



LUIZ PAULO MIRANDA PIRES

**PREDIÇÃO GENÔMICA DO DESEMPENHO DE HÍBRIDOS
DE MILHO CONSIDERANDO A INTERAÇÃO
GENÓTIPOS POR AMBIENTES**

LAVRAS - MG

2017

LUIZ PAULO MIRANDA PIRES

**PREDIÇÃO GENÔMICA DO DESEMPENHO DE HÍBRIDOS DE MILHO
CONSIDERANDO A INTERAÇÃO GENÓTIPOS POR AMBIENTES**

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

Orientador

Dr. Renzo Garcia Von Pinho

LAVRAS - MG

2017

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

Pires, Luiz Paulo Miranda.

Predição genômica do desempenho de híbridos de milho considerando a interação genótipos por ambientes / Luiz Paulo Miranda Pires. - 2017.

77 p.

Orientador(a): Renzo Garcia Von Pinho.

Tese (doutorado) - Universidade Federal de Lavras, 2017.
Bibliografia.

1. AMMI-Bayesiano. 2. Parentesco genético. 3. Ensaios multiambientes. I. Von Pinho, Renzo Garcia. . II. Título.

LUIZ PAULO MIRANDA PIRES

**PREDIÇÃO GENÔMICA DO DESEMPENHO DE HÍBRIDOS DE MILHO
CONSIDERANDO A INTERAÇÃO GENÓTIPOS POR AMBIENTES**

**GENOMIC PREDICTION OF THE MAIZE HYBRIDS PERFORMANCE
CONSIDERING THE INTERACTION GENOTYPES BY ENVIRONMENTS**

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

Aprovada em 20 de abril de 2017

Dr. Marcio Balestre	UFLA
Dr. Adriano Teodoro Bruzi	UFLA
Dr. Flávio Sérgio Afféri	UFSCAR
Dr. Samuel Pereira Carvalho	UFLA

Dr. Renzo Garcia Von Pinho
Orientador

LAVRAS – MG
2017

A Deus, por sempre caminhar ao meu lado com alegria, por me ajudar, amparar e sustentar em todos os momentos, por compartilhar comigo infinita graça e bondade, por iluminar meus pensamentos, dando-me calma para encarar cada um dos vários desafios que esta jornada trouxe em seu bojo. Agradeço-Lhe por me conceder seu imenso amor, representado aqui por mais esta conquista.

OFEREÇO.

*Aos meus pais, Gelson Francisco Pires e Rute Aparecida Miranda Pires, pelo amor incondicional, apoio, dedicação e ensinamentos.
À minha avó, Ana (in memoriam) por todo carinho, atenção e pelas orientações. Creio que onde estiver estará feliz pelo que me tornei.
Às minhas irmãs Raquel e Aline, pela amizade e companheirismo.
Aos amigos que, aos poucos, tornaram-se tão especiais.*

DEDICO.

AGRADECIMENTOS

A Deus, por mais uma vitória.

À Universidade Federal de Lavras, em especial, ao Departamento de Agricultura, pela oportunidade de realização do doutorado, juntamente com a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES e Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo financiamento de diferentes etapas por ocasião desta titulação.

A todos os integrantes do Núcleo de Estudos em Milho e Sorgo – DAG/UFLA – onde tive a oportunidade de ser vice-presidente, pela valiosa experiência adquirida, em todos momentos que passamos juntos, organização de simpósios, visitas técnicas, ciclos de palestras, dentre outras atividades. As experiências que adquiri com vocês foram etapas muito significativas para a minha formação.

Ao Grupo do Milho – DAG/UFLA – representados pelo Luiz Yanes, Leandro Cancellier, Renato Barbosa, João Paulo, Helcio Pereira, Humberto Martins, Ewerton Resende, Ivan Vilela, Felipe Resende, Augusto Magre, Ana Freire, Vander e todos os demais integrantes do Grupo, em que fui um dos coordenadores durante todo o meu doutorado. Agradeço por todos os momentos inesquecíveis que vivemos juntos; foram diversos trabalhos técnicos e científicos, elaborados com o apoio de vocês e, certamente, esta história não teria sido tão bonita sem a participação mais do que especial de cada integrante.

Ao Núcleo de Estudos em Estatística Genômica – DEX/UFLA – onde tive a oportunidade de ser membro presidente e onde ocorreram meus maiores desafios científicos. Ali vivi momentos sensacionais com muita parceria e amizade. A Carlos Pereira, Andrezza Kellen, Jhonathan dos Santos, Renato Vasconcellos, Joel Nuvunga e todos os demais parceiros do DEX, muito obrigado.

Aos amigos da Iowa State University, representados pelos amigos do “*seed lab*” Anderson Verzeznazzi, Henrique Uliana Trentin, Gerald de la Fuente, Darlene Sanches, Elizabeth Bovenmyer, Songlin Hu, Arthur Silva, Yuegui Su, Jiao Jiao, Pryianca e demais amigos do Doubled Haploid Facility.

A meus orientadores no período de visiting scholar, Ursula Karoline Frei e Thomas Lubberstedt, pessoas de grande notoriedade, com competência científica e de trabalho espetaculares, humanidade incomparável. A oportunidade de convivência fornecida por eles

contribuiu, incomensuravelmente, com o meu crescimento pessoal e profissional. Obrigado por sempre me abrirem as portas e a vários outros cientistas com os mesmos sonhos.

Agradeço a todos os amigos do “CSF”, incluindo meus parceiros de moradia durante o período de intercâmbio - Danilo Hortis, Leonardo Azevedo, Leandro Antunes, Matheus Nunes, Marcelo Savoldi. Obrigado pelos momentos felizes, divertidos e formidáveis que vivemos juntos em Ames – IA.

Aos companheiros de república em Lavras, desde a minha chegada dividiram comigo diversos momentos, Vinícius Giroto, Kaio Olímpio, Álvaro Oliveira, João Brixner, Deny Teixeira, Jeffinho, Otávio, Leandro Pereira, Renan Carlos.

Meu respeito e agradecimento à contribuição dos integrantes da banca examinadora de qualificação de doutorado, Flávia Barbosa e José Airton Rodrigues Nunes, juntamente com a banca examinadora de defesa de tese, Adriano Teodoro Bruzi, Samuel Carvalho, Flávio Sergio Afférri. Com meus orientadores, Renzo e Marcio, contribuíram muito para a melhoria e publicação dos trabalhos associados a essas etapas.

Ao meu coorientador, Marcio Balestre, de uma inteligência e capacidade científica gigantescas, as quais poucas vezes observei semelhantes em alguém em todos os lugares por onde passei. Este doutorado não teria acontecido sem a sua participação. Agradeço-lhe pelas oportunidades que me proporcionou e pela contribuição direta em minha formação como melhorista.

Ao meu orientador e gestor no doutorado, Renzo Garcia Von Pinho, meus agradecimentos vão muito além de orientações científicas e de escrita, estendem-se, também, às orientações profissionais, pessoais e à confiança depositada em meu trabalho. Um ser de valiosa sensibilidade acerca das necessidades de cada pessoa, explora como ninguém as habilidades individuais, além de se esforçar, incessantemente, para conduzir seus orientados a caminho da realização de seus sonhos. Serei eternamente grato a você, pelo orientador que foi e pelo amigo que fiz. Temos a cultura do milho como uma paixão em comum e, em todos os demais aspectos. Espelhar-me-ei em você e em seu sucesso, ao contribuir com a comunidade como fez e faz. Graças ao seu talento, descobri minha capacidade. Meu muito obrigado.

Aos meus grandes amigos Adriano, Renan Rufino e Caio Chain, pelos incontáveis momentos felizes. Esse “quarteto fantástico” nos proporcionou momentos e desafios mais leves e prazerosos, tornando, assim, todos os demais.

Aos grandes amigos que desde o mestrado caminham junto comigo, Kaio Olímpio, José Maria Pádua, Samira Librelon, Fernanda Keiko, Paulo Henrique, Carlos Caixeta, Camila

Bastos, Laiane Corsini, Renato Camargos, Rafael Pereira, Samuel Fernandes, Ricardo Cabral, Régis Castro, Matheus Mendes.

Aos amigos da Riber – KWS – Wagner Fernandes, Aparecido Carvalho e, em especial, ao Weber Moreira Neves, pessoa com experiência singular e que divide comigo parte de sua capacidade como melhorista. Obrigado por me permitir explorar sua experiência em diversos aspectos relacionados ao melhoramento comercial e desenvolvimento de linhagens e cultivares. Muito obrigado pelo apoio e parceria.

Às minhas irmãs, Raquel e Aline, agradeço pela parceria, amizade e companheirismo.

AGRADECIMENTO ESPECIAL

Meu agradecimento especial e mais profundo é dedicado a duas pessoas, meus pais, Gelson e Rute. Ambos estiveram, incondicionalmente, ao meu lado em todos os momentos difíceis, que não foram raros, nestes quatro anos de doutoramento, mas sempre me fizeram acreditar que chegaria com sucesso ao final desta árdua, porém gratificante etapa. Cada gesto de incentivo me mostrou de maneira muito singular uma verdade sobre nossa família que, mesmo fisicamente distante em alguns momentos, espiritualmente, sempre fomos uma só. Sou grato por cada gesto de apoio, carinho e acolhimento e espero um dia retribuir parte de toda a grandeza que vocês representaram para o meu crescimento. Com muito amor, meu muito obrigado.

Nesse período, aprendi que uma tese, ou qualquer outro trabalho, é uma extensão da vida do autor e para que alguma coisa de valor seja produzida, é necessário criar o valor em nós mesmos. Em minha história, mais uma etapa foi vencida, mas eu sei que, ainda, é apenas o começo...

*De todos os meios que conduzem à sorte, os mais seguros
são a perseverança e o trabalho (Louis Reyband).*

RESUMO GERAL

No melhoramento de plantas, especialmente na cultura do milho, o efeito da interação genótipos por ambientes está presente em todas as etapas de avaliação de híbridos, desde as etapas iniciais até as pré-comerciais. Trata-se de uma dificuldade importante na recomendação de cultivares em múltiplos ambientes e também na predição de híbridos, pois este cenário corresponde a um dos principais desafios para os profissionais envolvidos no processo de obtenção de novas cultivares. Em programas de melhoramento de milho, genótipos de diversos *backgrounds* são naturalmente incorporados a cada ciclo, ao passo que tantos outros são descartados, dificultando ainda mais a realização de predições acuradas como resultado do aumento da complexidade do germoplasma comercial, além de perdas substanciais relacionadas à diminuição da herdabilidade por meio do aumento da variância fenotípica pelo componente da interação. Isso afasta circunstancialmente o valor fenotípico do valor genético real dos híbridos em avaliação. Nesse cenário, vários estudos têm sido realizados no sentido de aumentar as contribuições da predição genômica no contexto da interação genótipos por ambientes. Predizer com eficiência híbridos no contexto da interação, pode aproximar o que é produzido pela pesquisa com as necessidades de programas de melhoramento comerciais, uma vez que os custos relacionados à genotipagem de materiais de interesse têm decrescido nos últimos anos, ao passo que os gastos relacionados à fenotipagem são crescentes e limitantes para o crescimento dos programas de melhoramento. Adicionalmente, diversos fatores promovem o desbalanceamento em ensaios multiambientais, não apenas relacionados à perda de parcelas, mas também, perda de ambientes inteiros de avaliação por fatores climáticos, quantidades insuficientes de sementes de determinados híbridos para que os mesmos sejam testados em todos os ambientes, ou mesmo inclusão e exclusão derivadas de critérios dos melhoristas durante as etapas de avaliação dos híbridos. Assim, novas propostas de modelos que sejam eficientes em condições de desbalanceamento e aplicações de processos de predição com informações de marcadores moleculares podem contribuir na redução de gastos financeiros e identificações de genótipos superiores em programas de melhoramento de milho. Com esse estudo procurou-se criar informações acerca da eficiência da incorporação de informações de parentesco genômico para predição de híbridos utilizando modelo GBLUP-AMMI com abordagem bayesiana, além de estudar a interação de genótipos com ambientes e realização de inferência acerca da contribuição dos genótipos para a interação.

Palavras-chave: AMMI-Bayesiano. Parentesco genético. Ensaio multiambiental.

GENERAL ABSTRACT

In plant breeding, especially in maize, the genotype interaction effect by environment is present in all stages of hybrid evaluation, from the initial to the pre-commercial stages. This is an important difficulty in the cultivar recommendation in multiple environments and also in the hybrids prediction. It means one of the main challenges for the professionals involved in the process of obtaining new cultivars. In maize breeding programs, genotypes from various backgrounds are naturally incorporated into each cycle, while so many others are discarded, making further accurate predictions is more difficult as a result of increased commercial germplasm complexity. In addition to substantial losses related to heritability decline, just because the phenotypic variance can increase by the interaction component. This circumstantially removes the phenotypic value of the actual genetic value of the hybrids under evaluate. At this scenario, several studies have been carried out in order to increase the contributions of genomic prediction in the genotypes interaction context by environments. Effectively predicting hybrids in the interaction context can approximate what are produced by the research with the needs of commercial breeding programs, since the costs related to genotyping of materials have declined in recent years, while costs related to phenotyping are increasing and limiting for increase breeding programs. In addition, several factors promote the umbalance in multienvironment trials, not only related to loss of plots, but also loss of whole environments of evaluation by climatic factors, insufficient amounts of seeds from some hybrids, so that they are tested in all environments, or even inclusion and exclusion derived from breeders criteria during the hybrids evaluation. Thus, new model proposals that are efficient in conditions of umbalance and applications of prediction processes with molecular markers information can contribute in the reduction of financial expenses and identifications of superior genotypes in maize breeding programs. With this study, we will try to create information about the efficiency of the genomic kinship information for hybrids prediction using GBLUP-AMMI model with Bayesian approach, besides studying the genotypes interaction with environments and realization of inference about the contribution of genotypes to the interaction.

Keywords: AMMI-Bayesian. Genetic kinship. Multi-environment trials.

LISTA DE ILUSTRAÇÕES

LISTA DE FIGURAS

Figura 1	Comportamento de dois genótipos (G1 e G2) em duas condições ambientais (A1 e A2), com respostas genotípicas caracterizando ausência de interação (A), interação simples (B) e com interação complexa (C e D)..	22
Figura 2	Representação do avanço de um programa de melhoramento por meio de diferentes estratégias de seleção, até a etapa relacionada aos experimentos multiambientes.....	24

SEGUNDA PARTE – ARTIGO

Figura 1	Comparação das correlações dos modelos AMMI Bayesiano sem parentesco e com parentesco, obtidas pelas predições dos híbridos eliminados durante o desbalanceamento gerado no processo de validação cruzada.....	56
Figura 2	Comparação das correlações médias entre os modelos AMMI Bayesiano sem parentesco e com parentesco, obtidas pelas predições dos híbridos eliminados durante o desbalanceamento gerado no processo de validação cruzada.....	57
Figura 3	Somas de quadrados dos erros residuais preditos (PRESS) dos modelos AMMI Bayesiano sem parentesco e com parentesco, obtidas pelas predições dos híbridos eliminados durante o desbalanceamento gerado no processo de validação cruzada.....	58
Figura 4	Intervalos de credibilidade (95% credibilidade) dos valores genéticos dos híbridos avaliados no ensaio multi-ambientes. A: Modelo AMMI Bayesiano sem parentesco. B: Modelo AMMI Bayesiano com parentesco...	59
Figura 5	Heatmap do parentesco genômico dos 86 híbridos utilizados para verificação da capacidade preditiva do modelo AMMI Bayesiano com adição da informação de parentesco aditivo. As regiões em verde implicam baixo parentesco e impossibilidade de formação de agrupamento dos híbridos muito aparentados; em amarelo, onde é possível observar pequenos agrupamentos, ocorre um nível intermediário de parentesco.....	61
Figura 6	Intervalos de credibilidade (95% credibilidade) dos valores genéticos dos híbridos avaliados no ensaio multiambientés identificados com maior parentesco genético aditivo no Heatmap.....	62
Figura 7	Biplot com os valores médio a posteriori dos escores genotípicos $(\lambda_1^{1/2}\alpha_{i1}, \lambda_2^{1/2}\alpha_{i2})$ e ambientais $(\lambda_1^{1/2}\gamma_{j1}, \lambda_2^{1/2}\gamma_{j2})$, considerando o primeiro componente principal (PCA1) e o segundo componente principal (PCA2)...	64
Figura 8	A: Biplot com as regiões bivariados de 95% de credibilidade dos escores genotípicos $(\lambda_1^{1/2}\alpha_{i1}, \lambda_2^{1/2}\alpha_{i2})$ que diferiram de zero. B: Biplot com os intervalos bivariados de 95% de credibilidade dos escores ambientais $(\lambda_1^{1/2}\gamma_{j1}, \lambda_2^{1/2}\gamma_{j2})$ que diferiram de zero. Os dois biplots foram construídos considerando o primeiro componente principal (PCA1) e o segundo componente principal (PCA2).....	65

Figura 9	Biplot com as regiões bivariados de 95% de credibilidade dos escores genotípicos $(\lambda_1^{1/2} \alpha_{i1}, \lambda_2^{1/2} \alpha_{i2})$ e ambientais $(\lambda_1^{1/2} \gamma_{j1}, \lambda_2^{1/2} \gamma_{j2})$ que diferiram de zero, considerando o primeiro componente principal (PCA1) e o segundo componente principal (PCA2).....	66
-----------------	---	----

LISTA DE TABELAS

Tabela 1	Conjunto de municípios, latitude, longitude, altitude, em que foram avaliados os híbridos do programa de melhoramento da Universidade Federal de Lavras, e produção média de grãos ($t\ ha^{-1}$) e coeficiente de variação (CV) dos ensaios multiambientais.....	50
Tabela 2	Estatísticas descritivas da distribuição à posteriori dos autovalores obtidos a partir do ajuste do modelo AMMIB-P e AMMIB-SP, em que SD corresponde ao desvio padrão, LI ao limite inferior e LS ao limite superior.	63

SUMÁRIO

	PRIMEIRA PARTE	16
1	INTRODUÇÃO GERAL	16
2	REFERENCIAL TEÓRICO	18
2.1	Histórico e Evolução do Melhoramento de Milho	18
2.2	Interação Genótipos por Ambientes no Melhoramento de Plantas	21
2.3	Predição Genômica no Contexto Atual do Melhoramento de Milho	24
2.4	Additive Main Effects and Multiplicative Interaction Analysis – AMMI	29
2.5	Utilização da Informação de Parentesco	32
2.6	Estatística Bayesiana	34
	REFERÊNCIAS	37
	SEGUNDA PARTE - ARTIGO	44
	RESUMO	45
1	INTRODUÇÃO	47
2	MATERIAL E MÉTODOS	50
2.1	Dados Fenotípicos	50
2.2	Dados Moleculares	51
2.3	Modelo AMMI Bayesiano	51
2.4	Prioris Para os Parâmetros do Modelo	52
2.5	Distribuição Conjunta a Posteriori para os Parâmetros do Modelo	52
2.6	Distribuições Condicionais Completas a Posteriori	53
2.7	Validação Cruzada	55
3	RESULTADOS	56
3.1	Análise da Capacidade Preditiva dos Modelos	56
3.2	Estimativas Pontuais, Regiões de Credibilidade	58
3.3	Estimativas Pontuais e Regiões de Credibilidade para os Scores Genotípicos e Ambientais	63
3.4	Inferência de Biplot com Intervalos Bivariados de Credibilidade Para os Scores Genotípicos e Ambientais	65
4	DISCUSSÕES	67
5	CONCLUSÕES	73
	REFERÊNCIAS	74

PRIMEIRA PARTE

1 INTRODUÇÃO GERAL

Existe, atualmente, 315 cultivares de milho disponíveis no Brasil para a safra 2016/2017 (PEREIRA FILHO; BORGHI, 2016); no entanto, o plantio concentra-se em um pequeno número dessas cultivares, mesmo o país possuindo ambientes de produção distintos. O sucesso de uma cultivar no mercado está associado ao elevado desempenho produtivo, às características agronômicas desejáveis e à sua elevada estabilidade produtiva. Apesar disto, a recomendação de novas cultivares tem se baseado fundamentalmente na produtividade média nos ambientes avaliados (FRITSCHÉ-NETO et al., 2010).

Um híbrido comercial poderia ser considerado ideal se preservasse a máxima produção dentro de um mega ambiente ou entre mega ambientes, entretanto, sabe-se que as cultivares, modernas ou não, têm comportamento variável conforme o ambiente de cultivo. Assim, surge um fator determinante para o desempenho final das cultivares, que é a interação dessas com os ambientes. Logo, é de interesse conhecê-la e quantificá-la para que recomendações seguras possam ser feitas. Nos programas de melhoramento almejam-se obter novos genótipos com ampla adaptabilidade aos ambientes de produção e elevada estabilidade ao longo dos anos e locais. Especificamente no caso da cultura do milho, que se caracteriza pela forte atuação de empresas privadas no melhoramento genético e grande dinamismo no lançamento de novas cultivares, é fundamental o conhecimento da interação dos genótipos com os ambientes, pois isso irá direcionar os lançamentos para regiões e mercados específicos.

