâncias entre os marcadores parece não afetar significativamente as respostas à seleção, como verificaram Bernardo e Yu (2007). Esses autores

constatarem uma perda média em resposta à seleção, levando em conta diferentes números de QTL's e herdabilidades, de apenas 2% ao se comparar marcadores com iguais variâncias ou variância específica por loco. Salientam ainda que para os esquemas de seleção genômica procedimentos bayesianos para modelar locos com variâncias específicas teriam, por ventura, pequena vantagem. Isso aconteceu no presente trabalho, já que nenhum modelo bayesiano destacou-se em acurácia em qualquer das situações.

Na mesma tendência apontada pelos resultados desse trabalho, Lorenzana e Bernardo (2009) concluíram a respeito da superioridade em acurácia do modelo G-BLUP sobre procedimentos bayesianos ao estudarem vários caracteres em quatro culturas. Inclusive, os autores constataram que incluir efeitos epistáticos em modelos bayesianos empíricos por eles desenvolvidos levou às piores estimativas de valores genotípicos. Essa última situação não foi aqui abordada, mas vale ressaltar pela importância da dificuldade em modelar interações epistáticas na seleção genômica.

Na mesma linha de pensamento, Meuwissen, Hayes e Goddard (2001) ao encontrarem ligeiras diferenças entre G-BLUP e modelos bayesianos de predição afirmaram que o efeito de considerar uma incorreta distribuição de probabilidade para a variância dos marcadores parece ser pequeno. Technow et al. (2014) afirmaram que os modelos bayesianos exigem grandes populações para estimação dos efeitos dos marcadores a fim de que possam manifestar alguma potencial vantagem, condição essa atendida no trabalho em questão visto que 5.000 linhagens foram utilizadas nas estimativas.

Ao referir-se aos resultados já encontrados no melhoramento genético de milho e trigo pelo centro de pesquisa CIMMYT, Crossa et al. (2014) afirmaram que em termos de algoritmos usados para predição nenhum modelo ajusta-se a todas as situações. No entanto, evidências empíricas têm mostrado que modelos que incluem funções não lineares de marcadores, quando usados em caracteres complexos, tendem a aumentar a

capacidade de predição quando comparados a modelos lineares, assim com também apontou Pérez-Rodríguez et al. (2012). Ainda segundo os primeiros autores para caracteres menos complexos em que predominam efeitos aditivos modelos lineares resultam nas melhores acurácias de predição.

De fato, Heslot et al. (2012) em um amplo trabalho sobre comparação de modelos de predição genômica em diversas culturas e caracteres encontraram a tendência geral de melhor acurácia de modelos RKHS. Esse método não paramétrico e não linear tendeu a superajustar o modelo (“*overfitting*”), ainda assim proporcionou maiores acurácias, o que indica maior capacidade de capturar tanto informação genética como não genética (“*noise*”). Essa última observação pode ser verificada ao analisarem-se os valores da soma de quadrados dos erros de predição (PRESS) nas Tabelas 3 e 4 em que o modelo RKHS apresentou as maiores estimativas de erros sem, no entanto, ter sido prejudicado na capacidade preditiva. Para Guo et al. (2013) quanto mais informação genética o modelo for capaz de usar melhor e mais estável predição é esperada.

A melhor capacidade preditiva de um modelo pode ter várias causas, uma delas é a habilidade de alguns modelos não lineares em capturar efeitos epistáticos (PÉREZ-RODRÍGUEZ et al., 2012). Para Howard, Carriquiry e Beavis (2014) a epistasia implica em dependência entre os marcadores, o que viola os pressupostos de modelos paramétricos.

Corroborando as afirmações de Crossa et al. (2014) e Heslot et al. (2012) Howard, Carriquiry e Beavis (2014) encontraram a tendência de que os métodos paramétricos por eles estudados (Bayes A, Bayes B, Bayes C, Bayes LASSO, BRR, G-BLUP, quadrados mínimos e outros) desempenharam-se ligeiramente melhor, em termos de acurácia e erro de predição, na situação em que apenas efeitos aditivos estavam presentes. Já na presença de epistasia os métodos não-paramétricos por eles estudados (RKHS, Redes neurais e outros) superaram os paramétricos nos dois critérios anteriormente mencionados. Os autores salientam que a arquitetura genética

do caráter foi a principal responsável pelas diferenças em acurácia de predição entre os 14 métodos de seleção genômica por eles avaliados.

