



NARJARA FONSECA CANTELMO

**CARACTERIZAÇÃO E ASSOCIAÇÃO
GENÉTICA DE LINHAGENS E PREDIÇÃO
GENÔMICA DA PRODUTIVIDADE DE GRÃOS
DE HÍBRIDOS DE MILHO UTILIZANDO
MARCADORES DARTS-GBS**

LAVRAS – MG

2016

NARJARA FONSECA CANTELMO

**CARACTERIZAÇÃO E ASSOCIAÇÃO GENÉTICA DE LINHAGENS E
PREDIÇÃO GENÔMICA DA PRODUTIVIDADE DE GRÃOS DE
HÍBRIDOS DE MILHO UTILIZANDO MARCADORES DARTS-GBS**

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

Orientador

Dr. Renzo Garcia Von Pinho

Coorientador

Dr. Márcio Balestre

LAVRAS – MG

2015

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

Cantelmo, Narjara Fonseca.

Caracterização e associação genética de linhagens e predição genômica da produtividade de grãos de híbridos de milho utilizando marcadores DARTS-GBS / Narjara Fonseca Cantelmo. – Lavras : UFLA, 2016.

128 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2015.

Orientador: Renzo Garcia Von Pinho.

Bibliografia.

1. Melhoramento Genético. 2. Milho. 3. Predição Genômica Ampla. I. Universidade Federal de Lavras. II. Título.

NARJARA FONSECA CANTELMO

**CARACTERIZAÇÃO E ASSOCIAÇÃO GENÉTICA DE LINHAGENS E
PREDIÇÃO GENÔMICA DA PRODUTIVIDADE DE GRÃOS DE
HÍBRIDOS DE MILHO UTILIZANDO MARCADORES DARTS-GBS**

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

APROVADA em 06 de Novembro de 2015

Dr. Dr. Marcio Balestre	UFLA
Dr. João Cândido de Souza	UFLA
Dra. Flavia Barbosa Silva Botelho	UFLA
Dr. Lauro J. M. Guimarães	EMBRAPA Milho e Sorgo

Dr. Renzo Garcia Von Pinho
Orientador

LAVRAS – MG

2015

*Aos meus pais Helder e Cidinha, pelo apoio, exemplo e ajuda. Sem eles
nada disso seria possível.*

A minha filha, Maria Antônia, pela compreensão, amor e carinho.

*Ao meu marido, Luciano, por caminhar ao meu lado e por torcer pelas
minha vitórias.*

Aos meus amigos, pela força e pelo apoio.

Aos meus familiares, pela torcida e amizade.

*A Deus pelo dom da vida, pelas oportunidades e por colocar tantas
pessoas especiais em meu caminho.*

DEDICO

AGRADECIMENTOS

À Universidade federal de Lavras e, em especial, ao Departamento de Biologia – Genética e Melhoramento de Plantas, por todo suporte e por tornar real o meu sonho de me tornar Doutora.

Ao professor Renzo Garcia Von Pinho, pela orientação durante todos estes anos, pelos ensinamentos e por acreditar no meu potencial.

Ao professor Marcio Balestre, pela coorientação, por todo suporte, pelos exemplos de competência e pela amizade.

À Capes e CNPq pela bolsa de doutorado no Brasil e nos Estados Unidos, durante meu período de vivência internacional.

Aos alunos de iniciação científica, que me ajudaram durante a condução deste trabalho.

Ao Núcleo de Estudos em Milho e Sorgo (G-Milho) e ao Grupo do Milho pelo convívio e aprendizagem.

As amigas de caminhada no doutorado e na vida, em especial Sheila Roberta Guilherme, Lais Pererira, Marcela Carvalho, Jessica Gentil e Cinthia Rodrigues.

Aos funcionários do Departamento de Biologia, em especial Rafaela, D. Ironcina, Zelia e Lilian pelo suporte, convívio e amizade.

Aos professores do Departamento de Biologia, em especial Prof. Magno, Prof. João Cândido, Prof. Cesar, Prof. José Ayrton e Prof. Flavia pelos ensinamentos e por serem exemplos de profissionalismo.

RESUMO

A cultura do milho é extremamente importante para a alimentação humana e animal em todo o mundo. Por isso, seu melhoramento, visando híbridos mais produtivos, é essencial frente à crescente demanda por alimentos. Para tanto, faz-se necessário o profundo conhecimento da coleção de linhagens usadas no melhoramento e o conhecimento das melhores combinações entre elas. Assim, na primeira parte deste trabalho, objetivou-se a exploração das informações obtidas a partir da genotipagem de um banco de linhagens proveniente de uma empresa de sementes visando o detalhamento do mesmo; a avaliação da diversidade genética e a estrutura populacional da coleção e a investigação do potencial da coleção como fonte para estudo da arquitetura genética do caráter peso de espigas. Na segunda parte do trabalho, objetivou-se a predição genômica ampla utilizando um conjunto de marcadores Darts-seq e o modelo GBLUP com dominância para produtividade de grãos de híbridos simples de milho avaliados em diferentes safras e locais. Foi realizada extração de DNA e posterior genotipagem dos acessos da coleção. Foi realizado o agrupamento genético das linhagens endogâmicas e a análise de associação genômica para o caráter peso de espigas. Do cruzamento das linhagens foram obtidos 838 híbridos simples avaliados em seis locais na safra de inverno do ano agrícola de 2013 e 797 híbridos simples avaliados em quatro locais na safra de verão do ano agrícola de 2013/2014. Para a predição da produção de grãos, foi utilizado o modelo GBLUP com dominância e validação cruzada utilizando diferentes níveis de desbalanceamento. Também foram utilizados os híbridos em comum nas duas safras para cálculo das correlações entre os mesmos. Observou-se que o banco de germoplasma não apresentou frequências alélicas fixadas, evidenciado assim o potencial das linhagens para seleção e melhoramento. Foi possível fazer a separação das linhagens em grupos heteróticos distintos, sendo, em sua maioria, condizente com as informações de *pedigree*. Contudo, o conjunto de dados utilizado para o estudo de associação não foi adequado para a identificação de marcas de efeitos significativos para o caráter peso de espigas. Houve coincidência entre VGGs nas safras de verão e de inverno. O método GBLUP foi capaz de gerar elevadas correlações entre híbridos preditos e observados, mesmo em elevados níveis de desbalanceamento e em diferentes locais e safras.

Palavras-chave: Melhoramento Vegetal. Diversidade Genética. QTLs. Seleção Genômica. GBLUP.

ABSTRACT

Corn culture is extremely important for human and animal feeding in the entire world. Thus, its breeding, aiming at obtaining more productive hybrids, is essential when facing increasing demands for food. For this, it is necessary to understand in depth the strain collection used for breeding and the best combinations between them. Therefore, in the first part of this work, we aimed at exploring the information obtained from genotyping a germplasm bank derived from a seed company, aiming at its detailing; at evaluating the genetic diversity and population structure of the collection; and investigate the potential of the collection as source for studying the genetic architecture of corn ear weight. In the second part of this work, the objective was the genome wide selection using a set of Darts-seq markers and the GLUP model with dominance for grain productivity of corn hybrids evaluated in different seasons and locations. We conducted DNA extraction and posterior genotyping of accesses from the collection. We performed genetic grouping of the inbred lines of the collection and genomic association analysis of corn ear weight. From the crosses of the inbred lines, 838 hybrids were obtained, evaluated at six locations in the winter season of 2013, and 797 hybrids, evaluated at four locations in the summer season of 2013/2014. We also used the common hybrids from both seasons to calculate the correlations between them. We verified that the germplasm bank presented no fixed allele frequencies, demonstrating the selection and breeding potential of the genotypes. It was possible to separate the inbred lines into distinct heterotic groups, with the majority consistent with the pedigree information. However, the set of data used for the association study was inadequate in identifying significant effect marks for corn ear weight. Coincidences occurred between the VGGs of the summer and winter seasons. The GBLUP method was capable of generating high correlations between predicted and observed hybrids, even with high levels of unbalance and different locations and years.

Keywords: Plant Breeding. Genetic Diversity. QTLs. Genomic Selection. GBLUP.

LISTA DE FIGURAS

SEGUNDA PARTE – ARTIGOS

ARTIGO 1

- Figura 1 Distribuição dos marcadores Darts a cada 100 kb ao longo de nove grupos de ligação nas 470 genótipos de milho oriundas do banco de linhagens de uma empresa de produção e melhoramento de milho.....58
- Figura 2 Distribuição da frequência alélica dos marcadores Darts ao longo dos nove grupos de ligação e também no DNA plasmídico e mitocondrial presentes no conjunto de linhagens de milho.....60
- Figura 3 a) *Call Rate* nos diferentes genótipos para os marcadores Darts (linha azul) com sua respectiva média (linha preta). b) Na segunda parte, são apresentados os *call rates* dos grupos heteróticos individualmente: A: SSS (variância=0.0021); B: Amarillo Dentado (variância = 0,0004); C: Fêmea DK-Arg (variância = 0,0007); D: Iodente (variância = 0,0039); E: Iodente (variância = 0,0033); F: Lancaster (variância = 0,0037); G: Macho NK-Arg (variância = 0,0054); H: NSSS (variância = 0,0016); I: SSS (variância = 0,0030); J: Suwan (variância = 0,0021).....62
- Figura 4 Decréscimo do desequilíbrio de ligação (DL) entre as linhagens de milho para os marcadores Darts nos nove grupos de ligação genotipados64
- Figura 5 Distribuição da frequência dos alelos idênticos por estado (IBS) entre as 447 linhagens do banco de acesso utilizado65

Figura 6	Separação dos grupos heteróticos do banco de germoplasma do programa de melhoramento de milho do Departamento de Agricultura – UFLA, pela decomposição espectral da matriz de parentesco A e plotagem dos dois primeiros componentes principais, que absorveram 18% da variância	67
Figura 7	Separação dos grupos heteróticos do banco de linhagens de uma empresa de melhoramento e produção de sementes de milho pela análise de agrupamento hierárquico.....	68
Figura 8	Efeitos das marcas para peso de espigas utilizando o método RRBLUP, com a linha de significância pelo teste de Wald (3,84).....	71
Figura 9	Efeitos das marcas para peso de espigas utilizando o método SSV, com a linha de significância pelo teste de Wald (3,84).....	71

ARTIGO 2

Figura 1	Regressão entre VGGs preditos na safra de inverno e médias ajustadas para efeitos fixos pelo método REML observadas na safra de verão para o caráter produtividade de grão de híbridos de milho.....	109
Figura 2	Agrupamento das linhagens dentro dos grupos de origem e correlação entre os VGGs preditos na safra de inverno e médias observadas na safra de verão	111

LISTA DE TABELAS

SEGUNDA PARTE – ARTIGOS

ARTIGO 1

Tabela 1	Número médio de marcas por grupo de ligação e seus respectivos comprimentos parciais presentes no conjunto de linhagens de milho	59
----------	--	----

ARTIGO 2

Tabela 1	Componentes de variância genotípico (σ_g^2), interação genótipos por ambientes ($\sigma_{g \times e}^2$) e residual (σ_r^2) e herdabilidade (h_a^2) e seus respectivos erros padrão.....	104
Tabela 2	Distribuição das marcas por grupo de ligação, mitocôndrias e plastídios das linhagens utilizando 100% e 50% dos marcadores Darts utilizados para predição da produtividade de grãos dos híbridos simples de milho pelo método GBLUP.....	105
Tabela 3	Correlação entre VGGs preditos e observados ($r_{\hat{g}g}$) com diferentes níveis de desbalanceamento e médias das variância aditiva (VA), variância de dominância (VD), variância residual (VE) e erros padrão para a safra de inverno (K-Fold)....	106
Tabela 4	Correlação entre VGGs preditos e observados ($r_{\hat{g}g}$) com diferentes níveis de desbalanceamento e médias das variância aditiva (VA), variância de dominância (VD), variância residual (VE) e erros padrão para a safra de verão (K-Fold)	108

SUMÁRIO

	PRIMEIRA PARTE	
1	INTRODUÇÃO E JUSTIFICATIVAS	11
2	REFERENCIAL TEÓRICO	15
2.1	Heterose em milho	15
2.2	Grupos Heteróticos e Diversidade	18
2.3	Predição do desempenho de híbridos de milho	23
2.4	Modelo GBLUP	28
	REFERÊNCIAS	33
	SEGUNDA PARTE - ARTIGOS	45
	ARTIGO 1 Análise genômica e associação genética entre linhagens de milho	45
1	INTRODUÇÃO	47
2	MATERIAL E MÉTODOS	51
2.1	Caracterização genética do banco de germoplasma	51
2.2	Obtenção e Análise dos dados Fenotípicos	52
2.3	Análise de Associação Genômica	53
2.3.1	Modelo GWAS	53
2.3.2	Modelo Bayesian Stochastic Search Variable (BSSV)	53
3	RESULTADOS	57
4	DISCUSSÃO	72
5	CONCLUSÕES	81
	REFERÊNCIAS	83
	ARTIGO 2 Predição genômica ampla para produtividade de grãos de híbridos simples de milho utilizando modelo GBLUP	91
1	INTRODUÇÃO	93
2	MATERIAL E MÉTODOS	96
2.1	Obtenção dos Híbridos Simples e Avaliação fenotípica	96
2.2	Genotipagem e Construção da Matriz de Marcadores	97
2.3	Predição Utilizando o Modelo GBLUP	98
2.4	Validação cruzada	101
2.5	Agrupamento de grupos heteróticos	102
3	RESULTADOS	103
4	DISCUSSÃO	112
5	CONCLUSÃO	121
	REFERÊNCIAS	122

PRIMEIRA PARTE

1 INTRODUÇÃO E JUSTIFICATIVAS

A produção de milho fornece alimentos para uma imensa população animal e combustível para mais de seis bilhões de pessoas no mundo e a população mundial continua a crescer em face do aumento dos preços de comida, diminuição das áreas agricultáveis, escassez de água, ameaça do aquecimento global e aumento dos níveis de dióxido de carbono (TROYER; WEELIN, 2009). Diante da demanda por alimentos e combustíveis, são necessários programas de melhoramento de milho mais eficientes e mais ágeis para suprir estas necessidades.

Neste contexto, o conhecimento da diversidade existente na coleção de germoplasma é importante para a pesquisa e melhoramento de culturas como o milho (XU, 2010; XU; ISHII; MCCOUCH, 2003; MELCHINGER et al., 1990; BERNARDO, 2010; HALLAUER; CARENA; MIRANDA, 2010; TROYER, 1990). A diferenciação de cultivares e a classificação das linhagens em grupos heteróticos; identificação de falhas e redundâncias nas coleções; monitoramento de perdas que possam ocorrer durante a conservação do banco de germoplasma, regeneração, domesticação e melhoramento; identificação de novos alelos que possam ter ligação com caracteres agronômicos, construção de conjuntos representativos ou coleções núcleo (LU et al, 2009) e modelos de seleção genômica aplicadas à população são informações importantes a serem conhecidas e estudadas nas coleções de germoplasma.

Uma das aplicações da análise de diversidade genética em germoplasma de milho é a classificação de acessos dentro de grupos heteróticos. Linhagens dentro de um mesmo grupo devem ter um alto nível de similaridade e alto nível de dissimilaridade entre grupos (LU et al., 2009). O conhecimento da

diversidade genética, da relação e da estrutura do germoplasma é importante no melhoramento de plantas por diferentes razões, entre elas para o auxílio na escolha das melhores combinações parentais para o desenvolvimento de progênies com máxima variabilidade genética (BARRETT; KIDWELL, 1998) e desempenho superior. No melhoramento de milho, esse conhecimento é essencial para o desenvolvimento de melhores linhagens e a criação de híbridos fenotipicamente superiores, principalmente para o caráter produtividade de grãos (DUVICK, 2001; HALLAUER; CARENA; MIRANDA, 2010).

Mesmo conhecendo-se os grupos heteróticos, também é necessário o conhecimento dos cruzamentos com maiores potenciais. Nos programas de melhoramento de milho, a seleção e a avaliação dos melhores genótipos são extremamente importantes para a obtenção de híbridos mais produtivos e adaptados a diferentes condições de cultivo (FERREIRA et al., 2010). Assim, a identificação de híbridos superiores é ponto chave para o sucesso do programa. Contudo, as avaliações ao nível de campo de todos os cruzamentos possíveis entre linhagens requerem muitos recursos (SHRAG et al., 2010).

Atualmente, nos programas de melhoramento de milho, existe a separação das populações em grupos heteróticos para extração de híbridos interpopulacionais. É calculada a capacidade específica de combinação (CEC) dessas populações, que é um dos parâmetros utilizados pelo melhorista para escolher os indivíduos para a obtenção de híbridos. Os pares que apresentam maiores capacidades específicas de combinações (CEC) e maiores valores de heterose são considerados os mais promissores na obtenção de híbridos. Normalmente, a capacidade geral de combinação (CGC) e a CEC auxiliam na predição dos melhores cruzamentos (BALESTRE et al., 2008). Porém, o processo de cruzamento entre os grupos heteróticos e verificação das melhores CGC e CEC em experimentos é muito oneroso e demorado.

Nesse contexto e, com o advento dos marcadores genéticos na década de 80, surgiu a ideia do uso de marcadores para o estudo da diversidade genética nos bancos de germoplasma e no auxílio da escolha das melhores combinações sem testá-las ao nível de campo, utilizando primeiramente distância genética para a seleção (LEE; LAMKEY; WOODMAN, 1989; BERNARDO, 1993; CHARCOSSET, 1994; FALCONER; MACKAY, 1990; LANZA et al., 1997; BETRAN et al., 2003; AMORIM; AMORIM; SANTOS, 2006).

Uma alternativa à predição de cruzamentos via distância genética foi a utilização de modelos mistos associados à informação de *pedigree* (BERNARDO, 1994). Os marcadores moleculares nesse caso apresentam estimativas mais acuradas do parentesco quando a informação de *pedigree* é escassa ou inexistente.

Com a diminuição do custo da genotipagem, tornou-se então possível a utilização de milhares de marcas (BERNARDO; YU, 2007), cobrindo todo o genoma, aumentando a quantidade de informações genéticas para a predição dos híbridos. Surgiu assim o desafio de incorporar essa grande quantidade de informações disponíveis aos modelos estatísticos (CAMPOS et al., 2009).

Visando solucionar esse problema, Meuwissen, Hayes e Goddard (2001), por meio de simulações, foram os pioneiros na incorporação de uma grande quantidade de marcas ao modelo, estimando todas simultaneamente. Assim, a Seleção Genômica Ampla (GWS) incorpora todas as marcas ao longo do genoma no modelo de predição simultaneamente, reduzindo o risco da identificação errônea ou imprecisa dos QTLs de efeitos menores. Cada marca é considerada um possível QTL e todas as marcas são combinadas para predizer o valor genético do genótipo (GUO et al., 2013).

Entre os modelos utilizados para a GWS, o GBLUP (Genomic Best Linear Unbiased Prediction), proposto por VanRaden (2008), vem sendo amplamente utilizado no melhoramento vegetal (CROSSA et al., 2013). O

GBLUP utiliza a relação genômica para estimar o mérito genético de um indivíduo. Para tal, uma matriz de correlações genéticas é utilizada e estimada a partir de informações de marcadores moleculares. Essa matriz define a covariância entre indivíduos baseada na similaridade ao nível genômico, em contrapartida da similaridade estimada por meio de *pedigree*, sendo assim mais preciso (CROSSA et al., 2013; HAYES et al., 2009; VANRADEN, 2008).

Diante do exposto, objetiva-se com este trabalho a caracterização genética de um banco de germoplasma, bem como a alocação das linhagens em grupos heteróticos distintos e identificação de genes de maiores efeitos para o caráter peso de espiga. Na segunda parte do trabalho, objetivou-se a predição da produtividade de grãos de híbridos simples de milho, utilizando validação cruzada, informações de marcadores Darts e resultados de ensaios conduzidos em vários locais a partir da abordagem GBLUP.

2 REFERENCIAL TEÓRICO

2.1 Heterose em milho

O fenômeno genético da heterose, ou vigor híbrido, vem beneficiando agricultores e fascinando geneticistas por mais de 100 anos (SCHNABLE; SPRINGER, 2013). A primeira citação da superioridade do híbrido com relação aos parentais foi feita por Shull em seu histórico artigo de 1908 (SHULL, 1908). Posteriormente ele propôs a palavra heterose, em uma palestra na Alemanha, em 1914, visando unificar a terminologia para esse fenômeno (SHULL, 1948). A heterose, desde então, revolucionou as práticas agrícolas e a indústria sementeira (SCHNABLE; SPRINGER, 2013).

Dada a importância econômica e científica da heterose, pesquisadores têm usado a genética quantitativa, a fisiologia vegetal e as técnicas moleculares para melhor entender este fenômeno (SCHNABLE; SPRINGER, 2013).

Várias hipóteses têm sido propostas para explicar a heterose como: a dominância (BRUCE, 1910; JONES, 1917) que atribui à heterose um efeito cumulativo de alelos que apresentam dominância completa ou parcial (Jones, 1918); sobredominância (SHULL, 1908; EAST, 1936) que atribui ao genótipo heterozigoto a superioridade em relação a ambos os parentais homozigotos e não apenas à média dos homozigotos (SHULL, 1908; EAST, 1936); epistasia (POWERS, 1944; WILLIAMS, 1959) que seria dada pela interação alélica nos diferentes locos controladores do caráter (POWERS, 1944) e a complementação gênica em que, no cruzamento dos parentais para formação do híbrido, um parental possui genes que o outro parental não possui, assim, a heterose se daria por meio dessa complementação (SCHNABLE; FREELING, 2011).

Molecularmente, a heterose envolve uma série de eventos, como modificações na cromatina, no controle transcricional, na tradução e

processamento de proteínas e diferentes interações entre e dentro de rotas bioquímicas. Existem diversos mecanismos moleculares que traduzem o DNA em fenótipo, e são esses mecanismos, entre vários genes, que produzem a heterose em caracteres complexos (KAEPLER, 2012). Mesmo com tantos estudos, as causas genéticas e moleculares da heterose ainda não são totalmente conhecidas (TANG et al., 2010).

O desenvolvimento de marcadores moleculares e os mapas de ligação têm se mostrado ferramentas úteis para a análise das bases genéticas da heterose. A heterose exibe uma distribuição contínua, sendo controlada por vários genes de pequeno efeito, por isso os pesquisadores estudam a heterose por meio dos *quantitative trait loci* (QTL) (TANG et al., 2010). Vários estudos têm sido realizados a fim de elucidar as causas da heterose em várias espécies (COCKRHAM; ZENG, 1996; LEPIERE et al., 2012; GARCIA et al., 2008; FRASCAROLI et al., 2007; MELCHINGER et al., 2007, SCHON et al., 2010).

Cockrham e Zeng (1996), estudando plantas F₃ de milho, a partir do delineamento III de Comstock e Robinson, utilizando marcadores moleculares, observaram que os resultados são, geralmente, a favor da hipótese de dominância de genes favoráveis para explicar a heterose na produção de grãos.

Frascaroli et al. (2007) utilizaram o cruzamento de 142 linhagens recombinantes com seus genitores e suas respectivas F₁ para o estudo da heterose em milho. Os autores constataram a importância do efeito de dominância para a produtividade de grãos por meio da alta significância da correlação entre o nível de heterozigosidade e a *performance* fenotípica. A estimativa da heterose para a produtividade foi a maior entre as heteroses de todos os caracteres estudados no trabalho. Esse resultado corrobora com Garcia et al. (2008) que também constataram o efeito de dominância na heterose para produtividade em milho. Verificaram também que para a produtividade, a epistasia foi não significativa.

O efeito de sobredominância em QTLs individuais não pode ser completamente descartado devido às limitações de sua estimação no delineamento III (FRASCAROLI et al., 2007; GARCIA et al., 2008).

Uma alta proporção de QTLs tem sido localizada próxima aos centrômeros, assim, pode-se admitir a hipótese de que a baixa proporção de recombinantes nessas regiões tem levado à detecção de QTLs em fase de repulsão, levando a uma aparente conclusão de sobredominância e quando esses QTLs estão ligados em fase de atração, a hipótese de efeito aditivo pode ser admitida (LARIEPE et al., 2012).

Schon et al. (2010), utilizando o cruzamento de genótipos de dois grupos heteróticos diferentes – Stiff Stalk Synthetic e Lancaster Sure Crop – verificaram que diferentes alelos foram fixados em cada grupo heterótico e por isso, a combinação de diferentes grupos heteróticos levava a uma maior heterose para o caráter produtividade.

Com relação à epistasia, vários trabalhos têm sido realizados, porém não obtendo resultados conclusivos no que se refere à participação da epistasia no controle genético da heterose (FRASCAROLI et al., 2007; MELCHINGER et al., 2007; GARCIA et al., 2008; SHON et al., 2010; LARIEPE, 2012). Diante da falta de uma conclusão clara a respeito da epistasia, restam duas dúvidas: o ajuste dos efeitos epistáticos apenas em QTLs de grande efeito pode não ser suficiente para quantificar a real contribuição desse fenômeno no controle genético da heterose, requerendo dessa maneira a aplicação de modelos complexos para o estudo da distribuição da epistasia ao longo do genoma (MOORE, 2005; HUGHES, 2005; LEHNER, 2011) e a epistasia pode não ser caráter-específica e poderia ser mais bem estudada dentro do conceito da multiplicidade de QTLs pleiotrópicos, uma vez que esses dois fenômenos (pleiotropia e epistasia) podem ser indissociáveis (SCHNELL; COCKERHAM,

1992; CHEVERUD et al., 2004; WOLF et al., 2005, 2006; PAVLICEV et al., 2008; MACKAY; STONE; AYROLES, 2009).

Um completo entendimento das bases genéticas da heterose por meio do mapeamento de QTLs é geralmente dificultado pelos pequenos tamanhos de populações utilizadas nos experimentos e pela frequência de recombinação de alguns genes que podem estar ligados, levando a conclusões errôneas de sobredominância (SCHNABLE; SPRINGER, 2013).

A heterose tem diminuído na medida em que é realizado o melhoramento de linhagens. A contribuição da heterose na produtividade em milho passou de 75% em 1916, em que os genitores eram derivados de cultivares de polinização aberta, para 50% na década de 60, quando já era realizado melhoramento nesses genitores. Em estudos de ganho de produção devido à heterose, verificou-se que, entre 1905 e 2005, a produtividade das linhagens cresceu de 1,9 a 3,5 vezes mais rápido que a contribuição da heterose. Observou-se ainda maior decréscimo da contribuição da heterose nas décadas mais recentes, devido ao melhoramento mais incisivo das linhagens. A tendência é que a contribuição da heterose tenda a diminuir ainda mais nos próximos anos (TROYER; WELLIN, 2009).