Nesse sentido diversos métodos já foram apresentados, dentre os quais citam-se os baseados em ANOVA, regressão linear uni e bissegmentada, regressão não linear, procedimentos multivariados analíticos e gráficos, procedimentos não paramétricos e, mais recentemente, métodos baseados na inferência Bayesiana. Dentre essas, destacam-se as abordagens AMMI (Additive Main Effects and Multiplicative Interaction Analysis) e GGE Biplot (CROSSA, 1990; GABRIEL, 1971; GAUCH; ZOBEL, 1996; YAN et al., 2000).

Essas análises exploratórias têm se mostrado eficientes na detecção de padrões de interação GE, identificação de mega ambientes, genótipos ideais e adaptações específicas. As vantagens e desvantagens de ambas as análises já foram discutidas de forma detalhada em Gauch Piepho e Annicchiarico (2008) e Yan et al. (2007). Entretanto, Figueiredo et al. (2015) relatam uma importante discussão, não acerca das particularidades de cada análise, mas sobre

um aspecto imprescindível, muitas vezes negligenciado, que é: como aplicar esses métodos frente ao dinamismo de um programa de melhoramento comercial?

É sabido que diversos genótipos são selecionados e descartados sequencialmente em todas as etapas de avaliação de híbridos pré-comerciais. Nesse sentido, um programa de melhoramento é, necessariamente, desbalanceado (FIGUEIREDO et al., 2015; OLIVEIRA et al., 2016). Sob essas condições, tanto a abordagem AMMI tradicional quanto a GGE-biplot são limitantes, pois dependem da decomposição de valor singular de uma matriz de covariâncias que, por sua vez, depende de matrizes balanceadas de GE ou da interação genótipo + genótipo por ambiente (G+GE). Assim, um procedimento de análise mais robusto tem contribuições potenciais aos programas de melhoramento no estudo da GE.

Nesse contexto, destaca-se a utilização do modelo AMMI com abordagem Bayesiana, com a vantagem da capacidade para lidar com dados altamente desbalanceados independentemente da natureza do desbalanceamento, heterogeneidade de variâncias e covariâncias residuais e genotípicas, ou seja, trata-se de um modelo fidedigno ao contexto dos programas de melhoramento (CROSSA et al., 2011; OLIVEIRA et al., 2015; SILVA et al., 2015). Outra vantagem particular é a possibilidade de incorporação do parentesco genômico entre os genótipos no contexto da predição de híbridos faltantes. Apesar de se tratar de uma ferramenta potencialmente útil, informações acerca da incorporação de coeficiente de parentesco no estado ou estimado via pedigree são, ainda, pouco abordados na literatura sob a ótica da interação genótipos por ambientes.

Visando aliar as vantagens do modelo AMMI bayesiano com métodos multivariados, o objetivo com o presente trabalho foi verificar a eficiência da incorporação das informações do parentesco estimado entre os genótipos na predição genômica em condição de desbalanceamento, propondo, assim, mudanças significativas na interpretação e inferências acerca da estabilidade e adaptabilidade dos genótipos avaliados.

2 REFERENCIAL TEÓRICO

Nesta seção serão apresentados os principais aspectos teóricos relacionados à importância dos experimentos multiambientais no contexto dos programas de melhoramento de milho, a predição genômica de híbridos simples e aspectos relevantes da interação GE. Será descrita a metodologia multivariada AMMI, assim como a relevância da informação de parentesco em modelos de predição do desempenho de híbridos no contexto da interação GE. Finalmente, apresentam-se os fundamentos da inferência bayesiana que subsidiarão o trabalho aqui desenvolvido.

2.1 Histórico e Evolução do Melhoramento de Milho

O milho ($2n=2x=20$) é uma gramínea da família Poaceae, tribo Maydeae, gênero *Zea* e espécie *mays* (*Zea mays* L.). Trata-se de um cereal cultivado em todo o mundo e originário da América, onde era o principal cultivo das civilizações pré-colombianas. A origem desta espécie despertou grande interesse por parte dos pesquisadores, o que culminou no desenvolvimento de três teorias.

A primeira delas, conhecida como teoria da origem divergente, foi proposta por Weatherwax (1954). Esta teoria diz respeito a um possível ancestral comum que deu origem ao milho, teosinto e ao tripsacum. A segunda teoria, proposta por Galinat (1977), relata que o milho como hoje é conhecido foi derivado diretamente do teosinto. A última discorre que o teosinto é originário do cruzamento entre o milho e o tripsacum (MANGELSDORF; MACNEISH; GALINAT, 1964). Entretanto, existem evidências importantes a serem consideradas acerca desta discussão. Tanto o milho cultivado quanto o teosinto possuem $n=10$ cromossomos homólogos. Como resultado do cruzamento entre ambos, é possível observar descendentes férteis e, por ocasião de obtenção das gerações seguintes, pode-se observar também tipos semelhantes ao milho e ao teosinto, o que indica grande proximidade genética entre as duas espécies. Não há na literatura relatos de um ancestral comum para o milho e o teosinto, o que inviabiliza a teoria proposta por Weatherwax (1954). Além disso, Mangelsdorf (1974) abandona a hipótese da origem do teosinto como resultado do cruzamento do milho com o tripsacum, de maneira que seja razoável aceitar a teoria de que o milho tenha derivado diretamente do teosinto.

Trata-se de um processo de domesticação obtido por meio de subseqüentes seleções de plantas que apresentavam características que favoreceram o aumento do tamanho da espiga

em detrimento do número de perfilhos. A exploração da variabilidade genética realizada desde então por meio da seleção realizada pelo homem, juntamente com a seleção natural, permitiu que a cultura se diversificasse e expandisse, ocupando todo o continente americano com inúmeras raças que apresentaram adaptações específicas às mais diversas regiões de cultivo (BRIEGER et al., 1958; WEATHERWAX, 1954).

Paterniani e Goodman (1977), em estudo acerca da distribuição desses genótipos, citam a altitude como um fator que justifica a adaptação de diferentes raças em regiões específicas. Os autores relatam que cerca de 50% das raças eram adaptadas a baixas altitudes (0 a 1000 m), quase 40% ocorriam acima de 2000 m e pouco mais de 10% são de altitudes intermediárias. Outros fatores ambientais também influenciaram a adaptação das raças em diversas regiões de cultivo, como diferentes condições de fotoperíodo, exigências de determinados mercados a diferentes tipos de grãos, dentre outros. É importante salientar que, independentemente dos fatores que promovem a interação, já é possível observar, desde o processo de expansão das raças, a capacidade dos genótipos em apresentarem adaptação específica aos ambientes em que são cultivados.

Falconer e Mackay (1996) relatam, geneticamente, a evolução como sendo uma mudança de frequência alélica das populações ao longo das gerações, predominando nas raças atuais alelos cuja expressão proporciona melhor adaptabilidade a condições específicas. Nesse contexto, Paterniani e Goodman (1977), antes mesmo da abordagem de Falconer e Mackay (1996), corroboraram essa teoria. Estes autores salientam que o alto nível de evolução do *Zea mays* L. tem como forte evidência a grande diversidade genética que caracteriza essa espécie, e, por sua vez, tal diversidade pode ter sido derivada por dois fatores principais que promovem alterações nas frequências alélicas das populações: a hibridação e a seleção. Inferências sobre qual dessas correspondeu à maior contribuição para o desenvolvimento das raças é uma tarefa difícil. É conhecido que a seleção praticada pelos índios e antigas civilizações, visando a obtenção de genótipos com preferências específicas, como tipo de grãos que atendiam determinadas populações, foi primordial para o surgimento de determinadas raças (BRIEGER et al., 1958; WEATHERWAX, 1954). Ainda assim, Wellhausen et al. (1952) atribuem a hibridação como principal fator no desenvolvimento das raças, além de ter se tornado base de importantes métodos de melhoramento e construção de germoplasma.

Neste sentido, é fato que o processo de domesticação do milho intensificou-se, ao passo que sua contribuição para a sociedade aumentava proporcionalmente à sua

domesticação. Atualmente, a produção de milho assume papel fundamental pela sua utilização em fábricas de ração animal e corresponde a importante matéria prima para a fabricação de combustível. Diversos tipos de produtos utilizados na alimentação humana são derivados do milho, de maneira que grande parte das pessoas no mundo seja diretamente dependente da produção desta cultura.

Entre as grandes culturas, o milho é considerada a mais importante do mundo e domina a produção global de alimentos; é cultivado desde o continente americano ao asiático, correspondendo a 37,83% da produção mundial de grãos (UNITED STATES DEPARTMENT OF AGRICULTURE - USDA, 2015). Por essas razões, provavelmente não exista outra espécie de importância econômica que tenha sido alvo de tão intensas pesquisas científicas, cujos resultados têm contribuído significativamente não só para o aperfeiçoamento de seu cultivo, mas também influenciado técnicas empregadas em outras culturas.

No caso do milho, as cultivares anteriormente utilizadas na produção agrícola eram majoritariamente variedades de polinização aberta. Jenkins (1936) relata a importância da seleção e posterior autofecundação das melhores cultivares para a síntese de outras novas cultivares de polinização aberta. Entretanto, com o advento de empresas de pesquisa agrícola e produção de sementes de milho híbrido, juntamente com a melhoria do vigor das linhagens obtidas por meio do melhoramento, os cruzamentos entre linhagens elite se tornaram uma estratégia eficiente para a obtenção de híbridos comerciais (TROYER, 1990). Todavia, antes da obtenção dos híbridos comerciais, o programa passa pelo melhoramento de linhagens.

Vários métodos de melhoramento podem ser implementados, como o método padrão com seleção entre e dentro de progênies, ainda largamente empregado e cuja base genética está descrita no trabalho de Shull (1909), método da cova única (JONES; SINGLETON, 1934), retrocruzamentos (LINDSTROM, 1939), método genealógico (ALLARD, 1971), ou mesmo por meio do processo de reciclagem de linhagens com a utilização de tecnologia duplo haploide (BATTISTELLI et al., 2013; BERNARDO, 2009; CHASE, 1952; ROBER; GORDILLO; GEIGER, 2005; SMITH et al., 2008). Com o advento desta última tecnologia, foi possível aumentar o número de linhagens obtidas dentro dos programas de melhoramento. Em consequência do aumento do número de linhagens, o número potencial de híbridos a serem testados ao final de cada ciclo de melhoramento aumentou exponencialmente. Desta maneira, tecnologias relacionadas à predição de híbridos ganharam cada vez mais espaço e importância nos programas de melhoramento em instituições públicas e privadas, no sentido de reduzir os custos atribuídos à obtenção e avaliação de novos híbridos, melhor utilização do

germoplasma disponível e para o conhecimento do desempenho de híbridos faltantes em determinados ambientes. Mais detalhes serão apresentados nos itens subsequentes.

2.2 Interação Genótipos por Ambientes no Melhoramento de Plantas

Em programas de melhoramento de plantas, a busca pela seleção e recomendação de genótipos que apresentam ampla adaptabilidade e estabilidade de produção aos mais variados ambientes é objetivo alvo dos profissionais envolvidos na obtenção de cultivares (SILVA; DUARTE, 2006). Devido ao número de alelos atuantes no controle genético dos principais caracteres agrônômicos, a probabilidade de se obter uma cultivar ideal é mínima, pois estes são, de maneira geral, altamente influenciados por variações ambientais, o que dificulta a seleção dos melhores indivíduos. Assim, um dos grandes desafios dos melhoristas consiste em obter novas combinações genótípicas agronomicamente superiores, que possam ser utilizadas como cultivares comerciais e que substituam com vantagens as preexistentes (BERNARDO, 2002). Adicionalmente, e não menos importante, para a permanência desse híbrido no mercado, é necessário que essas combinações híbridas apresentem também estabilidade produtiva em vários ambientes de cultivo.

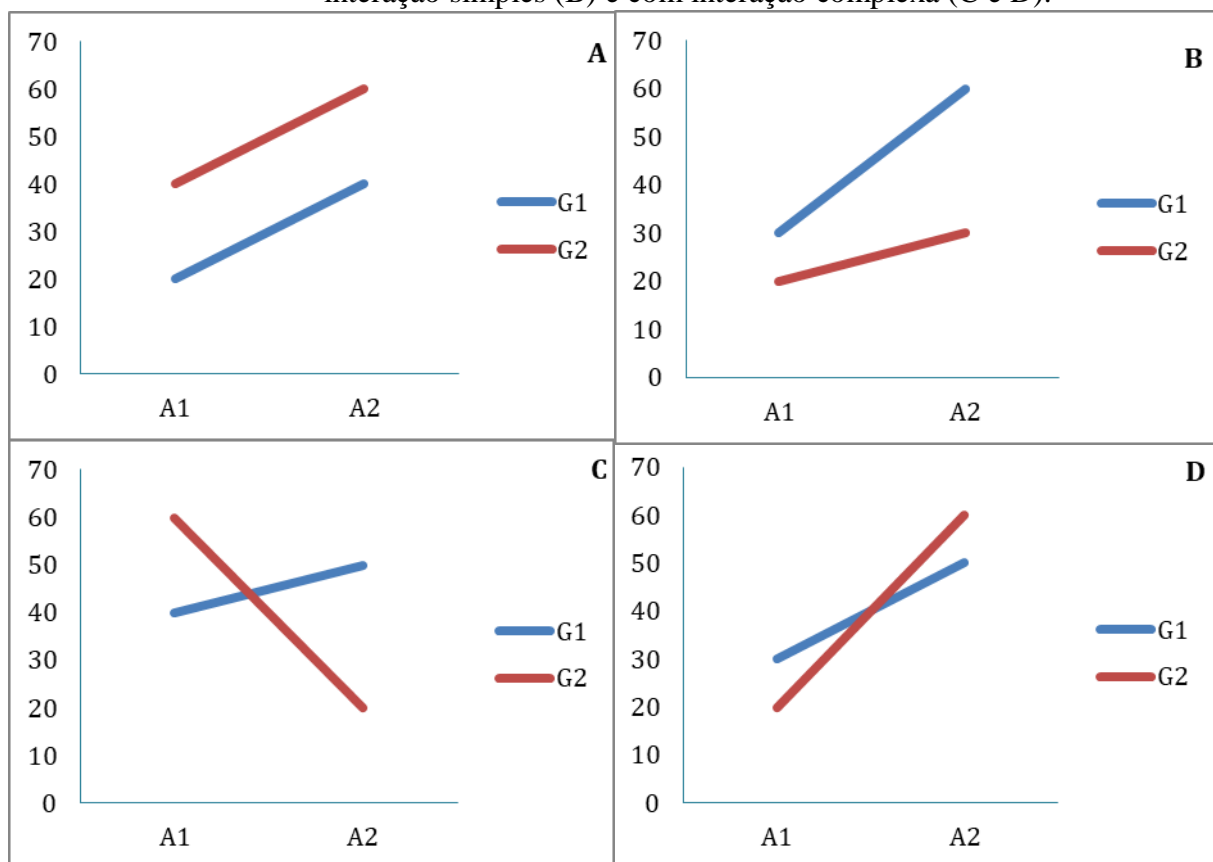
A interação GE pode ser percebida pelas respostas diferenciadas de genótipos quando avaliados em diferentes anos, locais, épocas de semeadura, dentre outros (CROSSA, 1990; YAN; KANG, 2003). Trata-se de uma situação decorrente de instabilidade nas manifestações genótípicas em diferentes condições ambientais, ou seja, ausência de correlação entre o comportamento dos genótipos em diferentes ambientes (CROSSA; YANG; CORNELIUS, 2004). É recorrente na literatura relatos de que a interação GE representa o principal motivo para que experimentos sejam conduzidos em vários ambientes em etapas pré-comerciais de avaliação (CROSSA, 1990). Assim, ainda que as diferenças genéticas e ambientais contribuam para a variação fenotípica, não é aceitável desconsiderar também a interação entre elas, o modo como essa interação pode ocorrer, bem como as formas com que ela pode ser detectada, medida e avaliada.

Entende-se por ambiente, todos os fatores intra e extracelulares que influenciam na expressão do genótipo (BREWBAKER; AQUILIZAN, 1965). No que diz respeito à variação ambiental, entende-se como toda e qualquer variação que ocorre de região para região dentro da área que abrange a distribuição de culturas. As variações ambientais dividem-se em variações de natureza previsível e imprevisível. Podem ser caracterizadas como variações

previsíveis características de clima e solo em uma mesma região que flutuam de maneira sistemática, como comprimento do dia, grau de insolação, e algumas práticas agrônômicas como: épocas de semeadura e colheita, doses e adubação, dentre outras (ALLARD; BRADSHAW, 1964; VENCOVSKY; BARRIGA, 1992). Já variações imprevisíveis são aquelas que não podem, *a priori*, serem preditas com alta confiabilidade, como quantidade e distribuição de chuvas e temperatura (RAMALHO et al., 2012).

Nesse contexto, pesquisadores e melhoristas têm grande interesse em conhecer a natureza da interação genótipos por ambientes e, desta forma, realizar predições mais precisas sobre os efeitos da mesma, viabilizando assim a seleção dos genótipos efetivamente superiores (YAN; KANG, 2003). Entretanto, na prática da seleção, é rotineiro o pesquisador realizar o ranqueamento dos genótipos avaliados, onde é possível verificar diferentes classificações quanto às respostas dos genótipos frente às variações ambientais. Essas classificações estão apresentadas na Figura 1.

Figura 1 Comportamento de dois genótipos (G1 e G2) em duas condições ambientais (A1 e A2), com respostas genotípicas caracterizando ausência de interação (A), interação simples (B) e com interação complexa (C e D).



Pode-se verificar, na Figura 1 (A), que quando os genótipos apresentam comportamentos semelhantes nos dois ambientes (A1 e A2), caracteriza-se um caso específico de ausência de interação. A diferença de resposta deve-se exclusivamente à diferença entre ambientes, de maneira que o melhor híbrido a ser recomendado no A1 também será o melhor em A2, ou seja, a recomendação do melhor genótipo não se altera. No segundo exemplo (Figura 1B), ocorre uma resposta diferenciada dos dois genótipos (G1 e G2) em relação a A1 e A2, pois G1 apresenta resposta mais acentuada à melhoria do ambiente. Neste caso, porém, ainda que a interação dos genótipos com os ambientes de avaliação seja evidente, não ocasiona dificuldades adicionais na seleção dos genótipos; a classificação dos mesmos não é alterada de acordo com a variação do estímulo ambiental e, por esta razão, é então denominada interação simples (CRUZ; CARNEIRO, 2003). Na Figura 1 (C e D) observa-se uma situação de interação complexa, cuja principal característica é a inversão no ranqueamento dos genótipos avaliados em função de variações nos estímulos ambientais (DUARTE; VENCOVSKI, 1999). Existe uma concordância entre os pesquisadores de que a interação do tipo complexa é a mais importante e de maior ocorrência nos ensaios multiambientais (GAUCH, 2013; YAN; KANG, 2003).