Deve-se ter cautela ao interpretar estimativas obtidas pela metodologia RKHS, pois como já enfatizado o que esse modelo estima não é um efeito genético “puro”, independente de efeitos de outra natureza, mas sim uma estimativa, em geral de maior magnitude, de um conjunto de efeitos confundidos. Esse modelo é exaltado na literatura justamente por essa “capacidade” de estimar melhor efeitos não lineares como os epistáticos, porém ele não permite conclusões desse nível, ou seja, da exata natureza genética do que está sendo estimado. Isso não significa que não tenha valor no contexto preditivo, basta comparar os resultados de acurácia proporcionados por esse modelo (Tabelas 7, 8 e 9) com os demais. Diferentemente, modelos como o G-BLUP, fundamentados em uma base genética de acordo com a natureza do efeito envolvido, permitem conclusões mais sólidas sobre a estimativa obtida pelo fato dessa ser melhor separada da variação total.

Por sua vez, trabalhando com dados reais da cultura do milho e modelos aditivos, Sun, Ma e Mumm (2012) verificaram que o modelo RKHS superou o Bayes B e G-BLUP na predição de duas características (produtividade de grãos e intervalo entre florescimento masculino e feminino) em híbridos de milho. O modelo Bayes A foi superado pelos demais também para as duas características. Entretanto, quando os autores simularam vários cenários, diferindo em herdabilidade e importância relativa da epistasia, os modelos não se comportaram de maneira consistente, indicando que a arquitetura genética do caráter é determinante para o desempenho do modelo. Um tratamento não paramétrico dos marcadores pode considerar efeitos epistáticos que não são capturados por modelos aditivos de regressão linear (PÉREZ-RODRÍGUEZ et al., 2012).

Riedelsheimer, Technow e Melchinger (2012) concluíram que pequenos ganhos em acurácia podem ser obtidos se o modelo de predição genômica é escolhido de acordo com a arquitetura genética do caráter.

Entretanto, Heffner, Sorrells e Jannink (2009) fazem uma importante observação a esse respeito: simplesmente não conhecemos, para caracteres complexos, qual é a arquitetura genética básica e portanto, não temos conhecimento *a priori* para tal procedimento.

Diferentemente do encontrado no presente trabalho, Charmet e Storlie (2012) avaliando modelos de predição genômica de rendimento na cultura do trigo encontraram superioridade na acurácia preditiva do modelo Bayes LASSO e piores valores de acurácia para o modelo G-BLUP, em diferentes situações de herdabilidade com dados simulados, considerando apenas efeitos aditivos. Já na comparação de acurácia com dados fenotípicos reais os métodos equivaleram-se na média de 6 locais de teste. Deve-se levar em conta que os autores mencionados trabalharam com uma população de estudo pequena (318 linhagens), menor densidade de marcadores (2121) e outra cultura, inclusive com modo de reprodução e provável arquitetura genética diferente do milho, o que pode explicar a discordância de resultados para as simulações dos dois trabalhos.

Também diferindo dos resultados aqui encontrados, Riedelsheimer, Technow e Melchinger (2012) avaliaram, com dados reais de linhagens de milho (289), a capacidade preditiva de modelos de seleção genômica para duas classes de caracteres que eles classificaram como agronômicos e metabólicos, sendo cada uma controlada por grande número de QTL's de pequeno efeito e poucos QTL's de grande efeito, respectivamente. Verificaram que o modelo RKHS equivaleu-se ao G-BLUP na condição de considerar efeitos aditivos e os caracteres ditos agronômicos. No entanto, a acurácia preditiva desses dois modelos superou a dos modelos Bayes LASSO e Bayes B, diferindo do aqui encontrado. Já para os caracteres metabólicos o Bayes LASSO destacou-se dos demais com o G-BLUP não diferindo do Bayes B. Os autores atribuíram às pequenas diferenças em acurácia preditiva dos modelos testados ao elevado nível de desequilíbrio de ligação entre as linhagens elite de milho trabalhadas.

