JENIFFER SANTANA PINTO COELHO EVANGELISTA

OTIMIZANDO A SELEÇÃO DE MILHO TROPICAL RESISTENTE A FUMONISINA: IMPLEMENTAÇÃO DA ABORDAGEM SINGLE-STEP

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

Orientador: Leonardo Lopes Bhering

Coorientadores: Maria Marta Pastina
Kaio Olímpio das Graças Dias
Camila Azevedo

VIÇOSA – MINAS GERAIS
2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
E92o
2023

Evangelista, Jeniffer Santana Pinto Coelho, 1992-
Otimizando a seleção de milho tropical resistente a fumonisina: implementação da abordagem single-step / Jeniffer Santana Pinto Coelho Evangelista. – Viçosa, MG, 2023.
1 tese eletrônica (56 f.): il. (algumas color.).

Texto em português e inglês.
Orientador: Leonardo Lopes Bhering.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Biologia Geral, 2023.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2023.482
Modo de acesso: World Wide Web.

1. Milho - Seleção. 2. Milho - Melhoramento genético. 3. Milho - Resistência a doenças e pragas. 4. Genômica. 5. *Fusarium verticillioides*. 6. Zea mays. I. Bhering, Leonardo Lopes, 1980-. II. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 633.152

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

JENIFFER SANTANA PINTO COELHO EVANGELISTA

OTIMIZANDO A SELEÇÃO DE MILHO TROPICAL RESISTENTE A FUMONISINA: IMPLEMENTAÇÃO DA ABORDAGEM SINGLE-STEP

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 4 de julho de 2023

Assentimento:

gov.br Documento assinado digitalmente
JENIFFER SANTANA PINTO COELHO EVAN(
Data: 20/08/2023 16:30:31-0300
Verifique em https://validar.iti.gov.br

Jeniffer Santana Pinto Coelho Evangelista
Autora

gov.br Documento assinado digitalmente
LEONARDO LOPES BHERING
Data: 21/08/2023 18:55:22-0300
Verifique em https://validar.iti.gov.br

Leonardo Lopes Bhering
Orientador

Aos meus pais, Marlene e Valdeli, minha irmã Dryelle e Juliano meu marido e parceiro de vida, que sempre me apoiaram nos meus sonhos.

AGRADECIMENTOS

Primeiramente a Deus por me guiar em toda minha trajetória, por ter dado oportunidades no meu caminho, que me fez chegar aonde estou hoje.

Aos meus pais que são um espelho, de honestidade e força, que me ensinaram a lutar pelos meus sonhos e nunca desistir deles.

A minha irmã Dryelle que sempre me motivou e comemorou minhas conquistas como se fosse sua.

Ao meu esposo Juliano, pelo seu apoio incondicional e pela dedicação em tornar meus sonhos realidade, me inspirando diariamente, mostrando que posso alcançar tudo o que almejo. Agradeço de todo o coração por estar ao meu lado em cada passo dessa jornada.

Ao meu orientador Leonardo Lopes Bhering, pelos seis anos de orientação científica, seu apoio e ensinamentos foram essenciais para o meu crescimento profissional e pessoal, serei eternamente grata.

Ao meu coorientador, Kaio Olímpio das Graças Dias, pelos valiosos conselhos, ensinamentos e constante incentivo a buscar conhecimentos além do horizonte.

Ao Diego Jarquin pelo acolhimento em seu laboratório, ensinamentos e amizade e, principalmente, por abrir portas, que foi o divisor de águas para eu chegar onde eu cheguei.

A todos amigos do laboratório de biometria, pelo dia a dia agradável, leve e também pelas excelentes parcerias em pesquisas, me fizeram aprender e crescer muito.

A Embrapa Milho e Sorgo, em especial Maria Marta Pastina, por ter confiado no meu trabalho e ter disponibilizados os dados da minha tese.

Ao Programa de Pós-Graduação em Genética e Melhoramento e seus funcionários, no qual disponibilizaram toda a estrutura e suporte para essa grande realização da minha vida.

A Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela concessão da bolsa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

MUITO OBRIGADA!

"Agradeço todas as dificuldades que enfrentei. Não fosse por elas, eu não teria saído do lugar. As facilidades nos impedem de caminhar. Mesmo as críticas nos auxiliam muito."

(Chico Xavier)

RESUMO

EVANGELISTA, Jeniffer Santana Pinto Coelho, D.Sc., Universidade Federal de Viçosa, julho de 2023. **Otimizando a seleção de milho tropical resistente a fumonisina: implementação da abordagem single-step**. Orientador: Leonardo Lopes Bhering. Corientadores: Maria Marta Pastina, Kaio Olímpio das Graças Dias e Camila Azevedo.

O milho é uma cultura de grande importância para o Brasil, devido a isso programas de melhoramento genético visam utilizar estratégias para desenvolver cultivares mais produtivas e adaptadas às condições locais de cultivo. No entanto, em ambientes tropicais, como no Brasil, surtos de doenças são comuns e um dos problemas mais frequentes é a infecção pelo fungo *Fusarium verticillioides*, que produz micotoxinas chamadas fumonisinas, prejudiciais à saúde humana e animal. Portanto, além de produtividade, é crucial desenvolver híbridos de milho resistentes à fumonisina. A seleção de híbridos com resistência a fumonisina é desafiadora, visto que essa é difícil e cara para realizar a fenotipagem. Neste contexto, o presente trabalho propôs a implementação da abordagem "single-step" (matriz B), abordagem que combina a matriz de parentesco com a matriz genômica, para otimizar duas fases distintas do programa de melhoramento de milho tropical: a fase intermediária e a fase final. Resultados mostram que, para programas de melhoramento que utilizam seleção genômica nas etapas intermediarias, notou-se que ao inserir a matriz B nos cinco modelos lineares testados, teve um aumento na capacidade preditiva, quando comparado estes mesmos modelos utilizando as matrizes de pedigree ou a genômica. Na fase final, que envolve a seleção dos melhores híbridos para o avanço nos ensaios de valor de cultivo e uso (VCU), a matriz B juntamente com o modelo de melhor ajuste aumentou a acurácia de seleção dos híbridos resistentes à fumonisinas em grãos. Mostrando que a abordagem de “single-step” é uma importante estratégia para ser utilizada na otimização da seleção de híbridos com resistência a fumonisina no pipeline de programas de melhoramento.

Palavras-chave: *Fusarium verticillioides*. Zea mays L. Seleção genômica. Melhoramento de plantas.

ABSTRACT

EVANGELISTA, Jeniffer Santana Pinto Coelho, D.Sc., Universidade Federal de Viçosa, July 2023. **Optimizing the fumonisin-resistant tropical maize selection: implementation of the single-step approach**. Adviser: Leonardo Lopes Bhering. Co-advisers: Maria Marta Pastina, Kaio Olímpio das Graças Dias and Camila Azevedo.