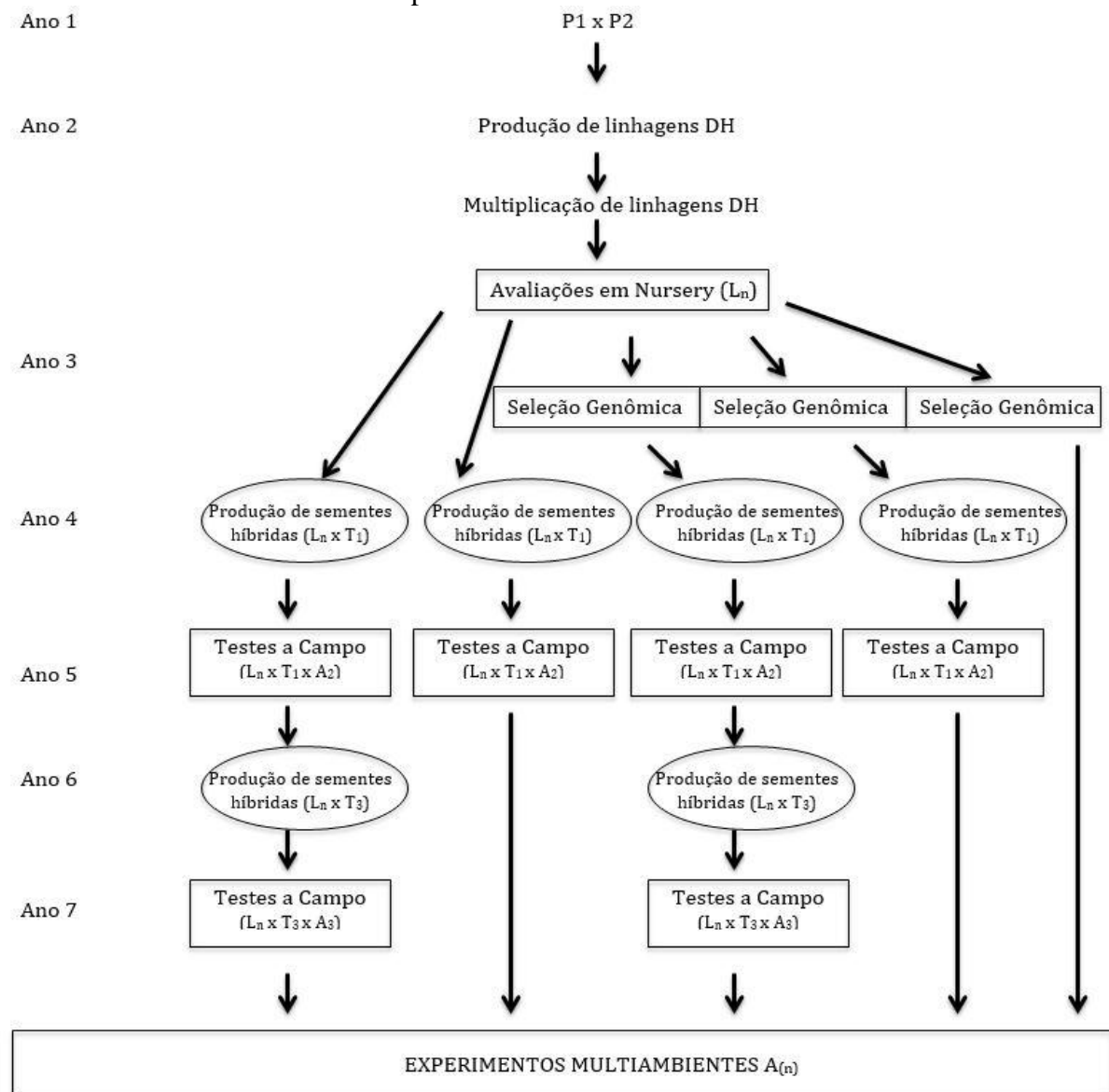
Desta forma, é importante que os genótipos possam ser recomendados para o maior número de ambientes possível (locais, épocas de semeadura, níveis tecnológicos dos produtores, dentre outros), representando assim a variação ambiental que caracteriza parte das áreas utilizadas para produção comercial de milho no Brasil. Adicionalmente, a identificação e utilização comercial de genótipos com ampla estabilidade e adaptabilidade tem interferência direta na amplitude do portfólio das empresas produtoras de sementes, aumentando diretamente a competitividade das mesmas no mercado.

Uma vez relatada a importância acerca da condução de ensaios multiambientais, alguns objetivos principais devem ser considerados, tais como: o estudo da interação GE, a análise da adaptabilidade e estabilidade genotípica dos genótipos em avaliação, bem como a realização de predições precisas de forma a garantir máxima eficiência na seleção (CROSSA et al., 2011). Nesse contexto, utilizar metodologias para lidar com essa interação, e que permitam a predição fidedigna dos melhores genótipos a serem recomendados em cada ambiente é um desafio para melhoristas, estatísticos e demais profissionais envolvidos na diminuição dos riscos intrínsecos ao processo de recomendação de cultivares.

2.3 Predição Genômica no Contexto Atual do Melhoramento de Milho

Com objetivo de obter as melhores combinações híbridas, o número de cruzamentos a serem realizados, bem como o número desses a serem testados em ensaios multiambientes, demanda custos elevados, o que pode inviabilizar o programa de melhoramento, além de dificultar a identificação dos híbridos efetivamente superiores (LAMKEY; HALLAUER; KAHLER, 1987). Na Figura 2 é mostrada a condução de um programa de melhoramento através de diversas estratégias de avanço de gerações.

Figura 2 Representação do avanço de um programa de melhoramento por meio de diferentes estratégias de seleção, até a etapa relacionada aos experimentos multiambientes.



Uma das etapas mais importantes no melhoramento de milho é a escolha das linhagens a serem utilizadas em blocos de cruzamentos para obtenção dos híbridos. Existem critérios que auxiliam os melhoristas no que diz respeito a essa tomada de decisão. Entretanto, por muitos anos, esse direcionamento foi realizado quase que empiricamente, baseado na média das linhagens e no conhecimento prático do germoplasma por parte do melhorista. Adicionalmente, é importante refletir acerca do novo cenário com a utilização da tecnologia duplo haploides. Uma das importantes consequências corresponde ao aumento do número potencial de cruzamentos entre linhagens de diferentes grupos heteróticos, o que incrementa também o número de híbridos obtidos nos programas de melhoramento (BATTISTELLI et al., 2013; CHASE, 1952).

Nesse contexto, é essencial considerar que, antes mesmo da descoberta e potencial utilização da tecnologia de duplos haploides por Chase (1952), também ocorreu o desenvolvimento de práticas que contribuem no sentido de reduzir o número total de híbridos testados, mantendo o foco naqueles que se mostrarem mais promissores. Em 1934, época em que nos Estados Unidos os genótipos comerciais mais utilizados eram os híbridos duplos, Jenkins iniciou os trabalhos de predição do desempenho dos mesmos por meio da *performance* dos híbridos simples não parentais (JENKINS, 1934). Posteriormente, Cockerham (1967) e Eberhart, Russel e Penny (1964) foram responsáveis por aprimorar as ideias de Jenkins, no sentido de criar modelos eficientes também para a predição de híbridos triplos, mas não eram úteis na predição de híbridos simples. É conveniente argumentar que, até então, pouca ou nenhuma informação era discutida acerca da predição de híbridos simples, cenário condicionado pela própria impossibilidade de utilização comercial desse tipo de genótipo, uma vez que o baixo vigor apresentado pelas linhagens correspondia a um grande gargalo na produção de sementes do mesmo.

Entretanto, registra-se que o melhoramento genético vem passando por importantes alterações que podem incrementar a eficiência de condução dos programas. Como exemplo, tem-se a utilização de marcadores de DNA, que possibilitou não só o estudo do genoma de várias espécies, mas também a identificação de regiões genômicas associadas a expressão de determinados caracteres. Nota-se que uma justificativa relevante para a utilização de informações moleculares é a expectativa de que as informações em nível de DNA tragam maior ganho genético e incremento na eficiência de condução do programa de melhoramento do que quando se utilizam apenas os dados fenotípicos, como relatado por Meuwissen, Hayes e Goddard (2001). Não obstante, a disponibilidade de marcadores moleculares em milho

permitiu que os melhoristas questionassem, a princípio, de que modo esses marcadores poderiam ser utilizados para alcançar melhores ganhos de seleção (BERNARDO; YU, 2007).

Nos últimos anos, o potencial de contribuição da genética molecular foi beneficiado por meio do desenvolvimento de técnicas de genotipagem em larga escala, como a Genotyping-by-sequencing (GBS) (ELSHIRE et al., 2011), que trouxeram em seu bojo a diminuição do custo de genotipagem. A partir deste cenário, associado à forte tendência de aumento de gastos com fenotipagem, houve grande interesse na utilização de técnicas visando explorar informações genômicas, focando em minimizar os custos totais de condução de programas de melhoramento (HESLOT et al., 2014). Além disso, uma segunda contribuição significativa é permitir ao melhorista ter acesso a informações de um maior número de híbridos, aumentando a utilização da variabilidade potencial presente no germoplasma. Desta forma, na prática da seleção feita para caracteres de importância econômica, antes realizada somente a partir de dados fenotípicos, passou-se, também, a considerar dados moleculares.

Neste contexto, foi possível incorporar com maior eficiência a estatística genômica ao melhoramento de plantas, pois metodologias estatísticas adequadas para lidar com banco de dados massivo se fizeram necessárias com o advento e integração dessas novas tecnologias nos programas de melhoramento. Para a cultura do milho, Eathington et al. (2007) relataram o uso de marcadores em programas comerciais de melhoramento, como exemplificado pela corporação Monsanto. No intervalo de tempo de 2000 a 2006, o número de *datapoints* gerados pela empresa em seu programa de melhoramento aumentou mais de 40 vezes, enquanto seu custo por *datapoint* diminuiu mais do que seis vezes.

A proposta da utilização massiva de marcadores moleculares em técnicas de seleção baseada no genótipo foi, inicialmente, proposta por Meuwissen, Hayes e Goddard (2001), que definiram seleção genômica ampla (GWS – Genome Wide Selection) como a seleção baseada em marcadores sem a necessidade de testes de significância em nível de marca ou identificação de um subconjunto de marcadores associados com o caráter em estudo. Por meio de estudo de simulação, estimaram o efeito de aproximadamente 50.000 marcadores haplótipos em desequilíbrio de ligação com QTLs a partir de um número limitado de registros fenotípicos, e utilizaram esses efeitos na predição dos valores genéticos genômicos dos animais para seleção. Os resultados indicam que a seleção genômica pode levar a altas correlações entre o valor genético predito e o verdadeiro, considerando um caráter quantitativo, e que a seleção baseada em valores genéticos genômicos preditos pode aumentar substancialmente as estimativas de progresso genético com a seleção em programas de

melhoramento animal e vegetal, especialmente se metodologias preditivas forem combinadas com técnicas eficientes para encurtar o intervalo entre gerações. Adicionalmente, Meuwissen, Hayes e Goddard (2001) compararam as técnicas dos quadrados mínimos, BLUP, Bayes A e Bayes B e 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. Assim, é importante salientar que esses autores também já verificaram que modelos bayesianos podem proporcionar maiores resultados de acurácia preditiva.

No âmbito dos programas de melhoramento de milho, antes mesmo do estabelecimento das diretrizes para aplicação da predição genômica, parecia comum a ideia de que a maior contribuição neste cenário seria a capacidade de prever com alta acurácia a performance de híbridos simples de milho. Bernardo (1994) explicitou que métodos de predição do rendimento de híbridos simples atuariam no sentido de facilitar a identificação de cruzamentos individuais superiores. O autor propôs uma alternativa para a predição genômica que leva em consideração informações de marcadores das linhagens parentais e os dados de produtividade de uma população de cruzamentos testes por meio do melhor preditor linear não viesado (BLUP). No caso, o estudo foi baseado em marcadores RFLP, ainda que o autor tenha relatado que, na ausência de informações moleculares, é possível a utilização da informação de parentesco obtida por meio das informações derivadas do pedigree entre as linhagens avaliadas. Assim, pode-se perceber que o autor lida com os híbridos a serem preditos como dados faltantes, pois na expressão $y_M = CV^{-1} y_P$, y_M corresponde ao vetor de híbridos simples cujos desempenhos serão preditos (não há dados de produtividade para os mesmos), $C = m \times n$ corresponde à matriz de covariância genética entre os híbridos faltantes e os híbridos cuja informação fenotípica é disponível, $V = n \times n$ é a matriz de variâncias e covariâncias fenotípicas entre os híbridos, e $y_P = n \times 1$ corresponde ao vetor de resposta dos híbridos preditores corrigidos para o efeito dos ensaios. Com estimativas consideravelmente altas de correlações entre os valores preditos e observados obtidos neste estudo, pode-se sugerir que BLUP pode ser útil para identificar combinações híbridas superiores. Outros trabalhos também corroboram os resultados obtidos por Bernardo (1994, 1995, 1996a, 1996b) e Massman et al. (2013).

A abordagem de predição por BLUP assume a mesma variância para todas as marcas, o que é irreal, e pode levar a uma redução da acurácia de predição para caracteres oligogênicos, como já havia sido demonstrado por Meuwissen, Hayes e Goddard (2001)

desde a primeira publicação relacionada à proposta da seleção genômica. Massman et al. (2013), por meio da avaliação de 479 híbridos experimentais derivados do cruzamento entre genótipos dos grupos Iowa Stiff Stalk Synthetic e non Stiff Stalk Synthetic, realizaram a predição do desempenho para algumas características agronômicas. Em geral, não verificaram diferenças expressivas entre as acurácias preditivas quando o parentesco foi estimado via marcadores SNP (RR-BLUP) ou via informações oriundas da genealogia entre os híbridos (BLUP). Por outro lado, resultados de simulação (ZHONG et al., 2009) demonstraram que RR-BLUP e BLUP diferem em suas precisões de predição sob diferentes níveis de desequilíbrio de ligação entre marcadores e QTL. Adicionalmente, Zhong et al. (2009) compararam a acurácia de diferentes modelos de predição genômica em vários cenários, variando a densidade de marcadores, o nível do desequilíbrio de ligação, o número de QTL e o tamanho da amostra. Os autores observaram que o Bayesian Shrinkage Regression Method pode ser mais eficiente em capturar o desequilíbrio de ligação entre marcador e QTL, e que métodos derivados de modelos mistos podem recuperar melhor a informação de parentesco.

Assim, algumas reflexões são importantes acerca do referencial supracitado. Uma delas é a importância da escolha do modelo mais adequado dentre uma gama de modelos disponíveis, em um cenário que caracteriza-se pelo aumento da densidade de marcadores nas populações de interesse para o melhorista de maneira proporcional ao decréscimo do custo por *datapoint*. Zhong et al. (2009) relataram uma pequena superioridade de métodos bayesianos em condições de alto grau de associação entre marcador e QTL. Além do mais, Habier et al. (2007) demonstraram, por simulação computacional, o decréscimo na acurácia preditiva quando é negligenciada a informação de parentesco entre os indivíduos da população estudada, relatando que o valor genético genômico das progênes pode ser em grande parte resultado do parentesco genômico capturado pelos marcadores moleculares. Em geral, isso é extrapolado para todos os métodos que estimam efeitos de marcadores para predição do *breeding value*, ainda que o impacto do parentesco negligenciado seja mais acentuado em métodos baseados em modelos mistos (RR-BLUP) do que em métodos bayesianos (Bayes B) (HABIER; FERNANDO; DEKKERS, 2007).

Este cenário, associado ao elevado número de ambientes em que os genótipos precisam ser avaliados antes da potencial utilização comercial, faz com que a elaboração e utilização de métodos que abordem de maneira fidedigna o estudo da interação entre genótipos e ambientes, associados ao acréscimo das estimativas de acurácia preditiva ou correlação entre os valores preditos e observados, assumam importância fundamental e

possam levar a um aumento significativo da eficiência dos programas de melhoramento de plantas. Maiores detalhes serão abordados nos tópicos subsequentes.

2.4 Additive Main Effects and Multiplicative Interaction Analysis – AMMI

Para tirar conclusões mais abrangentes sobre os resultados das avaliações de experimentos multiambientes, bem como explorar os efeitos positivos da interação GE, é preciso dispor de métodos eficazes que sejam capazes de captar tais informações (DUARTE; VENCOVSKY, 1999). Essa interação tem sido modelada à luz de diversos procedimentos envolvendo análises uni e multivariadas. Os últimos são considerados mais eficientes para explorar as informações contidas nos dados, como a análise de componentes principais (PCA), análises de agrupamento, análises discriminantes e o modelo AMMI, que integra análise univariada com PCA.

O modelo AMMI se configura em um procedimento que incorpora componentes aditivos para os efeitos principais de genótipos e de ambientes e, ainda, componentes multiplicativos para o efeito da interação entre esses fatores (GAUCH, 1988). Trata-se de um modelo proposto por Gollob (1968) e Mandel (1969, 1971), e que possui aplicações em diversas áreas de estudo; entretanto, é um modelo de efeitos fixos.

O ajuste do modelo AMMI é realizado em duas etapas. Primeiramente, ajustam-se os efeitos principais pelo método de mínimos quadrados aplicado à matriz de médias. A segunda etapa do ajuste corresponde à decomposição de valores singulares (DVS) da matriz de interações que é obtida como resíduo do ajuste aos efeitos principais pela análise de variância. Trata-se de uma metodologia que tem sua capacidade de predição proporcional à capacidade de recuperar fatores ambientais e genéticos associados à interação, conhecido como padrão, e descartar uma porção não associada à interação, denominada ruído. O ruído não tem interpretação e nem valor preditivo, e é considerado a variabilidade geralmente inexplicável entre as repetições (GAUCH, 1988; OLIVEIRA et al., 2016; ZOBEL; WRIGHT; GAUCH, 1988).

O procedimento de análise AMMI permite o ajuste de um número pequeno de componentes multiplicativos sem, no entanto, perder o poder explicativo, ou seja, permite a obtenção de um modelo parcimonioso. Existem diferentes critérios para se estabelecer o número ideal de termos bilineares a serem retidos no modelo. Os métodos mais utilizados são os testes F aproximados como o teste de Gollob (1968), o teste Fr proposto por Cornelius,

Seyedsadr e Crossa (1992) e aqueles baseados em reamostragem via computação intensiva, como o procedimento de validação cruzada (GAUCH, 1988). Existe ainda a possibilidade de utilização de estimadores com efeito de encolhimento “shrinkage” (CORNELIUS; CROSSA; SEYEDSADR, 1996; CORNÉLIUS; SEYEDSADR, 1997).

Uma das etapas da análise AMMI é a representação gráfica dos escores genotípicos e ambientais por meio de biplots (GAUCH, 1988; YAN; TINKER, 2006). O método biplot, proposto por Gabriel (1971), consiste na representação gráfica das linhas e colunas de uma matriz de forma que cada elemento dela seja exatamente o produto interno dos vetores correspondentes (vetores linha e coluna). Os escores plotados nesse gráfico referem-se às coordenadas dos vetores singulares multiplicados pelos respectivos valores singulares. O princípio fundamental subjacente ao biplot assenta-se na representação de indivíduos e variáveis de forma simultânea, num espaço de baixa dimensão, de forma a facilitar a observação das relações existentes entre eles.

Uma das grandes virtudes de um biplot decorre do fato do produto interno entre dois vetores ser de fácil interpretação visual. Essa interpretação é realizada pela observação do comprimento de cada vetor e do sinal do produto interno, que depende apenas do ângulo formado entre eles (GABRIEL, 1971; YAN; KANG, 2003).

Pode-se notar que podem ocorrer k componentes principais para o modelo multiplicativo e, com isso, inúmeros modelos AMMI são possíveis a depender do número de componentes principais que apresentaram significância. No modelo AMMI podem ocorrer desde o AMMI0 (modelo aditivo, ou seja, nenhum PC significativo) até AMMIF (todos os PCs significativos) (ANNICCHIARICO, 2002), sendo F igual ao menor posto da matriz de dados da interação.

Sabe-se que uma análise AMMI com mais de três componentes apresenta grande dificuldade de interpretação; sendo assim, os pesquisadores normalmente optam pela apresentação utilizando também somente dois componentes. Resumidamente, a diferença entre as equações do modelo AMMI e GGE é que AMMI subtrai o efeito genotípico, e GGE não o faz, antes de realizar a análise DVS. Conseqüentemente, acarretará em diferentes valores para os parâmetros multiplicativos encontrados nas análises DVS por utilizar diferentes matrizes de dados (GAUCH; PIEPHO; ANNICCHIARICO, 2008).

O modelo completo para a metodologia AMMI, inicialmente proposto por Mandel (1971), apresentado em várias literaturas, é dado por:

$$Y_{ij} = \mu + g_i + a_j + \sum \lambda_k \gamma_{ik} \delta_{jk} + \rho_{ij} + \varepsilon_{ij}, k=1$$

em que: Y_{ij} é o valor médio observado do i -ésimo genótipo no j -ésimo ambiente; μ é a média geral dos experimentos; g_i e a_j são os efeitos genotípicos e ambientais, respectivamente; $(g_a)_{ij}$ é a interação genótipos por ambientes, e é representada por $\sum \lambda_k \gamma_{ik} \delta_{jk} + \rho_{ij}$, $k=1$; k é o rank do componente principal, sendo $k = 1 \dots a$; λ_k é o valor singular do k -ésimo componente principal (PC), com $\lambda_1 > \lambda_2 > \dots > \lambda_a$, sendo que λ_k ao quadrado resulta na soma de quadrados explicada pelo k -ésimo componente principal; γ_{ik} e δ_{jk} são os vetores singulares do genótipo i e ambiente j , respectivamente, do k -ésimo componente principal; ρ_{ij} é o ruído do modelo multiplicativo; e ε_{ij} é o erro experimental médio.

Apesar das vantagens desse modelo para estudar padrões de desempenho de genótipos a diferentes ambientes e fazer previsões a respeito do desempenho médio de genótipos a ambientes específicos, o modelo AMMI, bem como outros modelos lineares-bilineares (CORNELIUS; SEYEDSADR, 1997), possui diversas limitações, dentre as quais se destaca o fato de não existir estatística inferencial para os escores genotípicos e ambientais no gráfico biplot (CROSSA et al., 2011; YANG et al., 2009).

Oliveira et al. (2015) citam alguns estudos conduzidos com o intuito de minimizar este problema, como os trabalhos sobre a Teoria assintótica de Denis e Gower (1994, 1996) e Denis e Pazman (1999), e os trabalhos que utilizam amostragem bootstrap de Lavoranti, Dias e Kraznowski (2007) e Yang et al. (2009). Porém, essas metodologias são baseadas em suposições restritivas, como normalidade assintótica ou processos de reamostragem problemáticos (YAN et al., 2010). No sentido de solucionar esse problema, Crossa et al. (2011) e Oliveira et al. (2015) relatam a possibilidade de utilização da abordagem bayesiana associada à implementação de regiões de máxima densidade a posteriori (HPD), que podem ser definidas para escores associados a genótipos e a ambientes, facilitando, assim, a identificação de grupos homogêneos para os efeitos em questão. Os autores ainda relatam que a associação entre a abordagem bayesiana para o modelo AMMI com regiões de máxima densidade a posteriori, proporcionam um método natural para construção de regiões de credibilidade no biplot.