Do exposto, fica evidente, que mesmo com todas as técnicas avançadas de genética molecular para elucidar as causas da heterose, não foi possível ainda obter uma conclusão consistente sobre os efeitos que integram este fenômeno (TROYER, 2006; HALLAUER; CARENA, 2012).

2.2 Grupos Heteróticos e Diversidade

O uso da diversidade genética é a base do avanço genético via melhoramento. Em milho, o germoplasma grande e diversificado permite a manipulação de diferentes genótipos que leva a uma melhora da *performance*

dos híbridos em termos de produtividade, resistência a doenças e outras características agronômicas (LABORDA et al., 2005).

As cultivares de milho eram, no passado, cultivares de polinização aberta. Por exemplo, na década de 30, nos Estados Unidos, as melhores cultivares foram autofecundadas e produziram mais de 100 cultivares de polinização aberta (JENKINS, 1936 citado por LU; BERNARDO, 2001). Porém, posteriormente com o desenvolvimento das indústrias de sementes de milho híbrido, os cruzamentos entre genótipos elite se tornaram fontes preferidas de novas linhagens (HALLAUER et al., 1988; TROYER, 1990). Essa reciclagem, que ficou conhecida como ciclo avançado de melhoramento de linhagens, envolve cruzamentos entre linhagens relacionadas como fonte de populações (LU; BERNARDO, 2001).

O germoplasma temperado tem sido explorado por diversas décadas de melhoramento e várias gerações de seleção foram realizadas, resultando em um germoplasma altamente produtivo (HALLAUER et al., 2010). Em geral, as populações temperadas são sintéticas, gerados a partir de um pequeno grupo de linhagens, enquanto que o germoplasma tropical é normalmente formado por compostos gerados a partir do cruzamento de várias populações. Por isso, as populações temperadas devem apresentar uma base genética mais estreita que as populações tropicais (MUMM; DUDLEY, 1994).

O conhecimento da diversidade genética, relação e ~~da~~ estrutura do germoplasma são importantes no melhoramento de plantas por diferentes razões, entre elas para o auxílio na escolha das melhores combinações parentais para o desenvolvimento de progênies com máxima variabilidade genética (BARRETT; KIDWELL, 1998) e para a separação do germoplasma em grupos heteróticos. No melhoramento de milho, essas duas utilidades são essenciais para o desenvolvimento de melhores linhagens e a criação de híbridos fenotipicamente superiores, principalmente para o caráter produtividade (DUVICK, 2001).

O germoplasma de milho é organizado em grupos heteróticos (HALLAUER et al., 2010). Grupo heterótico é uma coleção de linhagens altamente correlacionadas que tende a resultar em vigorosos híbridos quando ocorre o cruzamento dessas linhagens vindas de diferentes grupos heteróticos, mas não quando cruzadas dentro do mesmo (LEE, 1995). Assim as linhagens dentro de um mesmo grupo devem ter um alto nível de similaridade e alto nível de dissimilaridade entre grupos (LU et al., 2009).

Experiências nos Estados Unidos têm demonstrado que algumas populações de melhoramento (como por exemplo, Reid, Lancaster, Minnesota 13, e Stiff Stalk Synthetic) geram maior quantidade de linhagens promissoras que outras (por exemplo, Hickory King, Krug, e Corn Borer Synthetic 1). Diferenças entre as populações ocorrem devido à origem dos genes e ao processo de seleção que gerou conjuntos de genes em frequências favoráveis que são desejadas em híbridos modernos. A seleção pode ter sido tão intensa e eficiente quanto nas outras populações, mas provavelmente o germoplasma original não incluía os genes necessários para os padrões exigidos pelo melhorista. Se os genes não estão presentes, os esforços dos melhoristas serão em vão apesar das técnicas e seleções realizadas (HALLAUER; CARENA; MIRANDA, 2010).

Com relação ao germoplasma tropical, Wellhausen (1978) identificou quatro raças complexas que são utilizadas para o melhoramento de milho nos trópicos: Tuxpeno e seus relativos Caribenho e Dentados Americanos; Cuban Flint, Coastal Tropical Flint e ETO. Tuxpeno é o único *flint* puro sendo considerado superior aos demais devido a sua alta capacidade produtiva, vigor excepcional e resistência às doenças comuns em milho (Hallauer et al., 2010).

A separação das linhagens em diferentes grupos heteróticos pode ser realizada por meio de relações de *pedigrees*, *performance* fenotípica para determinado caráter, adaptabilidade e estabilidade, *top-crosses*, cruzamentos

dialélicos e distância genética estimada com caracteres morfológicos ou marcadores moleculares (BERTAN et al., 2007).

Os melhoristas de milho geralmente desenvolvem novas linhagens parentais selecionando progênies de linhagens inter cruzadas dentro de um mesmo grupo heterótico. O grupo heterótico é normalmente baseado na capacidade específica de combinação, geradas a partir de cruzamentos dialélicos que fornecem estimativas dos efeitos genéticos para a separação desses grupos (LAUDE; CARENA, 2014). A informação de *pedigree* também é normalmente empregada para alocar novas linhagens dentro de grupos heteróticos (MESSMER et al., 1993).

A fim de gerar informações de grupos heteróticos mais precisas, vários autores têm sugerido o uso de marcadores moleculares para agrupamento das linhagens relacionadas (FLINT-GARCIA et al., 2009). O desenvolvimento de marcadores moleculares de DNA gerou ferramentas para acessar a diversidade genética entre linhagens e organizá-las nesses diferentes grupos (SMITH et al., 1990; DUDLEY; SAGHAI MAROOF; RUFENER, 1991; MESSMER et al., 1993).

Durante a década de 90, os RFLPs eram os principais marcadores utilizados para determinar a diversidade genética em germoplasmas de milho europeu (BOPPENMAIER et al., 1992) e americano (MELCHINGER et al., 1990) e organizar os genótipos em grupos heteróticos.

Benchimol et al. (2000), utilizando germoplasma tropical para análise de divergência genética e separação da população em cluster, com o auxílio de marcadores RFLP, verificaram a separação em dois grupos, sendo que a mesma estava de acordo com o *pedigree* conhecido.

Contudo, como o uso deste marcador é de alto custo e de baixa repetibilidade, o desenvolvimento de outras tecnologias fez com que outros marcadores apresentassem vantagens atrativas (LABORDA et al., 2005). Com

isso, RAPDs e SSRs foram rapidamente adotados para utilização em análises genéticas devido aos seus menores custos e maior eficiência (LABORDA et al., 2005). Como consequência, vários estudos foram realizados a fim de acessar a diversidade genética em várias espécies, como cana-de-açúcar (LIMA et al., 2002), trigo (MANIFESTO et al., 2001), algodão (LIU et al., 2000), soja (PRIOLLI et al., 2002) e sorgo (GHEBRU et al., 2002). A cultura do milho também tem sido extensivamente estudada. Linhagens dos Estados Unidos têm sido utilizadas em grande número de investigações (LABORDA et al., 2005).

Com a diminuição dos custos de genotipagem (HESLOT; JANNINK; SORRELLS, 2014), marcadores mais modernos e precisos passaram então a ser utilizados. Devido ao baixo custo por marca, alta abundância no genoma, especificidade de locos, potencial de alta densidade de marcas, baixos níveis de erros (RAFALSKI, 2002) e facilidade na automatização (VAN INGHELANDT et al., 2010) os marcadores SNPs e Darts têm emergido como poderosas ferramentas para várias aplicações dentro da genética, incluindo estudos de diversidade genética (ZHU et al., 2003).

Lu et al. (2009) caracterizaram 770 linhagens de milho, entre tropicais, subtropicais e temperadas utilizando 1034 SNPs. Os autores puderam observar uma estrutura clara da população e a divergência genética entre as linhagens temperadas e tropicais.

Yan et al. (2009) estudaram 632 linhagens tropicais, temperadas e subtropicais vindas de diferentes programas de melhoramento e reportaram presença de uma clara estrutura entre as linhagens tropicais e temperadas e também uma complexa relação familiar entre as diferentes coleções ao longo do mundo.

É correto afirmar que o conhecimento detalhado da relação entre linhagens de milho é importante não só para a seleção de parentais, mas também para análises genéticas, *design* de sistemas de melhoramento mais eficientes (LU

et al., 2009) e para a manutenção da variabilidade para o uso futuro (HALLAUER; CARENA; MIRANDA, 2010). Contudo, apesar do extensivo conhecimento do germoplasma temperado de milho, relativamente poucos estudos de diversidade envolvendo grande número de linhagens de germoplasma tropical têm sido publicados (OLIVEIRA et al., 2004; REIF et al., 2004), especialmente em programas de melhoramento comercial.

2.3 Predição do desempenho de híbridos de milho

Os programas de melhoramento de milho buscam híbridos com alta produtividade de grãos, melhores aos que já existem no mercado, tentando assim, ganhar mercado em uma área tão concorrida, como é a área de sementes de milho em todo o mundo. Contudo, atingir esse objetivo não é algo simples. O melhoramento de milho envolve sistemas de cruzamentos, que buscam identificar as melhores combinações de genitores, tendo como produto final os híbridos. Para identificar as melhores combinações, são realizados milhares de cruzamentos com as linhagens obtidas no programa (LAMKEY; HALLAUER; KAHLER, 1987). O dialelo tem sido o principal método para encontrar essas melhores combinações, podendo também existir variações deste método.

Fica evidente que um dos maiores problemas do melhoramento de milho é a grande quantidade de cruzamentos a ser realizada a partir de um conjunto relativamente pequeno de linhagens.

Os programas de melhoramento de milho fazem cruzamentos entre diferentes grupos heteróticos, buscando a complementariedade dos locos. Mesmo utilizando esses grupos heteróticos, a quantidade de cruzamentos pode ser tão elevada que a realização de todos os cruzamentos torna-se inviável. Assim, faz-se necessário o uso de metodologias de predição que direcionem os

melhores cruzamentos a serem realizados, aumentando a eficiência dos programas (BERNARDO, 1994; SCHRAG et al., 2006).

O direcionamento dos cruzamentos normalmente é feito de forma empírica, o melhorista utiliza de sua experiência para direcionar os cruzamentos que ele acredita ter maior potencial. Este direcionamento pode ser realizado também com base nos dados dos genitores, normalmente a média, que, apesar de ser um bom estimador, não garante uma boa complementariedade dos cruzamentos. Levando-se em conta que o caráter produtividade de grãos apresenta baixa herdabilidade fica ainda mais difícil realizar esses direcionamentos somente com base no fenótipo, que se torna um pobre indicador do genótipo. Assim, quanto mais cruzamentos são descartados erroneamente, mais híbridos com alto potencial são perdidos e mesmo melhoristas experientes são incapazes de obter sucesso com direcionamentos baseados em seleção visual ou mesmo com base em média.

Já em 1934, Jenkins observou a necessidade de predição dos híbridos. Ele previu o desempenho dos híbridos duplos por meio da *performance* dos híbridos simples não parentais (JENKINS, 1934). Na década de 60, Eberhart (1964) e Cockerham (1967) aprimoraram as ideias de Jenkins, lapidando seus modelos genéticos. Seus modelos visavam à predição de híbridos duplos e triplos, mas não eram úteis na predição de híbridos simples.

Com o advento dos marcadores moleculares vários genomas foram estudados e, com o passar dos anos, essa técnica foi se tornando mais utilizada e mais acessível em termos técnicos e econômicos. A disponibilidade de marcadores abundantes e economicamente viáveis em milho possibilitou que os melhoristas questionassem de que modo esses marcadores poderiam ser utilizados para alcançar melhores ganhos de seleção (BERNARDO; YU, 2007).

A seleção feita para caracteres de importância econômica, como produtividade, que é um caráter quantitativo, antes realizado somente a partir de

dados fenotípicos, passou a utilizar dados moleculares. Uma justificativa para a utilização de informações moleculares é a expectativa de que as informações a nível de ~~sobre~~ DNA tragam maior ganho genético do que quando se utilizam apenas os dados fenotípicos (MEUWISSEN et al., 2001).

Utilizando modelos simples de predição, com marcadores menos informativos, como RFLP e AFLP, pequeno número de marcas, conhecimento de *pedigree* e localização de QTLs, vários trabalhos tentaram prever o comportamento de híbridos simples de milho (SMITH et al., 1990; GODSHLAK; LEE; LAMKEY, 1990; SCRHAG et al., 2006; MESSMER, 1993), porém resultados frágeis e inconclusivos foram obtidos, não permitindo que essas técnicas pudessem ser utilizadas em grande escala em empresas produtoras de sementes híbridas de milho.

Visando resolver os problemas de estimação dos métodos de predição, surgiu no melhoramento animal uma poderosa ferramenta que é a predição genômica ampla. Essa ferramenta permite que sejam utilizados simultaneamente todos os dados genômicos de uma população de treino fenotipada para prever os valores genéticos de outros genótipos sem necessidade de coleta de dados fenotípicos dos mesmos. Essa predição genômica ampla tem mostrado resultados promissores no melhoramento de gado, e pode ter impacto similar no melhoramento vegetal (RIEDELSCHELMER et al., 2012).

Assim, a predição genômica tem se mostrado uma ferramenta alternativa que utiliza a informação genômica na seleção das melhores combinações visando realizar o melhoramento de forma mais eficiente. Ao invés de utilizar somente as marcas moleculares que apresentam significância para o caráter em questão, a predição genômica ampla utiliza todas as marcas simultaneamente na construção do modelo de predição (WINDHAUSEN et al., 2012). Sendo utilizados também marcadores mais informativos, como é o caso do SNP, com

alta densidade visando à cobertura de grande parte do genoma, aumentando assim a eficiência do método.

É esperado que a predição genômica aumente o ganho de seleção por meio do aumento da intensidade de seleção e da aceleração do ciclo de melhoramento. Para o uso da predição de híbridos em milho, faz-se necessário uma definição clara do cenário de melhoramento, uma análise detalhada da estrutura da população antes de prosseguir com a validação cruzada e grandes populações de treinamento com forte relação genética com a população de validação. A acurácia da predição genômica é estimada pela avaliação da população treino e a população de validação em um mesmo ambiente ou em ambientes diferentes (WINDHAUSEN et al., 2012), para isso faz-se necessário dados fenotípicos altamente precisos e alta densidade de marcadores moleculares (MEUWISSEN et al., 2001; RIEDELSHELMER et al., 2012, TECHNOW et al., 2012).

Contudo a implementação da predição genômica ampla impõe vários desafios estatísticos e computacionais, como por exemplo, como os modelos agem frente à dimensionalidade, à colinearidade entre as marcas e à complexidade de caracteres quantitativos (CROSSA et al., 2010).

O uso de um grande número de marcas inutiliza o método de quadrados mínimos, pois os efeitos dos haplótipos não podem ser estimados simultaneamente. Mesmo quando são incluídas somente as marcas de grande efeito, os efeitos dos haplótipos são subestimados diminuindo assim a acurácia de predição.

Segundo Bernardo (1994), uma alternativa para a predição genômica seria o preditor linear não viesado (BLUP), que leva em consideração informações dos marcadores moleculares das linhagens parentais e os dados de produtividade de uma população de cruzamentos testes. Apesar da correlação entre o desempenho genotípico do cruzamento predito e o seu valor fenotípico

ser baixa em muitos casos, essa abordagem apresenta vantagens (BERNARDO, 1994). A utilização do BLUP tem mostrado eficiência considerável para tal finalidade (BERNARDO, 1994; 1995; 1996a, 1996b; MASSMAN et al., 2013), contudo esse método assume a mesma variância para todas as marcas, o que é irreal, podendo levar a uma redução da acurácia de predição (MEUWISSEN et al., 2001) para caracteres qualitativos.

Métodos que assumem uma distribuição a priori, para as variâncias associadas com cada marca, demonstram melhor acurácia de predição, mesmo quando a priori é incorreta (MEUWISSEN et al., 2001). Por isso, a abordagem bayesiana também vem sendo utilizada na predição genômica em melhoramento vegetal. Na estatística Bayesiana, parâmetros como as variâncias das marcas, assumem uma distribuição a priori. A variância não é a mesma para todas as marcas que é então combinada com a distribuição a priori sendo estimadas a partir dos dados de produção. Essa abordagem bayesiana, em que as variâncias das marcas não são fixas, é mais realística que a abordagem utilizada no BLUP (MEUWISSEN et al., 2001).

Meuwissen et al. (2001) comparam as técnicas dos quadrados mínimos, BLUP, Bayes A e Bayes B para predição do valor genético em animais. Os autores verificaram que a predição por meio do método dos quadrados mínimos teve acurácia muito baixa, o que leva a uma pobre predição. Os métodos Bayes A e Bayes B apresentaram, respectivamente, acurácias 9% e 16% acima da acurácia do método BLUP.

Crossa et al. (2010), comparando métodos de predição de híbridos em milho, verificaram que o LASSO Bayesiano (LB) superou o BLUP na acurácia de predição. Os autores atribuem o melhor desempenho do LB ao fato de que o BLUP utiliza os mesmos efeitos para todas as marcas, enquanto o LB utiliza um efeito de encolhimento nas marcas de pequeno efeito e ao fato de que no LB a variância dos parâmetros e o efeito das marcas são inferidos ao mesmo tempo,

enquanto no BLUP estima-se primeiro a variância dos parâmetros e posteriormente estima-se o efeito das marcas.

Technow et al. (2012) comparando o método bayesiano Bayes B com o BLUP observaram diferença significativa na acurácia de predição, sendo que o Bayes B apresentou maior eficiência.

Sem dúvidas, a predição genômica ampla mudou a eficiência e as metodologias utilizadas no melhoramento de plantas, por isso a predição genômica está passando por um período de intensa pesquisa científica (JANNINK; LORENZ; IWATA, 2010). Vários trabalhos têm sido realizados com dados de simulação (MEUWISSEN et al., 2001; HABIER; FERNANDO; DEKKERS, 2007; ZHONG et al., 2006), porém fazem-se necessários mais trabalhos com dados reais para que a superioridade da predição genômica ampla seja comprovada.

2.4 Modelo GBLUP

Na década de 70, Henderson (1973), visando solucionar problemas na área de melhoramento animal, criou um modelo de seleção, denominado Modelo Misto, que contempla o método do “Melhor Preditor Linear Não Viesado” – *Best Linear Unbiased Prediction* (BLUP). Os candidatos à seleção são variáveis aleatórias e o mérito de cada candidato é a soma da média da população mais o valor predito da variável aleatória associado ao candidato. Neste caso, a seleção depende também de efeitos fixos desconhecidos. Henderson derivou o método BLUP por meio da maximização da função densidade de probabilidade conjunta de y (valores fenotípicos) e g (valores genéticos) (HENDERSON, 1973 citado por RESENDE et al., 2012).

O BLUP passou então a ser utilizado primeiramente no melhoramento animal, onde o alto custo de fenotipagem bem como a impossibilidade da

replicação de indivíduos, fez com que essa metodologia se tornasse conveniente. O método, assumindo normalmente um modelo genético aditivo, foi utilizado extensivamente em avaliações de gado de leite. Contudo até a década de 90, as avaliações do BLUP em espécies vegetais foram bem restritas (BERNARDO, 1994).

Assim, a partir da década de 90, Bernardo (1994) propôs a utilização do BLUP para predição da *performance* de híbridos simples de milho baseado em dados fenotípicos e coeficientes de coancestria estimados por *pedigree* ou marcadores RFLPs. Desde então, a predição no melhoramento genético vegetal tem sido extensivamente estudada e novos modelos vêm sendo testados. Em seguida a metodologia foi estendida para a inclusão do genótipo das marcas para identificação de QTLs (BERNARDO, 1999).

Posteriormente Meuwissen et al. (2001), por meio de simulações, incorporaram uma grande quantidade de marcas ao modelo, estimando todas simultaneamente, sendo esta nova metodologia chamada de Seleção Genômica Ampla (GWS). A GWS incorpora todas as marcas simultaneamente ao longo do genoma no modelo de predição, reduzindo o risco da identificação errônea ou imprecisa dos QTLs com efeitos menores. Cada marca é considerada um possível QTL e todas as marcas são combinadas para predizer o valor genético da progênie (GUO et al., 2013).

Com o aumento da disponibilidade de marcas e a redução dos custos de genotipagem, as técnicas de GWS se tornaram uma realidade para os programas de melhoramento (HESLOT; JANNINK, SORRELLS, 2014; ZHAO et al., 2015). Contudo, existe um grave problema estatístico no uso de alta densidade de marcadores, tem-se um número muito maior de marcas do que de observações. Uma solução é modelar os efeitos de SNPs como aleatórios e realizar suposições a priori sobre a variância explicada devido aos seus efeitos. Alguns métodos como Bayes A, Bayes B (MEUWISSEN et al., 2001), Bayes

$C\pi$ e $D\pi$ (HABIER et al., 2011) e Lasso Bayesiano (PARK; CASELLA 2008) entre outros dão ênfase maior a algumas regiões do genoma permitindo variação das variâncias entre os locos das marcas. Já o método BLUP atribui variância homogênea para todos os locos e essencialmente trata os mesmos com igual importância, contribuindo igualmente para a variância genética (CLARK; VAN DER WERF, 2013).

Assim, GWS pode ser implementada utilizando métodos paramétricos ou não paramétricos. Entre os métodos paramétricos está o GBLUP (Genomic Best Linear Unbiased Prediction), que vem sendo amplamente utilizado no melhoramento vegetal. O GBLUP utiliza a relação genômica para estimar o mérito genético de um indivíduo. Para tal, uma matriz de correlações genéticas é utilizada, estimada a partir de informações de marcadores de DNA. Essa matriz define a covariância entre indivíduos baseada na similaridade ao nível genômico, em contrapartida da similaridade estimada por meio de *pedigree*, sendo assim mais preciso (CROSSA et al., 2013).

Clark, Hickey e Van der Werf (2011) demonstrou que predições utilizando o método Bayes B podem ser mais acuradas na medida em que se verificam poucos QTLs de grandes efeitos controlando o caráter, porém a acurácia tende a se igualar ao GBLUP à medida que são observados vários QTLs de pequenos efeitos (CLARK; VAN DER WERF, 2013). Daetwyler (2009) observou que a população e a arquitetura genética do caráter têm grande influência na *performance* dos métodos de seleção. Uma das maiores vantagens da predição genômica é que todo polimorfismo que afeta o caráter é modelado, independente da proporção de seu efeito (TECHNOW et al., 2013).

O uso do BLUP tem permitido aumento dos ganhos genéticos em vários programas de melhoramento de gado. O método BLUP tradicional utiliza uma informação de *pedigree* para definir a covariância entre genótipos conhecidos. Essa covariância pode ser também definida utilizando grandes quantidades de

informação de marcadores de DNA, mais comumente, os marcadores SNPs. Essa matriz é conhecida como matriz de relação genética (GRM) (CLARK; VAN DER WERF, 2013).

Já o GBLUP combina informação genética dentro do BLUP utilizando a matriz GRM como substituto do numerador da matriz de relação (CROSSA et al., 2013). O método GBLUP foi introduzido por VanRaden (2008). Na prática, o método utilizado para implementar o GBLUP é:

$$y = Xb + Zg + e$$

Em que y é o vetor de observações do fenótipo, X é a matriz de incidência dos efeitos fixos, b é o vetor os efeitos fixos, Z é a matriz de incidência dos efeitos aleatórios, g é o vetor dos efeitos aditivos de cada indivíduo e e é o vetor das variâncias residuais (σ_e^2), em que a matriz de variâncias residuais é dada por $R = I\sigma_e^2$ e da $\text{var}(g) = G\sigma_g^2$, em que G é a matriz de relação genética e σ_g^2 é a variância genética para o modelo (HESLOT; JANNINK; SORRELLS, 2014). Os indivíduos com dados fenotípicos constituem a população de treinamento e os indivíduos sem essa informação formam a população de validação, que são os genótipos a serem preditos (CLARK; VAN DER WERF, 2013), neste caso algumas colunas da matriz Z possuirão somente elementos 0 (HESLOT; JANNINK; SORRELLS, 2014).

Quando são incluídos os efeitos de dominância no modelo GBLUP, tem-se:

$$y = X\hat{\beta} + Z\alpha + Z\delta + e$$

Em que \mathbf{y} o vetor das observações fenotípicas $n \times 1$, em que t é o número de locais e n é o número de observações fenotípicas dos híbridos simples obtidas em cada local; \mathbf{X} é a matriz de incidência dos efeitos fixos de locais e média dos híbridos confundidas em uma mesma covariável $n \times p$, em que p é o número de efeitos fixos; $\hat{\boldsymbol{\beta}}$ é o vetor de efeitos fixos dos locais e médias confundidos $p \times 1$; \mathbf{Z} é a matriz de incidência dos híbridos simples avaliados em cada local $m \times n$; $\boldsymbol{\alpha}$ é o vetor de efeitos aditivos $n \times 1$; $\boldsymbol{\delta}$ é o vetor de efeito de dominância $n \times 1$; \mathbf{e} é o vetor dos resíduos $m \times 1$.

O método GBLUP tem três importantes características que o distingue e o torna preferível ao RRBLUP: as dimensões dos efeitos genéticos nas equações de modelo misto são reduzidas de $m \times m$ (m é o número de marcas) no RRBLUP a $n \times n$ (n é o número de indivíduos da população) no GBLUP, o que torna o método computacionalmente mais eficiente; a acurácia do valor genético genômico (VGG) do indivíduo pode ser calculada da mesma maneira que no BLUP baseada em informações de *pedigree* e informação do GBLUP pode ser incorporada à informação de *pedigree* utilizando um único método (MISZTAL; LEGARRA; AGUILAR, 2009).

Resultados indicam claramente que a GWS tem maior potencial de predição que a seleção assistida por marcadores e experimentos com conjunto de dados empíricos e todas as metodologias de predição têm semelhantes potenciais preditivos. Assim, o GBLUP demonstra vantagens, visto sua relativa simplicidade, menor tempo computacional e as já bem conhecidas propriedades ótimas dos modelos mistos de seleção, assim este método tem se mostrado eficiente na maioria das situações (HESLOT; JANNINK; SORRELLS, 2014).

REFERÊNCIAS

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

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

BARRETT, B. A.; KIDWELL, K. K. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. **Crop Science**, Madison, v. 38, n. 5, p. 1261-1271, Nov. 1998.