The maize (Zea mays L.) is a crop of great importance to Brazil, due to this, breeder genetic programs aim to devlop more productive cultivars adapted to local growing conditions. However, in tropical environments like Brazil, disease outbreaks are common, and one of the most frequent problems is infection by the fungus Fusarium verticillioides, which produces mycotoxins called fumonisins, harmful to human and animal health. Therefore, in addition to productivity, it is crucial to develop corn hybrids resistant to fumonisins. The selection of hybrids with fumonisin resistance is challenging due to the difficulty and cost of phenotyping. In this context, the present study proposed the implementation of the single-step approach (B matrix), which combines the pedigree matrix with the genomic matrix, to optimize two distinct phases of the tropical maize breeding program: the intermediate phase and the final phase. Results showed that for breeding programs using genomic selection in the intermediate stages, inserting the B matrix in the five tested linear models increased the predictive capacity compared to when these models used pedigree or genomic matrices. In the final phase, which involves the selection of the best hybrids for advancing in the value for cultivation and use (VCU) trials, the combination of the best-fit model with the B matrix increased the selection accuracy of fumonisin-resistant hybrids. Showing that the single-step approach is an important strategy to be used for optimizing the selection of hybrids with fumonisin resistance in the breeding program pipeline.

Keywords**:** *Fusarium verticillioides*. Zea mays L. Genomic selection. Plant breeding.

SUMÁRIO

1 INTRODUÇÃO GERAL 9
2 REFERÊNCIAS 14
3 ARTIGO 19
4 INTRODUCTION 20
5 MATERIALS AND METHODS 22
5.1 PHENOTYPIC DATA 22
5.2 PEDIGREE AND GENOTYPIC DATA 23
5.3 RELATIONSHIP MATRICES 24
5.4 PREDICTION MODELS 24
5.5 CROSS-VALIDATION AND B MATRIX OPTIMIZATION 29
5.6 PREDICTIVE ABILITY WITHIN AND ACROSS ENVIRONMENTS 31
5.7 SELECTION OF THE TOP HYBRIDS 31
6 RESULTS 35
6.1 MAIZE BREEDING PIPELINE: INTERMEDIATE STAGES 35
6.1.1 Optimization of the B matrix 35
6.1.2 Prediction models 36
6.2 MAIZE BREEDING PIPELINE: FINAL STAGES 40
6.2.1 Selection of the top hybrid 40
7 DISCUSSION 44
7.1 MAIZE BREEDING PIPELINE: INTERMEDIATE STAGES 44
7.1.1 B matrix optimization 44
7.1.2 Prediction models 44
7.2 MAIZE BREEDING PIPELINE: FINAL STAGES 46
7.2.1 Selection of the top hybrid 46
8 CONCLUSION 47
9 REFERENCES 48
10 SUPPLEMENT MATERIALS 55

1 INTRODUÇÃO GERAL

O milho (*Zea mays* L. spp.) é uma espécie pertencente à família Poaceae, tribo Maydeae, gênero Zea e espécie *Zea mays* (Linnaeus 1799). Trata-se de uma planta monoica e alógama, com inflorescências separadas na mesma planta (Paterniani e Campos 1999). O milho é uma cultura de grande importância para o Brasil, que é utilizado amplamente na alimentação humana e animal. Além disso, o milho é utilizado na fabricação de biocombustíveis, como o etanol, que tem papel estratégico no mercado brasileiro de combustíveis (Salla et al., 2010). De acordo com dados da Companhia Nacional de Abastecimento, a área total plantada no Brasil para a safra 2022/23 foi estimada em 77 milhões de hectares, representando um aumento de 3,3% em relação à safra anterior (CONAB 2022). Do total desta área, a cultura do milho ocupa 21,661 milhões de hectares, incluindo as três safras anuais, com uma produção esperada de 24.810,3 mil toneladas, volume 0,3% superior ao alcançado na safra 2020/21(CONAB 2022). Esse aumento é devido aos esforços dos programas de melhoramento genético, que visam desenvolver cultivares mais produtivas e adaptadas às condições locais de cultivo.

O programa de melhoramento genético do milho é um processo longo e complexo, que pode ser dividido em três fases distintas: i) fase inicial, onde são realizados cruzamentos entre plantas dentro do grupo heterótico, com o objetivo de aumentar a variabilidade e formar linhagens que possuem características desejáveis para avançar no programa; ii) fase intermediária, onde é feito o cruzamento entre grupos heteróticos, em que as linhagens geradas na fase inicial do programa são cruzadas com testadores, nesta etapa, o principal objetivo do melhorista é descartar linhagens que não apresentam potencial para originar híbridos superiores e iii) fase final, onde apenas as melhores combinações, passam pelos testes de valor de cultivo e uso (VCU), estes serão avaliadas e comparadas com as cultivares comerciais existentes. Nesta fase, o objetivo principal é selecionar o híbrido que apresente o melhor potencial produtivo e comercial, em relação as cultivares já existentes no mercado.

Entre os principais estados produtores de milho do país, destacam-se Mato Grosso, Paraná, Goiás e Pará (CONAB 2022). Entretanto devido a ampla extensão do Brasil, dentro do mesmo estado pode apresentar diferentes condições edafoclimáticas, ex.: solo e clima, proporcionado assim, diferentes desempenhos produtivos de uma mesma cultivar de milho entre as regiões. Esta resposta diferencial do genótipo em frente às condições edafoclimáticas é devido a interação genótipos por ambientes (GxE), sendo a performance fenotípica da cultivar moldada principalmente pelos efeitos de genótipo e da interação GxE (Sprague and Federer 1951; Tabery 2008). Existem diversos fatores que podem contribuir para a manifestação da

GxE ao longo das regiões e anos, por exemplo, os abióticos. As variações estáticas, longitude, latitude, solo e características sazonais, e as não estáticas, temperatura, precipitação e umidade, são as principais variações dos fatores abióticos, em que a variação não estática, a principal fonte para GxE na cultura do milho (Cullis et al. 2000; Kleinknecht et al. 2013; Dias et al. 2020; Krause et al. 2022). Geralmente a GxE aumenta a variância fenotípica, causando a redução na estimativa da herdabilidade (Krause et al. 2022), podendo ocorrer viés nas decisões do melhorista na seleção dos genótipos superiores, caso esta não seja considerada nas análises estatísticas. Portanto, os melhoristas avaliam genótipos candidatos em vários locais, que são amostrados de uma população-alvo de ambientes que irão representar a produção agrícola da região onde a cultivar será lançada, com o intuito de revelar os padrões da GxE (Smith et al. 2001; Oakey et al. 2016; Krause et al. 2022).