Algumas vantagens de destaque acerca do método supracitado devem ser citadas, tais como a flexibilidade inerente para lidar de maneira eficiente com conjunto de dados desbalanceados, considerar o efeito de genótipos como sendo aleatório associado à possibilidade de utilização de priori informativa para o mesmo e, também, a possibilidade de

incorporação de coeficiente de parentesco, o que pode conduzir a predições mais realistas (SMITH; CULLIS; THOMPSON, 2005).

2.5 Utilização da Informação de Parentesco

Uma das primeiras tentativas para melhorar as análises genéticas foi considerar o parentesco entre os indivíduos com vistas a explorar as conexões genéticas entre eles e, assim, refinar as predições obtidas. Henderson (1963) desenvolveu uma metodologia para análises genéticas no melhoramento animal exatamente explorando a covariância genética entre os indivíduos. Nesse momento, as estimativas de covariância genética proviam de informações do pedigree dos animais, conhecida como identidade por descendência (IBD). No entanto, o valor genético de um indivíduo tem dois componentes: a contribuição média do valor genético dos pais e um componente devido à segregação Mendeliana de gametas (DAETWYLER et al., 2013). Apenas o primeiro é considerado pelas estimativas IBD e, assim, um viés sempre acompanhará tais estimativas (POWELL; VISSCHER; GODDARD, 2010).

Por outro lado, marcadores podem ser usados para fornecer estimativas de relacionamento genético entre indivíduos, conhecida como identidade por estado (IBS), desconsiderando a proximidade genética e permitindo a recuperação de informação a partir de indivíduos mais distantes. Grande parte dos ganhos em acurácia de seleção devido ao uso de marcadores se dão por causa da exploração do componente de segregação Mendeliana. Atualmente, é conhecido que a confiabilidade das predições genômicas é fortemente influenciada pelo relacionamento genético entre os indivíduos envolvidos (CLARK et al., 2012; HABIER; FERNANDO; DEKKERS, 2007; WIENTJES; VEERKAMP; CALUS, 2013).

Desde a disponibilidade de plataformas de genotipagem de alta escala, os melhoristas têm demonstrado interesse no sentido de integrar essas novas ferramentas aos procedimentos clássicos já usados pelos programas. A tendência de redução dos custos com a genotipagem e a improvável redução de custos com a fenotipagem é um fator a favorecer um aumento no uso de marcadores moleculares no melhoramento de plantas (HILL, 2014).

Grande esforço tem sido direcionado no sentido de aumentar a eficiência de programas com a utilização de métodos analíticos genômicos. Atualmente, o emprego dessas técnicas vai desde a compreensão da arquitetura genética da inteligência de humanos, do entendimento de mecanismos genéticos associados à resistência de plantas contra patógenos, até o estudo da interação genótipos por ambientes (EDAE et al., 2014).

O modelo de regressão direta no genoma mais adotado em análises genômicas é o GBLUP (Genomic Best Linear Unbiased Predictor), devido à sua simplicidade interpretativa, flexibilidade de modelagem, ser pouco intensivo computacionalmente e intuitivo geneticamente (GIANOLA et al., 2014). Apesar de proposto recentemente (VANRADEN, 2008), já há muitos anos esse modelo teve seu fundamento elucidado (HENDERSON, 1984). Trata-se de um modelo semelhante ao modelo BLUP, com a diferença de o parentesco entre indivíduos não ser estimada por pedigree, e sim por marcadores moleculares.

O uso de marcadores moleculares permite a exploração de componentes aditivos e não aditivos, com informações de contrastes em nível genômico, que auxilia na superação das restrições comuns que são encontradas nos delineamentos tradicionais. Dessa maneira, seu uso tem grande potencial na estimação de parâmetros genéticos, levando a modelos mais acurados em relação àqueles baseados somente em pedigree (SU et al., 2012). Além disso, é esperado que a estimação de parâmetros genéticos considerando apenas informações do pedigree seja viesada e com maior erro por não considerar a segregação mendeliana e o relacionamento genético entre indivíduos com parentesco desconhecido. Deste modo, alguns trabalhos na literatura relatam maior acurácia na estimativa dos componentes de variância quando se utiliza a matriz de parentesco genômica (MUÑOZ et al., 2014).

Uma propriedade desejável durante a modelagem é a obtenção de modelos que promovam a decomposição de componentes de variância e efeitos de maneira ortogonal. Ou seja, no caso de modelos genéticos, que não exista confundimento entre os valores dos componentes obtidos pela decomposição da variância genética total ou do valor genético total. Existem três situações em que esse fenômeno pode ocorrer: (i) presença de desequilíbrio de ligação; (ii) emprego de matrizes de incidência sem ortogonalidade; e (iii) presença de interação genótipos por ambientes (GIANOLA et al., 2014; ZENG; WANG; ZOU, 2005).

Em geral, tem sido atribuída pouca utilidade prática, para fins seletivos, às informações de parentesco quando essas estão disponíveis aos melhoristas graças ao trabalho adicional de anotações genealógicas (RAMALHO et al., 2001). Atualmente, essas informações também podem ser obtidas por meio de marcadores moleculares, explorando a similaridade genética por estado entre os indivíduos, independentemente da proximidade genealógica entre os mesmos (POWELL; VISSCHER; GODDARD, 2010). Segundo Bearzoti (2002), o ordenamento de genótipos pode ser alterado, de acordo com a escolha do modelo estatístico, em situações nas quais houver parentesco entre as unidades de seleção e ou, desbalanceamento nos dados.

Bromley et al. (2000) investigaram se a inclusão ou não da informação de parentesco entre linhagens de milho afetava as estimativas dos componentes de variância. Os autores observaram que houve um acréscimo no componente de variância referente às linhagens quando a informação foi considerada, indicando que o ganho potencial com a seleção mediante inclusão dessa informação seria maior do que o predito, ignorando-a.

Panter e Allen (1995) compararam modelos BLUP, com e sem a inclusão da informação de parentesco genético das linhagens utilizadas no cruzamento, com o método dos mínimos quadrados para a predição de cruzamentos em soja. Os autores observaram que os erros padrões das estimativas obtidas por mínimos quadrados foram, em média, 6% maiores do que as obtidas com os modelos BLUP. Contudo, não foram observadas diferenças marcantes entre os modelos BLUP, sendo o modelo com informação de parentesco apenas ligeiramente superior.

2.6 Estatística Bayesiana

Em um modelo bayesiano, o parâmetro é considerado como escala ou vetor aleatórios (não observável), e não as estimativas como no caso do contexto frequentista. A inferência Bayesiana tem como filosofia que tudo o que é desconhecido é incerto e, portanto, toda incerteza deve ser quantificada em termos de probabilidade (PAULINO; TURKMAN; MURTEIRA, 2003). Isto se torna vantajoso do ponto de vista de seleção. Como se pretende identificar os melhores genótipos entre aqueles em avaliação, é interessante considerar seus efeitos como aleatórios, pois esta abordagem permite uma melhor predição dos verdadeiros valores genotípicos e, conseqüentemente, um melhor ranqueamento dos genótipos (SMITH; CULLIS; THOMPSON, 2005).

Os modelos bayesianos levam em conta uma informação inicial, chamada informação *a priori*, que possui uma distribuição de probabilidade (subjéitiva ou lógica) e que permite um processo de atualização do conhecimento em estudos posteriores (PAULINO; TURKMAN; MURTEIRA, 2003). Com as informações *a priori* e amostrais, é possível atualizar periodicamente a distribuição de probabilidade *a posteriori*, modificando as estimativas dos parâmetros (RESENDE, 2000).

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Desta forma, a informação *a posteriori* incorpora toda a informação disponível sobre o parâmetro, ou seja, a informação inicial já possuída e a nova informação obtida com os dados experimentais, incorporada na predição por meio da função de verossimilhança. Resende (2000) ressalta a importância da função de verossimilhança na transformação do conhecimento *a priori* sobre o parâmetro:

$$\text{distribuição a posteriori} \propto \text{verossimilhança} \times \text{distribuição a priori}$$

Gauch (2003) afirma que o teorema de Bayes constitui-se em uma simples forma de ajuste do conceito de probabilidade condicional e dá os seguintes significados para cada um de seus termos:

- Priori: este termo refere-se a uma probabilidade prévia aplicada a uma hipótese, proposição, modelo, evento ou crença, ou seja, é uma probabilidade inicial de uma proposição ser verdadeira ou um evento ocorrer antes de coletar novos dados.

- Verossimilhança: determina o impacto que os dados terão sobre a probabilidade da hipótese. A hipótese que fornece baixa probabilidade aos novos dados coletados tem pouca credibilidade, enquanto que a hipótese que fornece alta probabilidade a estes dados tem forte credibilidade. A credibilidade das hipóteses terá influência dos dados, variando conforme se ajustam aos mesmos.

- Posteriori: refere-se a uma probabilidade que, assim como a priori, é aplicada a uma hipótese, proposição, modelo, evento ou crença. É uma probabilidade obtida após a combinação das informações consideradas pela verossimilhança e pela priori.

Segundo Resende (2000), a análise Bayesiana permite obter estimativas mais precisas de componentes de variância, parâmetros genéticos, valores genéticos e ganhos genéticos. O autor afirma que a estimação Bayesiana permite a análise exata de amostras de tamanho finito, tendo importância em programas de melhoramento baseados em conjuntos de dados desbalanceados. Isto permite a realização de uma elegante análise de amostra finita, a qual não pode ser obtida pela metodologia frequentista de modelos mistos. Porém, o uso do teorema de Bayes no contexto dos ensaios multiambientais e seleção de cultivares é ainda limitado (CROSSA et al., 2011).

Josse et al. (2014), conduzindo trabalho utilizando o AMMI-Bayesiano, mostraram que esta abordagem permite responder questões práticas de importância para o melhorista, como a probabilidade de um genótipo produzir menos que um determinado valor em um ambiente

selecionado aleatoriamente. Os autores também ressaltam o fato de melhoristas já terem ideia do valor da média do ambiente em estudo, a magnitude das variâncias genética e ambiental, e a magnitude da interação quando possuem informações prévias, reforçando o interesse no uso da abordagem Bayesiana em estudos da interação GE.

A abordagem bayesiana permite o uso da estimação por regiões, em que um resumo da *posteriori*, mais informativo do que uma estimativa pontual, é obtido de uma região no espaço paramétrico, semelhantes aos intervalos de confiança da abordagem frequentista (PAULINO; TURKMAN; MURTEIRA, 2003). Esses autores afirmam que as regiões de credibilidade não são aleatórias e permitem um estudo probabilístico preciso, em contraste à região de confiança clássica.

Viele e Srinivasan (2000) foram os primeiros a desenvolver um método bayesiano para estimação de parâmetros no Modelo AMMI. Liu (2001), utilizando as mesmas distribuições *a priori* propostas por Viele e Srinivasan (2000), desenvolveu um conjunto de distribuições condicionais para os parâmetros do modelo de forma que a amostragem pôde ser conduzida utilizando apenas o amostrador de Gibbs. Assim, esse método evita a utilização do algoritmo Metropolis-Hastings, reduzindo tempo computacional e tornando o algoritmo mais estável.

Crossa et al. (2011) propuseram uma abordagem bayesiana com notação vetorial para inferência sobre os parâmetros do modelo AMMI. Nesta pesquisa são implementadas regiões de máxima densidade a posteriori (HPD) para os termos da interação do modelo “AMMI-bayesiano” representadas no biplot dos dois primeiros eixos de interação. De acordo com os autores, essas regiões oferecem uma inferência estatística sobre os parâmetros bilineares, o que permite identificar grupos homogêneos de ambientes e genótipos no que se refere ao efeito da interação.

REFERÊNCIAS

- ALLARD, R. W.; BRADSHAW, A. D. Implications of genotype-environmental interactions in applied plant breeding. *Crop science*, Madison, v. 4, n. 5, p. 503-508, 1964.
- ALLARD, R. W. **Princípios do melhoramento genético das plantas**. São Paulo: E. Blucher, 1971.
- ANNICCHIARICO, P. **Genotype x environment interactions**: challenges and opportunities for plant breeding and cultivar recommendations. Rome: Food and Agriculture Organization, 2002. (Paper, 174).
- BATTISTELLI, G. M. et al. Production and identification of doubled haploids in tropical maize. **Genetics and Molecular Research**, Ribeirão Preto, v. 12, n. 4, p. 4230-4242, 2013.
- BEARZOTI, E. Emprego de marcadores em modelos mistos. In: SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 4., 2002, Lavras. **Anais...** Lavras: UFLA, 2002. p. 21-43.
- 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 the performance of crosses between untested maize inbreds. **Crop science**, Madison, v. 36, n. 4, p. 872-876, 1996b.
- BERNARDO, R. **Breeding for quantitative traits in plants**. Woodbury: Stemma, 2002. 369 p.
- BERNARDO, R. Genetic models for predicting maize single-cross performance in unbalanced yield trial data. **Crop Science**, Madison, v. 35, n. 1, p. 141-147, 1995.
- BERNARDO, R. Genomewide selection for rapid introgression of exotic germplasm in maize. **Crop Science**, Madison, v. 49, n. 2, p. 419-425, 2009.
- BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science**, Madison, v. 34, n. 1, p. 20-25, 1994.
- BERNARDO, R.; YU, J. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, Madison, v. 47, n. 3, p. 1082-1090, 2007.
- BREWBAKER, J. L.; AQUILIZAN, F. Genetics of Resistance in Maize to a Mosaic-Stripe Virus Transmitted by Peregrins maidis. **Crop Science**, Madison, v. 5, n. 5, p. 412-415, 1965.
- BRIEGER, F. G. et al. **Races of maize in Brazil and other Eastern South American countries**. Washington: National Academy of Sciences, 1958. 283 p.
- BROMLEY, C. M. et al. Estimation of genetic variance from F1-performance with and without pedigree relationships among inbred lines. **Crop Science**, Madison, v. 40, p. 651-655, 2000.

CHASE, S. S. Production of homozygous diploids of maize from monoploids. **Agronomy Journal**, Madison, v. 44, n. 5, p. 263-267, 1952.

CLARK, S. A. et al. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. **Genetics Selection Evolution**, Paris, v. 44, n. 1, p. 1, 2012.

COCKERHAM, C. C. Prediction of double crosses from single crosses. **Der Züchter**, Cham, v. 37, n. 4, p. 160-169, 1967.

CORNELIUS, P. L.; CROSSA, J.; SEYEDSADR, M. S. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In: KANG, M. S.; GAUCH, H. G. (Org.). **Genotype-by-environment interaction**. Boca Raton: CRC, 1996. p. 199-234.

CORNELIUS, P. L.; SEYEDSADR, M. S.; CROSSA, J. Using the shifted multiplicative model to search for “separability” in crop cultivar trials. **Theoretical Applied Genetics**, New York, v. 84, n. 1/2, p. 161-172, 1992.

CORNELIUS, P. L.; SEYEDSADR, M. S. Estimation of general linear-bilinear models for two-way tables. **Journal of Statistical Computation and Simulation**, New York, v. 58, n. 4, p. 287-322, May 1997.

CROSSA, J. et al. Bayesian estimation of the additive main effects and multiplicative interaction model. **Crop Science**, Madison, v. 51, n. 4, p. 1458-1469, July 2011.

CROSSA, J. Statistical analyses of multilocation trials. **Advances in Agronomy**, San Diego, v. 44, p. 55-85, 1990.

CROSSA, J.; YANG, R. C.; CORNELIUS, P. L. Studying crossover genotype x environment interaction using linear-bilinear models and mixed models. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 9, n. 3, p. 362-380, 2004.

CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa, MG: UFV, 2003. 585 p.

DAETWYLER, H. D. et al. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, Austin, v. 193, n. 2, p. 347-365, 2013.

DENIS, J. B.; GOWER, J. C. Asymptotic confidence regions for biadditive models: interpreting genotype-environment interactions. **Journal of Applied Statistics**, Abingdon, v. 45, n. 4, p. 479-493, May 1996.

DENIS, J. B.; GOWER, J. C. Asymptotic covariances for parameters of biadditive models. **Utilitas Mathematica**, Winnipeg, v. 46, p. 193-205, 1994.

DENIS, J. B.; PAZMAN, A. Bias of LS estimators in nonlinear regression models with constraints: part II, biadditive models. **Journal of Applied Statistics**, Abingdon, v. 44, n. 5, p. 375-403, Oct. 1999.

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

EATHINGTON, S. R. et al. Molecular markers in a commercial breeding program. **Crop Science**, Madison, v. 47, p. S-154-S-163, 2007. (Suppl.).

EBERHART, S. A.; RUSSELL, W. A.; PENNY, L. H. Double cross hybrid prediction in maize when epistasis is present. **Crop Science**, Madison, v. 4, n. 4, p. 363-366, 1964.

EDAE, E. A. et al. Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. **Theoretical and Applied Genetics**, Berlin, v. 127, n. 4, p. 791-807, 2014.

ELSHIRE, R. J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS one**, San Francisco, v. 6, n. 5, 2011.

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

FIGUEIREDO, A. G. et al. Application of mixed models for evaluating stability and adaptability of maize using unbalanced data. **Euphytica**, Wageningen, v. 202, n. 3, p. 393-409, 2015.

FRITSCHÉ-NETO, R. et al. Factor analysis and SREG GGE biplot for the genotype × environment interaction stratification in maize. **Ciência Rural**, Santa Maria, v. 40, n. 5, p. 1043-1048, 2010.

GABRIEL, K. R. The biplot graphic display of matrices with application to principal components analysis. **Biometrika**, London, v. 58, p. 453-467, 1971.

GALINAT, W. C. The origin of corn. In: SPRAGUE, G. F. **Corn and corn improvement**. Madison: American Society of Agronomy, 1977. p. 1-47.

GAUCH, H. G. A simple protocol for AMMI analysis of yield trials. **Crop Science**, Madison, v. 53, p. 1860-1869, 2013.

GAUCH, H. G. Model selections and validation for yield trials with interactions. **Biometrics**, Washington, v. 44, p. 705-715, 1988.

GAUCH, H. G.; PIEPHO, H. P.; ANNICCHIARICO, P. Statistical analysis of yield trials by AMMI and GGE: further considerations. **Crop Science**, Madison, v. 48, n. 3, p. 866-889, 2008.

GAUCH, H. G. **Scientific method in practice**. Cambridge University, 2003.

GAUCH, H. G.; ZOBEL, R. W. AMMI analysis of yield trials. In: KANG, M. S.; GAUCH, H. G. (Ed.). **Genotype-by-environment interaction**. Boca Raton: CRC, 1996. p. 85-122.

GIANOLA, D. et al. Enhancing genome-enabled prediction by bagging genomic BLUP. **PLoS One**, San Francisco, v. 9, n. 4, 2014.

GOLLOB, H. F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**, New York, v. 33, p. 73-115, 1968.

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

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

HENDERSON, C. R. Selection index and expected genetic advance. In: _____. **Statistical genetics and plant breeding**. Washington: National Academy of Science Research Council, 1963. p. 141-163. (Publication 982).

HESLOT, N. et al. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. **Theoretical and Applied Genetics**, Berlin, v. 127, n. 2, p. 463-480, 2014.

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

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

JENKINS, M. T. The effect of inbreeding and selection within inbred lines of maize upon the hybrids made after successive generations of selfing. **Iowa State College Journal of Science**, Ames, v. 3, p. 429-450, 1936.

JONES, D. F.; SINGLETON, W. R. **Crossed sweet corn**. New Haven: Connecticut Agricultural Experiment Station, 1934. p. 489-536. (Bull. 361).

JOSSE, J. et al. Another look at Bayesian analysis of AMMI models for genotype-environment data. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 19, n. 2, p. 240-257, 2014.

LAMKEY, K. R.; HALLAUER, A. R.; KAHLER, A. L. Allelic differences at enzyme loci and hybrid performance in maize. **Journal of Heredity**, Cary, v. 78, n. 4, p. 231-234, 1987.

LAVORANTI, O. J.; DIAS, C. T. S.; KRAZNOWSKI, W. J. Phenotypic stability via AMMI model with bootstrap re-sampling. **Pesquisa Florestal Brasileira**, Colombo, n. 54, p. 45-52, Jan./June 2007.

LINDSTROM, E. W. Analysis of modern maize breeding principles and methods. In: INTERNATIONAL CONGRESS OF GENETICS, 7., 1939, Edinburg. **Proceedings...** Edinburg: [s. n.], 1939. p. 191-196.

LIU, G. **Bayesian computations for general linear-bilinear models**. 2001. 150 p. Thesis (Doctor of Philosophy) - University of Kentucky, Lexington, 2001.

MANDEL, J. A new analysis of variance model for non-additive data. **Technometrics**, Washington, v. 13, n. 1, p. 1-18, 1971.

MANDEL, J. The partitioning of interactions in analysis of variance. **Journal of Research of the National Bureau of Standards, Series B**, Washington, v. 73, p. 309-328, 1969.