Possíveis razões apontadas para o melhor desempenho do modelo Bayes LASSO sobre o G-BLUP, ambos os métodos paramétricos, são (CROSSA et al., 2010) (1): o “*shrinkage*” específico por marcador, que penaliza de maneira diferenciada os efeitos dos marcadores e (2): as variâncias dos parâmetros e os efeitos dos marcadores são estimados conjuntamente, enquanto nos modelos mistos isso envolve tipicamente duas etapas, gerando um viés nas estimativas.

4.5 Herdabilidades dos componentes de variância

Na Tabela 10 são representadas as estimativas de herdabilidade para os modelos considerando apenas efeitos aditivos na predição do peso de espigas dos híbridos. Como o único componente de variância presente nessa situação é a variância aditiva, as estimativas representadas na Tabela 10 são as herdabilidades dos efeitos aditivos e que, nesse caso, correspondem às estimativas da herdabilidade genética.

Tabela 10 Estimativas da herdabilidade dos efeitos da variância aditiva/variância genética total para os modelos Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR), Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) e G-BLUP A, na condição de modelos incluindo apenas efeitos aditivos e para as três faixas de herdabilidade paramétricas simuladas (h^2)

Modelos	$h^2 = 0,3$	$h^2 = 0,5$	$h^2 = 0,7$
Bayes A	0,2028	0,2849	0,4381
Bayes B	0,2178	0,2855	0,4663
Bayes C	0,2112	0,2784	0,4602
Bayes LASSO	0,2457	0,3042	0,4466
BRR	0,2002	0,2546	0,4320
RKHS	0,1819	0,2467	0,4389
G-BLUP A	0,1695	0,2705	0,6420

Observa-se que para a faixa de herdabilidade paramétrica de 0,3 o modelo que proporcionou a estimativa mais próxima de tal valor foi o Bayes LASSO, assim como para a herdabilidade paramétrica de 0,5, só que nesse caso a estimativa se apresentou bem mais distante do valor simulado.

Quanto mais próximo do valor paramétrico usado na simulação a estimativa apresentar-se mais o modelo conseguirá recuperar a informação genética “disponível” entre os dados. Isso é outra medida da capacidade preditiva dos modelos, além da acurácia, embora pouco discutida na literatura de predição genômica.

Na faixa de herdabilidade de 0,7 o modelo G-BLUP apresentou uma estimativa destacadamente superior aos demais. Interessante notar que na menor faixa de herdabilidade esse foi o modelo que teve a pior estimativa, demonstrando ter se beneficiado mais que os outros com o aumento da herdabilidade do caráter. Não há na literatura relatos de que um modelo em específico desempenhe-se bem apenas para determinadas faixas de herdabilidade. O que é geralmente investigado pela pesquisa é o comportamento dos modelos diante de cenários que incluem diferentes naturezas de efeitos genéticos, como foi a tônica principal desse trabalho também.

Na Tabela 11 são representadas as estimativas de herdabilidade ao incluir efeitos não aditivos na predição dos modelos para a herdabilidade genética simulada de 0,3.

Tabela 11 Estimativas de herdabilidade para os modelos Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR), G-BLUP dominante (G-BLUP AD) e G-BLUP epistático (G-BLUP EP) na condição de modelos incluindo efeitos não aditivos para a herdabilidade paramétrica de 0,3

Modelos	Natureza das herdabilidades					Genética total
	Aditiva	Dominante	Epistáticas*			
			A-A	A-D	D-D	
Bayes A	0,2047	0,1402	-	-	-	0,3449
Bayes B	0,2083	0,1374	-	-	-	0,3457
Bayes C	0,2025	0,1255	-	-	-	0,3280
Bayes LASSO	0,2055	0,1417	-	-	-	0,3472
BRR	0,1876	0,1050	-	-	-	0,2926
G-BLUP AD	0,1818	0,0809	-	-	-	0,2627
G-BLUP EP	0,1755	0,0717	0,0104	0,0051	0,0047	0,2674