BENCHIMOL, L. L. et al. Genetic diversity in tropical maize inbred lines: heterotic group assignment and hybrid performance determined by RFLP markers. **Plant Breeding**, [S.l.], v. 119, n. 6, p. 491-496, June 2000.

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

BERNARDO, R. Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. **Crop Science**, Madison, v. 36, n. 4, p. 862-866, 1996b.

BERNARDO, R. **Breeding for quantitative traits in plants**. Woodbury: Stemma, 2010.

BERNARDO, R. Estimation of coefficient of coancestry using molecular markers in maize. **Theoretical and Applied Genetics**, Berlin, v. 85, n. 8, p. 1055-1062, Fev. 1993.

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

BERNARDO, R. Marker-assisted best linear unbiased prediction of single-cross performance. **Crop Science**, Madison, v. 39, n. 5, p. 1277-1282, 1999.

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.; YU, J. Prospects for Genomewide Selection for Quantitative Traits in Maize. **Crop Science**, Madison, v. 47, n. 3, p. 1082-1090, May 2007.

BERTAN I.; CARVALHO, F. I. F.; OLIVEIRA, A. C. Parental selection strategies in plant breeding programs. **Jornal of Crop Science Biotechnology**, Seoul, v. 10, n. 4, p. 211-222, 2007.

BETRÁN, F. J. et al. Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and nonstress environments. **Crop Science**, Madison, v. 43, n. 3, p. 797-806, Apr. 2003.

BOPPENMAIER, J. et al. Genetic diversity for RFLPs in European maize inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. **Plant Breeding**, [S.l.], v. 84, n. 5, p. 217-226, Aug, 1992.

BRUCE, A. B. The Mendelian theory of heredity and the augmentation of vigor. **Science**, Washington, v. 32, n. 827, p. 627-628, Nov. 1910.

CAMPOS, G. et al. Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree. **Genetics**, Bethesda, v. 182, n. 1, p. 375-385, Apr. 2009.

CARENA, M. J. et al. Breeding Maize for Higher Yield and Quality Under Drought Stress. **Maydica**, Bergamo, v. 54, 287-296, July 2009.

CHARCOSSET A, E. L. The effect of population-structure on the relationship between heterosis and heterozygosity at marker loci. **Theoretical and Applied Genetics**, Berlin, v. 89, n. 2, p. 336-343, Oct. 1994.

CHEVERUD, J. et al. Pleiotropic effects on mandibular morphology II: differential epistasis and genetic variation in morphological integration. **Journal of Experimental Zoology**, Hoboken, v.302, n. 5, p.424-435, Jul. 2004.

CLARK, S.; HICKEY, J. M.; VAN DER WERF, J. H. J. Different models of genetic variation and their effect on genomic evaluation. **Genetics Selection Evolution**, 2011, v. 43, n. 18, p. 1-9.

CLARK, S. A.; VAN DER WERF, J. Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. In: GONDRO, C.; VAN DER WERF, J.; HAYES, B. (Ed.). **Genome-Wide Association Studies and Genomic Prediction**. New York: Humanas, 2013. (Methods in Molecular Biology, 2019).

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

COCKERHAM, C. C.; ZENG, Z. B. Design I11 With Marker Loci. **Genetics**, Bethesda, v. 143, n. 3, p. 1437-1456, July 1996.

CROSSA, J. et al. Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. **G3**, Bethesda, v. 3, n. 11, p. 1903-1926, Nov 2013.

CROSSA, J. et al. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. **Genetics**, Bethesda, v. 186, n. 2, p. 713-724, Oct. 2010.

DAETWYLER, H. D. **Genome-Wide Evaluation of Populations**. 2009. 192 p. Tese (Doutorado em Melhoramento Animal e Genômica)-Universidade de Wageningen, Wageningen, 2009.

DEVI, P.; SINGH N. K. Heterosis, molecular diversity, combining ability and their interrelationships in short duration maize (*Zea mays* L.) across the environments. **Euphytica**, Wageningen, v. 178, n. 1, p. 71-81, Mar. 2011.

DHLIWAYO, T. et al. Combining ability, genetic distances, and heterosis among elite cimmyt and iita tropical maize inbred lines. **Crop Science**, Madison, v. 49, n. 4, p. 1201-1210, Jul. 2009.

DUDLEY, J. W.; SAGHAI MAROOF, M. A.; RUFENER, G. K. Molecular markers and grouping of parents in a maize breeding program. **Crop Science**, Madison, v. 31, n. 3, p. 718-723, May 1991.

DUVICK, D. N. Biotechnology in the 1930s: the development of hybrid maize. **Nature Reviews Genetics**, London, v. 2, p. 69-74, Jan. 2001.

EAST, E. M. Heterosis. **Genetics**, Bethesda, v. 21, p. 375-397, July 1936.

EBERHART, S. A. Theoretical relations among single, three-way, and double cross hybrids. **Biometrics**, Washington, v. 20 n. 3, p. 522-39, Sept. 1964.

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

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, Dec. 2010.

FLINT-GARCIA, S. A. et al. Heterosis is prevalent for multiple traits in diverse maize germplasm. **PLOS ONE**, [San Francisco], v. 4, n. 10, p. 7433, Oct. 2009.

FRASCAROLI, E. et al. Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. **Genetics**, Bethesda, v. 176, n. 1, p. 625-644, May 2007.

GARCIA, A.A. et al. Quantitative trait mapping and the genetic basis of heterosis in maize and rice. **Genetics**, Bethesda, v. 180, n. 3, p. 1707-1724, Nov. 2008.

GHEBRU, B.; SCHMIDT, R. J.; BENNETZEN, J. L. Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. **Theoretical and Applied Genetics**, Berlin, v. 105, n. 2, p. 229-236, Aug. 2002.

GILKS, W. R., RICHARDSON, S.; SPIEGELHALTER, D. J. Markov chain Monte Carlo in practice. Boca Raton: CRC, 1996.

GODSHALK, E. E.; LEE, M.; LAMKEY, K. R. Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. **Theoretical and Applied Genetics**, Berlin, v. 80, n. 2, p. 273-280, Aug. 1990.

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, Jan. 2013.

HABIER, D. et al. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, Chichester, v. 12, n. 186, p. 1-12, May 2011.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, Bethesda, v. 177, n. 4, p. 2389-2397, Dec. 2007.

HALLAUER, A. R.; CARENA, M. J. Recurrent selection methods to improve germplasm in maize. **Maydica**, Bergamo, v. 57, n. 4, p. 266-283, 2012.

HALLAUER, A. R.; CARENA, M.; MIRANDA, J. B. F. **Quantitative genetics in maize breeding**. 3. ed. Ames: Iowa State University, 2010.

HALLAUER, A. R.; RUSSELL, W. A.; LAMKEY, K. R. Corn breeding. In: Sprague, G. F.; Dudley, J. W. (Eds.). **Corn and corn improvement**. 3. ed. Madison: American Society of Agronomy; Crop Science Society of America, 1988. p. 463-564. (Agronomy, 18).

HAYES, B. J. et al. Invited review: genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, Champaign, v. 92, n. 2, p. 433-443, 2009.

HESLOT, N.; JANNINK, J. L.; SORRELLS, M. E. Perspectives for Genomic Selection Applications and Research in Plants. **Crop Science**, Madison, v. 55, n. 1, p. 1-12, Nov. 2014.

HUGHES, T. R. Universal epistasis analysis. **Nature Genetics**, New York, v. 37, n. 5, p. 457-458, May 2005.

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

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

JONES, D. F. Dominance of linked factors as a means of accounting for heterosis. **Genetics**, Bethesda, v. 2, p. 466-479, Fev. 1917.

JONES, D. F. The effects of inbreeding and crossbreeding upon development. **Connecticut Agricultural Experimental Station Bulletin**, Washington, v. 297, p. 5-100, July 1918.

KAEPPLER, S. Heterosis: Many Genes, Many Mechanisms—End the Search for an Undiscovered Unifying Theory. **International Scholarly Research Network ISRN Botany**, [S.l.], v. 2012, p. 1-12, Oct. 2012.

- LABORDA, P. R. et al. Tropical maize germplasm: what can we say about its genetic diversity in the light of molecular markers? **Theoretical and Applied Genetics**, Berlin, v. 111, n. 7, p. 1288-1299, Nov. 2005.
- LAMKEY, K. R.; HALLAUER, A. R.; KAHLER, A. L. Allelic differences at enzyme loci and hybrid performance in maize. **Journal of Heredity**, Washington, v. 78, n. 4, p. 231-234, July 1987.
- LANZA, L. L. et al. Genetic distance of inbred lines and prediction of maize singlecross performance using RAPD markers. **Theoretical and Applied Genetics**, Berlin, v. 94, n. 8, p. 1023-1030, June 1997.
- 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**, Bethesda, v. 190, n. 2, p. 795-811, Feb. 2012.
- LAUDE, T. P.; CARENA, M. J. Diallel analysis among 16 maize populations adapted to the northern U.S. Corn Belt for grain yield and grain quality traits. **Euphytica**, Wageningen, v. 200, n. 1, p. 29-44, Nov. 2014.
- LEE M.; LAMKEY K. R.; WOODMAN, L. Association of restriction length polymorphism among maize inbreds with agronomic performance of their crosses. **Crop Science**, Madison, v. 29, n. 4, p. 1067-1071, Jul. 1989.
- LEE, M. DNA markers and plant breeding programs. **Advances in Agronomy**, San Diego, v. 55, p. 265-344, Dec. 1995.
- LEGESSE, B.W. et al. Relationship between hybrid performance and AFLP based genetic distance in highland maize inbred lines. **Euphytica**, Wageningen, v. 162, n. 3, p. 313-323, Aug. 2008.
- LEHNER, B. Molecular mechanisms of epistasis within and between genes. **Trends in Genetics**, v. 27, n. 8, June 2011.
- LIMA, M. L. A. et al. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum spp.*). **Theoretical and Applied Genetics**, Berlin, v. 104, n. 1, p. 30-38, Jan 2002.
- LIU, S. et al. Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. **Crop Science**, Madison, v. 40, n. 5, p. 1459-1469, Sept. 2000.

LU, H.; BERNARDO, R. Molecular marker diversity among current and historical maize inbreds. **Theoretical and Applied Genetics**, Berlin, v. 103, n. 4, p. 613-617, Sept. 2001.

LU, Y. et al. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 1, p. 93-115, Dec. 2009.

LÜBBERSTEDT, T. et al. Relationships among early European maize inbreds: IV Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree Data. **Crop Science**, Madison, v. 40, n. 3, p. 783-791, May 2000.

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

MANIFESTO, M. M. et al. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. **Crop Science**, Madison, v. 41, n. 3, p. 682-690, May 2001.

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

MELCHINGER, A. E. et al. Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. **Crop Science**, Madison, v. 31, n. 3, p. 669-678, Dec. 1990.

MELCHINGER, A. E. et al. The role of epistasis in the manifestation of heterosis: A systems-oriented approach. **Genetics**, Bethesda, v. 177, n. 3, p. 1815-1825, Nov. 2007.

MESSMER, M. M. et al. Relationship among early European maize inbreds. II Comparison of pedigree and RFLP data. **Crop Science**, Madison, v. 33, n. 5, p. 944-950, 1993.

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

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, Champaign, v. 92, n. 9, p. 4648-4655, Sept. 2009.

MOORE, J.H. A global view of epistasis. **Nature Genetics**, New York, v. 37, n. 1, p. 13-14, Fev. 2005.

MUMM, R. H.; DUDLEY, J. W. A classification of 148 U.S. maize inbreds: I. Cluster analysis based on RFLPs. **Crop. Science**, Madison, v. 34, n. 4, p. 842-851, Aug. 1994.

OLIVEIRA, K. M. et al. Evaluating genetic relationships between tropical maize inbred lines by means of AFLP profiling. **Hereditas**, Lund, v. 140, n. 1, p. 24-33, 2004.

PARK, T.; CASELLA, G. The Bayesian Lasso. **Journal of the American Statistical Association**, New York, v. 103, n. 482, p. 681-686, June 2008.

PAVLICEV, M. et al. Genetic variation in pleiotropy: differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. **Evolution**, Malden, v. 62, n. 1, p. 199-213, Jan. 2008.

POWERS, L. An expansion of Jones' theory for the explanation of heterosis. **American Naturalist**, Chicago, v. 78, n. 776, p. 275-280, 1944.

PRIOLLI, R. H. G. et al. Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, Ribeirão Preto, v. 25, p. 185-193, 2002.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, [Amsterdam], v. 5, n. 2, p. 94-100, Apr. 2002.

REIF, J. C. et al. Genetic diversity determined within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by SSR markers. **Crop Science**, Madison, v. 44, n. 1, p. 326-334, Fev. 2004.

RESENDE, M. D. V. et al. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial**. Viçosa, MG: Universidade Federal de Viçosa. 2012.

RIEDELSHEIMER, C. et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. **Nature Genetics**, New York, v. 44, n. 2, p. 217-222, Jan. 2012.

RIEDELSHEIMER, 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**, Chichester, v. 13, n. 452, Sept. 2012.

SCHNABLE, J. C.; FREELING, M. Genes Identified by Visible Mutant Phenotypes Show Increased Bias toward One of Two Subgenomes of Maize. **PLOS ONE**, [San Francisco], v. 6, n. 3, p. 1-6, Mar. 2011.

SCHNABLE, P. S.; SPRINGER, N. M. Progress Toward Understanding Heterosis in Crop Plants. **Annual Review Plant Biology**, Palo Alto, v. 64, p.71-88, Apr. 2013.

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

SCHÖN, C. C. et al. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 321-332, Jan. 2010.

SCHRAG, T. A. et al. Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. **Theoretical and Applied Genetics**, Berlin, v. 113, p. 1037-1047, Oct. 2006.

SCHRAG, T. A. et al. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. **Theoretical Applied Genetics**, Berlin, v. 120, n. 1, p. 451-461, Jan. 2010.

SCHRAG, T.A. et al. Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. **Theoretical Applied Genetics**, Berlin, v. 118, n. 4, p. 741-751, Feb. 2009.

SCHRAG, T. A. et al. Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. **Theoretical and Applied Genetics**, Berlin, v. 113, n. 6, p. 1037-1047, Oct. 2006.

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

SHULL, G. H. What is "heterosis"? **Genetics**, Bethesda, v. 33, p. 439-446, Sept. 1948.

SMITH, J. S. C. et al. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree. **Theoretical and Applied Genetics**, Berlin, v. 95, n. 1-2, p. 163-173, July 1997.

SMITH, O. S. et al. Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield, heterosis, and RFLPs. **Theoretical and Applied Genetics**, Berlin, v. 80, n. 6, p. 833-840, Dec. 1990.

TANG, J. et al. Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F2 population. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 333-340, Jan. 2010.

TECHNOW, F.; BÜRGER, A.; MELCHINGER, A. E. Genomic Prediction of Northern Corn Leaf Blight Resistance in Maize with Combined or Separated Training Sets for Heterotic Groups. **G3**, Bethesda, v. 3, n. 2, p. 197-203, Feb. 2013.

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, p. 1181-1194, Oct. 2012.

TROYER, A. F. A retrospective view of corn genetic resources. **Journal of Heredity**, Washington, v. 81, n. 1, p. 17-24, Jan. 1990.

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

TROYER, A. F.; WELLIN, E. J. Heterosis Decreasing in Hybrids: Yield Test Inbreds. **Crop Science**, Madison, v. 49, n. 6, p. 1969-1976, Sept. 2009.

VAN INGHELANDT, D. et al. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 7, p. 1289-1299, May 2010.

- VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, Champaign, v. 91, n. 11, p. 4414-4423, Nov. 2008.
- WELLHAUSEN, E. J. Recent developments in maize breeding in the tropics. In: D. B. Walden (Ed.) **Maize breeding and genetics**. New York: J. Wiley, 1978. p. 59-91.
- WILLIAMS, W. Heterosis and the genetics of complex characters. **Nature**, London, v. 184, p. 527-530, Aug. 1959.
- WINDHAUSEN, V. S. et al. Effectiveness of Genomic Prediction of Maize Hybrid Performance in Different Breeding Populations and Environments. **G3**, Bethesda, v. 2, n. 11, p. 1427-1436, Nov. 2012.
- WOLF, J. B. et al. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. **Genetics**, Bethesda, v. 171, n. 2, p. 1-12, Oct. 2005.
- WOLF, J. B. et al. The contribution of epistatic pleiotropy to the genetic architecture of covariation among polygenic traits in mice. **Evolution Development**, Malden, v. 8, n. 5, p. 468-476, Sept./Oct. 2006.
- XU, Y. Developing marker-assisted selection strategies for breeding hybrid rice. In: JANICK, J. (Ed.). **Plant Breeding Reviews**. New York: J. Wiley, 2010. v. 23. p. 73-174.
- XU, Y.; ISHII, T.; MCCOUCH, S. Marker-assisted evaluation of germplasm resources for plant breeding. In: INTERNATIONAL RICE RESEARCH CONFERENCE, 24., 2003, Beijing. **Proceedings...** Beijing: International Rice Research Institute, 2003. p. 213-229.
- YAN, J. et al. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. **PLOS ONE**, [San Francisco], v. 4, p. 8451, Dec. 2009.
- ZHAO, Y.; METTE, M. F.; REIF, J. C. Genomic selection in hybrid breeding. **Plant Breeding**, [S.l.], v. 134, n. 1, p. 1-10, Feb. 2015.
- ZHONG, S. et al. Molecular mapping and marker-assisted selection of genes for septoria speckled leaf blotch resistance in barley. **Phytopathology**, Saint Paul, v. 96, n. 9, p. 993-999, Apr. 2006.

ZHU, Y. L. et al. Single-Nucleotide Polymorphisms in Soybean. **Genetics**, Bethesda, v. 163, n. 3, p. 1123-1134, Mar. 2003.

SEGUNDA PARTE - ARTIGOS

ARTIGO 1 Análise genômica e associação genética entre linhagens de milho

**Artigo formatado de acordo com a NBR 6022 (ABNT, 2003), conforme
exige o manual de normalização da UFLA.**

RESUMO

O uso da diversidade é a base do avanço genético via melhoramento. O estudo e o entendimento profundo da diversidade genética nos bancos de germoplasma levam a uma melhora da *performance* dos híbridos em termos de produtividade, resistência a doenças e outras características agronômicas. A análise de diversidade genética em germoplasma de milho pode ser utilizada para classificar acessos dentro de grupos heteróticos e também para elucidar a arquitetura genética de importantes características. Neste contexto, objetivou-se a partir deste trabalho a exploração das informações obtidas a partir da genotipagem de um banco de germoplasma proveniente de uma empresa de sementes visando o detalhamento do mesmo; a avaliação da diversidade genética e estrutura populacional da coleção e a investigação do potencial da coleção como fonte para estudo da arquitetura genética do caráter peso de espigas. Para isto, foi utilizada, como fonte de material genético, uma coleção de linhagens de milho. As amostras para extração de DNA foram coletadas de um experimento instalado na fazenda experimental da UFLA, sediada no município de Lavras - MG. Foram realizados o agrupamento genético das linhagens endogâmicas e a análise de associação genômica. Para isto, o caráter peso de espigas foi avaliado em 238 linhagens de milho, as mesmas foram avaliadas no ano agrícola de 2012/2013 nos municípios de Lavras e Uberlândia, ambos localizados no estado de Minas Gerais. Para a análise de associação genômica, foram utilizados os modelos Mistos e Stochastic Search Variable. Observou-se que o banco de germoplasma não apresentou frequências alélicas fixadas, evidenciado assim o potencial das linhagens para o melhoramento. Apesar das diferentes origens dos genótipos estudados, todos os acessos possuem, pelo menos, 66% de alelos idênticos por estado em comum. Também foi possível fazer a separação das linhagens em grupos heteróticos distintos, sendo, em sua maioria, condizente com as informações de *pedigree*. Essas informações poderão ser utilizadas para direcionar os cruzamentos a serem realizados no programa de melhoramento. Contudo, a população utilizada para o estudo de associação não foi adequada para a identificação de marcas de efeitos significativos para o caráter peso de espigas.

Palavras-chave: *Zea Mays*. Melhoramento Vegetal. Diversidade Genética. Associação Genômica. Peso de Espigas.

1 INTRODUÇÃO

O uso da diversidade genética é a base do avanço genético via melhoramento. Em milho, o germoplasma grande e diversificado permite a manipulação de diferentes genótipos que leva a uma melhora da *performance* dos híbridos em termos de produtividade de grãos, resistência a doenças e outras características agronômicas (LABORDA et al., 2005).

A análise de diversidade na coleção de germoplasma é importante para a pesquisa e melhoramento de culturas como o milho (XU; ISHII; MCCOUCH, 2003; MELCHINGER et al., 1990; BERNARDO, 2010; TROYER, 1990). A diferenciação de cultivares e a classificação das linhagens em grupos heteróticos; a identificação de falhas e redundâncias nas coleções; monitoramento de perdas que possam ocorrer durante a conservação do banco de germoplasma, regeneração, domesticação e melhoramento; identificação de novos alelos que possam ter ligação com caracteres agronômicos e a construção de conjuntos representativos ou coleções núcleo (LU et al., 2009) são informações importantes a serem conhecidas e estudadas em qualquer coleção de germoplasma.

A análise de diversidade genética em germoplasma de milho pode ser utilizada para classificar acessos dentro de grupos heteróticos. Linhagens dentro de um mesmo grupo devem ter um alto nível de similaridade e alto nível de dissimilaridade entre grupos (LU et al., 2009). O conhecimento da diversidade genética, da relação e da estrutura do germoplasma é importante no melhoramento de plantas por diferentes razões, entre elas para o auxílio na escolha das melhores combinações parentais para o desenvolvimento de progênies com máxima variabilidade genética (BARRETT; KIDWELL, 1998). No melhoramento de milho, esse conhecimento é essencial para o desenvolvimento de melhores linhagens e a criação de híbridos fenotipicamente

superiores, principalmente para o caráter produtividade de grãos (DUVICK, 2001; HALLAUER; CARENA; MIRANDA, 2010).

Grupo heterótico é definido como uma coleção de acessos altamente correlacionados que tendem a resultar em vigorosos híbridos quando ocorre o cruzamento entre linhagens de diferentes grupos (Lee, 1995). O conhecimento detalhado da relação entre linhagens de milho é importante não só para a seleção de parentais, mas também para o design de sistemas de melhoramento mais eficientes (LU et al., 2009).

A separação das linhagens em diferentes grupos heteróticos pode ser realizada por meio de relações de *pedigrees*, *performance* fenotípica para determinado caráter, adaptabilidade e estabilidade, *top-crosses*, cruzamentos dialélicos e distância genética estimada com caracteres morfológicos ou marcadores moleculares (BERTAN; CARVALHO; OLIVEIRA, 2007). A criação de grupos heteróticos em algumas coleções ainda é de alguma forma subjetiva e cada programa de melhoramento individualmente necessita de uma metodologia que funcione melhor dentro de suas estratégias de seleção e seus objetivos específicos (LU et al., 2009).

Nesse contexto, o desenvolvimento de marcadores moleculares de DNA gerou ferramentas promissoras para se acessar a diversidade genética entre linhagens e organizá-las em diferentes grupos heteróticos (SMITH et al., 1990; DUDLEY; SAGHAI MAROOF; RUFENER, 1991; LIVINI et al., 1992).

Os RFLPs foram os primeiros marcadores utilizados para este fim durante a década de 90 (BOPPENMAIER et al., 1992; MELCHINGER et al., 1990). Marcadores RAPDs e SSRs foram em seguida rapidamente adotados para a utilização em análises genéticas por seus menores custos e maior eficiência (LABORDA et al., 2005). Como consequência, vários estudos foram realizados a fim de acessar a diversidade genética em várias espécies, como cana-de-açúcar (LIMA et al., 2002), trigo (MANIFESTO et al., 2001), algodão (ABDALLA et

al.; 2001), soja (PRIOLLI et al., 2002), sorgo (LABORDA et al., 2005) e milho (LU; BERNARDO, 2001; GETHI et al., 2002).

Mais recentemente surgiram os marcadores SNPs e Diversity Arrays Technology (Darts). Devido ao baixo custo por marca, alta abundância no genoma, especificidade de locos, potencial de alta densidade de marcas, baixos níveis de erros (RAFALSKI, 2002; CHAGNE et al., 2007) e facilidade de automação (VAN INGHELANDT et al., 2010) esses marcadores têm emergido como poderosas ferramentas para várias aplicações dentro da genética, incluindo estudos de diversidade (ZHU et al.; 2003) e de associação genômica.

Os marcadores Darts apresentam a vantagem de detecção da variação em milhares de locos gênicos sem a necessidade de sequenciamento prévio (WENZL et al., 2006). O polimorfismo detectado nos Darts incluem SNPs, inserções-deleções (InDels) e modificações de metilação herdáveis (JACCOUD et al., 2001). Características essas que têm feito com que esses marcadores se tornem ferramentas interessantes em estudos de diversidade e também de associação.

Os estudos de associação genômica são metodologias recentes para identificação de marcas correlacionadas a caracteres. Em contraste com o mapeamento baseado em populações biparentais segregantes, que explora a diferença alélica entre apenas dois parentais, a associação genômica utiliza a variação genética entre um vasto número de genótipos. Essencialmente, a associação genômica explora a recombinação histórica e evolutiva em nível populacional (THORNSBERRY et al., 2001; REMINGTON et al., 2001). Assim, espera-se que a variação desta característica ocorra entre os diferentes genótipos em contraposição à variação advinda de apenas dois parentais tendo a expectativa que possa ser mais bem mapeada (MATTHIES et al., 2012; YU; BUCKLER, 2006).

Contudo, apesar do extensivo conhecimento de coleções de germoplasmas norte-americanos de milho, relativamente poucos estudos de diversidade e de associação mais detalhados, envolvendo grande número de marcas e linhagens, têm sido realizados em coleções de germoplasmas tropicais (OLIVEIRA et al., 2004; REIF et al., 2004).

Assim, objetivou-se a partir deste estudo: 1) explorar as informações obtidas a partir dos marcadores Darts para detalhamento do banco de germoplasma; 2) avaliar a diversidade genética e a estrutura populacional da coleção e 3) investigar o potencial da coleção como fonte para estudo da arquitetura genética do caráter peso de espigas.