MANGELSDORF, P. C. **Corn its origin, evolution and improvement**. Cambridge: Mass, 1974.

MANGELSDORF, P. C.; MACNEISH, R. S.; GALINAT, W. C. Domestication of corn. **Science**, Washington, v. 143, n. 3612, p. 1310-1310, 1964.

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

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

MUÑOZ, P. R. et al. Unraveling additive from nonadditive effects using genomic relationship matrices. **Genetics**, Austin, v. 198, n. 4, p. 1759-1768, 2014.

OLIVEIRA, L. A. et al. Bayesian GGE biplot models applied to maize multi-environments trials. **Genetics and Molecular Research**, Ribeirão Preto, v. 15, p. 1-21, 2016.

OLIVEIRA, L. A. et al. Credible intervals for scores in the AMMI with random effects for genotype. **Crop Science**, Madison, v. 55, p. 1-12, 2015.

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

PATERNIANI, E.; GOODMAN, M. M. **Races of maize in Brazil and adjacent areas**. México: CIMMYT, 1977.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.

PEREIRA FILHO, I. A.; BORGHI, E. **Mercado de sementes de milho no Brasil: safra 2016/2017**. Sete Lagoas: Embrapa Milho e Sorgo, 2016. 28 p. (Documentos, 202).

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

RAMALHO, M. A. P. et al. **Aplicações da genética quantitativa no melhoramento de plantas autógamias**. Lavras: UFLA, 2012. 522 p.

- RAMALHO, M. A. P. et al. Melhoramento de espécies autógamas. In: NASS, L. L. **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, 2001. p. 201-230.
- RESENDE, M. D. V. **Inferência bayesiana e simulação estocástica na estimação de componentes de variância e valores genéticos em plantas perenes**. Colombo: Embrapa Florestas, 2000. 68 p. (Documentos, 46).
- ROBER, F. K.; GORDILLO, G. A.; GEIGER, H. H. In vivo haploid induction in maize-performance of new inducers and significance of doubled haploid lines in hybrid breeding. **Maydica**, Bergamo, v. 50, n. 3/4, p. 275, 2005.
- SHULL, G. H. A pure line method of corn breeding. **Report of the American Breeders Association**, Washington, v. 5, p. 51-59, 1909.
- SILVA, C. P. et al. A Bayesian shrinkage approach for AMMI models. **Plos One**, San Francisco, v. 10, n. 7, p. 1-27, 2015.
- SILVA, W. C. J.; DUARTE, J. B. Métodos estatísticos para estudo de adaptabilidade e estabilidade fenotípica em soja. **Pesquisa Agropecuária Brasileira**, Brasília, v. 41, n. 1, p. 23-30, 2006.
- SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **Journal of Agricultural Science**, Cambridge, v. 143, n. 6, p. 449-462, June 2005.
- SMITH, J. S. C. et al. Use of doubled haploids in maize breeding: implications for intellectual property protection and genetic diversity in hybrid crops. **Molecular Breeding**, Dordrecht, v. 22, n. 1, p. 51-59, 2008.
- SU, G. et al. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. **PloS one**, San Francisco, v. 7, n. 9, 2012.
- TROYER, A. F. A retrospective view of corn genetic resources. **Journal of Heredity**, Cary, v. 81, n. 1, p. 17-24, 1990.
- VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, Champaign, v. 91, n. 11, p. 4414-4423, 2008.
- VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 496 p.
- VIELE, K.; SRINIVASAN, C. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback-Leibler information. **Journal of Statistical Planning and Inference**, Amsterdam, v. 84, n. 1/2, p. 201-219, Mar. 2000.
- WEATHERWAX, P. **Indian corn in old America**. New York: The MacMillan, 1954. 253 p.

WELLHAUSEN, E. J. et al. **Races of maize in Mexico**. Cambridge: The Bussey Institute, Harvard University, 1952. 223 p.

WIJNTJES, Y. C. J.; VEERKAMP, R. F.; CALUS, M. P. L. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. **Genetics**, Austin, v. 193, n. 2, p. 621-631, 2013.

YANG, R. C. et al. Biplot analysis of genotype: environment interaction: proceed with caution. **Crop Science**, Madison, v. 49, n. 5, p. 1564-1576, Sept. 2009.

YAN, W. et al. Comment on "Biplot analysis of Genotype environment interaction: proceed with caution". **Crop Science**, Madison, v. 50, n. 4, p. 1121-1123, 2010.

YAN, W. et al. Cultivar evaluation and mega-environment investigation based on GGE biplot. **Crop Science**, Madison, v. 40, n. 3, p. 596-605, June 2000.

YAN, W. et al. GGE biplot vs. AMMI analysis of genotype-by-environment data. **Crop science**, Madison, v. 47, n. 2, p. 643-653, 2007.

YAN, W.; KANG, M. S. **GGE Biplot analysis**: a graphical tool for breeders, geneticists, and agronomists. Boca Raton: CRC, 2003. 286 p.

YAN, W.; TINKER, A. Biplot analysis of multi environment trial data: principles and applications. **Canadian Journal of Plant Science**, Ottawa, v. 86, n. 3, p. 623-645, Mar. 2006.

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

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

ZOBEL, R. W.; WRIGHT, M. J.; GAUCH, H. G. Statistical analysis of a yield trial. **Agronomy Journal**, Madison, v. 80, p. 388-393, 1988.

SEGUNDA PARTE - ARTIGO

**INCORPORAÇÃO DE PARENTESCO GENÔMICO NA PREDIÇÃO DO
DESEMPENHO DE HÍBRIDOS EM MILHO**

RESUMO

O efeito da interação entre genótipos e ambientes tem grande importância no processo de recomendação de cultivares. Para determinar o padrão desse efeito têm sido utilizados modelos lineares-bilineares, tais como AMMI e AMMI Bayesiano. Apesar de estudos mostrando a eficiência do modelo AMMI Bayesiano para determinar tal padrão, pouco se tem explorado acerca da possibilidade de acréscimo de informações adicionais ao modelo, tais como histórico de experimentos e parentesco genético. Assim, neste trabalho, teve-se como objetivo acrescentar informação de parentesco genético e comparar a capacidade preditiva do modelo AMMI-Bayesiano com e sem essa informação. Uma coleção de 113 híbridos de milho foi avaliada para o caráter produção de grãos ($t\ ha^{-1}$) em 14 ambientes representativos da região sudeste do Brasil, utilizando delineamento em blocos incompletos, e 23.153 DArT-seq foram obtidos para estimação da matriz de parentesco aditivo realizada. Os resultados evidenciaram que, gastos adicionais com genotipagem não são necessários para se utilizar o modelo AMMI Bayesiano com a informação de parentesco, já que este modelo apresentou capacidade preditiva similar à da modelagem à matriz de parentesco. Entretanto, considerando-se a seleção de cultivares, houve alteração na classificação dos híbridos (quanto aos seus efeitos sobre a produção de grãos), evidenciando a existência de mais híbridos com efeitos positivos quando se incorporou a informação de parentesco no modelo AMMI Bayesiano.

Palavras-chave: AMMI-Bayesiano. Inferência Bayesiana. Ensaio multiambientes. Parentesco genético. Milho.

ABSTRACT

The interaction effect between genotypes and environments has great importance in the cultivar recommendation process. To determine the pattern of this effect linear-bilinear models have been used, such as AMMI and Bayesian AMMI. Despite studies showing the efficiency of the Bayesian AMMI model to determine such a pattern, little has been explored about the possibility of adding extra information to the model, such as history of experiments and genetic kinship. Thus, in this work had as objective to add genetic kinship information and to compare the predictive capacity of the AMMI-Bayesian model with and without this information. A collection of 113 maize hybrids were evaluated for the grain yield character ($t\ ha^{-1}$) in 14 representative environments of the southeastern region of Brazil, using an incomplete block design, and 23.153 DArT-seq were obtained for estimation of the additive kinship matrix performed. The results showed that additional expenditures with genotyping are not necessary to use the Bayesian AMMI model with kinship information, since this model presented predictive capacity similar to that of modeling the kinship matrix. However, considering the cultivars selection, there was a change in hybrids classification (regarding their effects on grain production), evidencing the existence of more hybrids with positive effects when the kinship information was incorporated in the Bayesian AMMI model.

Keywords: AMMI-Bayesian. Bayesian Inference. Multi-environment trials. Genetic kinship. Maize.

1 INTRODUÇÃO

No melhoramento genético de plantas, o desenvolvimento e a recomendação de cultivares que apresentam produção satisfatória e estabilidade produtiva quando cultivadas em ambientes estratégicos e de importância socioeconômica estão entre os grandes desafios encontrados pelos melhoristas. Especialmente no melhoramento de milho, o cultivo de híbridos em ambientes com diferentes características climáticas, edafológicas, patogênicas, entre outras, desencadeia um fenômeno conhecido no melhoramento de plantas por interação genótipos por ambientes (GxE), o que justifica alterações expressivas no *ranking* de cultivares testadas em múltiplos ambientes (HESLOT et al., 2014). Esse efeito tem importância, principalmente, durante a avaliação de caracteres que apresentam mecanismos de herança infinitesimal ou quantitativa, ou seja, que são controlados por um grande número de genes não conhecidos e que interagem com o ambiente, apresentando pequenos efeitos acumulativos e atuantes na expressão de caracteres, alvo para a seleção, como a produção de grãos, altura de plantas, peso de espigas, dentre outros.

Vários trabalhos sobre o estudo da interação GxE estão disponíveis na literatura: propostas com o uso de análise de variância conjunta (YATES; CONCRAH, 1938), modelos baseados em regressão (FINLAY; WILKINSON, 1963), modelos lineares-bilineares (WILLIAMS, 1952), tais como modelo de efeitos principais aditivos e interação multiplicativa (AMMI), (GOLLOB, 1968), modelo de regressão por locais (SREG), (CROSSA; YANG; CORNELIUS, 2004) e com modelos mistos como os fatoriais analíticos (FA) (SMITH; CULLIS; THOMPSON, 2001).

Dentre os modelos citados, os lineares-bilineares (MLB) se popularizaram por combinarem métodos univariados (ANOVA) e multivariados (análise de componentes principais associado à decomposição por valores singulares e decomposição espectral), além de fornecerem medidas de estabilidade e adaptabilidade em uma única abordagem. Dentre os MLB, vale destacar o modelo AMMI, que proporciona uma análise para os efeitos principais e um detalhado estudo da interação aliado com análise gráfica biplot para identificação de padrões.

Apesar de reconhecidas as vantagens do modelo AMMI, tradicionalmente, ele é baseado em efeitos fixos e, por essa razão, apresenta diversas limitações, como a incapacidade de analisar com eficiência dados multiambientes (MET) em condições de desbalanceamento, presença de heterogeneidade de variância entre os ambientes ou blocos, erros correlacionados

e não incorporação da informação de parentesco genético entre os tratamentos na análise (CROSSA et al., 2011; YANG et al., 2009). Tentativas de contornar esses problemas já são reportadas na literatura com o modelo AMMI de efeito aleatório, que resolveram grande parte das deficiências do modelo AMMI de efeito fixo. Entretanto, além de não ocorrer um consenso até os dias atuais acerca de quais efeitos devem ser considerados fixos ou aleatórios (JOSSE et al., 2014; SMITH; CULLIS; THOMPSON, 2001), esse modelo ainda deve ser penalizado, por proporcionar apenas um estudo descritivo por meio do biplot da interação. Trata-se de problema prático relevante que está em desacordo com a necessidade do melhorista em realizar, com certo nível de credibilidade, inferências acerca do desempenho relativo dos híbridos quando avaliados em ensaios multiambientes (OLIVEIRA et al., 2015).

Propostas recentes para os MBL (CROSSA et al., 2011; OLIVEIRA et al., 2015, 2016; VIELE; SRINIVASAN, 2000) foram realizadas dentro da abordagem Bayesiana, de modo que todos os parâmetros do modelo são amostrados de forma direta, obedecendo às restrições impostas aos componentes da interação. Os estudos com base nesta proposta, têm se mostrado como uma boa maneira para descrever tanto os parâmetros lineares quanto os parâmetros da interação, sendo para estes uma forma prática de estabelecer regiões de credibilidades e incorporar incerteza ao biplot.

Nos últimos anos, a disponibilidade de marcadores moleculares baseados em protocolos de redução de complexidade genômica permite facilmente incorporar matrizes de parentesco em modelos Bayesianos (VANRADEN, 2008). Entretanto, não se tem encontrado na literatura trabalhos envolvendo matrizes de parentesco genético nos modelos MBL Bayesianos, apesar de já mencionada a facilidade em incorporá-la por meio de *prioris* (CROSSA et al. 2011; OLIVEIRA et al., 2015).

O fornecimento de modelos que incorporem ou com capacidade de prever a interação GxE estão entre os desafios que podem incrementar a contribuição da seleção genômica no contexto da interação GxE (JONAS; KONING, 2013). Trata-se de um tema que reflete uma das principais preocupações dos melhoristas, uma vez que, rotineiramente, os híbridos que são utilizados em ensaios multiambientes são avaliados desde o seu primeiro ano de teste até o seu lançamento, e a informação do desempenho dos híbridos é um dos critérios utilizados na geração de novas populações de melhoramento. Essa dinâmica gera o desenvolvimento de genótipos com amplo *background*, o que pode dificultar ainda mais a realização de predições acuradas acerca do desempenho dos híbridos.

Nesse contexto, a incorporação da matriz de parentesco é uma proposta para melhorar a qualidade de ajuste do modelo AMMI Bayesiano, principalmente para a análise de caracteres quantitativos. Assim, os principais objetivos com este trabalho foram: (i) avaliar a inclusão da informação de parentesco aditivo realizada no modelo AMMI Bayesiano por meio da capacidade preditiva na ausência e presença dessa informação; (ii) investigar a estabilidade de germoplasmas tropicais de híbridos de milho avaliados para o caráter produção de grãos em vários ambientes da região sudeste do Brasil.

2 MATERIAL E MÉTODOS

2.1 Dados Fenotípicos

Uma coleção de 113 híbridos simples foi avaliada para o caráter produção de grãos ($t\ ha^{-1}$) durante o ano agrícola de 2013 e 2014 em 14 locais representativos da região Sudeste e Sul do Brasil (Tabela 1). O *design* experimental utilizado foi um delineamento em blocos incompletos com número variável de repetições. Alguns híbridos do programa de melhoramento da Universidade Federal de Lavras (UFLA) foram avaliados com número superior de repetições, pois já estavam em ensaios de avaliação avançados para produção de grãos.

Tabela 2 Conjunto de municípios, latitude, longitude, altitude, em que foram avaliados os híbridos do programa de melhoramento da Universidade Federal de Lavras, e produção média de grãos ($t\ ha^{-1}$) e coeficiente de variação (CV) dos ensaios multiambientais.

Ambientes	Municípios	Latitude	Longitude	Altitude (m)	Produção* ($t\ ha^{-1}$)	CV* (%)
E1	Guarapuava - PR	25°23'S	51°27'W	1.120	10,163	16,451
E2	Vacaria - RS	28°30'S	50°56'O	971	5,289	20,867
E3	Ipiranga - PR	25°01'S	50°35'O	800	6,374	19,728
E4	Sananduva - RS	27°57'S	51°48'W	636	8,071	15,29
E5	Faxinal - PR	24°0'S	51°19'W	820	10,416	15,749
E6	Presidente Olegário - MG	18°25'S	46°25'W	947	8,985	13,507
E7	Uberaba - MG	19°44'S	47°55'W	823	9,509	13,895
E8	Capinópolis - MG	18°40'S	49°34'W	530	7,652	23,451
E9	Araguari - MG	18°38'S	48°11'W	1.013	8,116	14,496
E10	Madre de Deus - MG	21°28'S	44°19'O	990	9,485	17,875
E11	Nazareno - MG	21°12'S	44°36'W	930	7,627	17,638
E12	Boa Esperança - MG	21°05'S	45°33'O	775	6,349	24,871
E13	Lavras - MG	21°14'S	45°0'W	919	7,253	13,41
E14	Itapeva - SP	23°58'S	48°52'W	684	7,892	16,992

A produtividade média de grãos e o coeficiente de variação (CV) foram calculados via análise individual dentro de cada local, com um modelo misto atribuindo efeito de genótipos como aleatório e de repetições como fixo.

2.2 Dados Moleculares

Um processo de extração de DNA de 182 linhagens de milho foi realizado, seguindo o protocolo de extração da empresa *Diversity Arrays Technology*. As linhagens genotipadas correspondem aos parentais dos híbridos de milho fenotipados nos ensaios multiambientais. As amostras de DNA foram submetidas ao protocolo de genotipagem DArT/GBS (ELSHIRE et al., 2011; JACCOUD et al., 2001), até a obtenção de uma biblioteca final de leituras de DNA com complexidade genômica reduzida. Subsequentemente, as bibliotecas foram sequenciadas com a plataforma IlluminaHiSeq 2000. Os processos de preparo e alinhamento das leituras, chamada de SNPs, e filtragem de SNPs de baixa qualidade foram terceirizados e realizados pela empresa *Diversity Arrays Technology*. Ao final desses processos foram obtidos 23.153 SNPs de alta qualidade.

2.3 Modelo AMMI Bayesiano

O modelo AMMI adotado para o presente estudo é dado por:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}.$$

sendo \mathbf{y} o vetor de observações; $\boldsymbol{\beta}$ o vetor de efeitos de bloco + ambiente (confundidos); \mathbf{g} o vetor de efeitos principais de genótipos; $\boldsymbol{\alpha}_k$ e $\boldsymbol{\gamma}_k$ os vetores singulares (relacionados aos genótipos e ambientes); λ_k o valor singular relacionado aos k -ésimos vetores, com $k = 1, \dots, t$, em que $t = \min(r, c)$ denota o posto da matriz $\mathbf{G} \times \mathbf{E}$; e \mathbf{X}_1 , \mathbf{X}_2 e \mathbf{Z} as matrizes de delineamento. O vetor contendo os erros aleatórios é denotado por $\boldsymbol{\varepsilon}_{n \times 1}$, com $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$, sendo $\mathbf{0}$ o vetor de zeros, σ_e^2 a variância do resíduo e \mathbf{I}_n a matriz identidade.

A distribuição condicional dos dados tem densidade normal multivariada $\mathbf{y} | \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{g}, \boldsymbol{\beta}, \sigma_e^2 \sim N(\boldsymbol{\theta}, \mathbf{I}_n \sigma_e^2)$ em que o vetor de médias é dado por:

$$\boldsymbol{\theta} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_1\boldsymbol{\gamma}_k.$$

2.4 Prioris Para os Parâmetros do Modelo

Para os efeitos principais de β e g foram atribuídas prioris normais multivariadas, sendo a priori de β proporcional a uma constante e $g \sim N(0, \mathbf{G}\sigma_g^2)$ em que \mathbf{G} é a matriz de parentesco genético. A matriz \mathbf{G} em trabalhos como Crossa et al. (2011) e Oliveira et al. (2015), foi assumida como uma identidade, que seria o mesmo que assumir que não existe correlação entre os genótipos. Para σ_g^2 atribuiu-se uma priori não informativa de modo que a incerteza em relação à estimação de g fica determinada basicamente pela função de verossimilhança. Para σ_e^2 também foi atribuída uma priori não informativa.

Sendo os parâmetros da GxE sujeitos às restrições provenientes da decomposição por valores singulares, tem-se que os valores singulares são sempre positivos e estão em ordem decrescente em relação à magnitude ($\lambda_k \geq \lambda_{k+1}$) e, com isso, optou-se por uma priori normal truncada $\lambda_k \sim N^+(0, \sigma_{\lambda_k}^2)$. Para os vetores singulares α_k e γ_k foram atribuídas distribuições uniformes esféricas no subespaço corrigido, dado que os vetores devem possuir norma unitária, caso particular da distribuição Von Mises-Fisher quando o parâmetro de concentração tem valor igual a zero. A amostragem é realizada no espaço corrigido onde o vetor não possui restrições impostas pelo modelo (CROSSA et al., 2011; OLIVEIRA et al., 2015; SILVA et al., 2015; VIELE; SRINIVASAN, 2000).

2.5 Distribuição Conjunta a Posteriori para os Parâmetros do Modelo

Dadas a função de verossimilhança e as prioris para os parâmetros do modelo e aplicando o teorema de Bayes, estabelece-se a distribuição conjunta a posteriori para os parâmetros do modelo como:

$$p(y | \dots) \propto p(y | \theta, \mathbf{I}_n \sigma_e^2) p(\beta | \mu_\beta, \mathbf{I} \sigma_\beta^2) p(g | \mu_g, \mathbf{G} \sigma_g^2) \prod_{k=1}^t \left(p(\lambda_k | \mu_{\lambda_k}, \sigma_{\lambda_k}^2) p(\alpha_k) p(\gamma_k) \right) \times \\ \times p(\sigma_g^2 | \nu_g, S_g^2) p(\sigma_e^2 | \nu_e, S_e^2)$$

Como não é possível realizar amostragem dos parâmetros diretamente da distribuição conjunta, obtém-se as distribuições condicionais a posteriori para cada parâmetro para realizar o processo de amostragem MCMC.