Ao longo das décadas, várias técnicas estatísticas têm sido utilizadas para compreender a interação GxE, com o objetivo de desenvolver cultivares de milho que sejam estáveis e responsivas a ambientes específicos de interesse para o programa de melhoramento (Zea et al. 2010). Um exemplo é o método AMMI (do inglês, *Additive Main Effects and Multiplicative Interaction*), proposto por (Crossa 1990), utilizado por Oliveira et al. no ano de 2003, para selecionar genótipos de milho com alta estabilidade para produtividade de grãos em uma população-alvo de ambientes no Brasil Central. Porém, ao passar do tempo, novas estratégias começaram a ser implementadas, por exemplo, predição dos valores genéticos utilizando modelos mistos. Kleinknecht et al. (2013), comparam o desempenho do BLUE (do inglês, *BestLlinear Unbiased Estimation*) e BLUP (do inglês, *Best Linear Unbiased Prediction*) como estimadores para os efeitos de genótipos de milho, em mega-ambientes localizados na Índia, concluindo que o BLUP tende a fornecer estimativas mais precisas do desempenho dos híbridos. É comum ocorrer desbalanceamento dos dados, delineamentos não ortogonais e heterogeneidade de variância em ensaios multi-ambientes, o que pode gerar viés nas estimativas dos valores genotípicos e prejudicar a seleção dos melhores genótipos se forem ignorados (Gauch and Zobel 1997). O BLUP leva em consideração a variabilidade das interações entre os genótipos e os ambientes, minimizando o viés na estimativa do efeito genotípico (Piepho et al. 2008), sendo uma ferramenta importante e amplamente utilizada para seleção de genótipos em programas de melhoramento genético (Santiago et al. 2020; Zhang et al. 2020; Haug et al. 2021; Chaves et al. 2023; Evangelista et al. 2023).

Além dos fatores abióticos, outro fator importante que contribui para a manifestação da GxE, são os fatores bióticos, como pragas e doenças. Em ambientes tropicais, como no Brasil, os surtos de pragas e doenças são mais frequentes e podem apresentar grandes variações

ambientais entre locais, anos e estações dentro do ano, afetando diretamente a produtividade das culturas (Oliveira et al. 2014; Sandhu and Dhillon 2021; Chaves et al. 2023). Entre as doenças mais frequentes na cultura do milho, destaca-se a fusariose causada pelo fungo *Fusarium verticillioides* (Jorge et al. 2022). Além de causar perdas significativas na produtividade, o fungo produz micotoxinas prejudiciais à saúde humana e animal, sendo as fumonisinas as mais preocupantes (Blacutt et al. 2018). As fumonisinas são uma classe de micotoxinas que afetam principalmente os grãos de milho e podem contaminar produtos derivados, como a farinha de milho (Butoto et al. 2022). Em vista disso, o Brasil estabeleceu um limite máximo de tolerância para a contaminação de 2 μg/g destas micotoxinas em grãos (Agência Nacional de Vigilância, 2011). Este limite muitas vezes pode impedir a exportação e a comercialização nacional dos lotes de milho, consequentemente, gerando grandes prejuízos aos produtores. Dessa forma, é de extrema importância que os programas de melhoramento de milho considerem a resistência das plantas à contaminação por fumonisina como um critério de seleção (Lanubile et al. 2014; Holland et al. 2020).

No entanto, a fenotipagem dessa característica pode ser cara e demorada (Bush et al. 2004). Em vista disso, a seleção genômica (SG) é uma estratégia importante para a seleção de cultivares de milho mais resistentes à contaminação de fumonisina, pois permite identificar, de forma precisa e rápida, indivíduos com baixo teor de fumonisina com base em suas informações genéticas (Holland et al. 2020). Como resultado, a SG pode aumentar os ganhos obtidos na seleção, ao mesmo tempo em que reduz o tempo e os custos necessários para a identificação dos melhores materiais (Heslot et al. 2015). A SG teve início com trabalhos pioneiros realizados por Bernardo em 1994 e Meuwissen et al. em 2001. Bernardo (1994) propôs o uso da SG para aprimorar a eficiência da seleção em programas de melhoramento, enquanto Meuwissen et al. (2001) desenvolveram um método para predizer o mérito genético de indivíduos com base em seus valores genômicos. Desde então, a SG é amplamente utilizada em programas de melhoramento, principalmente após o surgimento da genotipagem de nova geração, que possibilitou reduzir o custo e o tempo necessário para a genotipagem (Bhat et al. 2016).

Dessa forma, a SG tem sido difundida, apresentando como uma das principais vantagens a capacidade de realizar previsões sobre o desempenho de genótipos em diferentes cenários, sem a necessidade de testá-lo em campo. Isso inclui prever genótipos que ainda não foram observados em nenhum dos ambientes (CV1), previsão de ensaios de campo incompletos onde alguns genótipos foram observados em alguns ambientes,

mas não em outros (CV2), e até mesmo prever genótipos em ambientes não observados (CV0) e genótipos não testados em ambientes não observados (CV00) (Persa et al. 2021). Isso é possível porque as informações genômicas fornecem uma estimativa do potencial genético de um indivíduo, permitindo selecionar aqueles com maior probabilidade de apresentar desempenho genético superior, em diferentes cenários do programa de melhoramento (Crossa et al. 2017).

Uma das ferramentas utilizadas na SG, com o objetivo de aumentar a acurácia de predição, é a matriz de relacionamento construída a partir da matriz genômica (matriz G) e da matriz de pedigree (matriz A), metodologia denominada single-step (matriz B) (Legarra et al. 2014; Ukrainetz and Mansfield 2020). A matriz G é construída a partir das informações de marcadores moleculares presentes no genoma dos indivíduos, que são utilizados para inferir o grau de parentesco entre eles (VanRaden et al. 2008). Por sua vez, a matriz A é gerada a partir das informações genealógicas, com base na probabilidade de compartilhamento de alelos em comum devido à descendência dos ancestrais. (Henderson 1976). Combinar estas informações é importante, uma vez que nem todos os marcadores genéticos estão em desequilíbrio de ligação com as características de interesse. Construindo uma matriz de relacionamento, a partir das matrizes A e G, as informações de pedigree podem contribuir para capturar associações entre alelos causais devido à identidade ancestral comum (Velazco et al., 2019). Consequentemente, o uso da matriz B aumenta a acurácia de predição dos valores genômicos, o que resulta em uma predição mais eficiente dos indivíduos não testados (Crossa et al. 2013; Basnet et al. 2019; Velazco et al. 2019; de Oliveira et al. 2020).

Em vista disso, o uso da metodologia *single-step* pode ser importante estratégia para as etapas intermediarias do programa de melhoramento, para auxiliar na seleção de genótipos de milho mais resistentes à contaminação de fumonisinas em grãos. No entanto, nas etapas finais do programa de melhoramento, onde o foco é o lançamento de uma cultivar no mercado, é necessário que todos os genótipos candidatos sejam testados em campo. No Brasil, para a cultura do milho, são exigidos que todos os genótipos candidatos sejam testados no mínimo, em três locais por região edafoclimática de importância para a cultivar e em 2 anos (Benin et al. 2013). Estes ensaios são denominados de valor de cultivo e uso (VCU’s) que tem o intuito de certificar o valor agronômico do material candidato a ser lançado no mercado (Uate et al. 2019). A seleção dos genótipos superiores para o avanço dos ensaios de VCU's requer precisão, o que torna fundamental a escolha do modelo estatístico adequado para a previsão acurada dos valores genéticos. Para aumentar a confiabilidade da seleção, uma estratégia eficiente é utilizar modelos mistos juntamente com a modelagem da estrutura de covariância dos efeitos genéticos

e não genéticos do modelo estatístico (Pereira et al. 2018; Melo et al. 2020; Chaves et al. 2022; Evangelista et al. 2023).