2 MATERIAL E MÉTODOS

2.1 Caracterização genética do banco de germoplasma

Foi utilizada como fonte de material genético uma coleção de linhagens de milho proveniente de uma empresa privada de melhoramento de produção de sementes de milho. As amostras para extração de DNA foram coletadas de um experimento instalado na safra de verão do ano de 2013 na fazenda experimental da UFLA, sediada no município de Lavras - MG. O experimento foi constituído de 470 linhagens identificadas previamente, provenientes de diferentes grupos heteróticos em delineamento de blocos incompletos em 47 blocos com 10 linhagens cada, com duas repetições e a parcela constituída de duas fileiras de 3 m de comprimento com espaçamento de 0,6 m entre fileiras. Foram semeadas 25 sementes por linha e, posteriormente, feito o desbaste para 15 plantas por linha. As amostras de folhas foram coletadas de três plantas em estádio V2 na parcela.

A extração de DNA foi realizada seguindo protocolo específico da empresa Diversity Arrays Technology. As amostras de DNA, devidamente diluídas, foram enviadas para a empresa Diversity Arrays Technology (<http://www.diversityarrays.com/>) para genotipagem. A genotipagem das 470 linhagens foi realizada utilizando-se os marcadores Darts, porém devido a não amplificação de algumas amostras enviadas, foram obtidos os dados genotípicos para 447 linhagens. As marcas foram avaliadas como “1” para a presença e “0” para ausência do alelo e “-” quando não foi possível a leitura com confiabilidade pelo *software*. Marcadores perdidos foram imputados pela função *A.mat*, método *mean*, do pacote *rrBLUP* (ENDELMAN, 2011) do *software* R.

Para capturar as relações genéticas entre os indivíduos, foi elaborada a matriz de parentesco aditiva (**A**) seguindo a proposta de Vitezica, Varona e Legarra (2013):

$$A = \frac{W_A W'_A}{\sum pq}$$

Em que p é a frequência do alelo favorável e q a frequência do alelo desfavorável; W_A é a matriz de desvio dos marcadores.

O agrupamento genético das linhagens endogâmicas foi promovido pela decomposição espectral da matriz de parentesco A e, em seguida, foram plotados os dois primeiros componentes principais. Posteriormente foi observada a consistência entre o agrupamento genético utilizando os marcadores moleculares e os grupos heteróticos, definidos de acordo com as informações de *pedigree*.

Para o cálculo do desequilíbrio de ligação (r^2) foi utilizado o pacote *LdcorSV* no *software* R (R CORE TEAM, 2015).

2.2 Obtenção e Análise dos dados Fenotípicos

O peso de espigas foi avaliado em 238 linhagens de milho e em quatro testemunhas. As linhagens e as testemunhas foram avaliadas no ano agrícola 2012/2013 em dois locais: no município de Lavras (910 m, 21°14'S e 45°00'W) e de Uberlândia (863 m, 18°55'S e 48°16') ambos localizados no estado de Minas Gerais, Brasil. Foi utilizado o delineamento experimental de blocos incompletos com testemunhas comuns. Cada bloco foi formado por dez tratamentos (oito regulares e dois comuns) e três repetições com parcelas experimentais compostas por uma fileira de três metros de comprimento com espaçamento de 0,7 m entre fileiras.

A metodologia de modelos mistos foi utilizada para ajuste das médias, via máxima verossimilhança restrita (REML), utilizando o algoritmo EM (expectation- maximization) (PATTERSON; THOMPSON, 1971). As médias

foram corrigidas considerando os efeitos de repetições, ambientes, interação genótipos x ambientes e resíduo, adotando-se o modelo:

$$y = X\beta + Tg + \Omega b + W\delta + e$$

Em que y é o vetor de observações $n \times 1$; X é uma matriz $n \times p$ de incidência dos efeitos fixos (repetição dentro de local e local); T é uma matriz $n \times q$ de incidência dos efeitos genéticos; Ω é uma matriz de incidência dos efeitos aleatórios de blocos dentro de locais e repetições; W é uma matriz de incidência dos efeitos de interação linhagens x ambientes; e g , b , δ , são vetores dos efeitos relacionados a X, T, Ω, W respectivamente e e são os efeitos residuais. As distribuição dos efeitos g , b , δ e e são assumidas como $N(0, \sigma_g^2)$, $N(0, \sigma_b^2)$, $N(0, \sigma_\delta^2)$ e $N(0, \sigma_e^2)$, respectivamente.

2.3 Análise de Associação Genômica

2.3.1 Modelo GWAS

O modelo GWAS (YU et al., 2006) foi utilizado para análise de associação genômica. A análise foi realizada com a função *GWAS* do pacote *rrBLUP* (ENDELMAN, 2011) do *software* R. A matriz de parentesco de Vitezica, Varona e Legarra (2013) foi utilizada como estruturadora populacional. O *threshold* para significância foi dado pela correção de FDR a 5% utilizando o algoritmo de Benjamini e Hochberg (1995).

2.3.2 Modelo Bayesian Stochastic Search Variable (BSSV)

Dentre os modelos bayesianos propostos na literatura, o Bayesian Stochastic Search Variable (BSSV) foi empregado neste trabalho devido à sua grande capacidade de selecionar marcas de grande efeito em modelos com múltiplas marcas. Algumas modificações no modelo original proposto por Yi, George e Allison (2003) foram realizadas. O modelo é dado por:

$$\tilde{y} = \mu + \sum_{j=1}^m z_j a_j + e \quad (3)$$

Onde \tilde{y} é um vetor das médias correlacionadas; μ é o intercepto da amostra; z_{ij} é o genótipo do marcador j do indivíduo i ; a_j é o efeito do marcador j e e_i é o erro da observação i seguindo a distribuição $\mathbf{N}(0, \sigma_e^2)$.

A aceitação de um efeito de marca depende de uma mistura de prioris condicionadas a um conjunto de variáveis latentes ou indicadoras. Assim, pode-se assumir que os efeitos aditivos das marcas podem ser dados a priori por:

$$a_j | \rho, \Delta_j, \delta \sim (1 - \rho) N(0, \sigma_{a_j}^2) + \rho N(0, \delta), j = 1, \dots, K$$

Em que Δ_j e δ são valores de alta e baixa magnitude. Neste estudo assumiu-se que:

$$\Delta_j | a, b \sim \text{inversa-escalada} - \chi^2(v = 4, s^2 = 0.002) \text{ e } \delta = 10^{-6}$$

Os hiperparâmetros v e s^2 são relacionados ao método de Bayes A. O $\delta = 10^{-6}$ corresponde à herdabilidade individual do marcador considerando 1% de variância fenotípica.

Outra modificação no método SSVS original foi assumir que o parâmetro $\rho | a, b \sim Beta(a = 1, b = 1)$ ao invés de 0,5 como em Yi, George e Allison (2003).

A distribuição a priori para os efeitos da média populacional foi assumida como constante e para a variância residual σ_e^2 assumiu-se a mesma priori de Δ_j .

A integração das distribuições de condicionais posterioris foi realizada utilizando algoritmo Monte Carlo Cadeias de Markov via amostragem de Gibbs (CASELLA; GEORGE, 1992), descrito pelos seguintes passos:

Amostrar μ da distribuição a posteriori condicional completa:

$$p(\mu | \dots) \sim N \left[\sum_{i=1}^n \left(y_i - \sum_{j=1}^k x_{ij} a_j \right), \frac{\sigma_e^2}{n} \right]$$

Amostrar a_j da distribuição a posteriori condicional completa:

$$p(a_j | \dots) \sim N \left[\left(\sum_{i=1}^n x_{ij}^2 + \frac{\sigma_e^2}{v_k} \right)^{-1} \sum_{i=1}^n x_{ij} \left(y_i - \mu - \sum_{j' \neq j}^k x_{ij'} a_{j'} \right), \left(\sum_{i=1}^n x_{ij}^2 + \frac{\sigma_e^2}{v_k} \right)^{-1} \sigma_e^2 \right]$$

Em que:

$v_k = \eta_k \Delta_k + (1 - \eta_k) \delta$, sendo $\eta: \{1, 0\}$ e $p(\eta) \sim Bernoulli(\rho)$ são todos os efeitos a_j , exceto o efeito em amostragem.

1. Amostrar Δ_k de uma inversa qui-quadrado com os seguintes parâmetros:

$$p(\Delta_j | \dots) \sim \text{inversa-escalada} - \chi^2(v + 1, a_j^2 + v s^2)$$

2. Amostrando η_j de uma distribuição Bernoulli dada por:

$$p(\eta_j = 1 | \dots) \sim \frac{\rho N(a_j | 0, \Delta_j)}{\rho N(a_j | 0, \Delta_j) + (1 - \rho)N(a_j | 0, \delta)}$$

3. Amostrando ρ_j de distribuição BETA utilizando a seguinte condicional:

$$p(\rho | \dots) \sim \text{Beta} \left[1 + \sum_{j=1}^k \eta_j, 1 + \left(k - \sum_{j=1}^k \eta_j \right) \right]$$

4. Amostrando a variância residual:

$$p(\sigma_e^2 | \dots) \sim \text{inversa-escalada} - \chi^2(v + n, \text{SSR} + v s^2)$$

Em que SSR é a soma de quadrados do resíduo.

5. Repetir as etapas descritas até atingir a convergência.

Testes de significância dos efeitos dos marcadores foram executados pelo Teste de Wald. A estatística desse teste $W(\lambda)$, em condições de hipótese de nulidade, segue uma distribuição assintótica χ^2 com um grau de liberdade. Os valores do teste podem ser obtidos por:

$$W(\lambda) = \frac{a_j^2}{\sigma_{a_j}^2}$$

O nível de significância mínimo para uma marca apresentar valor diferente de zero foi dado como de 5%.

3 RESULTADOS

O conjunto de germoplasma avaliado neste experimento foi composto de 447 linhagens de milho oriundos de um banco de germoplasma de propriedade de uma empresa de melhoramento e produção de sementes instalada no Brasil. A maioria destas linhagens tem origem em áreas tropicais do mundo, incluindo o Brasil e outros países. Algumas linhagens são provenientes de áreas subtropicais e temperadas de vários países.

Os acessos foram sequenciados uma vez com uma amostra representativa de três plantas para extração de DNA, resultando em uma amostra de cada linhagem para a análise genotípica.

A análise genotípica realizada nas linhagens gerou 23.153 marcas Darts distribuídas ao longo de nove cromossomos além de ser analisado também o DNA mitocondrial e dos plastídios. A análise genotípica no décimo cromossomo não foi realizada pela empresa Dart Arrays Technology.

A distribuição dos marcadores variou de acordo com cada cromossomo e a reprodutibilidade média das leituras foi de 99,5%. Na maioria dos casos, as marcas estão distribuídas ao longo de nove grupos de ligação do milho. Pode-se observar que existe maior frequência das marcas nas regiões teloméricas dos cromossomos e menor concentração na região dos centrômeros (Figura 1).

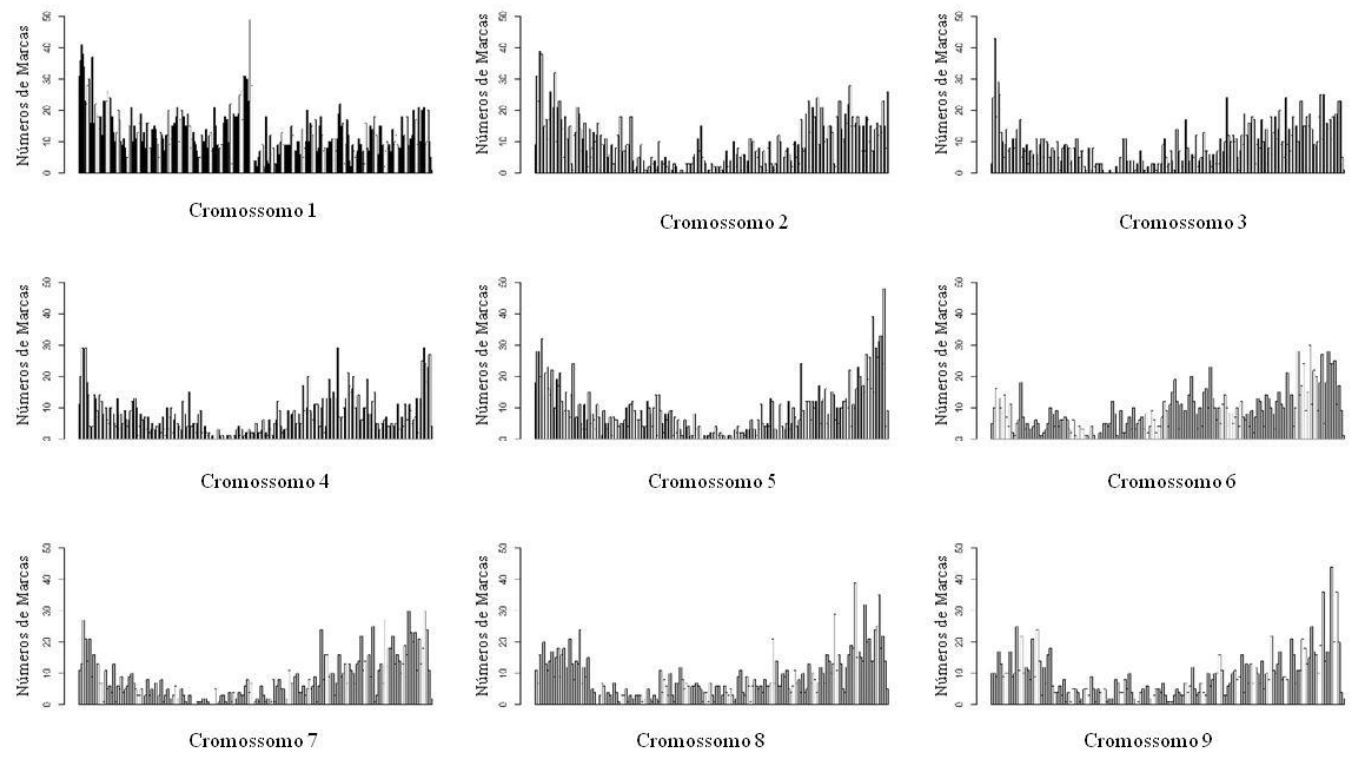


Figura 1 Distribuição dos marcadores Darts a cada 100 kb ao longo de nove grupos de ligação nas 470 genótipos de milho oriundas do banco de linhagens de uma empresa de produção e melhoramento de milho

O número de marcas por cromossomos variou de 3.612 a 1.386 para os cromossomos 1 e 7, respectivamente (Tabela 1). O cromossomo 1 apresentou maior quantidade de marcas devido ao seu maior tamanho em relação aos demais (http://ensembl.gramene.org/Zea_mays/Location/Genome). Também foi analisada uma pequena quantidade de marcas nas mitocôndrias e nos plastídios. Calculou-se o comprimento relativo das marcas para cada cromossomo, utilizando para isto a relação entre a quantidade de bases de todas as marcas pela quantidade de bases em cada cromossomo. A cobertura variou de 0,07 a 0,04%. A maior cobertura foi observada no cromossomo 1, o que pode ser indicativo de maior precisão, maior capacidade de identificar as variações no genoma.

Tabela 1 Número médio de marcas por grupo de ligação e seus respectivos comprimentos parciais presentes no conjunto de linhagens de milho

Cromossomo	Dart	
	Marcas	Comprimento parcial (%)
1	3612	0,07
2	2107	0,06
3	1896	0,05
4	1706	0,04
5	1904	0,05
6	1459	0,05
7	1386	0,05
8	1522	0,05
9	1415	0,06
Mitocôndria	4	0,04
Plastídios	1	0,05
Desconhecidos	6141	
Total/Média	23153	0,05

Observou-se boa distribuição da frequência alélica do marcador Dart (Figura 2), com grande parte das marcas apresentando frequência alélica entre

0,2 e 0,4, sendo aproximadamente 32% do total. Vale salientar que não foram observados alelos fixados, o que demonstra a variabilidade para seleção em todos os locos considerando as linhagens do banco de germoplasma. Com relação aos alelos raros na população, observou-se aproximadamente 6% com frequência inferior a 0,1.

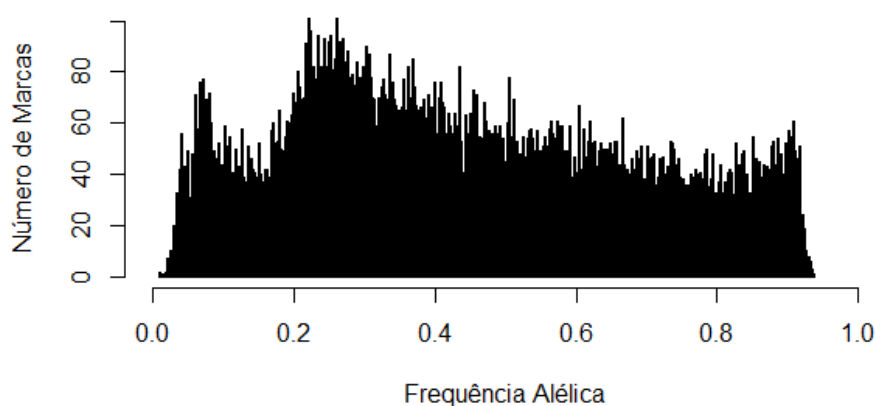


Figura 2 Distribuição da frequência alélica dos marcadores Darts ao longo dos nove grupos de ligação e também no DNA plasmídico e mitocondrial presentes no conjunto de linhagens de milho

O *call rate* é a quantidade de marcas úteis, onde foi possível a identificação dos alelos, excluindo-se as marcas onde não foi realizada leitura, quer seja devido à falha no processo de PCR, quer seja pela ausência da sequência no genoma. Os marcadores Darts apresentaram *call rate* médio de 0,94, com valores variando de 0,63 a 0,98. Esse valor pode ser devido à natureza dos marcadores e a fonte de variabilidade utilizada como referência para a criação do conjunto de marcas utilizadas (Figura 3a). Não foi observada grande variação do *call rate* entre os grupos heteróticos (Figura 3b). Esta informação

pode ser indicativa de que as marcas sem leitura devam ser devidas à falha da metodologia e à ausência da sequência nas linhagens.

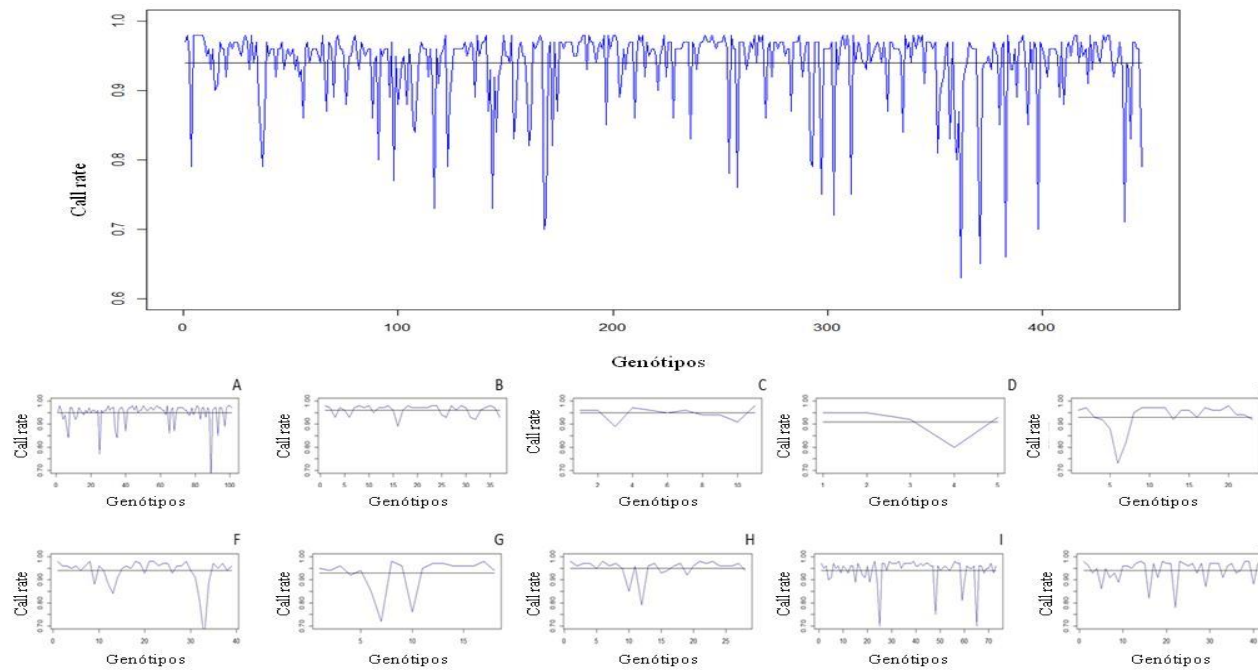


Figura 3 a) *Call Rate* nos diferentes genótipos para os marcadores Darts (linha azul) com sua respectiva média (linha preta). b) Na segunda parte, são apresentados os *call rates* dos grupos heteróticos individualmente: A: SSS (variância=0,0021); B: Amarillo Dentado (variância = 0,0004); C: Fêmea DK-Arg (variância = 0,0007); D: Iodente (variância = 0,0039); E: Iodente (variância = 0,0033); F: Lancaster (variância = 0,0037); G: Macho NK-Arg (variância = 0,0054); H: NSSS (variância = 0,0016); I: SSS (variância = 0,0030); J: Suwan (variância = 0,0021)

Nos gráficos apresentados na Figura 4, pode-se observar o rápido decréscimo do desequilíbrio de Ligação (DL) com o distanciamento das marcas. O DL cai praticamente pela metade já na distância de 30 kpb, se estabilizando a partir deste ponto. Esta informação indica a necessidade de grande número de marcas no mapeamento do genoma de milho para aumentar a chance de uma das marcas estar ligada com genes de interesse, devido ao baixo DL observado em pequenas distâncias.

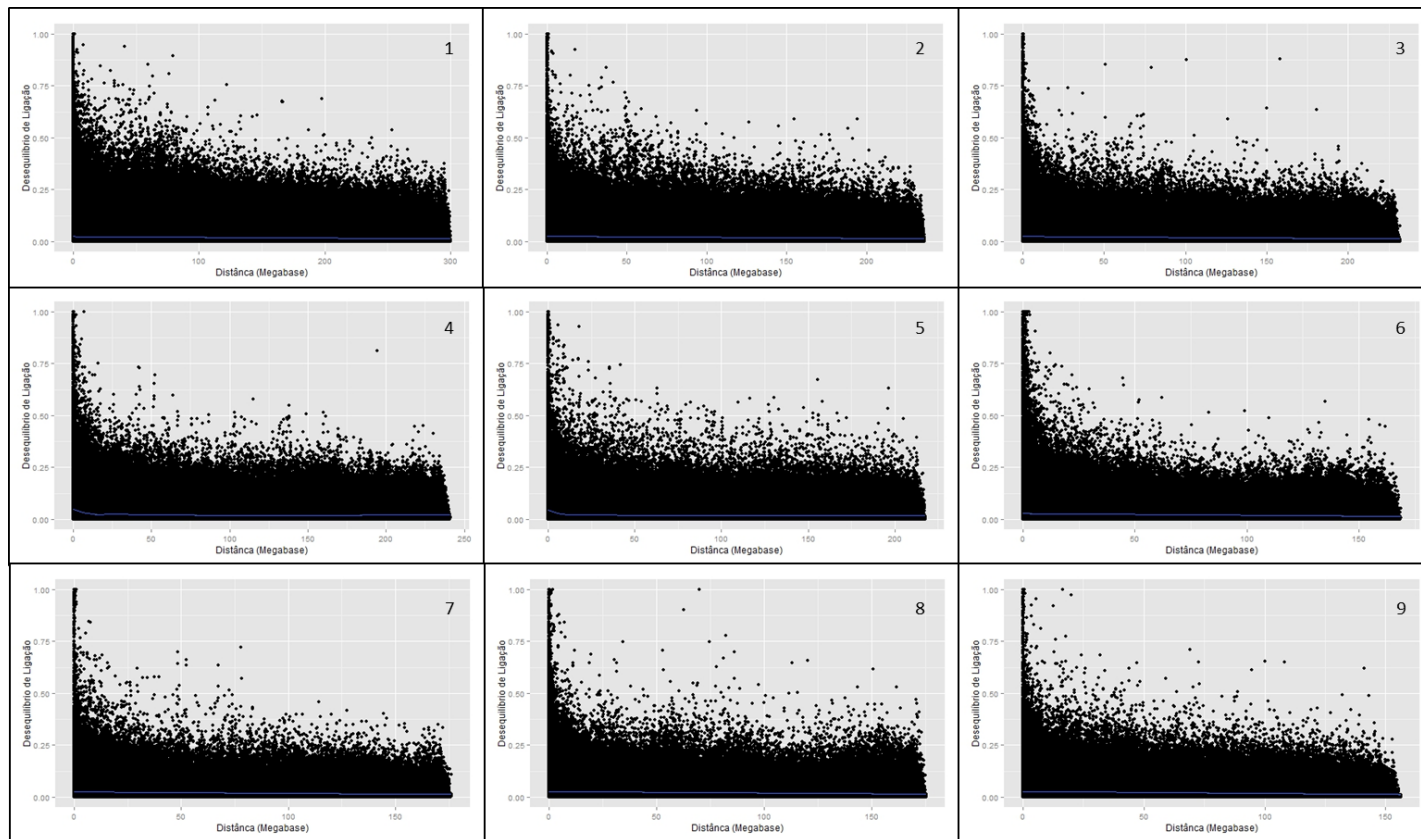


Figura 4 Decréscimo do desequilíbrio de ligação (DL) entre as linhagens de milho para os marcadores Darts nos nove grupos de ligação genotipados

Calculou-se também a proporção de marcas idênticas por estado (IBS) para todos os pares de linhagens (Figura 5). Os dados dos marcadores Darts demonstraram que todas as linhagens compartilham pelo menos 66% das marcas. Foi também observado que 85 das 447 linhagens genotipadas compartilham 95% ou mais dos alelos com pelo menos um genótipo. Este resultado demonstra que mesmo linhagens de diferentes grupos heteróticos tendem a compartilhar os mesmos alelos, evidenciando o impacto do melhoramento e seleção de alelos nos genótipos ao longo do tempo.