2.6 Distribuições Condicionais Completas a Posteriori

Distribuição condicional $\boldsymbol{\beta} \mid \dots$, (distribuição para $\boldsymbol{\beta}$ dados os demais parâmetros)

$$\boldsymbol{\beta} \mid \dots \sim N \left[\left(\mathbf{X}_1' \mathbf{X}_1 \right)^{-1} \mathbf{X}_1' \mathbf{A}, \left(\mathbf{X}_1' \mathbf{X}_1 \right)^{-1} \sigma_e^2 \right]$$

em que $\mathbf{A} = \mathbf{y} - \mathbf{Z}\boldsymbol{\delta} - \boldsymbol{\Theta} \mathbf{e} \boldsymbol{\Theta} = \sum_{k=1}^t \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}\boldsymbol{\gamma}_k$.

Distribuição condicional $\sigma_e^2 \mid \dots$

$$\sigma_e^2 \mid \dots \sim \text{Inv-escalada} - \chi^2 \left(n, \left[(\mathbf{y} - \boldsymbol{\Theta})' (\mathbf{y} - \boldsymbol{\Theta}) \right] / n \right)$$

Distribuição condicional $\sigma_g^2 \mid \dots$

$$\sigma_g^2 \mid \dots \sim \text{Inv-escalada} - \chi^2 \left(n_g, (\mathbf{g}' \mathbf{G}^{-1} \mathbf{g}) / n_g \right)$$

Distribuição condicional $\lambda_k \mid \dots$

$$\lambda_k \mid \dots \sim N^+ \left[\left(\boldsymbol{\phi}' \boldsymbol{\phi} + \frac{\sigma_e^2}{\sigma_{\lambda_k}^2} \right)^{-1} \boldsymbol{\phi}' \mathbf{C}, \left(\boldsymbol{\phi}' \boldsymbol{\phi} + \frac{\sigma_e^2}{\sigma_{\lambda_k}^2} \right)^{-1} \sigma_e^2 \right]$$

em que $\boldsymbol{\phi} = \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}\boldsymbol{\gamma}_k$, $\mathbf{C} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \mathbf{D}$ e $\mathbf{D} = \sum_{k' \neq k}^t \lambda_{k'} \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_{k'}) \mathbf{X}\boldsymbol{\gamma}_{k'}$ com

$$\lambda_1 \geq \lambda_2, \dots, \lambda_t \geq 0$$

Distribuição condicional $\boldsymbol{\alpha}_k \mid \dots$

$$\boldsymbol{\alpha}_k \mid \dots \sim \text{VMF} \left\{ \frac{\lambda_k}{2\sigma_e^2}, (\text{diag}(\mathbf{X}\boldsymbol{\gamma}_k)' \mathbf{Z}') (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \mathbf{D}) \right\}$$

em que $\mathbf{W} = \lambda_k \text{diag}(\mathbf{Z}\boldsymbol{\alpha}_k) \mathbf{X}\boldsymbol{\gamma}_k$ e $\mathbf{E} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \mathbf{D}$.

Distribuição condicional $\boldsymbol{\gamma}_k \mid \dots$

$$\boldsymbol{\gamma}_k \mid \dots \sim \text{VMF} \left\{ \frac{\lambda_k}{2\sigma_g^2}, [\text{diag}(\mathbf{Z}(\boldsymbol{\alpha}_k)' \mathbf{X}')] (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \mathbf{D}) \right\}$$

A diferença entre o modelo AMMI Bayesiano com e sem parentesco está vinculada à priori dos valores genéticos \mathbf{g} . O modelo com a incorporação da informação de parentesco apresenta como principal característica a inclusão de informação genética na priori dos efeitos genéticos (\mathbf{g}) da seguinte forma:

$$\mathbf{g} \mid \mathbf{G}, \sigma_g^2 \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$$

em que \mathbf{G} corresponde à matriz de parentesco aditivo (VANRADEN, 2008), dada por:

$$\mathbf{G} = \frac{\mathbf{W}'\mathbf{W}}{2 \sum_{p=1}^q p_l (1-p_l)}$$

de tal forma que:

$$\mathbf{W} = \begin{cases} 2-2p_l & \text{para } AA \\ 1-2p_l & \text{para } Aa \\ 0-2p_l & \text{para } aa \end{cases}$$

no qual p_l é a frequência alélica do marcador l no loco l .

Conforme se pode observar, o parâmetro \mathbf{g} comporta-se como a contribuição acumulativa dos valores genéticos de cada marcador, para cada híbrido, e o produto \mathbf{Zg} como valor genético genômico esperado dos híbridos de milho em análise (KAO; ZENG, 2002). Desta forma, tem-se que a distribuição condicional a posteriori para \mathbf{g} é:

$$\mathbf{g}|\dots \sim N \left[\left(\mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1} \frac{\sigma_e^2}{\sigma_g^2} \right)^{-1} \mathbf{B}, \left(\mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1} \frac{\sigma_e^2}{\sigma_g^2} \right)^{-1} \sigma_e^2 \right]$$

em que $\mathbf{B} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Theta}$. Essa versão da distribuição a posteriori de \mathbf{g} é equivalente à posteriori da versão Bayesiana do modelo GBLUP para AMMI.

Como já apresentado em Crossa et al. (2011) e Oliveira et al. (2015), as distribuições condicionais a posteriori dos parâmetros do modelo são conhecidas e, assim, a amostragem pode ser realizada de forma direta por meio do amostrado de Gibbs. Realizado o processo de amostragem Monte Carlo via Cadeia de Markov (MCMC), foi realizada a avaliação da convergência das cadeias por meio dos métodos de Heidel-Berger e Welch (1983) e Raftery e Lewis (1995).

Com isso, pode-se realizar o estudo sobre os parâmetros lineares e bilineares do modelo e a representação biplot AMMI2 para regiões de credibilidade dos escores genotípicos e ambientais (essas regiões serão construídas conforme apresentado em Oliveira et al. (2015). Também, pode-se realizar o estudo da capacidade preditiva do modelo AMMI, como apresentado em Oliveira et al. (2015), e do modelo aqui posposto com a inclusão da informação de parentesco genético, a fim de compará-las.

2.7 Validação Cruzada

A capacidade preditiva dos modelos AMMI Bayesiano com e sem parentesco, foi avaliada utilizando um procedimento de validação cruzada *10-fold*. O processo de validação cruzada foi realizado com os dados fenotípicos de 86 híbridos amostrados do conjunto total de dados fenotípicos, sendo 86 o número mínimo de híbridos avaliados com, pelo menos, uma repetição em todos os locais. Foi realizado um processo de desbalanceamento sem reposição que resultou em dez subconjuntos de dados para ajuste dos modelos Bayesianos com e sem parentesco, a fim de avaliar a capacidade preditiva de ambos.

A correlação de Pearson foi calculada entre os valores fenotípicos preditos e observados (subconjunto fora da análise). Essa correlação foi a estatística utilizada para avaliar a capacidade preditiva dos modelos. Adicionalmente, a soma de quadrados dos erros residuais preditos (PRESS) foi calculada para verificar o ajuste das predições:

$$PRESS = \sum_{i=1}^{n_p} (y_i - y_{p_i})^2$$

em que, y_i são os valores observados do subconjunto fora do ajuste do modelo e n_p corresponde ao número de observações preditas.

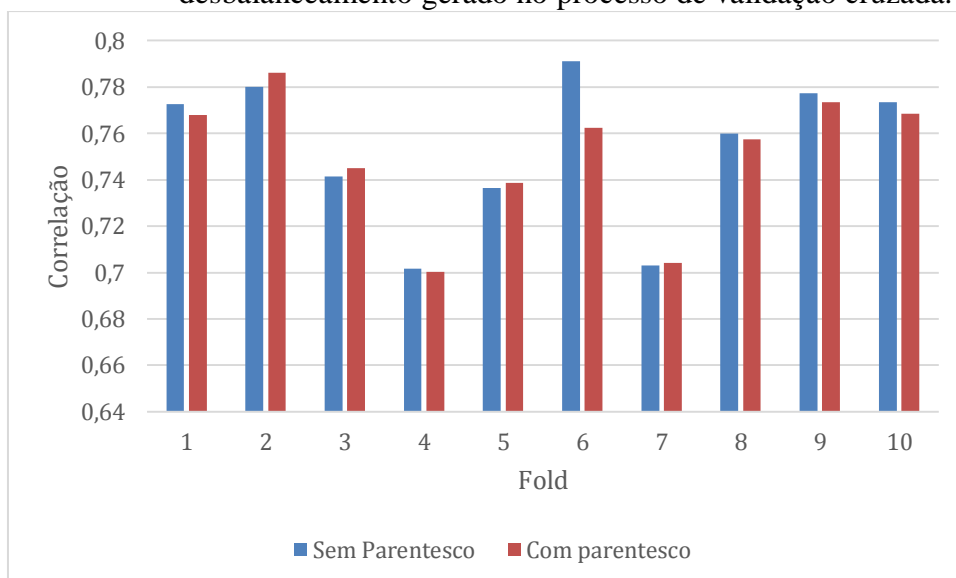
3 RESULTADOS

Foram simuladas por Monte Carlo via Cadeias de Markov (Markov Chain Monte Carlo, MCMC), 188 mil iterações para os parâmetros dos modelos com e sem informação de parentesco genético para todos os cenários deste trabalho, tanto para o estudo de avaliação preditiva como para o estudo da interação GxE. De modo a evitar problema na convergência durante a amostragem dos parâmetros, as primeiras oito mil observações foram excluídas da amostragem. Adicionalmente, com o objetivo de manter as observações como independentes, saltos a cada 20 observações foram realizados. Desta forma, de acordo com o teste de Heidelberg e Welch (1983) e Raftery e Lewis (1995), todas as cadeias apresentaram boas propriedades de convergência.

3.1 Análise da Capacidade Preditiva dos Modelos

As capacidades preditivas dos modelos AMMIB-SP e AMMIB-P, obtidas pelo processo de validação cruzada, podem ser observadas nas Figuras 1-3.

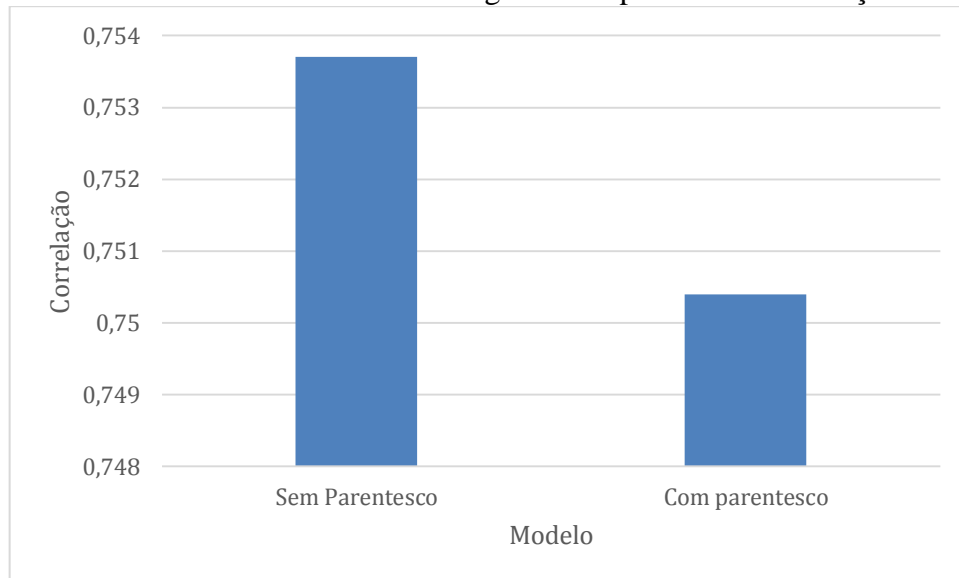
Figura 1 Comparação das correlações dos modelos AMMI Bayesiano sem parentesco e com parentesco, obtidas pelas previsões dos híbridos eliminados durante o desbalanceamento gerado no processo de validação cruzada.



Na Figura 1 é apresentada a correlação entre o valor observado e valor predito, nos dez cenários, para os modelos AMMIB-P e AMMIB-SP. Os resultados sugerem que não houve

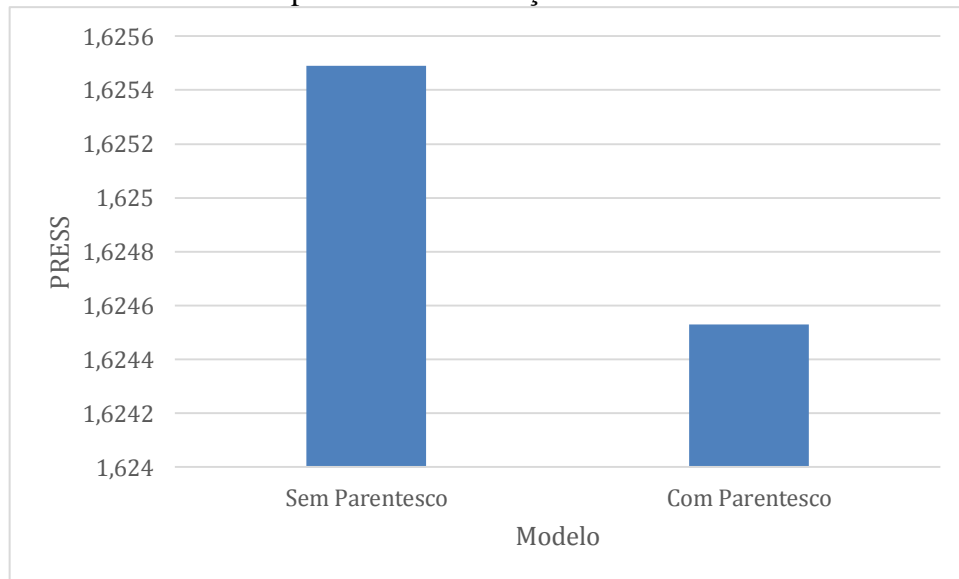
ganhos em relação à predição com o ajuste do modelo AMMI Bayesiano com parentesco, quando comparado ao modelo AMMI Bayesiano sem parentesco, pois em grande parte dos cenários ou ambos podem ser considerados iguais ou o AMMIB-SP torna-se melhor.

Figura 2 Comparação das correlações médias entre os modelos AMMI Bayesiano sem parentesco e com parentesco, obtidas pelas predições dos híbridos eliminados durante o desbalanceamento gerado no processo de validação cruzada.



Em termos médios (Figura 2), pode-se verificar, quanto à capacidade preditiva, que não se tem um melhor modelo, pois o ganho do AMMIB-SP foi marginal ($\bar{r} = 0.7536$) quando comparado aos resultados obtidos com o modelo AMMIB-P ($\bar{r} = 0.7504$), uma vez que essa diferença entre os modelos é apenas na terceira casa decimal.

Figura 3 Somas de quadrados dos erros residuais preditos (PRESS) dos modelos AMMI Bayesiano sem parentesco e com parentesco, obtidas pelas predições dos híbridos eliminados durante o desbalanceamento gerado no processo de validação cruzada.

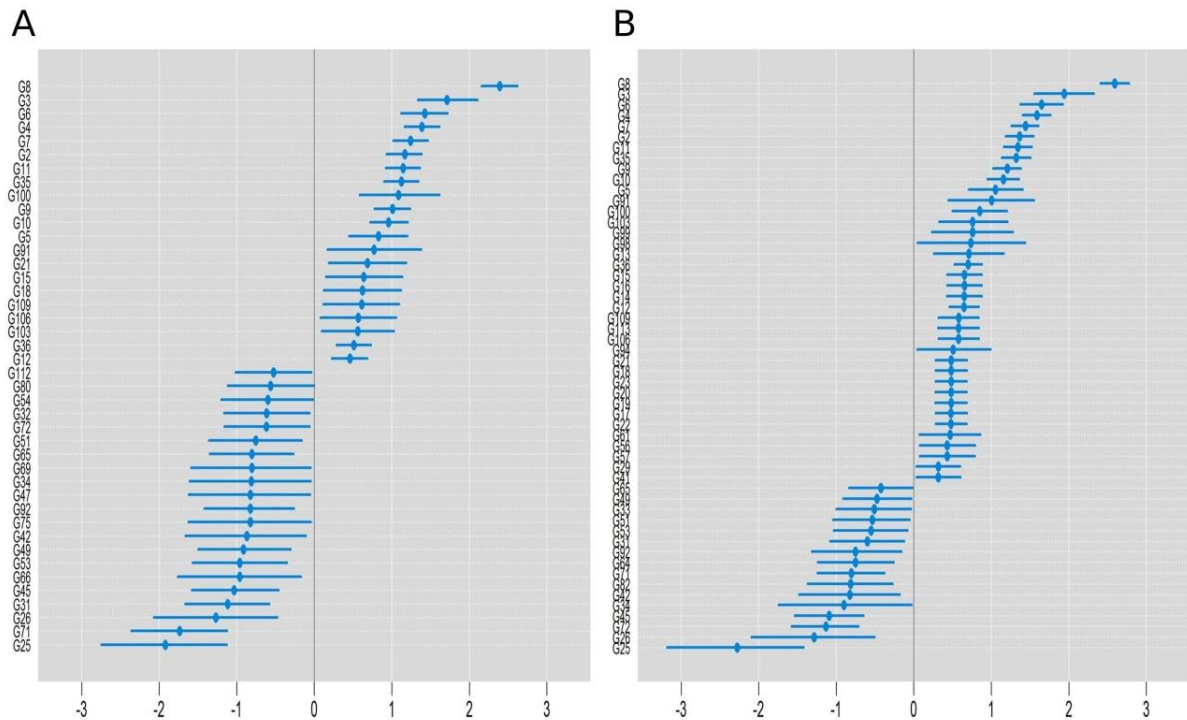


Os resultados de soma de quadrados dos erros residuais preditos PRESS (Figura 3) também sugerem que, com uma diferença apenas na terceira casa decimal, os dois modelos podem ser considerados equivalentes na eficiência das predições, porém o modelo com parentesco foi o mais eficiente.

3.2 Estimativas Pontuais, Regiões de Credibilidade

As médias e regiões de credibilidade com densidade máxima a posteriori (HPD) dos valores genéticos obtidos com os modelos AMMI Bayesiano com e sem parentesco podem ser observadas na Figura 4. Os valores genéticos estão em ordem crescente de acordo com a magnitude dos efeitos (esquerda para direita).

Figura 4 Intervalos de credibilidade (95% credibilidade) dos valores genéticos dos híbridos avaliados no ensaio multi-ambientes. A: Modelo AMMI Bayesiano sem parentesco. B: Modelo AMMI Bayesiano com parentesco.



Conforme se pode observar na Figura 4, os valores genéticos obtidos pelos modelos AMMI Bayesiano com e sem parentesco foram semelhantes. A principal diferença entre esses resultados foi a obtenção de regiões de credibilidade mais largas a partir do modelo sem parentesco (Figura 4A), quando comparado ao modelo que contempla informação de parentesco (Figura 4B). O modelo AMMI Bayesiano com parentesco (AMMIB-P) permitiu a identificação de uma maior fração de genótipos significativamente diferentes de zero (regiões que sobrepuseram zero não foram plotadas) do que os observados com o modelo AMMI Bayesiano sem parentesco (AMMIB-SP).