* epistasias aditiva-aditiva (A-A), aditiva-dominante (A-D) e dominante-dominante (D-D)

Observa-se pela Tabela 11 que para todos os modelos a variância aditiva contabilizou a maior porção. Para o modelo G-BLUP (EP), o único a considerar epistasias, as variâncias epistáticas foram de magnitude bem pequenas como esperado, já que são efeitos multiplicativos entre locos. A herdabilidade para a variância genética total foi melhor estimada pelo modelo BRR e pior estimada pelo modelo Bayes LASSO, diferente do ocorrido para a arquitetura genética com apenas efeitos aditivos para essa mesma faixa de herdabilidade em que esse modelo superou os demais.

Vale destacar o comportamento dos modelos G-BLUP AD e G-BLUP EP que demonstraram um expressivo acréscimo na qualidade da predição comparado ao G-BLUP A (Tabela 10), na herdabilidade de 0,3.

O mesmo comportamento com relação à participação da variância aditiva na variância genética total foi observado na herdabilidade paramétrica de 0,5 (Tabela 12). A classificação dos modelos agora inverteu-se com o modelo BRR apresentando as piores estimativas de herdabilidade

genética e os modelos G-BLUP pouco diferindo entre si, mas superando largamente os demais.

Tabela 12 Estimativas de herdabilidade para os modelos Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR,) G-BLUP dominante (G-BLUP AD) e G-BLUP epistático (G-BLUP EP) na condição de modelos incluindo efeitos não aditivos para a herdabilidade paramétrica de 0,5

Modelos	Natureza das herdabilidades					Genética total
	Aditiv a	Dominan te	Epistáticas*			
			A-A	A-D	D-D	
Bayes A	0,2074	0,1214	-	-	-	0,3288
Bayes B	0,1970	0,1566	-	-	-	0,3536
Bayes C	0,2064	0,1117	-	-	-	0,3181
Bayes LASSO	0,2147	0,1319	-	-	-	0,3466
BRR	0,1886	0,1163	-	-	-	0,3049
G-BLUP AD	0,2372	0,2257	-	-	-	0,4629
G-BLUP EP	0,2229	0,2015	0,0242	0,0143	0,0111	0,4739

* epistasias aditiva-aditiva (A-A), aditiva-dominante (A-D) e dominante-dominante (D-D)

Na maior faixa de herdabilidade simulada (Tabela 13) as diferenças entre as estimativas da herdabilidade genética total diminuiram e novamente os modelos BRR e Bayes LASSO voltaram a se destacar, porém com vantagem bem estreita em relação aos modelos G-BLUP.

Tabela 13 Estimativas de herdabilidade para os modelos Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayesian Ridge Regression (BRR,) G-BLUP dominante (G-BLUP AD) e G-BLUP epistático (G-BLUP EP) na condição de modelos incluindo efeitos não aditivos para a herdabilidade paramétrica de 0,7

Modelos	Natureza das herdabilidades					Genética total
	Aditiva	Dominante	Epistáticas*			
			A-A	A-D	D-D	
Bayes A	0,4133	0,3183	-	-	-	0,7316
Bayes B	0,4126	0,3204	-	-	-	0,7330
Bayes C	0,4145	0,2937	-	-	-	0,7082
Bayes LASSO	0,4148	0,2827	-	-	-	0,6975
BRR	0,4085	0,2912	-	-	-	0,6997
G-BLUP AD	0,4116	0,2913	-	-	-	0,7029
G-BLUP EP	0,3796	0,2473	0,0731	0,0062	0,0058	0,7121

* epistasias aditiva-aditiva (A-A), aditiva-dominante (A-D) e dominante-dominante (D-D)

Isso indica que todos os modelos beneficiaram-se de forma semelhante nesse cenário mais favorável de efeitos genéticos, diminuindo as discrepâncias, e logo, dificultando a discriminação dos mesmos, levando em conta esse critério.

O bom desempenho geral dos modelos G-BLUP para as estimativas de herdabilidade genética, com exceção para o G-BLUP A na herdabilidade de 0,3, reforça o ponto forte desses modelos em discriminar bem as estimativas de efeitos genéticos.