Existem diferentes tipos de estruturas de covariância que podem ser utilizadas na modelagem dos efeitos genéticos e não genéticos. As mais comuns são: identidade de variância (IDV), diagonal (DIAG), simetria composta (CS), simetria composta heterogênea (CORH), autoregressiva heterogenia (ARH) e fator analítico (FA). As estruturas de covariâncias IDV e CS assumem variâncias homogêneas, ou seja, variância igual para cada elemento da diagonal principal, enquanto a DIAG assume uma variância diferente para cada elemento da diagonal principal, considerando-os independentes (Evangelista et al. 2023). As estruturas ARH1 e CORH, também assumem diferentes variâncias para cada elemento da diagonal principal, porém, permitem algumas covariâncias desiguais multiplicadas pelo coeficiente de correlação (Faveri et al. 2015). A estrutura FA, permite a redução da dimensionalidade dos dados, utilizando variáveis latentes, tornando-se uma opção mais flexível para capturar a complexidade dos dados e pode ser particularmente útil em situações onde há heterogeneidade nos dados (Smith and Cullis 2018).

A escolha da estrutura de covariância adequada depende de uma série de fatores, incluindo a natureza dos dados e a população em estudo (Chaves et al. 2022). Para selecionar a estrutura de covariância de melhor ajuste, é comum realizar testes de adequação do modelo, como o critério de informação de Akaike (AIC), que é amplamente utilizado nos programas de melhoramento (Pereira et al. 2018; Melo et al. 2020; Chaves et al. 2022; Evangelista et al. 2023). O AIC avalia a qualidade do ajuste do modelo, considerando a complexidade do modelo e o número de parâmetros estimados, sendo o modelo que apresentar menor valor desta estatística é considerado o de melhor ajuste (Akaike 1974). Diversos estudos mostram que utilizar a estrutura de covariância adequada, para os efeitos genéticos e não genéticos, aumenta a precisão das estimativas dos valores genotípicos, permitindo assim, identificação dos melhores genótipos com maior confiabilidade (Chaves et al. 2022; Evangelista et al. 2023). Por outro lado, ignorar a modelagem de estrutura de covariância pode levar resultados equivocados e comprometer a eficiência da seleção genética, induzindo assim, o melhorista a tomar decisões errôneas. Assim, a seleção do modelo de melhor ajuste é uma estratégia importante para ser aplicada para a seleção nos estágios finais do programa de melhoramento.

Outra estratégia para aumentar a precisão das previsões dos BLUP's, é a inclusão de uma matriz de parentesco no modelo. A matriz de parentesco leva em conta as relações genéticas entre os indivíduos avaliados e ajusta as previsões do BLUP de acordo com a magnitude do parentesco, reduzindo assim o viés nas predições (Piepho et al. 2008). Dessa

forma, a inclusão da matriz de parentesco aumenta a acurácia das predições dos valores genéticos, tornando assim, a seleção mais assertiva. Em vista disso, utilizar a seleção do modelo de melhor ajuste juntamente com a inclusão da matriz de parentesco para a predição do BLUP's, são estratégias importantes para selecionar, com maior precisão, os genótipos candidatos para avançar nos ensaios de VCU's.

Assim, o objetivo principal do trabalho, foi investigar como o uso da matriz B pode contribuir para a otimização de duas fases distintas (intermediária e final) do programa de melhoramento de milho tropical para a seleção de híbridos com maior resistência a fumonisina em grãos, abordando os tópicos a seguir: i) Otimização da matriz B para inclusão nos modelos de predição; ii) comparar a capacidade preditiva de diferentes modelos de predição, utilizando três matrizes de parentesco diferentes (matrizes G, A e B); iii) como a matriz B juntamente com a modelagem da estrutura de covariância, pode auxiliar o melhorista na seleção dos melhores híbridos para o avanço nos ensaios de VCU's.

2 REFERÊNCIAS

Adebola FB, Fasoranbaku OA, Kupolusi JA (2020) On prediction error variance to determining optimal design for two variable quadratic logistic model. Cogent Math Stat 7:1853888. https://doi.org/10.1080/25742558.2020.1853888.

Jarquin D, de Leon N, Romay C, et al (2021) Utility of Climatic Information via Combining Ability Models to Improve Genomic Prediction for Yield Within the Genomes to Fields Maize Project. Front Genet 11:1–11. https://doi.org/10.3389/fgene.2020.592769.

Jorge K, Teixeira C, Sylvia G, et al (2022) A genome - wide association study investigating fumonisin contamination in a panel of tropical maize elite lines. Euphytica 1–12. https://doi.org/10.1007/s10681-022-03082-0.

Kamle M, Mahato DK, Devi S, et al (2019) Fumonisins: Impact on agriculture, food, and human health and their management strategies. Toxins (Basel) 11:328.

Khanna A, Anumalla M, Catolos M, et al (2022) Optimizing predictions in IRRI's rice drought breeding program by leveraging 17 years of historical data and pedigree information. Front Plant Sci 13:. https://doi.org/10.3389/fpls.2022.983818.

Kleinknecht K, Möhring J, Singh KP, et al (2013) Comparison of the performance of best linear unbiased estimation and best linear unbiased prediction of genotype effects from zoned Indian maize data. Crop Sci 53:1384–1391. https://doi.org/10.2135/cropsci2013.02.0073.

Krause MD, Dias KOG, Singh AK, Beavis WD (2022) Using large soybean historical data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors.

Lado B, Barrios PG, Quincke M, et al (2016) Modeling genotype× environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Sci 56:2165–2179.

Lanubile A, Maschietto V, Marocco A (2014) Breeding maize for resistance to mycotoxins. Mycotoxin Reduct grain Chain 37–58.

Legarra A, Christensen OF, Aguilar I, Misztal I (2014) Single Step, a general approach for genomic selection. Livest Sci 166:54–65.

Linnaeus C (1799) Species plantarum. Impensis GC Nauk.

Liu Z, Seefried FR, Reinhardt F, et al (2011) Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet Sel Evol 43:1–9.

Macedo FL, Christensen OF, Astruc JM, et al (2020) Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. Genet Sel Evol 52:1–10. https://doi.org/10.1186/s12711-020-00567-1.

Mäntysaari EA, Koivula M, Strandén I (2020) Symposium review: Single-step genomic evaluations in dairy cattle. J Dairy Sci 103:5314–5326. https://doi.org/10.3168/jds.2019-17754.

Martini JWR, Schrauf MF, Garcia-Baccino CA, et al (2018) The effect of the H -1 scaling factors τ and ω on the structure of H in the single-step procedure. Genet Sel Evol 50:1–9. https://doi.org/10.1186/s12711-018-0386-x.