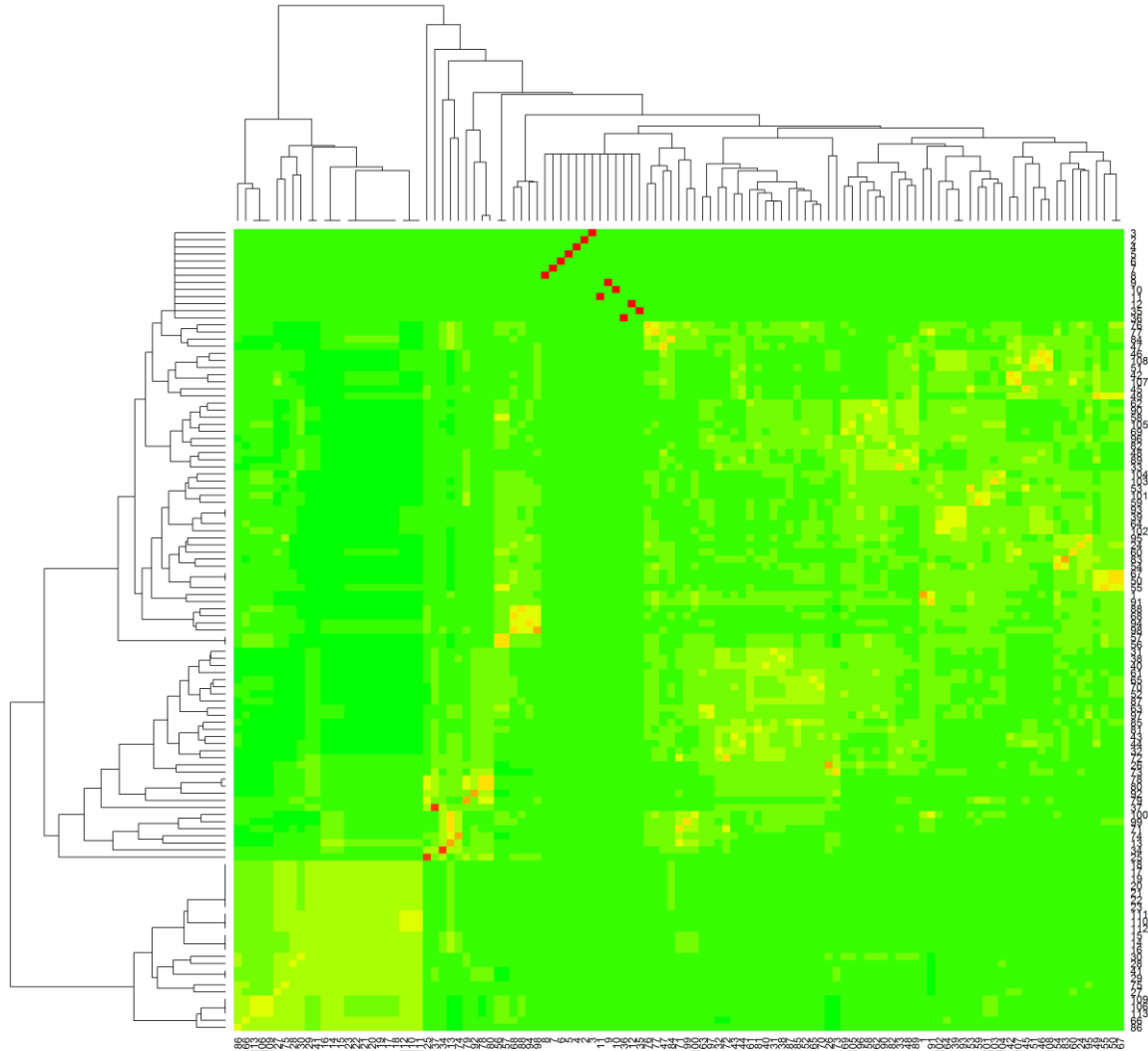
O híbrido de milho G8 apresentou o maior mérito genético quando comparado aos demais híbridos, independente da consideração da informação de parentesco no modelo (Figura 4) - não houve sobreposição com os demais intervalos de credibilidade. O híbrido G25 evidenciou a maior probabilidade de ser o pior híbrido avaliado para produção de grãos entre os híbridos testados, apesar de seu valor genético ainda sobrepor aos intervalos de credibilidade dos valores genéticos dos híbridos G71 ao G80 (sequência na região inferior à esquerda da Figura 4A) no caso do modelo AMMIB-SP, e os híbridos G26, G72, G45, G34 e G42 no caso do modelo AMMIB-P (região inferior à esquerda da Figura 4B). Os valores

genéticos obtidos pelo modelo AMMIB-SP apresentaram dois grupos com valores positivos intermediários no gráfico de mérito genético (Figura 4A). Em ordem crescente, da esquerda para direita, o primeiro grupo com os híbridos G12 ao G10, e o segundo grupo com os híbridos G9 ao G3. Já os valores genéticos obtidos com o modelo AMMIB-P evidenciou um grupo intermediário de valores negativos (G82 ao G65), e dois grupos com valores positivos, o primeiro grupo positivo que apresenta os híbridos G41 ao G91, e o segundo grupo positivo constituído pelos híbridos G5 ao G6 (Figura 4B).

De acordo com os resultados obtidos, o modelo AMMI Bayesiano sem parentesco aparenta ser uma escolha conveniente para a análise de dados obtidos em ensaios multiambientais em programas de melhoramento de milho. Apesar do modelo AMMIB-P permitir a obtenção de intervalos de credibilidade mais estreitos, e também apresentar maior poder durante a identificação de híbridos com valores genéticos significativos, os resultados sugerem que gastos financeiros adicionais com genotipagem em programas de melhoramento não justificam as vantagens da inclusão da informação de parentesco (estimado via marcadores) para aumentar a capacidade de predição de híbridos não testados, pois não houveram ganhos substanciais de correlação entre os valores preditos e observados com a utilização do modelo AMMIB-P. Entretanto, apesar de não ocorrer alteração clara de ranqueamento entre os genótipos, pode-se observar maior proporção de valores genéticos superiores a zero com a utilização da informação de parentesco.

Observa-se que o germoplasma utilizado nesta pesquisa não permitiu a obtenção de clusters bem definidos entre os híbridos avaliados (Figura 5). A formação de pequenos blocos dispersos no heatmap mostra que o agrupamento de híbridos foi dificultado em função do pequeno parentesco entre os mesmos.

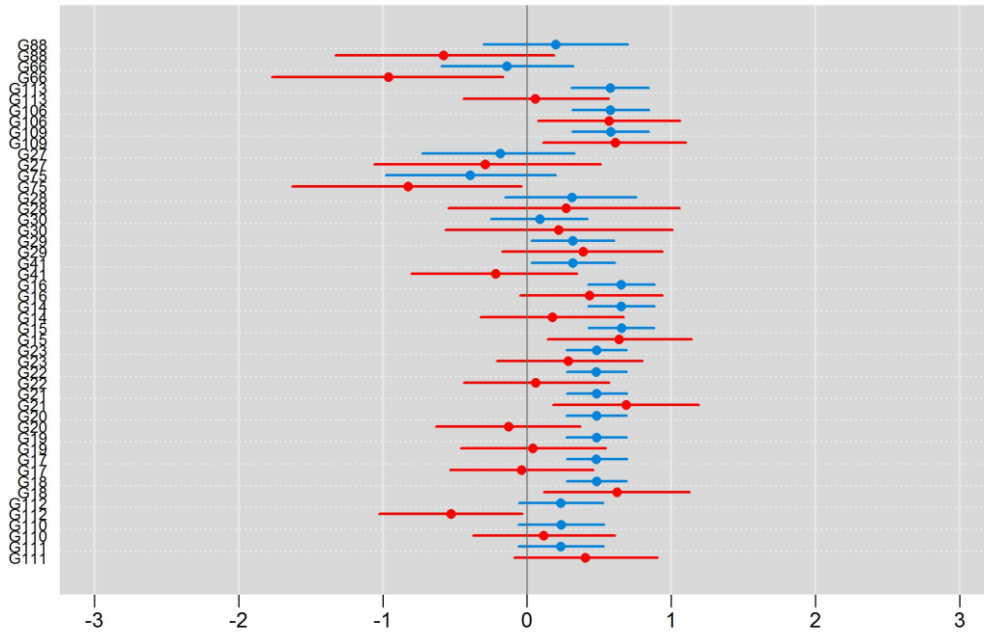
Figura 5 Heatmap do parentesco genômico dos 86 híbridos utilizados para verificação da capacidade preditiva do modelo AMMI Bayesiano com adição da informação de parentesco aditivo. As regiões em verde implicam baixo parentesco e impossibilidade de formação de agrupamento dos híbridos muito aparentados; em amarelo, onde é possível observar pequenos agrupamentos, ocorre um nível intermediário de parentesco.



Um agrupamento com uma definição mais nítida formado por 16 híbridos foi observado na base inferior esquerda da Figura 5, sugerindo que, dentre todos os híbridos do conjunto avaliado, esse grupo apresentou maior parentesco genômico entre os híbridos a eles pertencentes. Assim, os híbridos desse subgrupo foram utilizados também para a verificação dos intervalos de credibilidade nos dois cenários estudados. A partir disso, foi possível observar um intervalo mais estreito para todos os híbridos quando o parentesco foi incorporado (Figura 6). Por exemplo, para o híbrido 113, a maior precisão na avaliação do mesmo pode ser definida pelo menor intervalo de credibilidade na predição desse híbrido, além do fato de que não

negligenciar o parentesco entre os híbridos proporcionou valor genético diferente de zero, o que não pode ser observado para o mesmo híbrido na ausência de parentesco.

Figura 6 Intervalos de credibilidade (95% credibilidade) dos valores genéticos dos híbridos avaliados no ensaio multiambientes identificados com maior parentesco genético aditivo no Heatmap.



Essa mesma dinâmica pode ser observada para outros híbridos. Assim, apesar de não ocorrer um acréscimo substancial na correlação entre valores preditos e observados, esse resultado é mais uma evidência acerca da possibilidade de melhoria na seleção por parte do melhorista com a incorporação da informação de parentesco na predição de híbridos, pois mesmo que não ocorra uma mudança expressiva no *ranking* dos híbridos, os mesmos podem ser selecionados com maior confiabilidade.

A contribuição de cada valor singular obtido à posteriori a partir do ajuste dos modelos AMMIB-P e AMMIB-SP permitem a identificação do número ideal de efeitos multiplicativos que descrevem o componente de interação da GxE. Seis autovalores explicam aproximadamente 84,82% e 84,65% da variância total observada para os modelos AMMI-P e AMMI-SP, respectivamente (Tabela 2).

Tabela 2 Estatísticas descritivas da distribuição à posteriori dos autovalores obtidos a partir do ajuste do modelo AMMIB-P e AMMIB-SP, em que SD corresponde ao desvio padrão, LI ao limite inferior e LS ao limite superior.

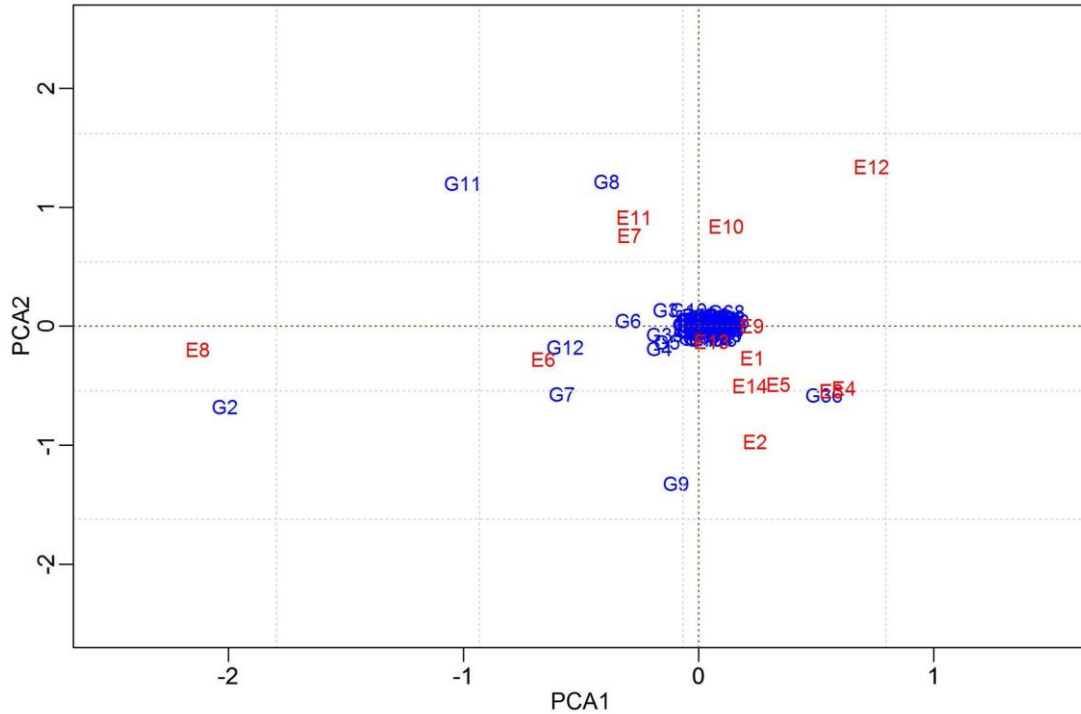
Valores singulares	Com Parentesco				Sem Parentesco			
	Média	SD	LI	LS	Média	SD	LI	LS
λ_1	6,7165	0,4372	5,8228	7,5425	6,7307	0,431	5,8499	7,5366
λ_2	6,1053	0,3625	5,4108	6,836	6,1169	0,3599	5,4412	6,8272
λ_3	4,7219	0,5317	3,6326	5,6879	4,7389	0,5426	3,6896	5,7959
λ_4	4,0993	0,4831	3,2328	5,1085	4,1038	0,4832	3,1736	5,0576
λ_5	3,5583	0,5042	2,5593	4,5387	3,5716	0,5048	2,5861	4,5609
λ_6	3,0475	0,5776	1,9919	4,2113	3,0655	0,5708	1,9483	4,118
λ_7	2,392	0,795	0,5572	3,7833	2,4338	0,7756	0,5858	3,7411
λ_8	1,3863	0,8336	0	2,7569	1,3959	0,8441	0	2,7897
λ_9	0,6864	0,6096	0	1,9055	0,6884	0,6063	0	1,8988
λ_{10}	0,3275	0,3807	0	1,1594	0,3333	0,3821	0	1,1446
λ_{11}	0,1529	0,2215	0	0,6212	0,1614	0,2262	0	0,6314
λ_{12}	0,0735	0,1274	0	0,331	0,0801	0,1369	0	0,3456
λ_{13}	0,0374	0,0781	0	0,1911	0,042	0,0871	0	0,1954

Devido aos resultados mencionados acerca dos ajustes dos modelos AMMIB-SP e AMMIB-P, a partir desse momento será reportado com maiores detalhes apenas os resultados obtidos com o modelo AMMIB-SP

3.3 Estimativas Pontuais e Regiões de Credibilidade para os Scores Genotípicos e Ambientais

Após o ajuste do modelo AMMB-SP, os dois primeiros componentes dos valores a posteriori dos parâmetros α_k e γ_k foram transformados em escores genotípicos $(\lambda_1^{1/2}\alpha_{i1}, \lambda_2^{1/2}\alpha_{i2})$ e escores ambientais $(\lambda_1^{1/2}\gamma_{j1}, \lambda_2^{1/2}\gamma_{j2})$. O gráfico biplot construído a partir dos escores genotípicos e ambientais pode ser observado na Figura 7.

Figura 7 Biplot com os valores médio a posteriori dos escores genotípicos $(\lambda_1^{1/2}\alpha_{i1}, \lambda_2^{1/2}\alpha_{i2})$ e ambientais $(\lambda_1^{1/2}\gamma_{j1}, \lambda_2^{1/2}\gamma_{j2})$, considerando o primeiro componente principal (PCA1) e o segundo componente principal (PCA2).



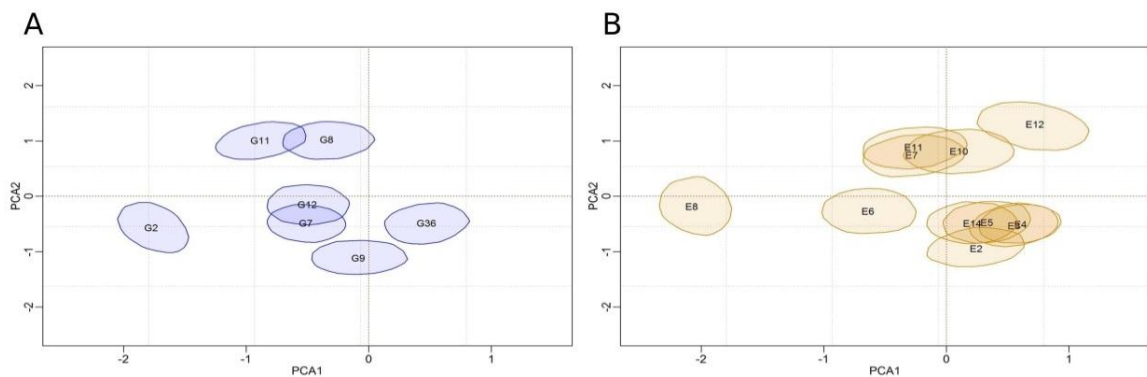
De acordo com a Figura 7, os híbridos mais próximos da origem (0,0) equivalem aos híbridos mais estáveis (pequeno efeito de GxE). Grande parte dos híbridos está agrupada densamente na região do biplot próxima da coordenada (0,0), o que sugere que no geral os híbridos apresentam alta estabilidade para o caráter produção de grãos.

Pode-se exemplificar o processo de recomendação regionalizada com o caso do híbrido G2, que aparenta ser o híbrido com a maior influência do efeito da interação GxE devido à maior distância da coordenada (0,0), apesar de ser o híbrido ranqueado com o sexto maior mérito genético (Figura 4). Esse resultado sugere que esse híbrido deve ser recomendado para ser cultivado no ambiente E8, que corresponde ao ambiente mais próximo ao híbrido G2 no biplot (Figura 7). Por outro lado, o híbrido G8 com o maior mérito genético entre os híbridos testados deve ser recomendado especialmente aos ambientes E11 e E7.

3.4 Inferência de Biplot com Intervalos Bivariados de Credibilidade Para os Scores Genotípicos e Ambientais

Apesar de biplots com informações pontuais de interação GxE serem informativos durante a avaliação do impacto da interação GxE na recomendação de híbridos de milho, estimativas bivariadas de credibilidade permitem inferências que contabilizam a incerteza vinculada ao fenômeno de interação GxE. A Figura 8 mostra os intervalos bivariados de 95% de credibilidade dos escores genotípicos ($\lambda_1^{1/2}\alpha_{i1}, \lambda_2^{1/2}\alpha_{i2}$) e ambientais ($\lambda_1^{1/2}\gamma_{j1}, \lambda_2^{1/2}\gamma_{j2}$) que diferiram de zero, considerando os dois primeiros componentes da interação GxE.

Figura 8 A: Biplot com as regiões bivariados de 95% de credibilidade dos escores genotípicos ($\lambda_1^{1/2}\alpha_{i1}, \lambda_2^{1/2}\alpha_{i2}$) que diferiram de zero. B: Biplot com os intervalos bivariados de 95% de credibilidade dos escores ambientais ($\lambda_1^{1/2}\gamma_{j1}, \lambda_2^{1/2}\gamma_{j2}$) que diferiram de zero. Os dois biplots foram construídos considerando o primeiro componente principal (PCA1) e o segundo componente principal (PCA2).

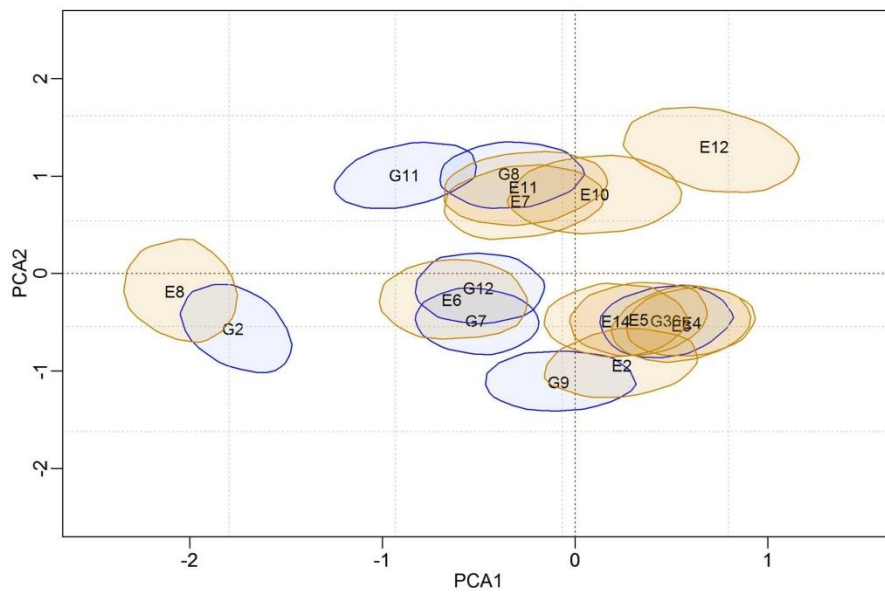


Conforme observado na Figura 8A, os híbridos G8 (1° maior mérito genético) e o híbrido G11 (7° maior mérito genético) apresentam sobreposição das regiões bivariadas a 95% de credibilidade, o que sugere que o componente de GxE entre esses dois híbridos não difere entre si. O mesmo fato é observado entre os genótipos G12 (21° maior mérito genético) e G7 (5° maior mérito genético), que também apresentam sobreposição das elipses. Os demais híbridos plotados não evidenciam sobreposição das regiões bivariadas de credibilidade e, portanto, apresentam efeitos de GxE que diferem entre si ao nível de 95% de credibilidade. Já em relação ao agrupamento dos escores ambientais (Figura 8B), pode-se observar

sobreposição dos ambientes E14, E5, E3, E4 e E2, bem como dos ambientes que constituem o E11, E7, E10 e E12. Os ambientes E8 e E6 não sobrepõem a nenhum outro ambiente.

As regiões bivariadas a 95% credibilidade dos escores genotípicos e ambientais unificados em apenas um biplot podem ser observados na Figura 9.

Figura 9 Biplot com as regiões bivariados de 95% de credibilidade dos escores genotípicos $(\lambda_1^{1/2}\alpha_{i1}, \lambda_2^{1/2}\alpha_{i2})$ e ambientais $(\lambda_1^{1/2}\gamma_{j1}, \lambda_2^{1/2}\gamma_{j2})$ que diferiram de zero, considerando o primeiro componente principal (PCA1) e o segundo componente principal (PCA2).