Outra observação é que a análise dos modelos pelo critério de herdabilidade das estimativas de efeitos genéticos não concordou inteiramente com a análise feita pelo critério de acurácia preditiva, evidenciando que, de fato, contribui com informação complementar para a escolha dos modelos, dependendo dos objetivos de seu uso.

5 CONCLUSÕES

Ocorre aumento na acurácia de predição e redução na Soma de Quadrados dos Erros de Predição (PRESS) dos modelos com o aumento da herdabilidade, mas de forma não consistente. Melhorias em acurácia também ocorrem com a inclusão de efeitos não aditivos.

Para a população e caráter considerados no trabalho os efeitos de epistasias não afetam as estimativas de acurácia de predição e a correlação entre os efeitos genéticos simulados e estimados.

A inclusão de efeitos não aditivos aos modelos G-BLUP melhora as estimativas da variância genética e atribuída aos efeitos aditivos.

Ao incluir efeitos genéticos não aditivos o G-BLUP não é superado em acurácia de predição pelos modelos bayesianos.

A inclusão de efeitos não aditivos aos modelos melhora as estimativas de herdabilidade dos efeitos genéticos.

REFERÊNCIAS

- ALBRECHT, T. et al. Genome-based prediction of testcross values in maize. **Theoretical and Applied Genetics**, Berlin, v. 123, n. 2, p. 339-350, 2011.
- ARNHOLD, E. et al. Prediction of genotypic values of maize for the agricultural frontier region in northeastern Maranhão, Brazil. **Crop Breeding and Applied Biotechnology**, Londrina, v. 12, n. 2, p. 151-155, 2012.
- ASHRAF, M.; FOOLAD, M. Crop breeding for salt tolerance in the era of molecular markers and marker-assisted selection. **Plant Breeding**, Berlin, v. 132, n. 1, p. 10–20, 2013.
- BALESTRE, M. et al. Applications of multi-trait selection in common bean using real and simulated experiments. **Euphytica**, Wageningen, v. 189, n. 2, p. 225-238, 2013.
- BALESTRE, M.; VON PINHO, R. G.; SOUZA, J. C. Prediction of maize double-cross hybrids using the best linear unbiased prediction with microsatellite marker information. **Genetics and Molecular Research**, Ribeirão Preto, v. 10, n. 1, p. 25-35, 2011.
- BERNARDO, R. Genome wide Selection when Major Genes Are Known. **Crop Science**, Madison, v. 54, n. 1, p. 68-75, 2014.
- BERNARDO, R. Prediction of Maize Single-Cross performance using RFLPs and information from related hybrids. **Crop Science**, Madison, v. 34, n. 1, p. 20-25, 1994.
- BERNARDO, R. Testcross additive and dominance effects in best linear unbiased prediction of maize single-cross performance. **Theoretical and Applied Genetics**, Berlin, v. 93, n. 7, p. 1098-1102, 1996.
- BERNARDO, R.; YU, J. Prospects for Genomewide Selection for Quantitative Traits in Maize. **Crop Science**, Madison, v. 47, n. 3, p. 1082-1090, 2007.
- BRADBURY, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, Oxford, v. 23, n. 19, p. 2633-2635, 2007.
- BRUCE, A. B. The Mendelian theory of heredity and the augmentation of vigor. **Science**, Washington, v. 32, p. 627-628, 1910.