Masuda Y, Tsuruta S, Bermann M, et al (2021) Comparison of models for missing pedigree in single-step genomic prediction. J Anim Sci 99:1–10. https://doi.org/10.1093/jas/skab019.

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem medica 22:276–282.

Melo DT, Marc DS, Silva R, Anjos P (2020) Modeling (co) variance structures for genetic and non- genetic effects in the selection of common bean progenies. 9:. https://doi.org/10.1007/s10681-020-02607-9.

Mesterhazy A, Lemmens M, Reid LM (2012) Breeding for resistance to ear rots caused by Fusarium spp. in maize–a review. Plant Breed 131:1–19.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92:4648–4655.

Oakey H, Cullis B, Thompson R, et al (2016) Genomic selection in multi-environment crop trials. G3 Genes, Genomes, Genet 6:1313–1326.

Oliveira CM, Auad AM, Mendes SM, Frizzas MR (2014) Crop losses and the economic impact of insect pests on Brazilian agriculture. Crop Prot 56:50–54.

Oliveira JP, Moreira Jr. WN, Duarte JB, et al (2003) Genotype-environment interaction in maize hybrids: an application of the AMMI model. Crop Breed Appl Biotechnol 3:185–192. https://doi.org/10.12702/1984-7033.v03n03a02.

Pádua JMV, Das Graças Dias KO, Pastina MM, et al (2016) A multi-environment trials diallel analysis provides insights on the inheritance of fumonisin contamination resistance in tropical maize. Euphytica 211:277–285. https://doi.org/10.1007/s10681-016-1722-2.

Paterniani E, CAMPOS MS (1999) Melhoramento do milho. Melhor espécies Cultiv Viçosa UFV 429–485.

Pereira FAC, De Carvalho SP, Rezende TT, et al (2018) Selection of coffea arabica L. Hybrids using mixed models with different structures of variance-covariance matrices. Coffee Sci 13:304–311. https://doi.org/10.25186/cs.v13i3.1444.

Pérez P, De Los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. Genetics 198:483–495. https://doi.org/10.1534/genetics.114.164442.

Persa R, Grondona M, Jarquin D (2021) Development of a Genomic Prediction Pipeline for Maintaining Comparable Sample Sizes in Training and Testing Sets across Prediction Schemes Accounting for the Genotype-by-Environment Interaction.

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. https://doi.org/10.1007/s10681-007-9449-8.

R Core Team (2022) R: A language and environment for statistical computing.

Richard JL (2007) Some major mycotoxins and their mycotoxicoses—An overview. Int J Food Microbiol 119:3–10.

Rocha JR do AS de C, Nunes KV, Carneiro ALN, et al (2019) Selection of superior inbred progenies toward the common bean ideotype. Agron J 111:1181–1189. https://doi.org/10.2134/agronj2018.12.0761.

Rosa Junior OF, Dalcin MS, Nascimento VL, et al (2019) Fumonisin production by Fusarium verticillioides in maize genotypes cultivated in different environments. Toxins (Basel) 11:215.

Salla DA, Furlaneto F de PB, Cabello C, Kanthack RAD (2010) Estudo energético da produção de biocombustível a partir do milho. Ciência Rural 40:2017–2022.

Samayoa LF, Cao A, Santiago R, et al (2019) Genome-wide association analysis for fumonisin content in maize kernels. BMC Plant Biol 19:1–11.

Sandhu S, Dhillon BS (2021) Breeding plant type for adaptation to high plant density in tropical maize—A step towards productivity enhancement. Plant Breed 140:509–518.

Santiago R, Cao A, Malvar RA, Butrón A (2020) Genomics of maize resistance to fusarium ear rot and fumonisin contamination. Toxins (Basel) 12:1–16. https://doi.org/10.3390/toxins12070431.

Silva KJ, Guimarães CT, Guilhen JHS, et al (2020) High-density SNP-based genetic diversity and heterotic patterns of tropical maize breeding lines. Crop Sci 60:779–787. https://doi.org/10.1002/csc2.20018.

Smith A, Cullis B, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138–1147.

Smith AB, Cullis BR (2018) Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. Euphytica 214:143.

Souza VF de, Ribeiro PC de O, Vieira Junior IC, et al (2021) Exploring genotype× environment interaction in sweet sorghum under tropical environments. Agron J 113:3005–3018.

Sprague GF, Federer WT (1951) A Comparison of Variance Components in Corn Yield Trials: II. Error, Year x Variety, Location x Variety, and Variety Components 1 . Agron J 43:535–541. https://doi.org/10.2134/agronj1951.00021962004300110003x.

Tabery J (2008) R. A. Fisher, Lancelot Hogben, and the origin(s) of genotype-environment interaction. J Hist Biol 41:717–761. https://doi.org/10.1007/s10739-008-9155-y.

Tiezzi F, de los Campos G, Parker Gaddis KL, Maltecca C (2017) Genotype by environment (climate) interaction improves genomic prediction for production traits in US Holstein cattle. J Dairy Sci 100:2042–2056. https://doi.org/10.3168/jds.2016-11543.

Uate JV, Nuvunga JJ, da Silva CP, et al (2019) Genetic progress, adaptability and stability of maize cultivars for value of cultivation and use trials. Acta Sci - Agron 41:1–11. https://doi.org/10.4025/actasciagron.v41i1.42624.

Ukrainetz NK, Mansfield SD (2020) Prediction accuracy of single-step BLUP for growth and wood quality traits in the lodgepole pine breeding program in British Columbia. Tree Genet Genomes 16:1–13.

USFDA Center for Food Safety and Applied Nutrition (2001) Guidance for industry: fumonisin levels in human foods and animal feeds. US Food Drug Adm.

VanRaden PM, Van Tassell CP, Wiggans GR, et al (2008) Reliability of genomic predictions for North American dairy bulls. J Dairy Sci 91:305.

Velazco JG, Malosetti M, Hunt CH, et al (2019) Combining pedigree and genomic information to improve prediction quality: an example in sorghum. Theor Appl Genet 132:2055–2067. https://doi.org/10.1007/s00122-019-03337-w.

Vitezica Z-G, Aguilar I, Misztal I, Legarra A (2011) Bias in genomic predictions for populations under selection. Genet Res (Camb) 93:357–366.

Zea L, Kandus M, Almorza D, et al (2010) Statistical models for evaluating the genotype-environment interaction in maize. 39–46.

Zhang M, Cui Y, Liu Y-H, et al (2020) Accurate prediction of maize grain yield using its contributing genes for gene-based breeding. Genomics 112:225–236.