Na Figura 9 observa-se que as regiões bivariadas de 95% de credibilidade para os scores genotípicos e ambientais confirmam as recomendações anteriores do híbrido G2 ao ambiente E8, e os híbridos G8 e G11 aos ambientes E7, E10, E11 e E12. Além disso, pode-se concluir que os híbridos G7 e G12 devem ser recomendados especificamente ao ambiente E6, e os híbridos G9 e G36 aos ambientes E2, E3, E4, E5 e E14.

4 DISCUSSÕES

A avaliação fenotípica de experimentos em melhoramento de plantas tem custo significativo. Isso, associado ao tempo necessário para a obtenção e avaliação em múltiplos locais e anos antes que um híbrido seja lançado, torna o programa oneroso. Estima-se que o custo mínimo de avaliação de uma parcela é de 50 dólares. Nesse contexto, o uso da seleção genômica tem despertado grande interesse para os melhoristas, pois tem sido considerada a hipótese que a utilização da seleção genômica pode auxiliar na economia de tempo e dinheiro por ciclo de melhoramento, oferecendo resultados mais confiáveis (HESLOT et al., 2014).

Jonas e Koning (2013) relatam que, para o sucesso na implementação da predição genômica no melhoramento, vários desafios devem ser superados, tais como captura de efeitos complexos causados por epistasia, alelos raros e, também, a incorporação de efeitos ambientais no processo de seleção. Neste estudo, avaliou-se pela primeira vez no contexto de modelo AMMI, a inclusão da informação de parentesco estimado com marcadores moleculares. Os resultados mostraram que, apesar da informação de parentesco permitir maior poder na discriminação de genótipos com méritos genéticos que diferenciam entre si a posteriori, mesmo assim, em nível de correlação entre valores preditos e observados não há vantagens em procedimentos de predição com a exploração de informações a priori de parentesco obtidas via marcadores moleculares.

Burgueño et al. (2012) incorporaram pela primeira vez informações de parentesco obtidas via pedigree, marcadores moleculares e pedigrees em conjunto com marcadores moleculares em modelos mistos multivariados destinados a análise em um único passo de dados de múltiplos ambientes. Esses autores mostraram que a inclusão da informação de parentesco na maior parte dos modelos testados incrementava a acurácia preditiva, e que os modelos que contemplavam informações de pedigree e marcadores moleculares apresentavam os melhores desempenhos preditivos. Apesar de o modelo utilizado no presente trabalho ter como principal característica a redução da dimensão do efeito multiplicativo (interação GxE), e diferir do modelo de Burgueño et al. (2012), que explora a covariância genética entre ambientes, não foi possível visualizar um resultado evidente de melhoria de capacidade preditiva. Os autores supracitados utilizaram várias formas de exploração de parentesco e compararam com um outro cenário, onde a informação de parentesco era ignorada. Por exemplo, no esquema de desbalanceamento semelhante ao utilizado neste trabalho (chamado pelos autores de CV1), Burgueño et al. (2012) mostraram que ao modelar a matriz **G** com

uma estrutura $cov(\mathbf{g}, \mathbf{g}') = (\mathbf{A}\mathbf{A}' + \boldsymbol{\psi}) \otimes \mathbf{I}_g$ e a \mathbf{R} como identidade, sem informação de parentesco, observaram a pior acurácia preditiva durante a predição dos méritos genéticos dos genótipos desbalanceados no ambiente codificado como E1 (0,22), e a melhor acurácia preditiva na predição dos genótipos desbalanceados no ambiente E2 (0,607).

No caso específico dos modelos em que os autores utilizavam as mesmas estruturas \mathbf{G} e \mathbf{R} , mas com informação de parentesco, esses obtiveram acurácias que oscilaram entre ambientes para os modelos com informação de pedigree, marcadores moleculares, e marcadores mais pedigree de 0,389 a 0,437, 0,404 a 0,532, e 0,432 a 0,542, respectivamente. Conforme pode-se observar nos resultados de Burgueño et al. (2012), é muito difícil afirmar que os modelos com informação de parentesco realmente foram superiores em nível de capacidade preditiva ao modelo que não contemplou informações de parentesco, corroborando os resultados obtidos com o presente estudo, onde foram encontradas diferenças marginais entre as predições realizadas com o AMMI-P e AMMI-SP.

Normalmente, procedimentos de validação cruzada são sensíveis ao número de subconjuntos utilizados durante o processo de subdivisão dos dados, e normalmente acurácias preditivas oscilam dependendo do nível de desbalanceamento (HESLOT et al., 2014). O problema de viés presente em procedimentos de validação cruzada dificulta ainda mais concluir pelos resultados de Burgueño et al. (2012), em que há vantagens na inclusão da informação de parentesco, especialmente nesse cenário em que os autores utilizaram dados fenotípicos coletados em apenas quatro ambientes e compararam diversos modelos diferentes.

Recentemente, Oakey et al. (2016) mostraram incrementos em 5,7% de acurácia preditiva ao acrescentar efeitos genéticos aditivos e não aditivos, utilizando modelos mistos multivariados e testando várias estruturas de VAR-COV. Os efeitos genéticos não aditivos, foram obtidos com a inclusão de um componente genético com distribuição

$\mathbf{g} \sim N\left(0, \mathbf{A}_{n \times n} \boldsymbol{\sigma}_g^2\right)$, além do efeito genético dos valores genéticos aditivos, ou seja,

$\mathbf{a} \sim MVN(0, \mathbf{A} \otimes \mathbf{G}_e)$, em que \mathbf{A} é a matriz de parentesco aditiva estimada com marcadores moleculares e \mathbf{G}_e é a matriz de VAR-COV genética entre ambientes. Apesar de paradoxalmente esses autores não terem comparado, em seu procedimento de validação cruzada, os modelos que utilizavam a informação dos marcadores moleculares com o modelo que não contemplava marcadores, novas investigações com a inclusão de componentes

genéticos não aditivos pode ser uma alternativa parcimoniosa para se contabilizar no modelo AMMI Bayesiano, e merecem novas investigações.

Em caminho semelhante ao adotado neste trabalho, Lado, Barrios e Quincke (2016), com o objetivo de comparar estratégias para explorar o efeito da interação genótipos por ambientes na capacidade preditiva do modelo GBLUP, avaliaram 1477 linhagens de trigo em 35 diferentes combinações com diferentes anos e locais de avaliação. Em síntese, os autores incluíram a correlação entre ambientes na predição dos valores genéticos $\mathbf{g} \sim \mathcal{N}\left[0, (\mathbf{A}_{nvn} \otimes \mathbf{P}_{kvt}) \boldsymbol{\Sigma}_g^2\right]$. A obtenção de informações de parentesco entre genótipos avaliados em múltiplos ambientes e a modelagem da matriz de correlação entre os ambientes representaram a melhor estratégia para a predição de genótipos, porém, com diferenças apenas marginais com relação às acurácias obtidas quando a correlação entre ambientes foi desprezada, $\mathbf{g} \sim \mathcal{N}\left(0, \mathbf{A}_{nvn} \boldsymbol{\Sigma}_g^2\right)$. Ganhos mais significativos foram obtidos apenas em cenários cuja participação do componente associado à interação era muito alta, especialmente quando a abordagem envolveu vários anos.

É possível que o processo de predição utilizando informações de parentesco torna-se ainda mais um grande desafio nas condições climáticas do Brasil, por se tratar de um país de clima tropical, com alta irregularidade de precipitação, alta pressão de pragas e doenças durante o desenvolvimento de plantas no campo (HALLAUER; CARENA; MIRANDA FILHO, 2010). Um exemplo claro da dificuldade de realização de ensaios em condições climáticas tropicais é enfatizado por Borgognone et al. (2016). Esses autores analisaram dados de 37 ensaios internacionais conduzidos durante 4 anos em mais de 17 localizações, conduzido em seis países diferentes. A complexidade da condução de ensaios em condições climáticas tropicais fica evidente ao observar as produtividades médias de trigo obtidas em experimentos conduzidos na Etiópia, que obteve grupos de ensaios com valores de produtividade entre 0,3 e 0,8 t ha⁻¹ e outros grupos com produtividade entre 2,3 a 3,6 t ha⁻¹ (BORGOGNONE et al., 2016). Possivelmente, o desenvolvimento de modelos que explorem informações de covariáveis climáticas, e que permitam a dissecação do padrão de interação GxE, podem ser novas opções que permitam a melhor utilização de informação de marcadores moleculares.

Um exemplo de tal estratégia foi realizado por Heslot et al. (2014), que propuseram e avaliaram várias versões de modelos mistos multivariados que contemplavam informações de covariáveis ambientais, que afetam o ciclo de desenvolvimento do trigo, em conjunto com

informação de marcadores moleculares. Ao final do trabalho, Heslot et al. (2014) concluíram que o melhor modelo testado teve redução de 11% na acurácia preditiva quando comparado ao modelo com o pior desempenho preditivo observado. Nesse mesmo contexto, Jarquin et al. (2014) mostram como realizar predições no contexto da interação genótipos por ambientes, utilizando modelo reação de norma, que permite explorar informações de marcadores associados a covariáveis ambientais, que podem ser calculadas com registros climáticos e características do solo. Os gradientes genéticos e ambientais foram descritos como funções lineares de marcadores e de covariáveis ambientais, respectivamente. Os autores conseguiram até 34% de melhoria precisão da predição realizada, quando comparada às predições baseadas apenas nos efeitos principais.

Novas propostas de modelo AMMI Bayesiano com a inclusão de informações de covariáveis ambientais poderão fornecer uma melhor compreensão sobre a base genética da interação GxE, e fornecer insights para o desenvolvimento de novos modelos que explorem informações de marcadores moleculares, com resultados práticos de acurácia preditiva que justifiquem gastos adicionais com genotipagem em programas de melhoramento.

Entre as principais vantagens da utilização da abordagem Bayesiana do modelo AMMI, em comparação às obtidas pelo modelo clássico, podem-se listar: (i) Permite a inclusão de informações históricas a priori; (ii) Permite a obtenção de estimativas intervalares dos parâmetros como resultado da própria abordagem Bayesiana, sem necessidade de pressupostos assintóticos da distribuição dos parâmetros que controlam o processo generativo dos dados; (iii) O processo de inferência permite a obtenção de elipses que auxiliam a identificação de grupos de genótipos e de ambientes estáveis (JARQUIN et al., 2016; OLIVEIRA et al., 2015; SILVA et al., 2015).

Em relação à vantagem (i), pode-se considerar que se trata da principal vantagem da utilização da abordagem Bayesiana, principalmente ao analisar dados de ensaios multiambientes de desempenho de caracteres de milho, que está entre as espécies mais estudadas e a mais cultivada no mundo (VARSHNEY et al., 2012). O aproveitamento de informações históricas na construção de distribuições a priori, especialmente utilizando informações bem estudadas sobre a arquitetura genética do milho, pode ser estratégia que pode melhorar ainda mais a capacidade preditiva de modelos AMMI e otimizar a eficiência do processo de identificação de genótipos superiores em populações segregantes, além de permitir com maior confiabilidade a recomendação de cultivares a mega-ambientes. Por exemplo, estratégias já utilizadas na elaboração de simulações realísticas utilizando

informações históricas de componentes de variância genética podem ser táticas eficientes para construção de distribuições a priori informativas, como as realizadas por Balestre et al. (2013) simulando dados de caracteres de feijão-comum, e Santos et al. (2015), utilizando informações históricas para simulação de caracteres de milho. Essas estratégias podem aperfeiçoar ainda mais a qualidade dos resultados obtidos pelo modelo AMMI Bayesiano. Novos estudos que exploram a utilização de distribuições a priori informativas a partir de informações históricas são necessários para avaliar o benefício da utilização da abordagem Bayesiana e o impacto de resultados obtidos por meio de processos de aprendizado com a atualização priori/posteriori.

Inúmeras pesquisas estão sendo realizadas na tentativa associar predição de híbridos no contexto da interação genótipos por ambientes com seleção genômica. Entretanto, uma segunda importante discussão não é considerada na maioria dessas pesquisas: além da importante questão associada à melhoria da capacidade preditiva, em que medidas a incorporação do parentesco genômico podem auxiliar no incremento da credibilidade na seleção e recomendação dos híbridos por parte dos melhoristas? No presente trabalho evidenciou-se uma redução considerável no intervalo de credibilidade dos híbridos preditos no cenário onde a informação de parentesco não foi negligenciada. Adicionalmente, o aumento da proporção de híbridos com valores preditos positivos vão de acordo com o melhorista, que, intrinsecamente, busca sempre a seleção de genótipos com maiores médias para produtividade.

Por ora, as vantagens das estimativas intervalares que o modelo AMMI Bayesiano proporciona para recomendação de cultivares ficam claras pela fácil identificação de híbridos que diferem entre si a posteriori. Esse procedimento não necessita do uso de técnicas de amostragem baseadas em teoria assintótica para construção de intervalos de credibilidade, que podem corromper a confiabilidade da inferência da importância dos componentes GxE (SILVA et al., 2015). Adicionalmente, o uso de gráficos com intervalo de credibilidade também permite agilidade durante a recomendação de cultivares com base no componente GxE, que podem ser sugeridas pela sobreposição das elipses de credibilidade dos escores genotípicos e ambientais.

Apesar do grande número de trabalhos desenvolvidos nos últimos anos que mostram a vantagem da utilização do modelo AMMI Bayesiano (JARQUIN et al., 2016; OLIVEIRA et al., 2015; SILVA et al., 2015; VIELE; SRINIVASAN, 2000), poucos trabalhos exploraram o modelo AMMI Bayesiano com grande número de genótipos e de ambientes, como

investigado nesta pesquisa. Alguns exemplos de estudos que utilizaram dados reais são: (i) Crossa et al. (2011) - 9 genótipos em 20 ambientes; (ii) Josse et al. (2014) - 16 genótipos em 16 ambientes; (iii) e (iv) Oliveira et al. (2015) e Silva et al. (2015) - 55 genótipos em 9 ambientes, e (vi) Jarquin et al. (2016) - 20 genótipos em 12 ambientes. A principal justificativa para o número reduzido de genótipos e ambientes utilizados em trabalhos que envolvem o modelo AMMI Bayesiano pode ser devido à alta demanda computacional necessária para o ajuste desse modelo. O desenvolvimento de novos modelos AMMI Bayesianos que viabilizam a utilização de banco de dados em larga escala trata-se de um tópico que necessita de novos avanços, e este assunto deve ser investigado intensivamente. Principalmente, levando em consideração a evolução de plataformas de fenotipagem em larga escala, que atualmente já possibilitam a coleta de milhões de dados fenotípicos em programas de melhoramento (ARAUS; CAIRNS, 2014). Futuramente, dados advindos de plataformas robóticas poderá ser a solução e, ao mesmo tempo, oferecer novas desafios para o desenvolvimento de novos modelos para análise de ensaios multiambiente. Principalmente, modelos que revelem padrões da interação GxE em cenários *Big Data*, com as vantagens dos resultados do modelo AMMI Bayesiano, serão imperativos, e possivelmente serão comuns em programas de melhoramento no futuro (CABRERA-BOSQUET et al., 2012).

5 CONCLUSÕES

A inclusão da matriz de parentesco realizada aditiva no modelo AMMI Bayesiano possibilitou aumentar a precisão na seleção de híbridos com maiores valores genético. O modelo AMMI Bayesiano permitiu a identificação de dois mega-ambientes que representam a região sudeste do Brasil. O modelo AMMI Bayesiano sem a informação de parentesco mostrou ser uma opção eficiente para a identificação de híbridos de milho com alta estabilidade para o caráter produção de grãos em ensaios multiambiente.

REFERÊNCIAS

- ARAUS, J. L.; CAIRNS, J. E. Field high-throughput phenotyping: the new crop breeding frontier. **Trends Plant Science**, Oxford, v. 19, p. 52–61, 2014.
- BALESTRE, M. et al. Applications of multi-trait selection in common bean using real and simulated experiments. **Euphytica**, Wageningen, v. 189, p. 225–238, 2013.
- BORGOGNONE, M. G. et al. Molecular marker information in the analysis of multi-environment trials helps differentiate superior genotypes from promising parents. **Crop Science**, Madison, v. 56, p. 2612-2628, 2016.
- BURGUEÑO, J. et al. Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. **Crop Science**, Madison, v. 52, p. 707-719, 2012.
- CABRERA-BOSQUET L. et al. High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge. **Journal of Integrative Plant Biology**, Medford, v. 54, p. 312–320, 2012.
- CHEN, M. H.; SHAO, Q. M. Monte Carlo estimation of Bayesian credible and HPD intervals. **Journal of Computational and Graphical Statistics**, Alexandria, v. 8, p. 69–92, 1999.
- CROSSA, J. et al. Bayesian estimation of the additive main effects and multiplicative interaction model. **Crop Science**, Madison, v. 51, p. 58–69, 2011.
- CROSSA, J.; YANG, R. C.; CORNELIUS, P. L. Studying crossover genotype x environment interaction using linear-bilinear models and mixed models. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 9, p. 362-380, 2004.
- ELSHIRE, R. J. et al. A robust, Simple Genotyping-by-Sequencing (GBS) approach for high diversity species. **PLoS One**, San Francisco, v. 6, 2011.
- FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, Victoria, v. 14, p. 742-754, 1963.
- GOLLOB, H. F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**, New York, v. 33, p. 73–115, 1968.
- HALLAUER, A. R.; CARENA, M. J.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. Heidelberg: Springer Science & Business Media, 2010.
- HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, Baltimore, v. 31, p. 1109-1144, 1983.

- HESLOT, N. et al. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. **Theoretical and Applied Genetics**, Berlin, v. 127, p. 463-480, 2014.
- JACCOUD, D. et al. Diversity arrays: a solid state technology for sequence information independent genotyping. **Nucleic Acids Research**, Oxford, v. 29, p. e25, 2001.
- JARQUIN, D. et al. A hierarchical Bayesian estimation model for multi-environment plant breeding trials in successive years. **Crop Science**, Madison, v. 56, p. 2260-2276, 2016.
- JARQUIN, D. et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. **Theoretical and Applied Genetics**, Berlin, v. 127, p. 595-607, 2014.
- JONAS, E.; KONING, D. J. Does genomic selection have a future in plant breeding? **Trends in Biotechnology**, Amsterdam, v. 31, p. 497-504, 2013.
- JOSSE, J. et al. Another look at Bayesian analysis of AMMI models for genotype-environment data. **Journal of Agricultural, Biological, and Environmental Statistics**, Bethesda, v. 19, p. 40-57, 2014.
- KAO, C. H.; ZENG, Z. B. Modeling epistasis of quantitative trait loci using Cockerham's model. **Genetics**, Austin, v. 160 p. 1243-1261, 2002.
- LADO, B.; BARRIOS, P. G.; QUINCKE, M. Modeling genotype by environment interaction for genomic selection with unbalanced data from a wheat breeding program. **Crop Science**, Madison, v. 56, p. 2165-2179, 2016.
- OAKLEY, H. et al. Genomic selection in multi-environment crop trials . **G3 Gene, Genomes, Genetics**, New York, v. 6, p. 1313-1326, 2016.
- OLIVEIRA, L. A. et al. Bayesian GGE biplot models applied to maize multi-environments trials. **Genetics and Molecular Research**, Ribeirão Preto, v. 15, p. 1-21, 2016.
- OLIVEIRA, L. A. et al. Credible Intervals for Scores in the AMMI with Random Effects for Genotype. **Crop Science**, Madison, v. 55, p. 465-476, 2015.
- RAFTERY, A. E.; LEWIS, S. M. The number of iterations, convergence diagnostics and generic Metropolis algorithms. In: GILKS, W. R.; SPIEGELHALTER, D. J.; RICHARDSON, S. (Ed.). **Practical markov chain Monte Carlo**. London: Chapman and Hall, 1995.
- SANTOS, J. P. R. et al. Genome-wide prediction of maize single-cross performance, considering non-additive genetic effects. **Genetics and Molecular Research**, Ribeirão Preto, v. 14, p. 18471-18484, 2015.
- SILVA, C. P. et al. A Bayesian Shrinkage approach for AMMI models. **PLoS One**, San Francisco, v. 10, 2015.

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, Washington, v. 57, p. 1138-1147, 2001.

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

VARSHNEY, R. K. et al. Can genomics boost productivity of orphan crops? **Nature Biotechnology**, New York, v. 30, p. 1172–1176, 2012.

VIELE, K.; SRINIVASAN, C. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback–Leibler Information. **Journal of Statistical Planning and Inference**, Amsterdam, v. 84, p. 201–219, 2000.

WILLIAMS, E. J. The interpretation of interactions in factorial experiments. **Biometrika**, London, v. 39, p. 65-81, 1952.

YANG, R. C. et al. Biplot analysis of genotype x environment interaction: proceed with caution. **Crop Science**, Madison, v. 49, p. 1564-1576, 2009.

YATES, F.; COCHRAN, W. G. The analysis of groups of experiments. **Journal of Agricultural Science**, Cambridge, v. 28, p. 556-580, 1938.