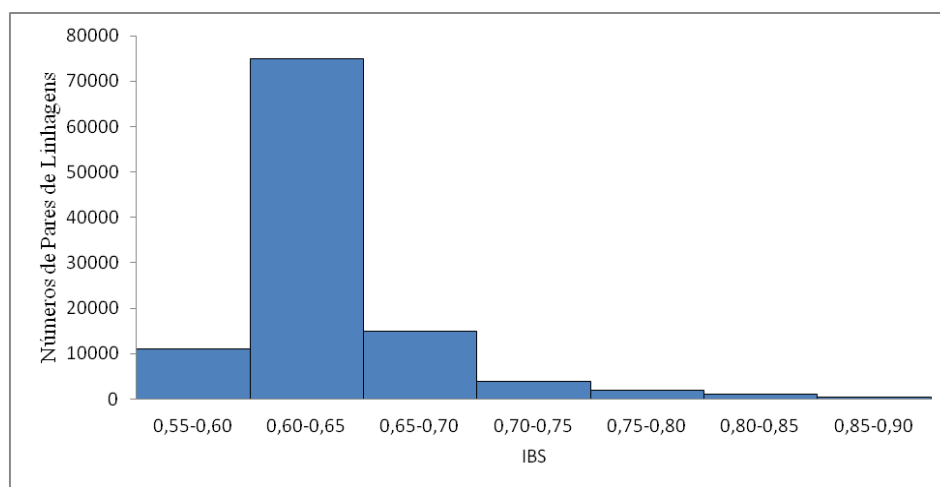


Figura 5 Distribuição da frequência dos alelos idênticos por estado (IBS) entre as 447 linhagens do banco de acesso utilizado

Este programa de melhoramento conta com diferentes linhagens de grupos heteróticos diversos, isso pode ser observado claramente na estratificação do germoplasma (Figura 6). As linhagens foram agrupadas de acordo com a similaridade entre as mesmas, tomando como limite inferior para as análises a similaridade de 0,3.

Na Figura 6, pode-se observar que a análise de componentes principais, utilizando os marcadores Darts, foi capaz de separar os grupos heteróticos, em

sua maior parte, de acordo com seu *pedigree*. Foi possível observar a separação de 10 grupos heteróticos no germoplasma avaliado: Iodent, Lancaster, Suwan, Non Stiff Stalk Synthetic (NSSS), Stiff Stalk Synthetic (SSS), Tropical Dent, Tropical Flint, Amarillo Dentado, Fêmeas DK-Argentina e Machos NK-Argentina. Vale ressaltar que foram utilizados grandes grupos heteróticos e que dentro de cada um deles existem subdivisões de acordo com o país de origem ou mesmo devido ao genótipo que originou a linhagem. Como exemplo disso podem ser citadas as linhagens B18_2, B29_1, B8_9, B1_3 que formam um grupo disperso do grupo heterótico SSS, porém todas apresentam em comum a mesma linhagem genitora (B73).

Foi possível também observar a separação do germoplasma em duas partes (Figura 7). A primeira parte contém os grupos heteróticos tropicais (Tropical Dent e Flint, Suwan e Amarillo Dentado). A outra parte contém os grupos heteróticos temperados (SSS, NSSS, Iodent, Lancaster, F-Arg e M-Arg), demonstrando a eficiência do método na separação de genótipos distintos. Pode-se observar ainda a separação de alguns grupos heteróticos de forma mais eficiente, como por exemplo, os grupos SSS, F-Arg, M-Arg, Amarillo dentado e Iodent.

As linhagens com as cores brancas são genótipos nos quais não existem informações de *pedigree*, neste sentido a análise ajuda a classificá-las dentro de grupos heteróticos e direcionar os cruzamentos a serem realizados com as mesmas.

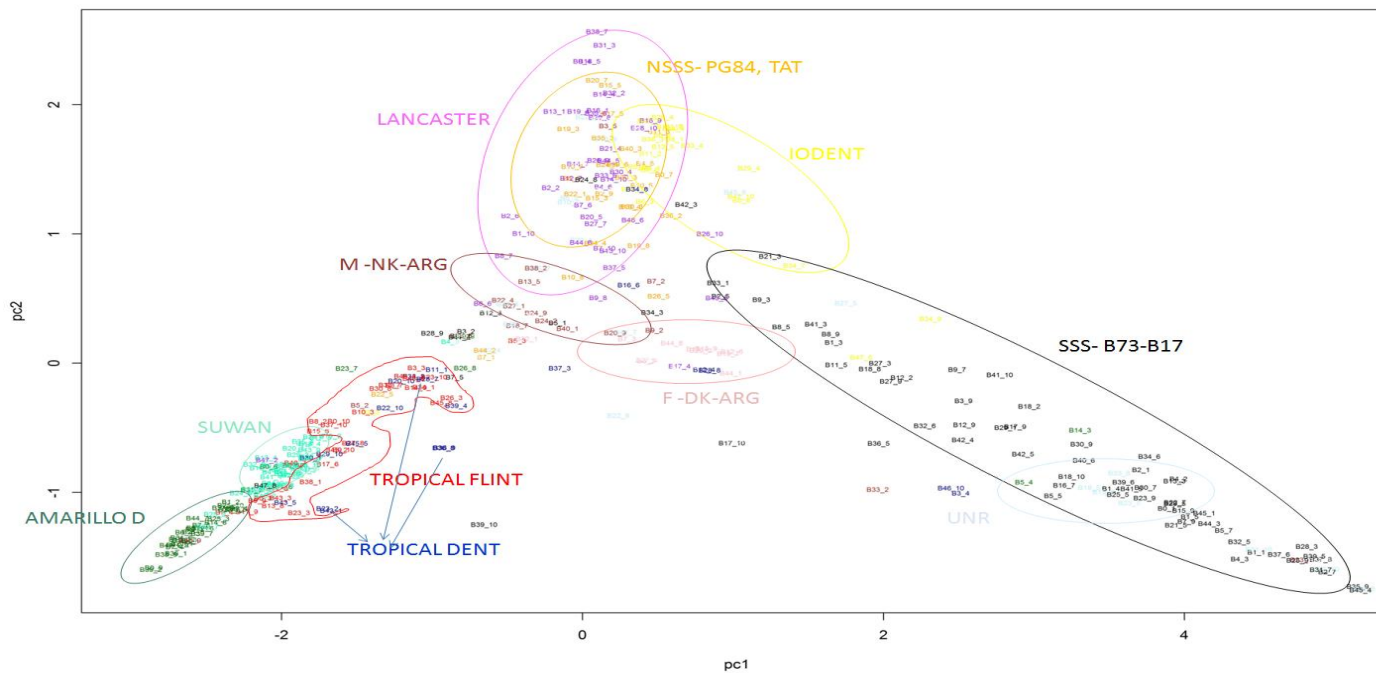


Figura 6 Separação dos grupos heteróticos do banco de germoplasma do programa de melhoramento de milho do Departamento de Agricultura – UFLA, pela decomposição espectral da matriz de parentesco **A** e plotagem dos dois primeiros componentes principais, que absorveram 18% da variância

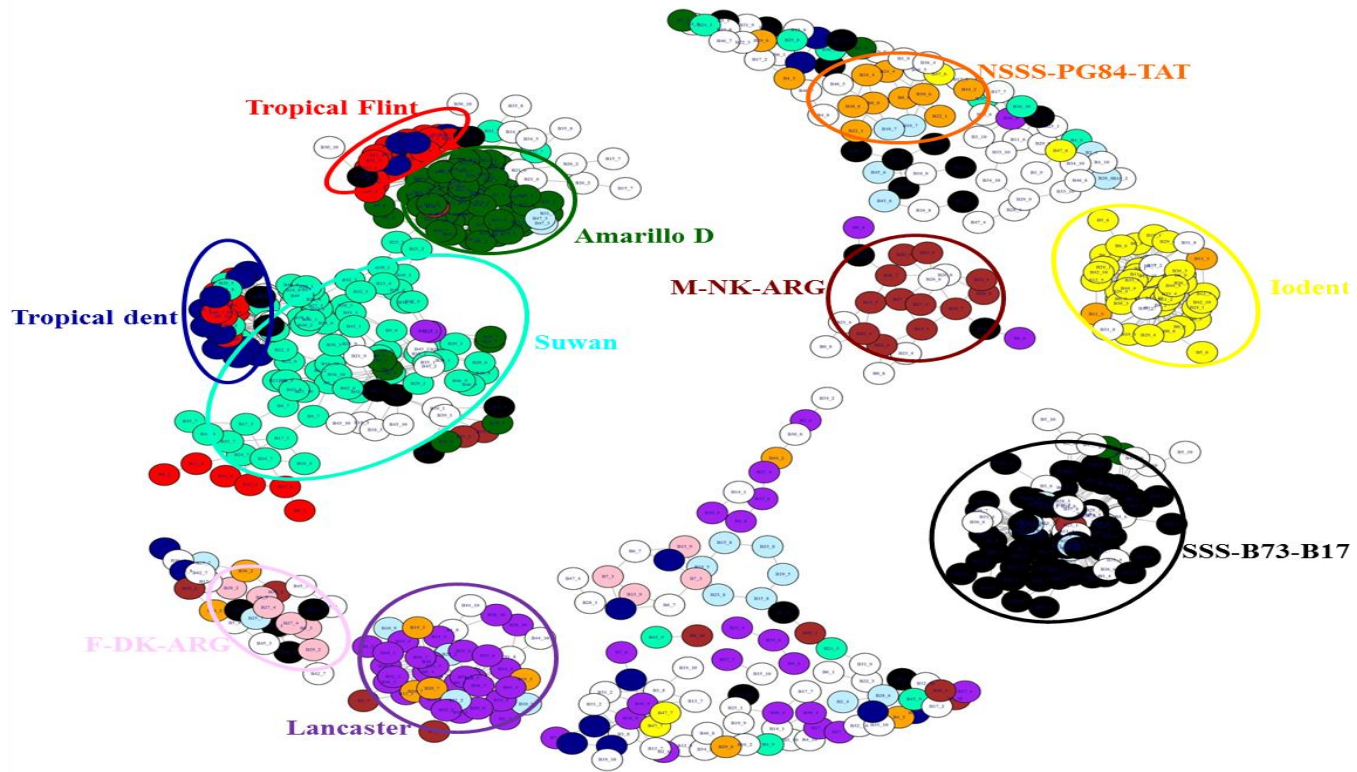


Figura 7 Separação dos grupos heteróticos do banco de linhagens de uma empresa de melhoramento e produção de sementes de milho pela análise de agrupamento hierárquico

Neste estudo também foi realizada a associação genômica ampla para o caráter peso de espigas. Foram utilizados dois modelos para a análise de associação genômica: Bayesian Stochastic Search Variable (SSV), que é uma metodologia bayesiana, que considera variâncias individuais por marcas e o RR-BLUP que considera variâncias homogêneas das marcas. Quando se utilizou o modelo RRBLUP nenhuma marca significativa para o caráter peso de espigas foi identificada (Figura 8). Duas considerações importantes devem ser ressaltadas: a primeira é o efeito de encolhimento que é inerente ao método RRBLUP, visto a adoção de uma única variância para todas as marcas utilizando assim, esta variância comum, para realizar o teste de Wald. Outro fato a ser considerado é a estrutura do caráter estudado. A produtividade é considerada uma característica poligênica infinitesimal com poucos genes de grandes efeitos e muitos genes de pequenos efeitos (LYNCH; WALSH, 1998).

O modelo utilizado para a identificação de QTLs foi o aditivo, contudo é sabido que a dominância tem forte impacto na produtividade de espigas. Considerando a estrutura de população utilizada no estudo, o número de marcas e o modelo aditivo, não foi possível a identificação de marcas de grandes efeitos.

Já quando se utilizou o modelo SSV para o estudo de associação genômica, foi possível a identificação de uma marca significativa, localizada no cromossomo 5, na posição 17227941pb. O efeito dessa marca foi de aproximadamente -190g/espiga. A identificação de somente uma marca de efeito significativo de acordo com o teste de Wald pode ter sido devido a várias razões. Uma delas é que o teste utilizado em conjunto com a estrutura da população (linhagens) pode ter tido poder suficiente para a identificação de somente uma marca com grande ligação com o caráter, considerando mais uma vez o efeito poligênico infinitesimal do mesmo. Outro motivo pode ser devido à utilização do teste de Wald para verificar a significância da marca. O teste de Wald é conservativo e, por vezes, dificulta a identificação de marcas significativas.

Outra questão é o tamanho da população utilizada para análise, apenas 232 linhagens. Por fim, pode-se também levantar a hipótese da presença de “outlier” que fez com que somente uma marca apresentasse significância.

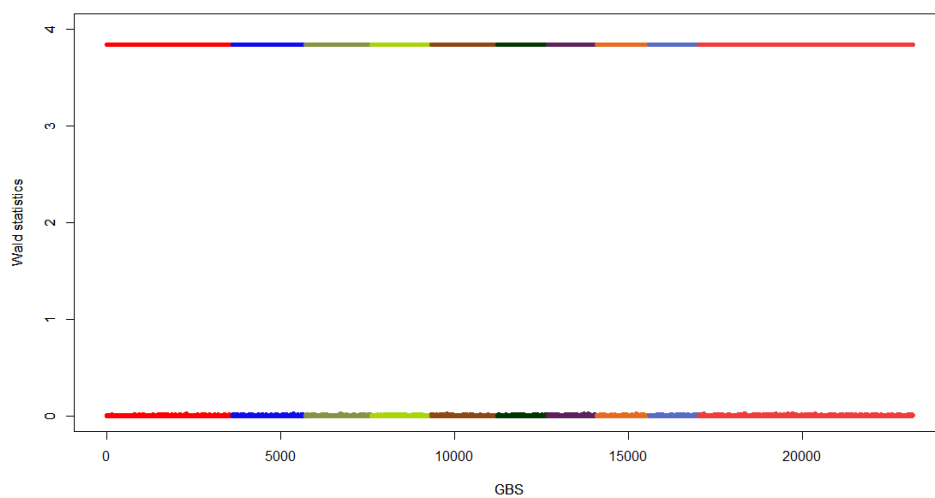


Figura 8 Efeitos das marcas para peso de espigas utilizando o método RRBLUP, com a linha de significância pelo teste de Wald (3,84)

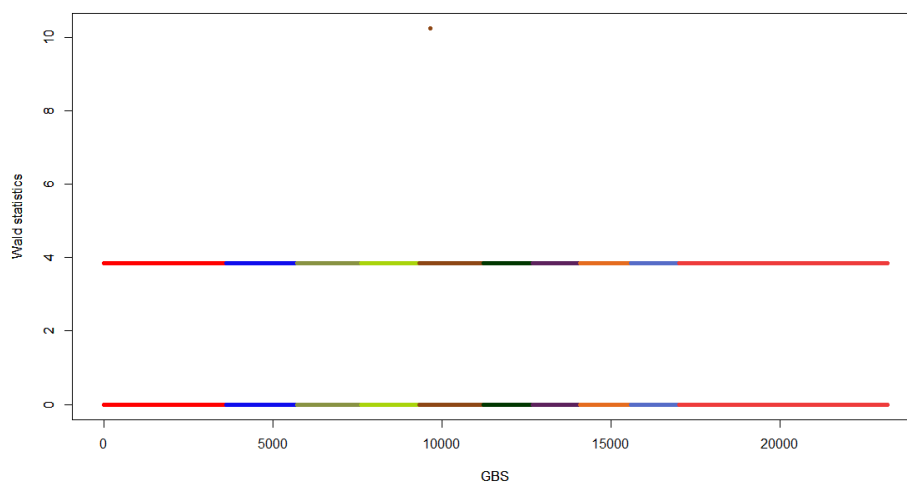


Figura 9 Efeitos das marcas para peso de espigas utilizando o método SSV, com a linha de significância pelo teste de Wald (3,84)

4 DISCUSSÃO

O conhecimento profundo da estrutura genética do germoplasma faz-se importante para direcionar os cruzamentos e aumentar o ganho genético nos programas de melhoramento. Para isso, genotipagens com grande número de marcas vêm sendo cada vez mais adotadas, influenciadas pela diminuição de seus custos, permitindo assim a exploração de diversas informações referentes ao germoplasma estudado, como separação de grupos heteróticos, frequência alélica, similaridade de alelos, estudos de associação, dentre outros (COOPER et al., 2014).

Foram gerados 23.153 marcadores Darts para genotipagem das 447 linhagens do banco de germoplasma. A posição dessas marcas nos cromossomos indica boa qualidade das informações dos dados genotípicos visto que a maioria delas se encontra nas extremidades dos cromossomos, onde é observada a maior quantidade dos genes no genoma do milho (SCHENABLE et al., 2009).

É importante examinar não só a estrutura genética das populações, mas também a diversidade entre elas (VAN INGHELANDT et al., 2010), pois o ganho de seleção em longo prazo depende da variabilidade genética existente. O fato de não ter sido evidenciado a fixação de alelos apesar da reciclagem de linhagens, que é um processo amplamente utilizado em todos os programas de melhoramento de milho e que tende a reduzir a diversidade existente (GOODMAN, 1999), demonstra que a variabilidade vem sendo mantida e que a seleção para qualquer marca é possível.

Durante o processo de genotipagem, pôde ser observada a presença de leituras nulas. Chamou-se de *call rate* a proporção de leituras válidas, ou seja, um menos a quantidade de leituras nulas. Essas leituras nulas podem ser tanto devido à ausência de certas regiões do genoma em algumas linhagens analisadas, visto que para o desenvolvimento do conjunto de marcas utiliza-se um genótipo

base, que muitas vezes possui sequências únicas para si, ou mesmo devido a erros durante o processo de genotipagem (Romay et al., 2013).

Schenable et al. (2009) relata que o genoma do milho apresenta amplas variações estruturais e uma variação importante é a presença/ausência de sequências entre diferentes genótipos. Como exemplo pode ser citado o cromossomo 6, que apresenta uma região de aproximadamente 2 Mb intacta que contém cópias únicas de genes que são presentes em uma linhagem e não em outras. Esses haplótipos específicos contribuem para a heterose e para a grande variação fenotípica entre as linhagens (SPRINGER et al., 2009).

Foi observado que a cobertura média ou *call rate* foi superior a encontrada por Romay et al. (2013) utilizando marcadores SNPs, que foi de 35%, quando as amostras foram sequenciadas uma vez. Para o mesmo cenário, porém com a utilização de marcadores Dart, o *call rate* médio neste trabalho foi de 94%, demonstrando a qualidade das análises quando da utilização deste marcador.

O estabelecimento de grupos heteróticos para direcionar os cruzamentos em um programa de melhoramento de milho é crucial para o sucesso no desenvolvimento de novos híbridos (BARATA; CARENA, 2006). As estimativas de diversidade genética auxiliam na classificação do germoplasma em grupos heteróticos para o melhoramento genético de híbridos (MENZ et al., 2004). A separação em grupos heteróticos maiores utilizada neste estudo foi escolhida pela maior simplicidade de organização do germoplasma como um todo, podendo se estender à análise para subgrupos visando facilitar o cruzamento para geração de novas linhagens.

A concordância dos resultados obtidos utilizando o método de componentes principais com as informações de *pedigree*, revelando grupos bem definidos, neste caso dez grupos heteróticos, está de acordo com os resultados de Maurer et al. (2006) e Van Inghelandt et al. (2010) que também relataram boa

concordância entre informações de *pedigree* e de marcas. Este resultado indica que as informações de *pedigree* deste programa de melhoramento descrevem confiavelmente a estrutura da população para a maioria dos acessos. Assim, esses grupos heteróticos são a base da diversidade genética do programa de melhoramento (VAN INGHELANDT et al., 2010).

Romay et al. (2013), utilizando esta mesma metodologia de separação dos grupos heteróticos em um conjunto de 2.815 linhagens de milho e mais de 600.000 marcadores SNPs, não obtiveram um agrupamento consistente como pôde ser observado neste estudo, mesmo utilizando maior densidade de marcas. Este fato pode ser devido à utilização da matriz euclidiana de distância genética pelos autores em contrapartida à matriz de parentesco unificado A , utilizada neste estudo como fonte de informação para decomposição espectral, demonstrando assim a aplicabilidade e superioridade da metodologia aplicada.

Em geral foi observada maior distância genética entre grupos que dentro dos grupos. Contudo, para uma parte das linhagens, a informação de *pedigree* não esteve em acordo com a análise de componentes principais. Este fato pode ser parcialmente explicado por informações de *pedigrees* errôneas ou mesmo incompletas, especialmente das linhagens advindas de autofecundação de híbridos comerciais de empresas de sementes e também por informações de *pedigree* de linhagens selecionadas em cruzamentos intergrupos nos quais a localização das mesmas em grupos heteróticos se torna incerta (VAN INGHELANDT et al., 2010).

O resultado obtido neste estudo corrobora com os resultados de Melchinger et al. (1990), Smith et al. (1997) e Van Inghelandt et al. (2010) que afirmam que os marcadores moleculares permitem uma classificação mais coerente dos grupos heteróticos que informações de *pedigrees*. Contudo, segundo Melchinger (1999), para genótipos de origem mista, a classificação baseada nos marcadores moleculares deve ser confirmada com dados de campo

de capacidade específica de combinação com testadores de diferentes grupos heteróticos.

A variação não genética, como o processo de extração de DNA, a reação de PCR ou análise dos dados dos marcadores (ROMERO-SEVERSON et al., 2001), também podem fazer com que o agrupamento sofra falhas. Outros fatores genéticos como heterozigidade residual na linhagem, seleção divergente, deriva genética, mutações ocasionais, contaminação na polinização, amostragem do genoma e número de marcas (NEI, 1987; TIVANG; NIENHUIS; SMITH, 1994) também podem influenciar nas falhas.

Para a produção de linhagens, é importante que se obtenha o cruzamento dentro de cada grupo heterótico. De modo oposto, para a produção de híbridos deve-se realizar o cruzamento entre os diferentes grupos heteróticos (HALLAUER; CARENA; MIRANDA, 2010) e, nesse sentido, orienta-se a escolha de cruzamento entre grupos heteróticos mais distantes, a fim de que se aumentem as chances de complementação gênica e o aumento da heterose (HALLAUER; CARENA; MIRANDA, 2010). Diante dos resultados observados neste trabalho, para produção de híbridos, os cruzamentos entre os grupos heteróticos SSS x NSSS; SSS x Lancaster; NSSS x Lancaster, M-Arg x F-Arg seria preferível aos cruzamentos F-Arg x Lancaster, Iodent x NSSS, Iodent x M-Arg, por apresentarem maior distância entre si.

Além do conhecimento do posicionamento dos genótipos dentro de cada grupo, o conhecimento da estrutura de importantes caracteres também é interessante. O propósito da associação de QTL é descobrir a arquitetura genética de características chaves no programa de melhoramento. Os mesmos podem ser mapeados em famílias ou progênes segregantes F_2 de cruzamentos entre parentais divergentes para o caráter em questão ou em indivíduos não aparentados de uma mesma população.

A recombinação nas progênes F_2 gera novos haplótipos e pode desconectar as marcas do loco causal (MACKAY; STONE; AYROLES, 2009). Matthies et al. (2012) ressaltam que o estudo de associação genômica com genótipos diversos pode apresentar vantagens frente ao mapeamento com populações biparentais, como: avaliação da variação genética em vasto número de genótipos frente a somente dois parentais, mapeamentos com melhores resoluções e uso direto da variação evitando-se assim o gasto de tempo com condução das populações. Contudo, as metodologias estatísticas necessárias para a realização das análises devem ser mais complexas, visto que falsos positivos (erro tipo I) ou negativos (erro tipo II) entre marcas e caracteres podem ocorrer com maior frequência (FALUSH; STEPHENS; PRITCHARD, 2003).

O erro tipo I pode ser gerado pela estrutura da população avaliada (PRITCHARD; STEPHENS; DONNELLY, 2000). Se a presença de grupos correlacionados não for incluída no modelo, os mesmos podem gerar covariâncias entre indivíduos e uma tendência na estimação dos efeitos dos alelos. Já o erro tipo II pode ser devido a três fatores: baixa correlação entre as marcas e os genes devido à queda de LD levando a uma baixa estimação das verdadeiras associações, delineamento desbalanceado advindo da presença de alelos em frequências alélicas diversas e múltiplos problemas de teste (MATTHIES et al., 2012).

Assim, estes podem ser os motivos pelos quais não foram observadas marcas significativas na população estudada. Neste estudo foram utilizadas linhagens muitas das vezes aparentadas. Estas linhagens, por terem sido selecionadas e melhoradas ao longo do tempo, possuem blocos de haplótipos, com marcas de efeitos contrários próximas, anulando o efeito uma da outra, dificultando assim a observação de marcas com efeitos significativos (YU; BUCKLER, 2006).

Como pode ser observado na análise de associação de QTLs, os dados utilizados não foram capazes de gerar mapeamento acurado de características complexas, como peso de espigas. Isso porque quando dados moleculares são utilizados para o estudo de associação de QTLs, a probabilidade de se encontrar marcas causais é altamente dependente do contexto em que são estudados (Romay et al., 2013). Contudo, os estudos de associação genômica têm sido estabelecidos com sucesso para caracteres de alta herdabilidade (YU; BUCKLER, 2006).

A dissecação genética de caracteres quantitativos se depara com dois desafios principais: o poder para detecção e localização de QTLs e o contexto biológico no qual se encontra as associações genótipos-fenótipos. São necessárias avaliações fenotípicas acuradas e alta densidade de marcas moleculares em alguns milhares de indivíduos para o mapeamento de QTLs, com alta resolução requerida para separar QTLs intimamente ligados. Para resolução deste problema, faz-se necessária a repetição do mapeamento em diferentes safras e locais (MACKAY; STONE; AYROLES, 2009). Assim, o fato de terem sido utilizadas somente 238 linhagens com dados fenotípicos gerados em somente dois locais em uma mesma safra para verificação de QTLs pode ter sido um problema.

Produtividade de grãos ou espigas é uma característica poligênica, com poucos genes de grandes efeitos e muitos genes de pequenos efeitos. Isto dificulta a localização desses genes de maiores efeitos (STANGE et al., 2013), principalmente quando se leva em consideração que esses genes de efeitos maiores são raros. Existem fortes argumentos a favor do modelo de alelos raros, que assume que características quantitativas são controladas por alelos raros de grandes efeitos (PARK et al., 2011; GIBSON, 2012). Estudos de associação têm demonstrado que a variação em alguns caracteres é devido a alelos raros e que os mesmos poderiam explicar a parte adicional da herdabilidade perdida

(MORREL; BUCKLER; ROSS-IBARRA, 2012). Evidências de estudos com humanos e *D. Melanogaster* indicam que variações raras com baixa frequência alélica (menores que 5%) estão frequentemente associadas com a variação de caracteres quantitativos (CARBONE et al., 2006; WANG et al., 2007; COHEN et al., 2004; JI et al., 2008 citados por MACKAY; STONE; AYROLES, 2009). Com isso, a proporção de alelos raros na população estudada pode ser outra causa de não ter sido encontrado QTLs associados à produtividade, visto que a maior parte da frequência alélica estava distribuída entre 0,2 e 0,4.