3 ARTIGO

Single-step approach: optimization of breeding program pipeline of tropical maize for reducing fumonisin concentration in grains

ABSTRACT

The mycotoxins called fumonisins are a serious problem for maize (Zea mays L.), mainly because are a risk to human and animal health. In general, this study aimed to use the single-step approach for the optimization of two different stages (intermediary and final) of the maize breeding program, with emphasis on the reduction of fumonisin in grains. We used a dataset with 392 single-cross tropical maize hybrids originating from 359 inbred lines, evaluated in three different years. For the intermediary stages of the program, we focused on the deployment of genomic prediction by considering five linear predictor models and three kinship matrices (genomic matrix, pedigree matrix, and, single-step matrix). For the final stages, we focused on the selection of the top hybrids candidates to be a cultivar, for this, we used models with different covariance structures for the genetic and residual effects plus these models with or without the single-step matrix, for the BLUP prediction. In the intermediary stages of the program, when we compare the three kinship matrices used, the single-step matrix improved the prediction ability for the five linear predictor models tested. In the final stages of the program, when used the best fit model plus the single-step matrix, increased selective accuracy of the hybrids with greater resistance to fumonisins in grains. Showing that the single-step approach is a good tool to optimize the breeding programs pipeline.

Keywords: *Zea mays* L., genomic selection, relationship matrices, covariance structures.

4 INTRODUCTION

Fusarium ear rot, a disease caused by fungi of the genus Fusarium, mostly *F. verticillioides* (Sacc), is a serious problem for maize (*Zea mays* L.) because it damages all parts of the plant, especially grains, and ears, causing up to 30% of yield losses on the most severe cases (Mesterhazy et al. 2012; Rosa Junior et al. 2019). The infected grains contain mycotoxins that are hazardous to human and animal health (Blacutt et al. 2018). These toxins are called fumonisins and are divided into four categories, A, B, C, and P, with the B type being the most frequent and toxic (Kamle et al. 2019). Several authors associate the presence of fumonisins with the occurrence of cancers in the gastrointestinal systems of humans and animals (Blacutt et al. 2018; Chen et al. 2021). This is unsettling given the fact that the plant can remain symptomless after *F. verticillioides* infection (Richard 2007). Exploring the genetic resistance of the plants is the most cost-effective solution to overcome the fumonisin contamination. Association studies have shown the presence of QTLs influencing the fumonisin content in maize grains (Samayoa et al. 2019; Santiago et al. 2020; da Silva et al. 2022). Given the dispersion of these QTLs across the maize's genome and the polygenic nature of the trait, a genome-wide selection approach can significantly contribute to distinguishing resistant candidates (Gaikpa and Miedaner 2019). Nevertheless, studies of this kind are severely lacking (Holland et al. 2020; Butoto et al. 2022).

Employing information available from thousands of markers across the genome using statistical models to predict the genomic estimated breeding values (GEBVs) is the core idea of genomic selection (GS) (Meuwissen et al. 2001). The use of GS provides an increase in the selection intensity, expediting the breeding cycles and increasing the genetic gains (Atanda et al. 2021; Beyene et al. 2021; Persa et al. 2021). Due to these advantages, many breeding companies are already using GS in the intermediate stages of their programs for predicting the performance of untested genotypes. This is possible because GS explores the realized genomic relationships between genotypes, i.e. the alleles they share, through kinship matrix (G) (VanRaden et al. 2008) whose entries describe the genomic similarities between pairs of individuals. This is a more realistic approach compared to the pedigree-based selection, which depend on relationship matrix (A) built based on the expected similarity between individuals (Hayes et al., 2009).

This does not mean that the pedigree is expendable or lack of value in data analysis. Frequently, not all the selection candidates or parental lines are genotyped, but their pedigree is registered (Callister et al., 2021). In this scenario, A can be enriched or complemented by G,

forming a single relationship matrix (Misztal et al. 2009; Aguilar et al. 2010). Thence, the full set of genotypes can be evaluated/analyzed using a standard genomic selection method, such as the Genomic Best Linear Unbiased Predictor (GBLUP) method by characterizing the single-step GBLUP, or ssGBLUP (Legarra et al. 2014). Studies highlight the potential of ssGBLUP compared to the traditional GBLUP and ABLUP (pedigree-based selection) for predicting untested genotypes (Ashraf et al. 2016; Ukrainetz and Mansfield 2020).

Another important stage to the breeding program is the final selection of genotypes. It consists of the hybrid's final evaluation and selection of the genotypes that are candidates to be launched/released in the market. The Brazilian government requires at least three years of field evaluation to launch/release a commercial maize cultivar, thus the use of genomic selection is unfeasible at this very last stage. However, the use of a model that returns the best fit modeling the structures of covariances increases the reliability of genetic selection (Chaves et al. 2022). In addition, using the relationship matrix could help to control the bias induced in the BLUP´s predictions (Piepho et al. 2008). In this way, both approaches can be used for a more accurate selection at the final stage of the breeding program.

In this study, we use a dataset with 392 single-cross tropical maize hybrids originated from 359 inbred lines and evaluated in three different years. For assessing the use of the single-step approach (here called B matrix) contributing in the two different stages (intermediary and final) of the maize breeding program, with emphasis on the reduction of fumonisin in grains. First, we focused on the intermediary stages of the program where the main goal is the discard of the inferior hybrids, based on the deployment of genomic prediction by considering the following objectives: *i*) i) optimization of the B matrix, selecting the best way to find the weighting factor (w) optimum for combining genomic information and pedigree data in the prediction models; ii) comparing the predictive ability of different models, using marker SNPs (G), pedigree information (A), and the single-step matrix (B matrix, in our case). In the second study, we focused on the final stages of the breeding program, where the main goal is the selection of the top hybrids, rather than their prediction. In this stage, we showed how to implement models with different covariance structures for the genetic and residual effects, with or without the B matrix, to evaluate how these might influence the breeders' decisions in the selection of the top hybrids to the advancement of the release stage of the breeding program.

5 MATERIALS AND METHODS

5.1 PHENOTYPIC DATA

The analyzed datasets correspond to three trials stablished at the Brazilian Agricultural Research Corporation (EMBRAPA) Maize and Sorghum headquarters, which is in Sete Lagoas city at the Minas Gerais state, Brazil (19º28'S, 44º15'W) in a lattice design, with two repetitions. These trials were conducted in three consecutive agronomical years (2014/2015, 2015/2016, and 2016/2017), and each trial was considered as a different environment. In total, 392 single-cross hybrids were evaluated: 156 in environment 1 (E1), 153 in E2, and 155 in E3. Furthermore, care was taken to allow connectivity between the environment: 33 hybrids in common between E1 and E2, 15 between E1 and E3, and 37 between E2 and E3 (**Figure 1**). The analyses performed to quantify fumonisin concentration in parts per million (ppm) were realize in the Laboratory of Food Safety at Embrapa Milho e Sorgo, using the Fluorometer VICAM according to the manufacturer's protocols.