- CAIXETA, D. S. et al. Relationship between heterosis and genetic divergence for phosphorus use efficiency and its components in tropical maize. **Ciência Rural**, Santa Maria, v. 43, n. 1, p. 60-65, 2013.
- CALINSKI, T. et al. Analyzing multi-environment variety trials using randomization-derived mixed models. **Biometrics**, Washington, v. 61, n. 2, p. 448-455, 2005.
- CHARMET, G.; STORLIE, E. Implementation of Genome Wide Selection in Wheat. **Russian Journal of Genetics**, New York, v. 2, n. 4, p. 298-303, 2012.
- COCKERHAM, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. **Genetics**, Austin, v. 39, n. 6, p. 859, 1954.
- COLLARD, B. C. Y.; MACKILL, D. J. Marker-assisted selection: an approach for precision plantbreeding in the twenty-first century. **Philosophical Transactions of the Royal Society B: Biological**, London, v. 363, n. 1491, p. 557-572, 2008.
- COLVIN, J. A.; DYKSTRA, R. L. Maximum likelihood estimation of a set of covariance matrices under Lowner order restrictions with applications to balanced multivariate variance components models. **The Annals of Statistics**, Hayward, v. 19, p. 850-869, 1991.
- CROSSA, J. et al. Genomic prediction in CIMMYT maize and wheat breeding programs. **Heredity**, London, v. 112, n. 1, p. 48-60, 2014.
- CROSSA, J. et al. Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. **G3: Genes/Genomes/Genetics**, Bethesda, v. 3, n. 11, p. 1903-1926, 2013.
- CROSSA, J. et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, Austin, v. 186, n. 2, p. 713-724, 2010.
- CROW, J. F. Mid-Century controversies in population genetics. **Annual Review of Genetics**, Palo Alto, v. 42, p. 1-16, 2008.
- DAETWYLER, H. D. et al. Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. **Genetics**, Austin, v. 193, n. 2, p. 347-365, 2013.

- DE LOS CAMPOS, G. et al. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. **Genetics Research**, Cambridge, v. 92, n. 4, p. 295-308, 2010.
- DE LOS CAMPOS, G. et al. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. **Genetics**, Austin, v. 193, n. 2, p. 327-345, 2013.
- DEMPSTER, A. P. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society**, London, v. 39, n. 1, p. 1-38, 1977.
- DESCHAMPS, S.; LLACA, V.; MAY, G. D. Genotyping-by-sequencing in plants. **Journal of Biology**, Santiago, v. 1, n. 3, p. 460-483, 2012.
- DESTA, Z. A.; ORTIZ, R. Genomic selection: genome-wide prediction in plant improvement. **Trends in Plant Science**, Oxford, v. 19, n. 9, p. 592-601, 2014.
- DONÁ, S. et al. Heterose e seus componentes em híbridos de populações F2 de milho. **Bragantia**, Campinas, v. 70, n. 4, p. 767-774, 2011.
- DUDLEY, J. W.; JOHNSON, G.R. Epistatic models and pre-selection of markers improve prediction of performance in corn. **Molecular Breeding**, Dordrecht, v. 32, n. 3, p. 585-593, 2013.
- EAST, E. M. The distinction between development and heredity in breeding. **American Naturalist**, Chicago, v. 43, p. 173-181, 1909.
- FALCONER, D. S.; MACKAY, T. F. **Introduction to quantitative genetics**. 4th ed. London: Longman, 1996. 464 p.
- FERREIRA, D. V. et al. Prediction of maize hybrid performance using similarity in state and similarity by descent information. **Genetics and Molecular Research**, Ribeirão Preto, v. 9, n. 4, p. 2381-2394, 2010.
- FRITSCH NETO, R. et al. Prediction of genotypic values of maize hybrids in unbalanced experiments. **Crop Breeding and Applied Biotechnology**, Londrina, v. 10, n. 1, p. 32-39, 2010.
- GARCIA, A. A. F. et al. Quantitative Trait Loci Mapping and The Genetic Basis of Heterosis in Maize and Rice. **Genetics**, Austin, v. 180, n. 3, p. 1707-1724, 2008.

GIANOLA, D.; MOROTA, G.; CROSSA, J. Genome-enabled prediction of complex traits with kernel methods: what have we learned? In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 10., 2014, Vancouver. **Proceedings...** Vancouver: [s. n], 2014. 1 CD ROM.

GODDARD, M. E.; HAYES, B. J. Genomic selection. Journal compilation. **Journal of Animal Breeding and Genetics**, Berlin, v. 124, n. 6, p. 323-330, 2007.

GODDARD, M. E.; HAYES, B. J.; MEUWISSEN, T. H. E. Genomic selection in livestock populations. **Genetics Research**, Cambridge, v. 92, n. 5/6, p. 413-421, 2010.