O único QTL identificado a partir do método SSV encontra-se na posição 17227941 no cromossomo cinco. Realizando-se a busca no banco de dados *Maize Data Base*, pode-se observar que o gene GRMZM2G110423 que transcreve duas proteínas (GRMZM2G110423_T01 e GRMZM2G110423_T02) se encontra dentro da sequência da marca de efeito significativo. Essas proteínas são classificadas como *ring zinc-finger*, que são encontradas em fatores de transcrição em plantas. Muitas dessas proteínas estão relacionadas à regulação de importantes processos biológicos, como desenvolvimento floral, morfogênese e resposta a patógenos (TAKATSUJI, 1998).

Segundo Sekhon et al. (2011), que analisaram a expressão de diversos genes ao longo do ciclo do milho e nos diferentes tecidos da planta, o gene GRMZM2G110423 apresenta maior expressão nas fases finais do ciclo, no estágio reprodutivo, apresentando maior expressão também em tecidos como grãos de pólen, sementes e palha da espiga.

O aumento na quantidade de marcas pode ajudar na identificação de QTLs de médios a pequenos efeitos. A precisão das estimativas dos efeitos, especialmente para QTLs menores, a precisão na localização bem como o poder para identificar QTLs ligados aumentam com o aumento da densidade de marcas. Esse fato levanta a questão de que somente QTLs de grandes efeitos estão segregando para a característica examinada nesta população ou, se QTLs

de pequenos efeitos estão presentes, os mesmos escaparam à detecção devido ao poder insuficiente da análise (STANGE et al., 2013).

Stange et al. (2013) ainda salientam que com a utilização de pequenas populações, que são rotineiramente utilizadas para análises de QTLs (de 100 a 200 genótipos), como foi o caso deste estudo, apenas QTLs principais são detectados. Isto é devido ao poder de detecção que não é suficiente para identificação de QTLs de médios e pequenos efeitos.

É importante ressaltar que em organismos que possuem alta diversidade de haplótipos, com alto decréscimo do desequilíbrio de ligação (DL) em pequenas distâncias, como foi observado neste estudo, mesmo utilizando-se grande quantidade de marcas, é possível que não se encontre uma marca em DL com um gene de razoável efeito na característica (ROMAY et al., 2013). Romay et al. (2013) ainda ressaltam que para germoplasmas de origem tropical, onde existe maior diversidade de haplótipos, são necessárias quantidades ainda maiores de marcas.

O DL é afetado pela taxa de recombinação, deriva genética, seleção, mistura de populações e parentesco (REMYNGTON et al., 2001). Estudos avaliaram diferentes níveis de DL em coleções de germoplasma gerando informações para estudos de associação. No presente estudo, o DL caiu rapidamente a partir da distância de 30 Kpb. Valor similar foi observado por Weng et al. (2011) que, analisando 284 linhagens de diversos *backgrounds*, observaram decréscimo do DL a partir de 27,7 kb. Valor ainda inferior foi observado por Yan et al. (2009), 5-10 Kb, analisando uma coleção global de milho contendo 632 linhagens.

Outro ponto importante a ser ressaltado é referente à qualidade dos dados experimentais utilizados para o estudo de associação genômica, visto que os dados de produtividade de espigas de linhagens de milho foram oriundos de ensaios inoculados com o fungo *Stenorcapella maydis*, que causa grão ardido,

afetando assim a produtividade e peso de espigas, o que resultou em dados de baixa qualidade, podendo assim ter influenciado na não associação de QTLs ao caráter.

A conclusão inevitável das últimas duas décadas de estudo de associação são que QTLs de grandes efeitos são raros e que a variação genética dos caracteres quantitativos é devido a muitos locos com efeitos individualizados ou agregados (com QTLs de efeitos opostos) muito pequenos para serem detectados (MACKAY; STONE; AYROLES, 2009).

Os bancos de germoplasma com seus acessos e informações genóticas podem ser melhores explorados, com caracterização mais profunda e utilização da associação genômica ampla para, assim, adquirir melhor entendimento dos genótipos presentes para que eles possam ser mais bem explorados nos programas de melhoramento (ROMAY et al., 2013).

5 CONCLUSÕES

O banco de germoplasma não apresenta frequências alélicas fixadas, evidenciado assim o potencial das linhagens para o melhoramento.

Apesar das diferentes origens das linhagens estudadas, todas possuem, pelo menos, 66% de alelos idênticos por estado em comum.

Foi possível fazer a separação das linhagens em grupos heteróticos distintos, sendo, em sua maioria, condizente com as informações de *pedigree*. Essas informações poderão ser utilizadas para direcionar os cruzamentos a serem realizados no programa de melhoramento.

O conjunto de dados utilizado para o estudo de associação não foi adequado para a identificação de marcas de efeitos significativos para o caráter peso de espigas de milho.

GENOMIC ANALYSIS A GENETIC ASSOCIATION BETWEEN CORN STRAINS

ABSTRACT

The use of diversity is the basis of genetic advancement by means of breeding. The profound study and understanding of genetic diversity in the germplasm banks leads to the improvement of hybrid performance in terms of productivity, resistance to diseases and other agronomic traits. The analysis of genetic diversity in corn germplasm can be used for classifying genotypes within heterotic groups, as well as to elucidate the genetic architecture of important traits. In this context, we aimed at exploring the information obtained from the genotyping of a germplasm collection originated from a seed company, aiming its detailing; at evaluating the genetic diversity and population structure of the collection; and at investigating the collections' potential as source for studying the genetic architecture of the corn ear weight trait. For this, we used this collection of corn inbred lines as source of genetic material. The samples for extracting DNA were collected from an experiment installed at the experimental farm at the Universidade Federal de Lavras (UFLA), located at the municipality of Lavras, state of Minas Gerais, Brazil. We performed the genetic grouping of the inbred lines as well as analysis of the genomic association. The trait corn ear weight was evaluated in 238 inbred lines, from 2012/2013, in the municipalities of Lavras and Uberlândia, both located in the state of Minas Gerais. For analyzing the genomic association, we used the Mixed and Stochastic Search Variable Models. We verified that the germplasm bank presented no fixed allele frequencies, demonstrating the breeding potential of the genotypes. Despite the different origins of the studied genotypes, all accesses presented, at least, 66% of identical alleles identical by state. It was also possible to separate the inbred lines into distinct heterotic groups, with the majority consistent with the pedigree information. This information can be used for directing the crosses to be conducted in the breeding program. However, the population used for studying the association was not adequate for identifying significant effect marks for the trait corn ear weight.

Keywords: *Zea mays*. Plant Breeding. Genetic Diversity. Genomic Association. Corn ear Weight.

REFERÊNCIAS

- ABDALLA, A. M. et al. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. **Theoretical and Applied Genetics**, Berlin, v. 102, n. 2, p. 222-229, Feb. 2001.
- BARATA, C.; CARENA, M. J. Classification of North Dakota maize inbred lines into heterotic groups based on molecular and testcross data. **Euphytica**, Wageningen, v. 151, n. 3, p. 339-349, Oct. 2006.
- BARRETT, B. A.; KIDWELL, K. K. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. **Crop Science**, Madison, v. 38, n. 5, p. 1261-1271, Nov. 1998.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society - Series B (Methodological)**, London, v. 57, n. 1, p. 289-300, 1995.
- BERNARDO, R. **Breeding for quantitative traits in plants**. Woodbury: Stemma, 2010.
- BERTAN, I.; CARVALHO, F. I. F.; OLIVEIRA, A. C. Parental selection strategies in plant breeding programs. **Jornal of Crop Science Biotechnology**, Seoul, v. 10, n. 4, p. 211-222, 2007.
- BOPPENMAIER, J. et al. Genetic diversity for RFLPs in European maize inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. **Plant Breeding**, [S.l.], v. 84, n. 5, p. 217-226, Aug, 1992.
- CARBONE, M. A. et al. Phenotypic variation and natural selection at Catsup, a pleiotropic quantitative trait gene in *Drosophila*. **Current Biology**, London, v. 16, n. 9, p. 912-919, May 2006.
- CASELLA, G.; GEORGE, E. I. Explaining the Gibbs Sampler. **The American Statistician**, Washington, v. 46, n. 3, p. 167-174, Aug. 1992.
- CHAGNE, D. et al. Single nucleotide polymorphisms genotyping in plants. In: ORAGUZIE, N. C. et al. (Ed.). **Association Mapping in Plants**. New York: Springer, 2007. cap. 5, p. 77-94.

COHEN, J. C. et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. **Science**, Washington, v. 305, n. 5685, p. 869-872, Aug. 2004.

COOPER, M. et al. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. **Crop and Pasture Science**, Victoria, v. 65, n. 4, p. 311-336, Apr. 2014.

DUDLEY, J. W.; SAGHAI MAROOF, M. A.; RUFENER, G. K. Molecular markers and grouping of parents in a maize breeding program. **Crop Science**, Madison, v. 31, n. 3, p. 718-723, May 1991.

DUVICK, D. N. Biotechnology in the 1930s: the development of hybrid maize. **Nature Reviews Genetics**, London, v. 2, p. 69-74, Jan. 2001.

ENDELMAN, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. **The Plant Genome**, Madison v. 4, n. 3, p. 250-255, Nov. 2011.

FALUSH, D.; STEPHENS, M.; PRITCHARD, J. K. Inference of population structure: extensions to linked loci and correlated allele frequencies. **Genetics**, Bethesda, v. 164, n. 4, p. 1567-1587, Aug. 2003.

GETHI, J. G. et al. SSR variation in important US maize inbred lines. **Crop Science**, Madison, v. 42, n. 3, p. 951-957, June 2002.

GIBSON, G. Rare and common variants: twenty arguments. **Nature Reviews Genetics**, London, v. 13, n. 2, p. 135-145, Jan. 2012.

GOODMAN, M. M. Broadening the genetic diversity in maize breeding by use of exotic germplasm. In: COORS, J. G.; PANDEY, S. (Ed.). **The Genetics and Exploitation of Heterosis in Crops**. Madison: American Society of Agronomy; Crop Science Society of America, 1999. p. 139-148.

HALLAUER, A. R.; CARENA, M.; MIRANDA, J. B. F. **Quantitative genetics in maize breeding**. 3^o ed. Iowa State University Press, Ames, 2010

HALLAUER, A. R.; RUSSELL, W. A.; LAMKEY, K. R. Corn breeding. In: Sprague, G. F.; Dudley, J. W. (Ed.). **Corn and corn improvement**. 3. ed. Madison: American Society of Agronomy; Crop Science Society of America, 1988. p. 463-564. (Agronomy, 18).

JACCOUD, D. et al. Diversity arrays: a solid state technology for sequence information independent genotyping. **Nucleic Acids Research**, Oxford, v. 29, n. 4, p. e25, Feb. 2001.

Ji, W. et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. **Nature Genetics**, New York, v. 40, n. 5, p. 592-599, Apr. 2008.

LABORDA, P. R. et al. Tropical maize germplasm: what can we say about its genetic diversity in the light of molecular markers? **Theoretical and Applied Genetics**, Berlin, v. 111, n. 7, p. 1288-1299, Nov. 2005.

LEE, M. DNA markers and plant breeding programs. **Advances in Agronomy**, San Diego, v. 55, p. 265-344, Dec. 1995.

LI, L. et al. Characterization of maize amylose-extender (*ae*) mutant starches. Part I: Relationship between resistant starch contents and molecular structures. **Carbohydrate Polymers**, Barking, v. 74, n. 3, p. 396-404, Nov. 2008.

LI, S. et al. Identification and characterization of the zinc-regulated transporters, iron-regulated transporter-like protein (ZIP) gene family in maize. **BMC Plant Biology**, Chichester, v. 13, n. 114, p. 1-14, Aug. 2013.

LIMA, M. L. A. et al. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum spp.*). **Theoretical and Applied Genetics**, Berlin, v. 104, n. 1, p. 30-38, Jan 2002.

LIVINI, C. et al. Genetic diversity of maize inbred lines within and among heterotic groups revealed by RFLP. **Theoretical and Applied Genetics**, Berlin, v. 84, n. 1-2, p. 17-25, June 1992.

LU, H.; BERNARDO, R. Molecular marker diversity among current and historical maize inbreds. **Theoretical and Applied Genetics**, Berlin, v. 103, n. 4, p. 613-617, Sept. 2001.

LU, Y. et al. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 1, p. 93-115, Dec. 2009.

LYNCH, M.; WALSH, B. **Genetics and Analysis of Quantitative Traits**. Sunderland: Sinauer Associates, 1998.

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

MANIFESTO, M. M. et al. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. **Crop Science**, Madison, v. 41, n. 3, p. 682-690, May 2001.

MATTHIES, I. E. et al. Population structure revealed by different marker types (SSR or DArT) has an impact on the results of genome-wide association mapping in European barley cultivars. **Molecular Breeding**, Dordrecht, v. 30, n. 2, p. 951-966, Aug. 2012.

MAURER, H. P. et al. Linkage disequilibrium between SSR markers in six pools of elite lines of an European breeding program for hybrid maize. **Maydica**, Bergamo, v. 51, p. 269-279, 2006.

MELCHINGER, A. E. et al. Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. **Crop Science**, Madison, v. 31, n. 3, p. 669-678, Dec. 1990.

MELCHINGER, A. E. Genetic diversity and heterosis. In: COORS, J. G.; PANDEY, S. (Ed.). **The Genetics and Exploitation of Heterosis in Crops**. Madison: American Society of Agronomy; Crop Science Society of America, 1999. p. 99-118.

MENZ, M. A. et al. Genetic diversity of public inbreds of Sorghum determined by mapped AFLP and SSR markers. **Crop Science**, Madison, v. 44, n. 4, p. 1236-1244, 2004.

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

MORRELL, P. L.; BUCKLER, E. S.; ROSS-IBARRA, J. Crop genomics: advances and applications. **Nature Reviews Genetics**, London, v. 13, p. 85-96, Feb. 2012.

NCBI: National Center for Biotechnology Information (website). Disponível em: <<http://www.ncbi.nlm.nih.gov/>>. Acesso em: 12 jul. 2015.

NEI, M. **Molecular evolutionary genetics**. New York: Columbia University, 1987.

OLIVEIRA, K. M. et al. Evaluating genetic relationships between tropical maize inbred lines by means of AFLP profiling. **Hereditas**, Lund, v. 140, n. 1, p. 24-33, 2004.

PARK, J. H. et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. **PNAS**, Washington, v. 108, n. 44, p. 18026-18031, Nov. 2011.

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.

PRIOLLI, R. H. G. et al. Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, Ribeirão Preto, v. 25, p. 185-193, 2002.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. J. Inference of population structure using multilocus genotype data. **Genetics**, Bethesda, v. 155, n. 2, p. 945-959, June 2000.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2012.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, [Amsterdam], v. 5, n. 2, p. 94-100, Apr. 2002.

REIF, J. C. et al. Genetic diversity determined within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by SSR markers. **Crop Science**, Madison, v. 44, n. 1, p. 326-334, Feb. 2004.

REMINGTON, D. L. et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. **PNAS**, Washington, v. 98, n. 20, p. 11479-11484, Sept. 2001.

ROMAY, M. C. et al. Comprehensive genotyping of the USA national maize inbred seed bank. **Genome Biology**, London, v. 14, n. 6, p. 1-18, June 2013.

ROMERO-SEVERSON, J. et al. Pedigree analysis and haplotype sharing within diverse groups of *Zea mays* L. inbreds. **Theoretical and Applied Genetics**, Berlin, v. 103, n. 4, p. 567-574, Sept. 2001.

SCHNABLE, P. S. et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. **Science**, Washington, v. 326, n. 5956, p. 1112-1115, Nov. 2009.

SEKHON, R. S. et al. Genome-wide atlas of transcription during maize development. **Plant Journal**, Oxford, v. 66, n. 4, p. 553-563, May 2011.

SMITH, J. S. C. et al. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. **Theoretical and Applied Genetics**, Berlin, v. 95, n. 1-2, p. 163-173, July 1997.

SMITH, O. S. et al. Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield, heterosis, and RFLPs. **Theoretical and Applied Genetics**, Berlin, v. 80, n. 6, p. 833-840, Dec. 1990.

SPRINGER, N. M. et al. Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content. **PLoS Genetics**, [San Francisco], v. 5, n. 11, Nov. 2009.

STANGE, M. et al. High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations. **Theoretical Applied Genetics**, Berlin, v. 126, n. 10, p. 2563-2574, Oct. 2013.

TAKATSUJI, H. Zinc-finger transcription factors in plants. **Cellular and Molecular Life Sciences**, Basel, v. 54, n. 6, p. 582-596, June 1998.

THORNSBERRY, J. M. et al. Dwarf polymorphisms associate with variation in flowering time. **Nature Genetics**, New York, v. 28, n. 3, p. 286-289, July 2001.

TIVANG, J. G.; NIENHUIS, J.; SMITH, O. S. Estimation of sampling variance of molecular-marker data using bootstrap procedure. **Theoretical Applied Genetics**, Berlin, v. 89, n. 2-3, p. 259-264, Oct. 1994.

TROYER, A. F. A retrospective view of corn genetic resources. **Journal of Heredity**, Washington, v. 81, n. 1, p. 17-24, Jan. 1990.

VAN INGHELANDT, D. et al. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers.

Theoretical and Applied Genetics, Berlin, v. 120, n. 7, p. 1289-1299, May 2010.

VITEZICA, Z. G.; VARONA, L.; LEGARRA, A. On the Additive and Dominant Variance and Covariance of Individuals within the Genomic Selection Scope. **Genetics**, Bethesda, v. 195, n. 4, p. 1223-1230, Dec. 2013.

WANG, P. et al. Association of polymorphisms in odorant-binding protein genes with variation in olfactory response to benzaldehyde in *Drosophila*. **Genetics**, Bethesda, v. 177, n. 3, p. 1655-1665, Nov. 2007.

WENG, J. et al. Genome-wide association study Identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) Inbred Lines. **PLOS ONE**, [San Francisco], v. 6, n. 12, 2011.

WENZL P. et al. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. **BMC Genomics**, Chichester, v. 7, p. 206, Aug. 2006.

WILSON, L. M. et al. Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Association. **The Plant Cell**, Rockville, v. 16, n. 10, p. 2719-2733, Oct. 2004.

XU, Y. Developing marker-assisted selection strategies for breeding hybrid rice. In: JANICK, J. (Ed.). **Plant Breeding Reviews**. New York: J. Wiley, 2010. p. 73-174. v. 23.

XU, Y.; ISHII, T.; MCCOUCH, S. Marker-assisted evaluation of germplasm resources for plant breeding. In: INTERNATIONAL RICE RESEARCH CONFERENCE, 24., 2003, Beijing. **Proceedings...** Beijing: International Rice Research Institute, 2003. p. 213-229.

YAN, J. et al. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. **PLOS ONE**, [San Francisco], v. 4, p. 8451, Dec. 2009.

YI, N.; GEORGE, V.; ALLISON, D. B. Stochastic Search Variable Selection for Identifying Multiple Quantitative Trait Loci. **Genetics**, Bethesda, v. 164, n. 3, p. 31129-31138, July 2003.

YU, J.; BUCKLER, E. S. Genetic association mapping and genome organization of maize. **Current Opinion in Biotechnology**, London, v. 17, n. 2, p. 155-160, Apr. 2006.

ZHU, Y. L. et al. Single-Nucleotide Polymorphisms in Soybean. **Genetics**, Bethesda, v. 163, n. 3, p. 1123-1134, Mar. 2003.

ARTIGO 2 Predição genômica ampla para produtividade de grãos de híbridos simples de milho utilizando modelo GBLUP

Artigo formatado de acordo com a NBR 6022 (ABNT, 2003), conforme exige o manual de normalização da UFLA.

RESUMO

O principal objetivo de um programa de melhoramento genético é gerar indivíduos mais produtivos dos já existentes. Contudo, no caso do melhoramento genético de milho, essa identificação exige grande número de cruzamentos entre linhagens. Na maioria dos casos, só é realizada uma parte desses cruzamentos, dificultando a identificação desses híbridos superiores. Assim, a predição genômica tem se mostrado uma ferramenta alternativa que utiliza a informação advinda dos marcadores moleculares para seleção das melhores combinações visando realizar o melhoramento de forma mais eficiente. Objetivou-se neste trabalho a predição genômica ampla utilizando um conjunto de marcadores Darts-seq e o modelo GBLUP com dominância para produtividade de grãos de híbridos simples de milho avaliados em diferentes safras e locais. Foi realizada a genotipagem com marcadores Darts-seq de 447 linhagens provenientes de um banco de germoplasma de uma empresa privada de melhoramento de produção de sementes de milho. Do cruzamento destas linhagens foram obtidos 838 híbridos simples avaliados em seis locais na safra de inverno do ano agrícola de 2013 e 797 híbridos simples avaliados em quatro locais na safra de verão do ano agrícola de 2013/2014. Para a predição, foi utilizado o modelo GBLUP com dominância e validação cruzada com quatro níveis de desbalanceamento – 10, 20, 30 e 50% e com 100 e 50% do conjunto de marcas geradas. Também foram utilizados os híbridos em comum nas duas safras para cálculo das correlações. A magnitude das correlações entre híbridos preditos e observados variou de 0,82 a 0,89 na safra de inverno e de 0,56 a 0,76 na safra de verão. As correlações entre combinações testadas em diferentes safras e locais foram encorajadoras. A coincidência entre os VGGs da safra de verão e inverno, em termos de descarte, mostra outra vertente da metodologia. Este resultado demonstra a possibilidade de utilização desta técnica em programas de melhoramento para descarte dos genótipos não promissores. O método GBLUP foi capaz de gerar elevadas correlações entre híbridos preditos e observados, mesmo em elevados níveis de desbalanceamento e em diferentes locais e safras.

Palavras-chave: Melhoramento Vegetal, Marcadores Darts, GBLUP, GWS, Predição Genômica.

1 INTRODUÇÃO

Os programas de melhoramento de milho buscam híbridos com alta produtividade, geralmente, melhores aos que já existem no mercado. Contudo, atingir esse objetivo não é algo simples. O melhoramento de milho envolve sistemas de cruzamentos, que buscam identificar as melhores combinações de genitores, tendo como produto final os híbridos (HALLAUER; CARENA; MIRANDA, 2010). Para testar todas as combinações híbridas possíveis seria necessária a realização de milhares de cruzamentos, o que se torna inviável nos programas de melhoramento. Por isso, somente uma parte dos cruzamentos são realizados diminuindo assim as chances de se encontrar as combinações mais promissoras (SHRAG et al., 2010).

Com o avanço do uso de marcas moleculares, a seleção feita para caracteres de importância econômica, como a produtividade de grãos, que é um caráter quantitativo, antes realizada somente a partir de dados fenotípicos, passou a utilizar dados moleculares. Uma justificativa para a utilização de informações moleculares é a expectativa de que informações sobre DNA tragam maior ganho genético do que quando se utilizam apenas os dados fenotípicos (MEUWISSEN; HAYES; GODDARD, 2001).

O primeiro modelo de predição de híbridos de milho foi proposto por Bernardo (1994) - *preditor linear não viesado (BLUP)* - que leva em consideração as informações dos marcadores moleculares das linhagens parentais e os dados de produtividade de uma população de cruzamentos testes. Apesar da correlação entre o desempenho genotípico do cruzamento predito e o seu valor fenotípico ser baixa em muitos casos, essa abordagem apresenta vantagens (BERNARDO, 1994). A utilização do BLUP tem mostrado eficiência considerável para tal finalidade (BERNARDO, 1994, 1995, 1996a, 1996b; MASSMAN et al., 2013).

A predição genômica tem se mostrado uma ferramenta alternativa que utiliza a informação genômica, advinda dos marcadores moleculares, para seleção das melhores combinações visando realizar o melhoramento de forma mais eficiente. Ao contrário de se utilizar somente as marcas moleculares que apresentam significância para o caráter em questão, a predição genômica ampla utiliza todas as marcas simultaneamente na construção do modelo (WINDHAUSEN et al., 2012).

Com o aumento da disponibilidade de marcas disponíveis e a redução dos custos de genotipagem, as técnicas de Predição Genômica Ampla (GWS) se tornaram uma realidade para os programas de melhoramento (HESLOT; JANNINK; SORRELLS, 2014; ZHAO; METTE; REIF, 2015). Contudo, existe um grave problema estatístico no uso de alta densidade de marcadores, observa-se um número muito maior de marcas do que de observações.

Meuwissen, Hayes e Goddard (2001), por meio de simulações, foram os pioneiros na incorporação de uma grande quantidade de marcas ao modelo, estimando todas simultaneamente. A GWS incorpora simultaneamente todas as marcas ao longo do genoma no modelo de predição, reduzindo o risco da identificação errônea ou imprecisa dos QTLs de efeitos menores. Cada marca é considerada um possível QTL e todas as marcas são combinadas para predizer o valor genético do genótipo (GUO et al., 2013a).

Entre os modelos utilizados para a GWS está o GBLUP (Genomic Best Linear Unbiased Prediction), proposto por VanRaden (2008) e que vem sendo amplamente utilizado no melhoramento vegetal (CROSSA et al., 2013). O GBLUP utiliza a relação genômica para estimar o mérito genético de um indivíduo. Para tal, uma matriz de correlações genéticas é utilizada, estimada a partir de informações de marcadores moleculares. Essa matriz define a covariância entre indivíduos baseada na similaridade em nível genômico, em

contrapartida da similaridade estimada por meio de *pedigree*, sendo assim mais preciso (CROSSA et al., 2013; HAYES et al., 2009; VANRADEN, 2008).

Resultados indicam claramente que o GWS tem maior potencial de predição que a seleção assistida por marcadores e experimentos como o conjunto de dados empíricos indicam que todos os modelos de predição têm semelhantes potenciais preditivos. Assim, o GBLUP demonstra vantagens visto que sua relativa simplicidade, menor tempo computacional e as já bem conhecidas propriedades ótimas dos modelos mistos de seleção fazem com que este método se mostre eficiente na maioria das situações (HESLOT; JANNINK; SORRELLS, 2014; VAN RADEN et al., 2009).