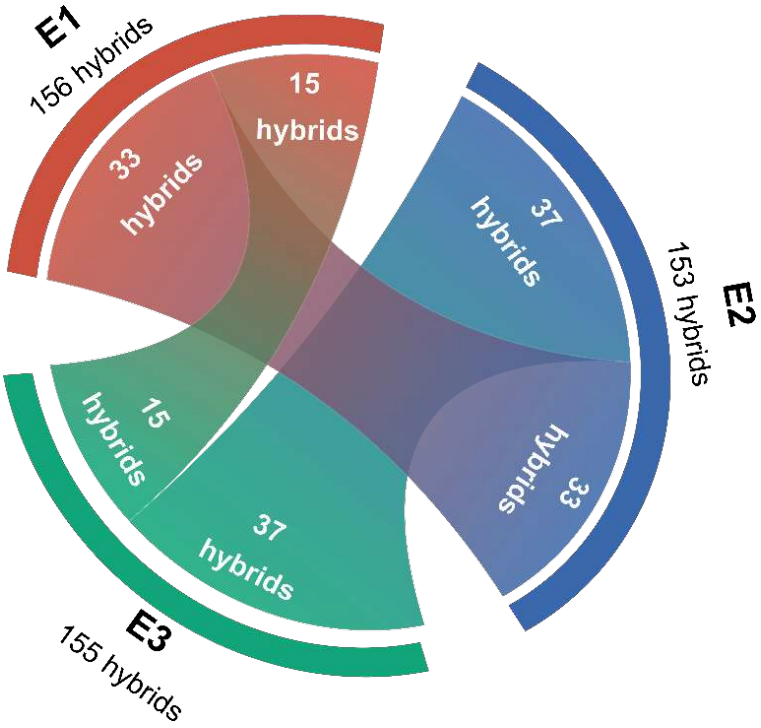


Figure 1. Representation of the allocation of the hybrids of maize in the evaluated three environments (E1, E2, and E3), inside the circle, presents the number of hybrids that were evaluated in common between environments.

5.2 PEDIGREE AND GENOTYPIC DATA

The hybrids originated from 356 unique inbred lines belonging to three heterotic groups: Flint, Dent, and C groups. The C group is intermediate to Flint and Dent, representing lines of several origins, when crossed with Dent and Flint testers, there is a good combining ability with both groups (Silva et al. 2020). The inbred lines were divided into two groups: P1 (inbred lines used as male) 145 lines; P2 (inbred lines used as female) 234 lines. Some inbred lines were used as male and female, thus these genotypes are allocated in P1 and P2 groups. Since only 392 hybrids were considered in this study, thus not all inbred genotypes of the P2 group were testcross with all genotypes P1 group (testers). In addition, for the 356 inbred lines there is also information regarding their ancestors. Thus, it was possible to compute the corresponding pedigree-based relationship A matrix for the parents P1 (A_{P1}), and P2 (A_{P2}), and for the hybrids (A_H).

The young leaves of 333 inbred lines out of the 356 were used for extracting genomic DNA via the CTAB method (Hexadecyltrimethylammonium bromide) (Saghai-Maroof et al., 1984). The DNA quantification was done using a fluorometer, following the manufacturer's instructions. The samples were shipped to the Genomic Diversity Facility of Cornell University (Ithaca, NY, USA) for genotyping-by-sequencing (GBS) (Elshire et al., 2011). Using the Burrows-Wheeler alignment (BWA) tool (Li and Durbin, 2009) the sequences were aligned with the B73 reference genome (AGPv3). A total 474,367 SNPs markers were available for analysis.

After applying quality control using TASSEL v.5.2.10 software (Bradbury et al., 2007), discarding those molecular markers with a Minor Allele Frequency (MAF)) greater than 10%, and a heterozygote's proportion per locus above 10%, 73,083 polymorphic SNPs remained in the analyses. In addition, the SNPs with missing information were imputed using Beagle software (Browning and Browning, 2016). The SNPs markers of the inbred lines were code as 0, 1, and 2, using as reference the allele with the minor allele frequency for 0, 1 for the heterozygous, and 2 for the allele with the major allele frequency. The genotypes of the hybrids were obtained through the combination of the markers from the respective parental inbred lines via expected value (mean allele dosage across parents for each maker).

5.3 RELATIONSHIP MATRICES

The pedigree data was used to build the additive relationship matrix for P1 (A_{P1}), P2 (A_{P2}), and the hybrids (A_H), according to Henderson (1976), using and the R package AGHmatrix (Amadeu et al., 2016). Using the genomic data, we the computed the realized genomic relationship matrix (or G matrix) for P1 (G_{P1}), P2 (G_{P2}), and the hybrids (G_H) following VanRaden (2008). In addition, the single-step relationship matrix was built for P1, P2, and hybrids, using the pedigree and genomic information. This matrix is commonly named the H matrix (reference); however, since this research involves the hybrid prediction, we reserve H for denoting the kinship matrix obtained from combining the parental information of the parents forming the crosses as the mean of the allele dosage of the parents for each marker position (these will be presented in the next section). Here, the resulting matrix from the single-step procedure is named B matrix. For example, E+H model (**Figure 2**) represent a linear predictor that includes the main effect of the environments and the main effect of the markers of the hybrids obtained as the mean across the marker information of the parents involved in the cross. This change was for a better understanding of the reader. The B matrix was built following Aguilar et al. (2010):

$$B = wA + (1 - w)G \qquad (1)$$

where A is the pedigree relationship matrix, G is the genomic relationship matrix and w is the weighting factor (ie, the fraction of total additive variance not addressed by the markers) (Velazco et al. 2019).

5.4 PREDICTION MODELS

A two-stage approach was considered to implement the prediction models. In the first stage, within environments adjusted means were obtained with best linear unbiased estimation (BLUE) implementing the following model:

$$y_{ikl} = \mu + g_i + r_k + b_{l(k)} + \varepsilon_{ikl} \qquad (2)$$

where y_{ijkl} is the phenotypic value of the i^{th} hybrid within block l^{th} t of replicate k^{th}; μ is the constant, r_k is the fixed effect of replication k, $b_{l(k)}$ is the random effect of the l^{th} block within replication k, with $b_{l(k)} \sim N(0, \sigma_b^2)$, where σ_b^2 is the block variance, g_i is the fixed effect of hybrid i, and ε_{ikl} is the residual random effect with $\varepsilon_{ikl} \sim N(0, \sigma_\varepsilon^2)$, with σ_ε^2 as the residual

variance. The first stage analysis was carried out using the statistical package ASReml-R v.4 (Butler et al. 2018) implemented in the statistical software R v.4.1.3 (R Core Team 2022).