GUIMARÃES, P. S. et al. Correlação da heterose de híbridos de milho com divergência genética entre linhagens. **Pesquisa Agropecuária Brasileira**, Brasília, v. 42, n. 6, p. 811-816, 2007.

GUO, Z. et al. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. **Theoretical and Applied Genetics**, Berlin, v. 124, n. 2, p. 261-275, 2012.

GUO, T. et al. Performance prediction of F1 hybrids between recombinant inbred lines derived from two elite maize inbred lines. **Theoretical and Applied Genetics**, Berlin, v. 126, n. 1, p. 189-201, 2013.

HEFFNER, E. L.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection for crop improvement. **Crop Science**, Madison, v. 49, n. 1, p. 1-12, 2009.

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

HENDERSON, C. R. Best linear unbiased prediction of non additive genetic merits in non inbreeds populations. **Journal of Animal Science**, Champaign, v. 60, n. 1, p. 111-117, 1985a.

HENDERSON, C. R. MIVQUE and REML estimation of additive and nonadditive genetic variances. **Journal of Animal Science**, Champaign, v. 61, n. 1, p. 113-121, 1985b.

HENDERSON, C. R. Prediction of the merits of single crosses. **Theoretical and Applied Genetics**, Berlin, v. 49, n. 6, p. 273-282, 1977.

HESLOT, N. et al. Genomic selection in plant breeding: a comparison of models. **Crop Science**, Madison, v. 52, n. 1, p. 146-160, 2012.

- HOWARD, R.; CARRIQUIRY, A. L.; BEAVIS, W. D. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. **G3:Genes/Genomes/Genetics**, Bethesda, v. 4, n. 6, p. 1027-1246, 2014.
- JENA, K. K.; MACKILL, D. J. Molecular Markers and Their Use in Marker-Assisted Selection in Rice. **Crop Science**, Madison, v. 48 n. 4, p. 1266-1276, 2008.
- JENKINS, M.T. Methods of estimating the performance of double crosses in corn. **Journal of the American Society of Agronomy**, Washington, v. 26, p. 199-204, 1934.
- JIA, Y.; JANNINK, J. L. Multiple-Trait genomic selection methods increase genetic value prediction accuracy. **Genetics**, Austin, v. 192, n. 4, p. 1513-1522, 2012.
- KAO, C. H.; ZENG, Z. B. Modeling epistasis of quantitative trait loci using Cockerham's model. **Genetics**, Austin, v. 160, n. 3, p. 1243-1261, 2002.
- LARIÈPE, A. 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**, Austin, v. 190, n. 2, p. 795-811, 2012.
- LIPPMAN, Z. B.; ZAMIR, D. Heterosis: revisiting the magic. **Trends in Genetics**, London, v. 23, n. 2, p. 60-66, 2007.
- LORENZANA, R. E.; BERNARDO, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 1, p. 151-161, 2009.
- MCMULLEN, M. D. et al. Genetics properties of the nested association mapping population. **Science**, Washington, v. 325, n. 5941, p. 737-740, 2009.
- MELCHINGER, A. E. et al. Optimum prediction of three-way crosses from single crosses in forage maize (*Zea mays* L.). **Theoretical and Applied Genetics**, Berlin, v. 74, n. 3, p. 339-345, 1987.
- MELCHINGER, A. E.; GUMBER, R. K. Overview of heterosis and heterotic groups in agronomic crops. In: LARNKEY, K. R.; STAUB, J. E. (Ed.). **Concepts and breeding of heterosis in crop plants**. Madison: CSSA, 1998. p. 29-44.

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

MOROTA, G.; GIANOLA, D. Kernel-based whole-genome prediction of complex traits: a review. **Frontiers in Genetics**, Lausanne, v. 5, p. 56, 2014.

NISHIO, M.; SATOH, M. Including Dominance Effects in the Genomic BLUP Method for Genomic Evaluation. **PloS one**, San Francisco, v. 9, n.1, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24416447>>. Acesso em: 22 ago. 2014.