Contudo, a maioria dos trabalhos publicados utilizando GBLUP incluem somente efeitos aditivos (CALUS, 2010). Número reduzido de pesquisas vem sendo realizado para expandir esse modelo para predizer o mérito genético levando em consideração os efeitos de dominância. Assim, mais recentemente o GBLUP, incluindo os efeitos de dominância, tem sido sugerido com a expectativa de melhorar a predição de híbridos, principalmente quando houver efeito de dominância pronunciado no caráter (NISHIO; SATOH, 2014).

Neste sentido, objetivou-se a partir deste trabalho a predição genômica ampla utilizando um conjunto de marcadores Darts e o modelo GBLUP com dominância para produtividade de grãos de híbridos simples de milho avaliados em diferentes safras e locais.

2 MATERIAL E MÉTODOS

2.1 Obtenção dos Híbridos Simples e Avaliação fenotípica

O material genético utilizado consistiu em 447 linhagens de diferentes *backgrounds* de várias regiões do mundo, que foram utilizadas nos cruzamentos para gerar os híbridos simples.

Do cruzamento das linhagens foram obtidos 838 híbridos simples avaliados na safra de inverno do ano de 2013 e 797 híbridos simples avaliados na safra de verão do ano de 2013/2014. Na safra de inverno, os híbridos simples de milho foram avaliados em seis locais: Primavera do Leste - MT, Sorriso - MT, Rio Verde - GO, Lucas do Rio Verde - MT, Campo Novo do Parecis - MT e Sapezal - MT localizados na região Centro-Sul do Brasil. Na safra de verão, os híbridos foram avaliados em quatro locais, sendo eles Araguari, Nova Ponte, Presidente Olegário e Uberlândia, todos localizados no estado de Minas Gerais.

Os experimentos foram conduzidos utilizando o delineamento experimental de blocos incompletos com duas repetições por local e parcela composta de quatro fileiras de 5 m de comprimento, com espaçamento de 0,7 m entre fileiras.

Foi avaliada a produção de grãos dos híbridos e o peso da parcela corrigido para umidade de 13% e convertido em $t \cdot ha^{-1}$. O preparo da área e o uso de fertilizantes de cobertura foram conduzidos de acordo com a recomendação para cada área experimental e os tratos culturais foram conduzidos para o controle da lagarta do cartucho (*Spodoptera frugiperda*) e da espiga (*Helicoverpa zea*) bem como o controle de plantas daninhas.

Dados de produtividade de grãos de todos os híbridos em todos os experimentos foram submetidos à análise utilizando modelos mistos:

$$y = X\beta + Zb + Wg + Ti + \varepsilon$$

Em que X corresponde à matriz dos efeitos fixos: repetição dentro de experimento dentro de local, experimento dentro de local e local; β é a matriz de incidência dos efeitos fixos; Z é a matriz dos efeitos aleatórios: bloco dentro de repetição e dentro de experimento e dentro de local; b é a matriz de incidência dos efeitos de blocos; W é a matriz de genótipos; g é a matriz de incidência dos genótipos; T é a matriz da interação genótipos por ambientes; i é a matriz de incidência da interação genótipos por ambiente e ε é a matriz dos resíduos.

As estimativas dos efeitos fixos, dos componentes da variância fenotípica e as predições dos efeitos aleatórios foram obtidas via máxima verossimilhança restrita (REML), utilizando o algoritmo EM (*expectation-maximization*). As médias foram ajustadas para utilização no modelo GBLUP. A herdabilidade no sentido amplo das análises entre os ambientes foi calculada utilizando a equação: $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{gxe}^2 / e + \sigma_r^2 / re)$, em que σ_g^2 , σ_{gxe}^2 e σ_r^2 (HALLAUER et al., 2010) correspondem à variância genotípica, variância da interação genótipos por ambientes e variância residual, respectivamente e r corresponde ao número de repetições e e ao número de ambientes.

2.2 Genotipagem e Construção da Matriz de Marcadores

A extração de DNA foi realizada seguindo protocolo específico da empresa Diversity Arrays Technology (DARTS, 2013) para a qual as amostras de DNA devidamente diluídas foram enviadas para análise. A genotipagem das 470 linhagens foi realizada utilizando-se os marcadores Darts, porém devido a não amplificação de algumas amostras enviadas, foram obtidos os dados genotípicos para 447 linhagens.

A matriz de informação de marcadores foi construída utilizando o código 1 para presença do alelo t na marca m na linhagem i e 0 para ausência do alelo e “-” quando não foi possível a leitura com confiabilidade pelo *software*. Essa codificação facilita a construção das matrizes aditivas e dominantes dos híbridos de maneira oposta à codificação usual de 2 e 0 que codifica linhagens homozigotas e diploides.

Marcadores perdidos foram imputados pela função *A.mat*, método *mean*, do pacote *rrBLUP* (ENDELMAN, 2011) do *software* R.

2.3 Predição Utilizando o Modelo GBLUP

O processo de predição dos valores genéticos dos híbridos simples avaliados nos ensaios de safra e safrinha foi realizado pelo modelo GBLUP com inclusão de efeitos aditivos e de dominância, o qual é definido por:

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\delta} + \mathbf{e}$$

Em que \mathbf{y} o vetor das observações fenotípicas $n \times 1$, em que t é o número de locais e n é o número de observações fenotípicas dos híbridos simples obtidas em cada local; \mathbf{X} é a matriz de incidência dos efeitos fixos de locais e média dos híbridos confundidas em uma mesma covariável $tn \times p$, em que p é o número de efeitos fixos; $\hat{\boldsymbol{\beta}}$ é o vetor de efeitos fixos dos locais e médias confundidos $p \times 1$; \mathbf{Z} é a matriz de incidência dos híbridos simples avaliados em cada local $tn \times n$; $\boldsymbol{\alpha}$ é o vetor de efeitos aditivos $n \times 1$; $\boldsymbol{\delta}$ é o vetor de efeito de dominância $n \times 1$; \mathbf{e} é o vetor dos resíduos $tn \times 1$.

A matriz de marcadores moleculares dos híbridos simples foi obtida pela soma das matrizes de marcadores das duas linhagens parentais genotipadas previamente.

Matrizes de parentesco aditiva e dominância foram construídas utilizando a métrica de Cockerham (1954):

$$W_{\alpha} \begin{cases} 2 - 2p_k \\ 1 - 2p_k \\ 0 - 2p_k \end{cases} \text{ para os genótipos } \begin{cases} A^1A^1 \\ A^1A^2 \\ A^2A^2 \end{cases}$$

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

Em que p_k é a frequência do alelo favorável no loco k , W_{α} é a matriz de incidência dos efeitos aditivos das marcas; W_{δ} é a matriz de incidência dos efeitos de dominância das marcas.

As matrizes de parentesco aditiva e de dominância foram obtidas por:

$$A = \frac{W_{\alpha}W_{\alpha}'}{2\sum p_k(1-p_k)} ; D = \frac{W_{\delta}W_{\delta}'}{4\sum(p_k(1-p_k))^2}$$

Em que A é a matriz de parentesco aditiva e D é a matriz de parentesco de dominância.

As distribuições dos efeitos aleatórios são consideradas como $\alpha \sim MVN(0, G_{\alpha})$, $\delta \sim MVN(0, G_{\delta})$ e $e \sim MVN(0, R)$, em que

$\mathbf{G}_\alpha = \mathbf{A}\hat{\sigma}_\alpha^2, \mathbf{G}_\delta = \mathbf{D}\hat{\sigma}_\delta^2, \mathbf{R} = \mathbf{I}_{n \times n} \hat{\sigma}_e^2$. As observações são assumidas como $\mathbf{y} \sim MVN(\mathbf{0}, \mathbf{V})$. A matriz \mathbf{V} é estruturada por:

$$Var(\mathbf{y}) = \mathbf{ZG}_\alpha\mathbf{Z}' + \mathbf{ZG}_\delta\mathbf{Z}' + \mathbf{R} = \mathbf{V}$$

Onde: \mathbf{G}_α é a matriz de variância-covariância genética aditiva, \mathbf{G}_δ é a matriz de variância-covariância genética de dominância, \mathbf{R} é uma matriz de variância-covariância residual e \mathbf{V} é a matriz de variância-covariância fenotípica.

A otimização dos componentes de variância da função de máxima verossimilhança restrita (REML) foi realizada pelo algoritmo *Fisher scoring* (HAN; XU, 2008).

Os efeitos fixos, aditivos e de dominância foram obtidos por:

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

$$\hat{\alpha} = \mathbf{G}_\alpha\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})$$

$$\hat{\delta} = \mathbf{G}_\delta\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})$$

As estimativas dos desvios padrões dos componentes de variância (CASELLA; MCCULLOCH, 2006) foram obtidas pela raiz quadrada dos elementos diagonais da matriz:

$$var \begin{pmatrix} \sigma_e^2 \\ \sigma_\alpha^2 \\ \sigma_\delta^2 \end{pmatrix} = 2 \begin{bmatrix} tr(\mathbf{PP}) & tr(\mathbf{PZAZ}'\mathbf{P}) & tr(\mathbf{PZDZ}'\mathbf{P}) \\ tr(\mathbf{PZAZ}'\mathbf{P}) & tr(\mathbf{PZAZ}'\mathbf{PZAZ}') & tr(\mathbf{PZAZ}'\mathbf{PZDZ}') \\ tr(\mathbf{PZDZ}'\mathbf{P}) & tr(\mathbf{PZAZ}'\mathbf{PZDZ}') & tr(\mathbf{PZDZ}'\mathbf{PZDZ}') \end{bmatrix}^{-1}$$

Em que $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$

Todo processo de análise foi executado na plataforma R (R Core Team, 2015).

Para a regressão realizada entre os VGGs obtidos na safra de inverno e verão, foram utilizadas as 262 linhagens comuns às duas safras e os 395 híbridos eram comuns às mesmas.

2.4 Validação cruzada

O método de validação cruzada empregado foi o K-Fold. O conjunto de observações foi subdividido aleatoriamente em grupos de dados por local, de acordo com a porcentagem de desbalanceamento adotada. No processo de análise, era eliminado sequencialmente um grupo para ser utilizado a população resultante como de treinamento, e o grupo excluído da análise como de validação, até que todos os grupos fossem utilizados como população de validação. Na predição dos valores genéticos dos genótipos integrantes do grupo de validação, eram realizadas duas modificações nas estruturas das matrizes X e Z : (i) eliminação das linhas da matriz X relacionadas aos genótipos integrantes do grupo de validação; (ii) eliminação das linhas e colunas da matriz Z relacionadas aos genótipos do mesmo grupo, sendo K dependente do desbalanceamento adotado.

Os níveis de desbalanceamento utilizados foram 10, 20, 30 e 50%. A eficiência foi estimada pela correlação de Pearson entre os valores genéticos preditos do grupo de validação, com os valores genéticos estimados na análise considerando os grupos. Essa correlação foi utilizada como parâmetro de validação cruzada.

2.5 Agrupamento de grupos heteróticos

A fim de capturar as relações genéticas entre os indivíduos, foi elaborada a matriz de parentesco aditiva (**A**) seguindo a proposta de Vitezica (2013):

$$A = \frac{W_A + W'_A}{\sum pq}$$

Em que **p** é a frequência do alelo favorável e **q** a frequência do alelo desfavorável; **W_A** é a matriz de desvio dos marcadores.

Posteriormente foi elaborado uma análise de agrupamento hierárquico por meio da função *hclust* do pacote *hclust* no *software* R (R CORE TEAM, 2014) configurado pelo método *wald*, utilizando como objeto uma matriz de distância euclidiana dos elementos da matriz **A**.

3 RESULTADOS

Para análise dos componentes de variância, foi utilizada a metodologia de modelos mistos, via máxima verossimilhança restrita (REML), por meio do algoritmo EM (*expectation-maximization*).

A produtividade média dos híbridos foi de 5,33 ton/ha na safra de inverno e 8,21 ton/ha na safra de verão. Na Tabela 1, podem ser observados os componentes de variância genética (σ_g^2), da interação genótipos por ambientes (σ_{gxe}^2) e do resíduo (σ_r^2), além da herdabilidade no sentido amplo na média dos híbridos.

Todos os componentes de variância foram estatisticamente diferentes de zero pelo teste z. A significância de genótipos indica a variabilidade existente entre os mesmos, permitindo assim a seleção dos mais promissores. A razão entre a variância da interação genótipos por ambientes e a variância genética variou de 0,99 para a safra de inverno e 0,13 para a safra de verão. Já a razão da variância residual e da variância genética foi estável entre as duas safras, com 3,94 para safra de inverno e 3,92 para a safra de verão.

A herdabilidade no sentido amplo foi similar entre as duas safras. Mesmo a safra verão apresentando menos ambientes, este fato foi compensado por sua maior variância genética e menor variância de interação, levando assim à herdabilidade de 0,66.

Tabela 1 Componentes de variância genotípico (σ_g^2), interação genótipos por ambientes (σ_{gxe}^2) e residual (σ_r^2) e herdabilidade (h_a^2) e seus respectivos erros padrão

Componentes de Variância	Safra Inverno	Safra Verão
σ_g^2	0,432 (0,037)**	1,387 (0,106)**
σ_{gxe}^2	0,430 (0,046)**	0,181 (0,071)*
σ_r^2	1,701 (0,042)**	5,438 (0,103)**
h_a^2	0,67	0,66

Significância ao nível de *0,05 e **0,01 de acordo com o teste Z

Pela análise genotípica das linhagens foram obtidos 23.153 marcadores Darts-GBS distribuídos ao longo de nove grupos de ligação, mitocôndrias, plastídios e alguns com posições desconhecidas. O número de marcas variou de 3.612, no cromossomo 1, a 1.386 no cromossomo 7 (Tabela 2). Não foram mapeados marcas no décimo cromossomo pela empresa Diversity Array Technology.

A fim de investigar a diminuição do número de marcas na predição dos híbridos, foram também utilizados para análise, um conjunto de 50% do total de marcas obtidas, ou seja, 11.576 marcas. A redução das marcas foi realizada visando aumentar a distância das mesmas entre si, ou seja, foi mantida uma marca e retirada uma, mantendo a distribuição homogênea das mesmas ao longo do genoma. Isto fez com que marcas muito próximas fossem retiradas.

Tabela 2 Distribuição das marcas por grupo de ligação, mitocôndrias e plastídios das linhagens utilizando 100% e 50% dos marcadores Darts utilizados para predição da produtividade de grãos dos híbridos simples de milho pelo método GBLUP

Cromossomo	Darts	
	100% Marcas	50% Marcas
1	3612	1806
2	2107	1053
3	1896	948
4	1706	853
5	1904	952
6	1459	730
7	1386	693
8	1522	761
9	1415	707
Mitocôndria	4	2
Plastídios	1	1
Desconhecidos	6141	3070
Total	23153	11576

Na predição genômica ampla, utilizando validação cruzada, a correlação entre VGGs preditos e observados na safra de inverno variou de 0,89 a 0,82, com discreta diminuição da correlação com o aumento do nível de desbalanceamento. Vale ressaltar que mesmo com o desbalanceamento de 50% puderam ser observadas correlações sempre acima de 0,80 (Tabela 3).

As variâncias aditivas (VA) e de dominância (VD) apresentaram a mesma magnitude. Tomando-se como base os componentes de variância obtidos

pelo método REML, que utiliza os dados fenotípicos, o modelo com marcadores permitiu a recuperação da variância genética total por meio das somas das variâncias aditiva e de dominância estimadas a partir do modelo GBLUP.

Como a predição foi realizada com as médias ajustadas para os efeitos fixos no método REML e essas médias foram utilizadas no GBLUP para predição, a variância residual que este modelo estima é, em sua maior parte, composta pela variância da interação genótipos por ambientes (VE), que na safra de inverno foi da ordem de 2,22 - 2,23.

Tabela 3 Correlação entre VGGs preditos e observados ($r_{\hat{g}g}$) com diferentes níveis de desbalanceamento e médias das variância aditiva (VA), variância de dominância (VD), variância residual (VE) e erros padrão para a safra de inverno (K-Fold)

Desb	100% MARCAS				50% MARCAS			
	$r_{\hat{g}g}$	VA	VD	VE	$r_{\hat{g}g}$	VA	VD	VE
10%	0,89	0,21 (0,04)	0,17 (0,04)	2,24 (0,05)	0,89	0,21 (0,04)	0,17 (0,04)	2,24 (0,05)
20%	0,88	0,20 (0,04)	0,17 (0,04)	2,23 (0,05)	0,87	0,20 (0,04)	0,17 (0,04)	2,23 (0,05)
33%	0,87	0,19 (0,05)	0,19 (0,05)	2,23 (0,06)	0,86	0,20 (0,05)	0,18 (0,05)	2,24 (0,06)
50%	0,82	0,19 (0,05)	0,18 (0,06)	2,23 (0,07)	0,83	0,19 (0,05)	0,20 (0,06)	2,22 (0,07)

O decréscimo da densidade de marcas pela metade, no conjunto de dados da safra de inverno, não influenciou a capacidade preditiva do modelo. A correlação ficou estável quando da utilização de todo o conjunto de marcas ou

quando foram utilizadas somente metade delas, mesmo nos maiores níveis de desbalanceamento.

Na safra de verão, foram observadas correlações inferiores às observadas na safra de inverno, variando de 0,73 a 0,56. A menor correlação e maior variação podem ser explicadas devido ao menor conjunto de dados e maior interação genótipos por ambientes estimados pelo modelo GBLUP. Pode-se observar também maior influência do grau de desbalanceamento nas mesmas, sendo que para o maior desbalanceamento (50%) houve forte queda da capacidade preditiva, o que não ocorreu na safra de inverno (Tabela 4).

Observou-se comportamento diferente da VD no conjunto de dados da safra de verão, que apresentou maior magnitude do que a VA. Essa superioridade demonstra a predominância dos efeitos não aditivos para o caráter produtividade nesta safra. Pode-se observar também maior interação genótipos por ambientes quando comparada à safra de inverno.

Com relação à diminuição das marcas moleculares para a realização da predição, só foi possível observar influência da mesma no desbalanceamento de 50%, que apresentou maior correlação quando da diminuição do número de marcadores. Nos outros desbalanceamentos, a correlação se manteve constante para as duas situações.

Tabela 4 Correlação entre VGGs preditos e observados ($r_{gg\hat{g}}$) com diferentes níveis de desbalanceamento e médias das variância aditiva (VA), variância de dominância (VD), variância residual (VE) e erros padrão para a safra de verão (K-Fold)

Desb	100% MARCAS				50% MARCAS			
	$r_{gg\hat{g}}$	VA	VD	VE	$r_{gg\hat{g}}$	VA	VD	VE
10%	0,73	0,36 (0,10)	1,00 (0,16)	3,46 (0,10)	0,73	0,35 (0,10)	1,00 (0,16)	3,47 (0,10)
20%	0,70	0,35 (0,10)	0,99 (0,17)	3,46 (0,11)	0,70	0,35 (0,16)	1,00 (0,17)	3,46 (0,11)
33%	0,68	0,36 (0,12)	0,99 (0,18)	3,45 (0,12)	0,69	0,33 (0,11)	1,00 (0,20)	3,44 (0,11)
50%	0,56	0,29 (0,12)	1,02 (0,21)	3,45 (0,14)	0,67	0,37 (0,18)	1,06 (0,26)	3,43 (0,14)

Entre as duas safras utilizadas neste estudo, 395 híbridos eram comuns aos dois ambientes. Foi realizado o ranqueamento dos genótipos de acordo com seus VGGs observados nas duas safras e foi utilizada uma intensidade de seleção e também de descarte de 20%. Verificou-se que as estimativas seriam melhores no sentido de descarte dos genótipos não promissores do que para a seleção dos híbridos superiores. Assim, no descarte dos genótipos 20% menos produtivos, ocorreu um erro de somente 11%, ou seja, somente nove genótipos superiores seriam descartados erroneamente tomando-se como base o VGG observado na safra de inverno para seleção na safra de verão.

Baseado nos dados fenotípicos da safra de inverno foi realizada a predição de 98.716 combinações híbridas possíveis, não fenotipadas, utilizando para o conjunto de linhagens genotipadas e os dados fenotípicos da safra de inverno. Dos híbridos preditos não avaliados na safra de inverno, 402 foram

plantados na safra de verão. A correlação entre os VGGs preditos na safra de inverno e as médias observadas na safra de verão foi de 0,53 (Figura 1). Vale ressaltar a discrepância das condições climáticas nas duas safras e também o diferente conjunto de ambientes utilizados, localizados em diferentes estados e regiões. Considerando novamente uma intensidade de descarte de 20% dos híbridos com menores VGGs preditos na safra de inverno e comparando-os com os mesmos genótipos na safra de verão, observou-se erro de somente 10% o que significa que somente dez genótipos promissores seriam descartados de forma errônea na safra de verão, tomando-se como base os dados de predição.

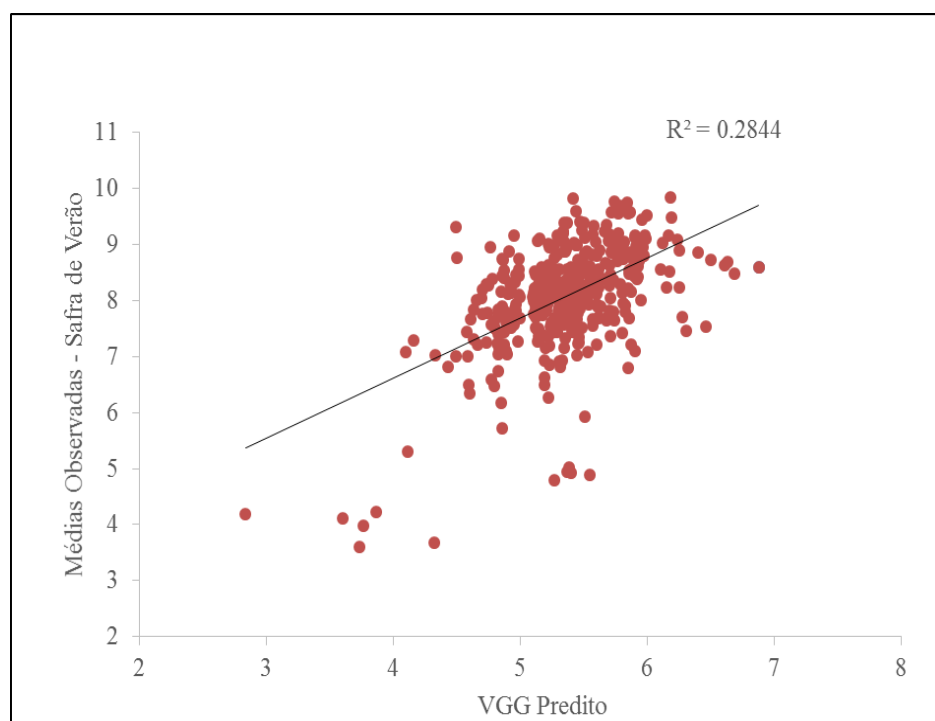


Figura 1 Regressão entre VGGs preditos na safra de inverno e médias ajustadas para efeitos fixos pelo método REML observadas na safra de verão para o caráter produtividade de grão de híbridos de milho

Utilizou-se também este conjunto de híbridos preditos na safra de inverno e testados na safra de verão para verificar se existem combinações entre grupos heteróticos mais previsíveis que outras (Figura 2).

As maiores correlações foram observadas nos cruzamentos entre os grupos heteróticos Amarillo Dentado x Lancaster e Tropical Flint x Suwan. Já as menores correlações foram entre os grupos F-DK-Arg x Suwan e Lancaster e NSSS quando cruzadas com Suwan. Puderam ser observadas também correlações altamente negativas nos cruzamentos entre Amarillo Dentado x Tropical Flint e Tropical Dent x Suwan.

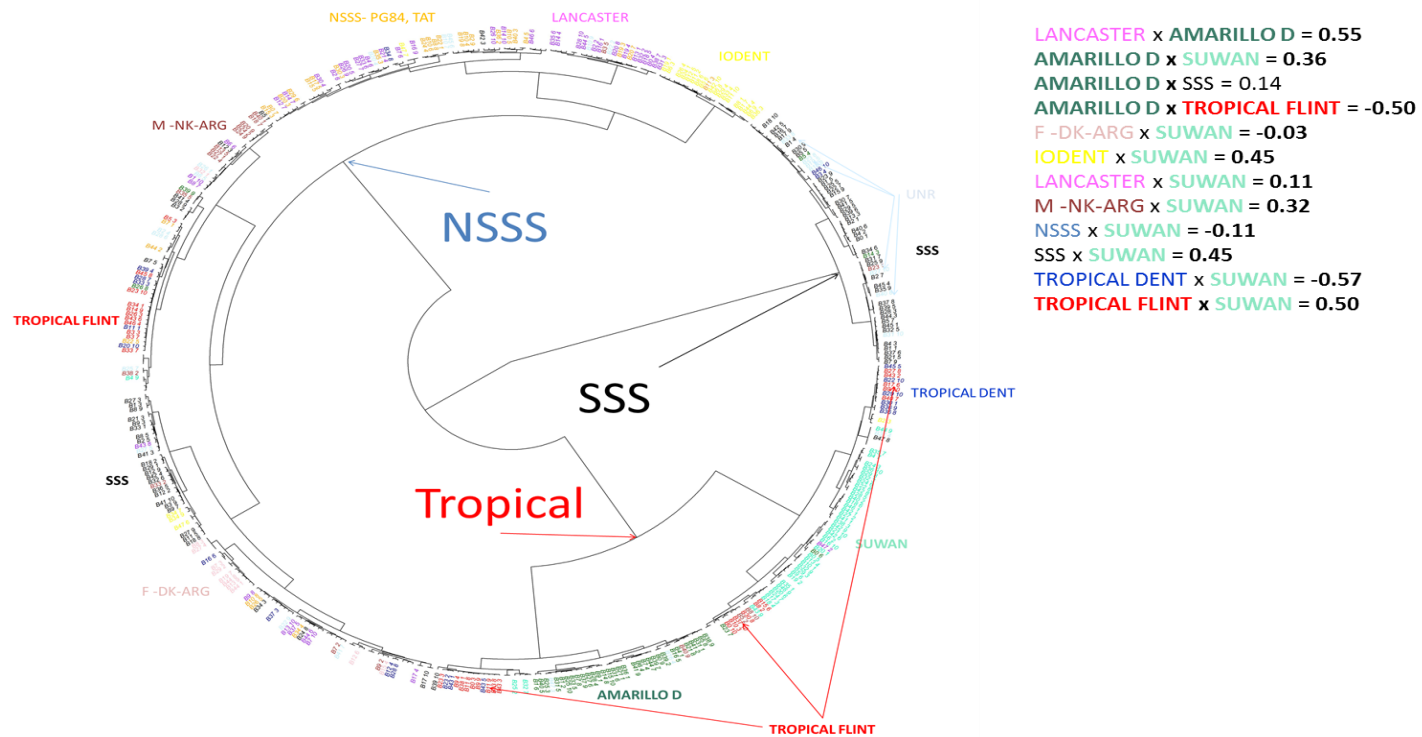


Figura 2 Agrupamento das linhagens dentro dos grupos de origem e correlação entre os VGGs preditos na safra de inverno e médias observadas na safra de verão

4 DISCUSSÃO

Resultados de validação cruzada obtidos em outros trabalhos indicam claramente que a seleção genômica ampla (GWS) é mais eficiente que a metodologia clássica de seleção assistida por marcadores (BERNARDO; YU, 2007) e que todos os métodos de predição têm poderes muito similares. Atualmente, um dos métodos mais utilizados na GWS é o GBLUP (HABIER; FERNANDO; DEKKERS, 2007).