PANZEA. **Data sets for download**. 2014. Disponível em: <http://www.panzea.org/lit/data_sets.html#NAM_map>. Acesso em: 22 ago. 2014.

PATERNIANI, M. E. A. G. Z. et al. Capacidade combinatória, divergência genética entre linhagens de milho e correlação com heterose. **Bragantia**, Campinas, v. 67, n. 3, p. 639-648, 2008.

PATTERSON, H. D.; THOMPSON, R. Recovery of Inter-Block information when Block Sizes are Unequal. **Biometrika**, London, v. 58, n. 3, p. 545-554, 1971.

PÉREZ, P.; DE LOS CAMPOS, G. Genome-Wide Regression & Prediction with the BGLR Statistical Package. **Genetics**, Austin, v. 114, p. 164-442, 2014.

PÉREZ-RODRÍGUEZ, P. et al. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. **G3.Genes/Genomes/Genetics**, Bethesda, v. 2, n. 12, p. 1595-1605, 2012.

POWELL, R. L.; NORMAN, H. D. Major Advances in Genetic Evaluation Techniques. **Journal of Dairy Science**, Champaign, v. 89, n. 4, p. 1337-1348, 2006.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <<http://www.Rproject.org>> Acesso em: 2 jun. 2014.

REIF, J. C.; HALLAUER, A. R.; MELCHINGER, A. E. Heterosis and heterotic patterns in maize. **Maydica**, Bergamo, v. 50, n. 3/4, p. 215, 2005.

RIEDELSEIMER, C.; TECHNOW, F.; MELCHINGER, A. E. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. **BMC Genomics**, London, v. 13, n. 1, p. 452, 2012.

RINALDI, D. A. et al. Correlação entre heterose e divergência genética estimadas por cruzamentos dialélicos e marcadores moleculares RAPD em populações de milho-pipoca. **Bragantia**, Campinas, v. 66, n. 2, p. 183-192, 2007.

ROMAY, M. C. et al. Comprehensive genotyping of the USA national maize inbred seed bank. **Genome Biology**, London, v. 14, n. 6, p. 55, 2013.

SCHRAG, T. A. et al. Marker-based prediction of hybrid performance in maize single-crosses envolvendo duplo-haplides. **Maydica**, Bergamo, v. 54, n. 2, p. 353, 2009.

SHULL, G. H. The composition of a field of maize. **Report of American Breeders' Association**, Washington, v. 4, p. 296-301, 1908.

SPRINGER, N. M.; STUPAR, R. M. Allelic variation and heterosis in maize: How do two halves make more than a whole? **Genome Research**, Cold Spring Harbor, v. 17, n. 3, p. 264-275, 2007.

SU, G. et al. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. **PloS one**, San Francisco, v. 7, n. 9, 2012. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045293#s1>>. Acesso em: 22 ago. 2014.

SUN, X.; MA, P.; MUMM, R. H. Nonparametric Method for Genomics-Based Prediction of Performance of Quantitative Traits Involving Epistasis in Plant Breeding. **PloS one**, San Francisco, v. 7, n. 11, 2012. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0050604>>. Acesso em: 21 ago. 2014.

TECHNOW, F. et al. Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. **Theoretical and Applied Genetics**, Berlin, v. 125, n. 6, p. 1181-1194, 2012.

TECHNOW, F. et al. Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. **Genetics**, Austin, v. 197, n. 4, p. 1343-1355, 2014.

TOLLENAAR, M.; LEE, E. A. Dissection of physiological processes underlying grain yield in maize by examining genetic improvement and heterosis. **Maydica**, Bergamo, v. 51, n. 2, p. 399, 2006.

VIANA, L. F. et al. Predição de médias de linhagens de HS de milho(zea mays). **Ciência e Agrotecnologia**, Lavras, v. 33, p. 1999-2004, 2009. Especial.

VITEZICA, Z. G.; VARONA, L.; LEGARRA, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. **Genetics**, Austin, v. 195, n. 4, p. 1223-1230, 2013.

XU, Y.; CROUCH, J. H. Marker-Assisted Selection in Plant Breeding: From Publications to Practice. **Crop Science**, Madison, v. 48, n. 2, p. 391-407, 2008.

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