A diferença entre os métodos estatísticos está na pressuposição dos efeitos das marcas, contudo, para características quantitativas, como no caso de produtividade de grãos, que apresentam efeito poligênico infinitesimal (HUANG et al., 2010; SCHÖN et al., 2004), essas pressuposições não têm influência nas correlações entre preditos e observados. Sendo assim, o método GBLUP tem se mostrado eficiente para prever caracteres quantitativos, mesmo considerando variâncias homogêneas de marcadores (XUA; ZHUB; ZHANG, 2014). Vários autores relatam eficiências comparáveis entre métodos bayesianos e o GBLUP (TECHOW et al., 2014; XUA; ZHUB; ZHANG, 2014), justificando assim a escolha da metodologia utilizada neste estudo.

Na análise dos componentes de variância utilizando a metodologia REML, observou-se a variância genética diferente de zero nas duas safras. Tal fato permitiu o prosseguimento das análises devido à comprovação da existência de variabilidade entre os híbridos possibilitando assim, a seleção no conjunto de dados. As herdabilidades no sentido amplo dos híbridos nas duas safras estão de acordo com as que são observadas na literatura (WINDHAUSEN et al., 2012).

A influência do tamanho da população de treinamento na capacidade de predição foi analisada nos conjuntos de dados oriundos das duas safras. Para este propósito, a população de treinamento variou entre 10 e 50% da população total. As correlações se mantiveram altas no conjunto de dados oriundo da safra de

inverno, já para o conjunto de dados da safra de verão, a correlação diminuiu consideravelmente com o aumento do desbalanceamento.

A capacidade preditiva do modelo GBLUP na safra de inverno foi elevada. Resultados obtidos em outros estudos evidenciam que com o aumento do tamanho da população de treinamento proporcional à população de validação, aumenta-se também a capacidade preditiva (ASORO et al., 2011; TECHNOW et al., 2013). Contudo, com esse conjunto de dados não foi possível observar tal fato. Uma explicação pode ser o alto nível de correlação já alcançado no desbalanceamento de 50%, não sendo possível o maior incremento com menores níveis de desbalanceamento (TECHOW et al., 2014).

As correlações observadas neste estudo, foram superiores às obtidas em estudos prévios utilizando BLUP, sem a utilização de marcadores para seleção genômica ampla, mas utilizando marcadores aleatórios para estimar o parentesco entre os indivíduos (BERNARDO, 1994). Este fato demonstra a aplicabilidade e superioridade do método.

Massman et al. (2013), utilizando validação cruzada, obtiveram acurácias variando de 0,75 a 0,87 para um desbalanceamento de 10%, utilizando RRBLUP e BLUP. Os resultados deste trabalho apresentaram correlações médias similares – 0,76 a 0,89 - para esse mesmo nível de perda.

Techow et al. (2014), utilizando as metodologias GBLUP e Bayes B, com um conjunto de 1.254 híbridos resultantes dos cruzamentos entre os grupos heteróticos ‘dent’ e ‘flint’ em vários anos e locais, com um conjunto de aproximadamente 35.000 marcadores SNPs, observaram correlações nas validações cruzadas variando de 0,75 a 0,92 para produtividade de grãos, valores esses similares aos resultados deste estudo. No estudo de Technow et al. (2014), a correlação variou em sua maior parte entre os números de parentais comuns na população treino e na população de validação.

Albrecht et al. (2011) verificaram na cultura do milho, aumento da correlação com o aumento do tamanho da população em todos os modelos testados para a predição da *performance* de *testcrosses* para a produtividade de grãos. Neste estudo, o decréscimo de 344 para 172 linhagens levou à forte redução da predição em todos os modelos testados. Neste sentido, na safra de verão, que foi avaliada em quatro ambientes, considerando os desbalanceamentos que ocorreram nos diferentes locais, a quantidade de dados final foi de 3.187. A safra de inverno foi avaliada em seis ambientes e também considerando que nem todos os genótipos foram avaliados em todos os locais, o conjunto foi composto por 5.016 dados. Essa diferença no número de dados afetou fortemente a correlação, principalmente nos maiores desbalanceamentos, como no de 50%, onde foi observada a maior discrepância entre as duas safras.

A influência do número de locais na habilidade de predição deve ser considerada. Em geral, Burgueño et al. (2011) e Guo et al. (2013) observaram ganho na capacidade preditiva em análises utilizando vários ambientes em contrapartida à análise em somente um ambiente. Resultados similares puderam ser observados neste estudo. Com 30% de redução do número de locais, reduziu-se expressivamente a correlação entre VGGs preditos e observados.

Albrecht et al. (2014) observou que a capacidade preditiva decresce com o decréscimo do número de locais. Para a produtividade de grãos, foi observada a redução de 0,65, com um conjunto de quatro locais e 698 linhagens, para 0,32, com um local e 87 linhagens.

Resultados similares foram observados por Burgeño et al. (2011) em que o aumento na capacidade preditiva ocorreu com o aumento do número de locais independentemente do uso de somente ambientes altamente correlacionados.

A qualidade experimental dos dados avaliados também é uma importante característica a ser levada em consideração para a predição. Observou-se o erro experimental mais pronunciado na safra de verão que na

safras de inverno. A qualidade das observações fenotípicas é ponto crucial para obtenção de estimativas fidedignas e de altas correlações entre preditos e observados. A heterogeneidade dos dados pode ser minimizada utilizando-se desenhos experimentais que maximizem a conectividade entre os dados através de ambientes comuns (ALBRECHT et al., 2014).

Na safra de inverno, as variâncias aditivas e de dominância apresentaram valores bem similares. Este fato vai ao encontro à preposição de Troyer e Wellin (2009) que observaram que os efeitos aditivos têm aumentado devido ao melhoramento das linhagens praticado pelas empresas de melhoramento de milho. Assim, devido ao melhoramento dessas linhagens, visando boa produtividade para viabilizar a produção de sementes, pode-se verificar na população de estudo analisada durante a safra de inverno, que tal fato também pode estar ocorrendo. Techow et al. (2014) também observou menor contribuição dos efeitos aditivos em seu conjunto de dados.

Já na safra de verão, observou-se a superioridade da VD com relação à VA. A VD foi aproximadamente três vezes superior à VA, o que demonstra a predominância dos efeitos de dominância no caráter produtividade de grãos como constatado em outros trabalhos na literatura (STUBER; EDWARDS; WENDEL, 1987; GUO et al., 2014).

A diferença entre VD nos dois experimentos pode ser explicada pelas condições favoráveis encontradas na safra de verão. Segundo Hamblin e Morton (1977), a dominância tende a se expressar mais em condições de cultivos favoráveis. Vale ressaltar que a média de produtividade na safra de verão foi de 8,21 ton/ha enquanto que na safra de inverno foi de 5,33 ton/ha e que as condições climáticas de cultivo na safra de verão são bem superiores as da safra de inverno.

A diferença das estimativas das variâncias aditivas e de dominância nas diferentes safras pode também ser justificada pelo diferente conjunto de

linhagens utilizadas. Na safra de verão, foram utilizadas 374 linhagens, resultando em 794 híbridos simples. Na safra de inverno, foram utilizadas 300 linhagens, resultando em 838 híbridos. Foram utilizadas 262 linhagens comuns às duas safras, enquanto que 395 híbridos eram comuns às mesmas. Assim, o fato de terem sido utilizadas diferentes linhagens, resultando em diferentes combinações híbridas, pode também ter causado a diferença das estimativas, visto que as variâncias são inerentes à população estudada (VISSCHER; HILL; WRAY, 2008).

Outro objetivo do estudo foi verificar o efeito da redução do número de marcas no modelo, para isto o número inicial de marcas foi reduzido à metade. O decréscimo da densidade de marcas pela metade, em geral, não influenciou a capacidade preditiva do modelo. Pode-se observar a estabilidade das correlações com a diminuição do número de marcadores independentemente do nível de desbalanceamento adotado.

Este resultado corrobora com Combs e Bernardo (2013) que verificaram que o aumento da correlação atinge um platô com moderado número de marcas, justificando tal fato devido à suficiente saturação do genoma a partir de certo número de marcas, saturação esta que depende da população estudada.

Albrecht et al. (2014) aumentando a densidade de marcas de 654 para 20.742 SNPs para um conjunto de 759 linhagens, observou aumento da capacidade preditiva de 0,59 para 0,62. Os mesmos autores também observaram que a capacidade preditiva pode diminuir com a utilização de maior número de marcas no modelo GBLUP.

Wong e Bernardo (2008) utilizando dados de simulação e o modelo RRBLUP em populações biparentais, verificou que baixa densidade de marcadores pode resultar em capacidades preditivas bem próximas das máximas, ou seja, o decréscimo no número de marcas não afeta fortemente a acurácia de predição e que existe uma quantidade máxima de marcas, a partir da qual a

acurácia tende a diminuir. Em outro estudo, Jannink et al. (2010) diminuindo a densidade de marcas em 75%, observou a diminuição na acurácia de predição de apenas 0,03. Estudos com gado de leite também têm demonstrado resultados semelhantes (HAYES et al., 2009).

Simulações sugerem ainda que, para certos modelos, a colinearidade entre marcadores ligados pode também diminuir a acurácia de predição (ZHONG et al., 2009), como foi observado no desbalanceamento de 50% na safra de verão. A necessidade de alta densidade de marcas na seleção genômica ampla também é reduzida se a população de treinamento for de alguma forma relacionada à população de validação (HAYES et al., 2009), como é o caso deste estudo.

Uma importante realidade em programas de melhoramento é a avaliação dos genótipos em localidades com diferentes condições ambientais, assim suas respostas são comparadas, a interação genótipos por ambientes estudada e os melhores genótipos dentro e entre ambiente são escolhidos (CROSSA et al., 2010). A predição é necessária no sentido de se usar os dados dos híbridos em uma localidade para prever em outra totalmente diferente. Porém grande parte dos trabalhos com predição genômica foi focada em um mesmo local/safra (GUO et al., 2013b). Essa informação é de extrema importância na tomada de decisão pelo melhorista, economizando recursos com o plantio de genótipos não promissores em localidades distintas.

Quando se realizou a análise considerando as diferentes safras, pôde ser evidenciado o potencial da metodologia para descarte dos híbridos menos promissores, o que pode ser considerada uma vantagem expressiva da metodologia.

Verificou-se também a correlação considerando os cruzamentos entre os grupos heteróticos distintos. Na maioria dos casos a maior correlação foi observada entre as combinações de grupos heteróticos mais presentes na

população de treinamento. A população de treinamento foi constituída pelos 838 híbridos avaliados na safra de inverno, os quais foram usados para predição de todas as demais combinações. O contrário também ocorreu, as correlações altamente negativas foram observadas nas combinações de grupos heteróticos menos presentes na população de treinamento. Pode-se inferir assim que a predição é mais acurada quanto mais indivíduos do cruzamento entre determinados grupos heteróticos existirem na população de treinamento.

Um dos fatores principais na acurácia de predição de características complexas é a relação existente entre a população de treinamento e a população de validação (ZHAO et al., 2015). Gowda et al. (2013) verificou redução de até 93% na acurácia de predição dependendo do cenário de parentesco entre a população de treinamento e de validação. Albrecht et al. (2011) também observou decréscimo na acurácia de predição da Capacidade Geral de Combinação (CGC) utilizando 1.380 *testcrosses* genotipados com 1.152 SNPs quando não foram utilizadas população de treinamento e de validação de alguma forma aparentadas.

Como consequência desta dependência entre acurácia de predição e relação entre observados e preditos, a utilização de alguns modelos, pode, por vezes, gerar correlações negativas (WANG et al., 2014) como pôde ser visto em algumas correlações entre híbridos preditos e observados que se encontravam em menor número na população de treinamento. Por isso, é importante que a população de treinamento seja representativa da população que será predita para que se aumente a capacidade preditiva (ALBRECHT et al., 2011).

Assim observa-se que, no caso de limites orçamentários nos programas de melhoramento, quando da alocação de recursos, o uso de maior quantidade de locais para avaliação fenotípica e uma população de treinamento representativa são preferíveis ao aumento do número de marcas, de acordo com os resultados deste estudo.

A utilização de recursos de forma a maximizar os ganhos de seleção é essencial nos programas de melhoramento. O conhecimento das etapas ou processos que devem receber maior porção destes recursos é de extrema importância para atingir tal objetivo (HESLOT; JANNINK; SORRELLS, 2014). Neste estudo, pode-se observar que o aumento do número de híbridos testados e locais utilizados exercem maior impacto na capacidade preditiva que o aumento do número de marcas. Esses devem ser então as etapas a receberem maior percentual de recursos do programa.

A utilização de modelos de predição para descarte e não para seleção vai de encontro ao que Heslot, Jannink e Sorrells (2014) evidencia em seu artigo e é uma forma de um maior controle de recursos nos programas. Considerar descarte de 20% dos híbridos menos produtivos em um programa de melhoramento consiste na diminuição de 20% dos custos com a fenotipagem ou mesmo o aumento em 20% de avaliações com diferentes combinações híbridas, considerando um orçamento fixo. Isto pode aumentar a probabilidade de se encontrar genótipos superiores.

Neste sentido, Shull, em 1908, definiu como principal papel do melhorista de milho a identificação das melhores combinações híbridas e reprodução das mesmas em grande escala. Contudo, no melhoramento de milho é realizada primeiramente a análise do desempenho da linhagem per se e posteriormente a análise da *performance* do *testcross*. A análise experimental das combinações híbridas é realizada somente nos últimos estágios do melhoramento, quando a variabilidade genética dentro foi fortemente exaurida. A predição genômica ampla da *performance* dos híbridos permite que se foque nessas melhores combinações desde os estágios iniciais. Com isso, o uso desta tecnologia permite uma mudança do paradigma do meio com que é conduzido o melhoramento atual, levando assim a uma direta implementação das ideias de Shull (TECHOW et al., 2014).

Contudo, a integração da predição genômica nos programas de melhoramento comerciais ainda tem desafios pela frente, e não é tão direto e viável quanto tem sido reportado para os outros programas de melhoramento, como no caso do melhoramento animal (JONAS; KONING, 2013; ALBRETCH et al., 2014). Os resultados deste trabalho demonstram que as acurácias preditivas podem ser elevadas ou melhoradas aumentando-se o número de locais e híbridos testados, utilizando população de treinamento adequada e levando-se em consideração a interação genótipos x ambientes para a predição. Mais estudos devem ser realizados a fim de se atestar a aplicabilidade da metodologia em diferentes safras, locais e conjunto de dados.

5 CONCLUSÃO

A magnitude das correlações entre híbridos preditos e observados tanto na safra de inverno quanto na de verão foram elevadas. Da mesma forma, essas correlações entre combinações testadas em diferentes safras e locais foram elevadas.

A coincidência entre os VGGs da safra de verão e inverno, em termos de descarte, mostra outra vertente da metodologia. Esse resultado demonstra a possibilidade de utilização desta técnica em programas de melhoramento para descarte dos genótipos não promissores.

A acurácia de predição pode ser incrementada com a otimização dos ensaios para coleta de dados fenotípicos como o aumento do número de dados e aumento do número de locais utilizados. Em contrapartida, verificou-se que o aumento do número de marcas não influenciou na capacidade preditiva do modelo.

Concluiu-se que o método GBLUP foi capaz de gerar elevadas correlações entre híbridos preditos e observados, mesmo em elevados níveis de desbalanceamento e em diferentes locais e safras.

GENOME WIDE SELECTION FOR GRAIN PRODUCTIVITY OF CORN HYBRIDS USING THE GBLUP MODEL

ABSTRACT

The main objective of a genetic breeding program is to generate individuals more productive than those existent. However, in the case of corn breeding, this identification demands a large number of crosses between inbred lines. In most cases, only a part of these crosses is performed, making it difficult to identify these superior hybrids. Therefore, genomic prediction has become an alternative tool that uses the information derived from molecular markers for the selection of the best combinations aiming at a more efficient breeding. In this work, the objective was to perform genome wide selection using a set of Darts-seq markers and the GBLUP model with dominance for grain productivity of corn hybrids evaluated in different years and locations. A genotyping was conducted with Darts-seq markers on 447 inbred lines derived from a germplasm bank of a private corn seed company. From the crosses of these inbred lines, 838 simple hybrids were obtained, evaluated at six locations in the winter season of 2013, and 797 simple hybrids, evaluated at four locations in the summer season of 2013/2014. For the prediction, we used the GBLUP model with dominance and cross validation with for levels of unbalancing – 10, 20, 30 and 50% and with 100 and 50% of the set of generated markers. The common hybrids from both seasons were used to calculate the correlations. The magnitude of the correlations between predicted and observed hybrids ranged from 0.82 to 0.89 in the winter season and from 0.56 to 0.76 in the summer season. The coincidences between the VGGs of the summer and winter seasons, in terms of disposal, show another dimension of the methodology. This result demonstrates the possibility of using this technique in breeding programs for disposing of non-promising genotypes. The GBLUP method was capable of generating high correlations between predicted and observed hybrids, even with high levels of unbalance and in different locations and years.

Keywords: Plant Breeding. Darts Markers. GBLUP. GWS. Genomic Prediction.

REFERÊNCIAS

- ALBRECHT T. et al. Genome-based prediction of testcross values in maize. **Theoretical Applied Genetics**, Berlin, v. 123, n. 2, p. 339-350, July 2011.
- ALBRECHT, T. et al. Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. **Theoretical Applied Genetics**, v. 127, n. 6, p. 1375-1386, June 2014.
- ASORO, F. G. et al. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. **The Plant Genome**, Madison, v. 4, n. 2, p. 132-144, Fev. 2011.
- BERNARDO, R Genetic models for predicting maize performance in unbalanced yield trial data. **Crop Science**, Madison, v. 35, n.1, p. 141-147, Jan. 1995.
- BERNARDO, R. Best linear unbiased prediction of maize single cross performance. **Crop Science**, Madison, v. 36, n.1, p. 50-56, Fev. 1996a.
- BERNARDO, R. Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. **Crop Science**, Madison, v. 36, n. 4, p. 862-866, June 1996b.
- 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, Apr. 1994.
- 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, Apr. 1994.
- BERNARDO, R.; YU, J. Prospects for Genomewide Selection for Quantitative Traits in Maize. **Crop Science**, Madison, v. 47, n. 3, p. 1082-1090, Nov. 2007.
- BURGUEÑO, J. et al. Prediction assessment of linear mixed models for multienvironment trials. **Crop Science**, Madison, 51, n. 3, p. 944-954, July 2011.

- BURGUEÑO, J. et al. Genomic prediction of breeding values when modeling genotype \times environment interaction using pedigree and dense molecular markers. **Crop Science**, Madison, v. 52, n. 2, p. 707-719, June 2012.
- CALUS, M. P. L. Genomic breeding value prediction: methods and procedures. **Animal**, Cambridge, v. 4, n. 2, p. 157-164, Feb. 2010.
- COCKERHAM, C. C. An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariances among Relatives When Epistasis Is Present. **Genetics**, Bethesda, v. 39, n. 6, p. 859-882, Mar. 1954.
- COMBS, E.; BERNARDO, R. Accuracy of Genomewide Selection for Different Traits with Constant Population Size, Heritability, and Number of Markers. **The Plant Genome**, Madison, v. 6, n. 1, p. 1-7, Mar. 2013.
- CROSSA J. et al. Predictions of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, Bethesda, v. 186, n. 2, p. 713-724, Oct. 2010.
- CROSSA, J. et al. Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. **G3**, Bethesda, v. 3, n. 6, p. 1903-1926, Nov. 2013.
- ENDELMAN, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. **The Plant Genome**, Madison, v. 4, n. 3, p. 250-255, Nov. 2011.
- GOWDA, M. et al. Best linear unbiased prediction of triticale hybrid performance. **Euphytica**, Wageningen, v. 191, n. 2, p. 223-230, May 2013.
- GUO, T. et al. Genetic basis of grain yield heterosis in an “immortalized F2” maize population. **Theoretical and Applied Genetics**, Berlin, v. 127, n. 10, p. 2149-2158, Oct. 2014.
- GUO, T. et al. Performance prediction of F1 hybrids between recombinant inbred lines derived from two elite maize inbred lines. **Theoretical Applied Genetics**, Berlin, v. 126, n. 1, p. 189-201, Jan. 2013a.
- GUO, Z. et al. Accuracy of Across-Environment Genome-Wide Prediction in Maize Nested Association Mapping Populations. **G3**, Bethesda, v. 3, n. 2, p. 263-272, Feb. 2013b.

HABIER, D.; FERNANDO, R.L.; DEKKERS, J.C.M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, Bethesda, v. 177, n. 4, p. 2389-2397, Dec. 2007.

HALLAUER, A. R.; CARENA, M.; MIRANDA, J. B. F. **Quantitative genetics in maize breeding**. 3. ed. Ames: Iowa State University, 2010.

HAMBLIN, J.; MORTON, J.R. Genetic interpretations of the effects of bulk breeding on four populations of beans (*Phaseolus vulgaris* L.). **Euphytica**, Wageningen, v. 26, n.1, p. 75-83, Feb. 1977.

HAN, L.; XU, S. A Fisher scoring algorithm for the weighted regression method of QTL mapping, **Heredity**, Oxford, v. 101, p. 453-464, Aug. 2008.

HAYES, B. J. et al. Invited review: genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, Champaign, v. 92, n. 2, p. 433-443, 2009.

HAYES, B. J.; VISSCHER, P. M.; GODDARD, M. E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research**, Cambridge, v. 91, n. 1, p. 47-60, Feb. 2009.

HESLOT, N.; JANNINK, J. L.; SORRELLS, M. E. Perspectives for Genomic Selection Applications and Research in Plants. **Crop Science**, Madison, v. 55, n. 1, p. 1-12, Nov. 2014.

HUANG, Y. F. et al. The genetic architecture of grain yield and related traits in *Zea mays* L. revealed by comparing intermated and conventional populations. **Genetics**, Bethesda, v. 186, n. 1, p. 395-404, Sept. 2010.

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

JONAS, E.; KONING, D.-J. de. Does genomic selection have a future in plant breeding? **Trend in Biotechnology**, Cambridge, v. 31, n. 9, p. 497-504, Sept 2013.

MASSMAN, J. M. et al. Genome wide predictions from maize single-cross data, **Theoretical Applied Genetics**, Berlin, v. 126, n. 1, p. 13-22, Jan. 2013.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-1829, Apr. 2001.

NISHIO, M.; SATOH, M. Including Dominance Effects in the Genomic BLUP Method for Genomic Evaluation. **PLOS ONE**, [San Francisco], v. 9, n. 1, p. 1-6, Jan. 2014.

SCHON, C. C. et al. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. **Genetics**, Bethesda, v. 167, n. 1, p. 485-498, May 2004.

SCHRAG, T. A. et al. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. **Theoretical Applied Genetics**, Berlin, v. 120, n. 2, p. 451-461, Jan. 2010.

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

STUBER, C. W.; EDWARDS, M. D.; WENDEL, J. F. Molecular Marker-Facilitated Investigations of Quantitative Trait Loci in Maize. II. Factors Influencing Yield and its Component Traits. **Crop Science**, Madison, v. 27, n. 4, p. 639-648, July 1987.

TECHNOW, F. et al. Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. **Genetics**, Bethesda, v. 114, n. 4, p. 1-42, Aug. 2014.

TECHNOW, F.; BÜRGER, A.; MELCHINGER, A. E. Genomic Prediction of Northern Corn Leaf Blight Resistance in Maize with Combined or Separated Training Sets for Heterotic Groups. **G3**, Bethesda, v. 3, n. 2, p. 197-203, Feb. 2013.

TROYER, A. F.; WELLIN, E. J. Heterosis Decreasing in Hybrids: Yield Test Inbreds. **Crop Science**, Madison, v. 49, n. 6, p. 1969-1976, Sept. 2009.

VANRADEN, P. M. et al. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, Champaign, v. 92, n. 1, p. 16-24, Jan. 2009.

VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, Champaign, v. 91, n. 11, p. 4414-4423, Nov. 2008.

VISSCHER, P. M.; HILL, W. G.; WRAY, N. R. Heritability in the genomics era — concepts and misconceptions. **Nature Reviews Genetics**, London, v. 91, n. 11, p. 255-266, Nov. 2008.

VITEZICA, Z. G.; VARONA, L.; LEGARRA, A. On the Additive and Dominant Variance and Covariance of Individuals within the Genomic Selection Scope. **Genetics**, Bethesda, v. 195, n. 4, p. 1223-1230, Dec. 2013.

WANG, Y. et al. The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of markerassisted selection and is equally augmented by multiple field evaluation locations and test years. **BMC Genomics**, Chichester, v. 15, July 2014.

WINDHAUSEN, V. S. et al. Effectiveness of Genomic Prediction of Maize Hybrid Performance in Different Breeding Populations and Environments. **G3**, Bethesda, v. 2, n. 11, p. 1427-1436, Nov. 2012.

WONG, C.; BERNARDO, R. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical Applied Genetics**, Berlin, v. 116, n. 6, p. 815-24, Apr. 2008.

XUA, S.; ZHUB, D.; ZHANG, Q. Predicting hybrid performance in rice using genomic best linear unbiased prediction. **PNAS**, Washington, v. 111, n. 3, p. 12456-12461, Aug. 2014.

ZHAO, Y.; METTE, M. F.; REIF, J. C. Genomic selection in hybrid breeding. **Plant Breeding**, [S.l.], v. 134, n. 1, p. 1-10, Fev. 2015.

ZHONG, S. et al. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, Bethesda, v. 182, n. 1, p. 355-364, Apr. 2009.