In the second stage, five linear predictors and three relationship matrices (G, A, and, B matrices) were combined in different ways for conducting the predictions, totalizing thirteen different training models (**Figure 2**). The first two linear predictors (1, and 2) were implemented to model the hybrid performance using the genetic relationship based on synthetics (mean of the allele content between the parents at each marker position) and the interaction of these synthetic markers with environments. In addition, for comparing which kinship matrices reaches a higher prediction accuracy, for linear models 1 and 2, the G and B matrices were considered in these models, totalizing four prediction models (Pred_m1, Pred_m2, Pred_m9, and Pred_m10 in **Figure 2**). The other three linear predictors (3 to 5) considered the genomic and/or the pedigree information of the parents via the general and specific combining ability (GCA and SCA) terms. For the investigation of which kinship matrix reach the best predictive ability, for the linear predictor models 3 to 5, the G, A, and B, matrices were considered, totalizing nine models (Pred_m3, Pred_m4, Pred_m5, Pred_m6, Pred_m7, Pred_m8, Pred_m11, Pred_m12 and Pred_m13 in **Figure 2**). Furthermore, the phenotypic information for the training of the thirteen prediction models was divided into two datasets: 1) named genomic only dataset (GOds), this consists of only the individuals that have genomic information, in total 246 hybrids, and; 2) named pedigree and genomic dataset (PGds), consisting of individuals that have might have genomics information or not, in total 392 hybrids. Below are the liner predictors models (i to v), and their corresponding components (**Figure 2**):

i. Linear predictor 1

Consider that $\bar{y}_{ij}$ represents the phenotypic response (FUMO) of the i^{th} hybrid observed in the j^{th} environment and it can be modeled as follows:

$$\bar{y}_{ij} = \mu + E_j + g_i + \varepsilon_{ij} \qquad (3)$$

where μ is the general mean (constant effect across genotypes and environments), E_j is the fixed effect of j^{th} environment, g_i is the random effect of i^{th} hybrid, where $\boldsymbol{g} = \{g_i\} \sim N(\mathbf{0}, \boldsymbol{K}\sigma_g^2)$. The g_i can be modeled considering different sources of information. For example, $\boldsymbol{K}$ might represent covariance structure computed with the SNPs markers G, or the B matrix of the hybrids, with σ_g^2 being the additive variance, and, $\varepsilon_{ij} \sim N(\mathbf{0}, \sigma_\varepsilon^2)$ where σ_ε^2 is the residual variance. Two training models were built using the G and B matrices and linear predictor 1 (**Figure 2**) varying the kinship matrix (either G or B).

ii. *Linear predictor 2*

This model is similar to liner predictor 1, but includes the interaction effect between hybrid and environment via the reaction norm model (Jarquin et al. 2014). Considering the specific response of the *i*th hybrid in the *j*th environment through the model term gE_{ij}, this linear predictor can be described as:

$$\bar{y}_{ij} = \mu + E_j + g_i + gE_{ij} + \varepsilon_{ji} \qquad (4)$$

where is the random effect of the hybrid by environment interaction such that $\boldsymbol{gE} = \{gE_{ij}\} \sim N[0, (\boldsymbol{Z_g K Z_g'})^{\circ} (\boldsymbol{Z_E Z_E'})\sigma^2_{ge}]$, Z_g and Z_E being the incidence matrices that connect phenotypes with genotypes and environments, respectively, σ^2_{ge} represents the variance component of the hybrid by environment interaction, and ° is the Hadamard or Shur product (the cell-by-cell product) between two matrices. The interaction term can take different values depending on the source of information (G, A, or B matrix). Two prediction models were built using G and B matrices with the linear predictor 2 (**Figure 2**).

iii. *Linear predictor 3*

Using the G, A, and B kinship matrices from the inbred lines via the GCA model terms this linear predictor was built such that the male and female effects for the prediction of the hybrids can be distinguished. Consider the *i*th hybrid originated by crossing parent 1(P1) and parent 2 (P2), with g_{P1i} and g_{P2i} representing the genomic random effects of the P1 and P2 respectively. The following linear predictor was implemented to model the performance of the *i*th hybrid in the *j*th environment via the GCA of the inbred lines:

$$\bar{y}_{ij} = \mu + E_j + g_{P1i} + g_{P2i} + \varepsilon_{ji} \qquad (5)$$

where $\boldsymbol{g_{P1}} = \{g_{P1i}\} \sim N(\boldsymbol{0}, \boldsymbol{K_{P1}}\sigma^2_{P1})$, with $\boldsymbol{K_{P1}}$ being either the G, A, or B kinship matrix of the P1 parent, σ^2_{P1} being the associated variance component; $\boldsymbol{g_{P2}} = \{g_{P2i}\} \sim N(\boldsymbol{0}, \boldsymbol{K_{P2}}\sigma^2_{P2})$, again with $\boldsymbol{K_{P2}}$ being a G, A, or B kinship matrix of the P2, and σ^2_{P2} being the associated variance component. Three different models were built considering the G, A, and, B matrices for the linear predictor 3 (**Figure 2**).

iv. *Linear predictor 4*

Similar to liner predictor M3, in addition to the GCA of the inbred lines this model also includes the SCA interaction effect of crossing a specific pair of parents (Acosta-Pech et al., 2017). Combining the assumptions from the linear predictor M3 with the SCA term, the resulting linear predictor is:

$$\bar{y}_{ij} = \mu + E_j + g_{P1i} + g_{P2i} + s_{P1iP2i} + \varepsilon_{ji} \qquad (6)$$

where $\boldsymbol{s_{P1P2}} = \{s_{P1iP2i}\} \sim N(\mathbf{0}, \boldsymbol{K_{P1P2}}\sigma_s^2)$, with $\boldsymbol{K_{P1P2}} = \boldsymbol{K_{P1}}^{\circ}\boldsymbol{K_{P2}}$, being $\boldsymbol{K_{P1}}$ and $\boldsymbol{K_{P2}}$ of either the G, A, or B kinship matrices of P1 and P2, respectively, and σ_s^2 is the associated variance component with this interaction term. Three training models were built using combining either G, A, and, B matrices and the linear predictor 4 (**Figure 2**).

v. Linear predictor 5

Similarly to M4, but also including the interaction between the GCA and SCA components with the environment via a reaction norm model. The resulting linear prediction can be described as:

$$\bar{y}_{ij} = \mu + E_j + g_{P1i} + g_{P2i} + s_{P1iP2i} + gE_{P1i} + gE_{P2i} + sE_{P1iP2i} + \varepsilon_{ji} \quad (7)$$

where $\boldsymbol{gE_{P1}} = \{gE_{P1i}\} \sim N[\mathbf{0}, (\boldsymbol{Z_{P1}K_{P1}Z'_{P1}})^{\circ}(\boldsymbol{Z_E Z'_E})\sigma^2_{P1e}]$,

$\boldsymbol{gE_{P2}} = \{gE_{P2i}\} \sim N[\mathbf{0}, (\boldsymbol{Z_{P2}K_{P2}Z'_{P2}})^{\circ}(\boldsymbol{Z_E Z'_E})\sigma^2_{P2e}]$,

$\boldsymbol{sE_{P1P2}} = \{sE_{P1iP2i}\} \sim N[\mathbf{0}, (\boldsymbol{Z_{P1}K_{P1}Z'_{P1}})^{\circ}(\boldsymbol{Z_{P2}K_{P2}Z'_{P2}})^{\circ}(\boldsymbol{Z_E Z'_E})\sigma^2_{se}]$, Z_{P1}, Z_{P2} and Z_E being the incidence matrices to connect g_{P1i}, g_{P2i} and E_j with $\bar{y}_{ij}$ (the vector of phenotypes); σ^2_{P1E}, σ^2_{P2E} and σ^2_{sE} as the corresponding variance components of the interaction between the GCA (P1 and P2) and SCA (P1 × P2) terms with the environments *E*. Three linear predictors were built using either the G, A, or B matrices (see **Figure 2**).

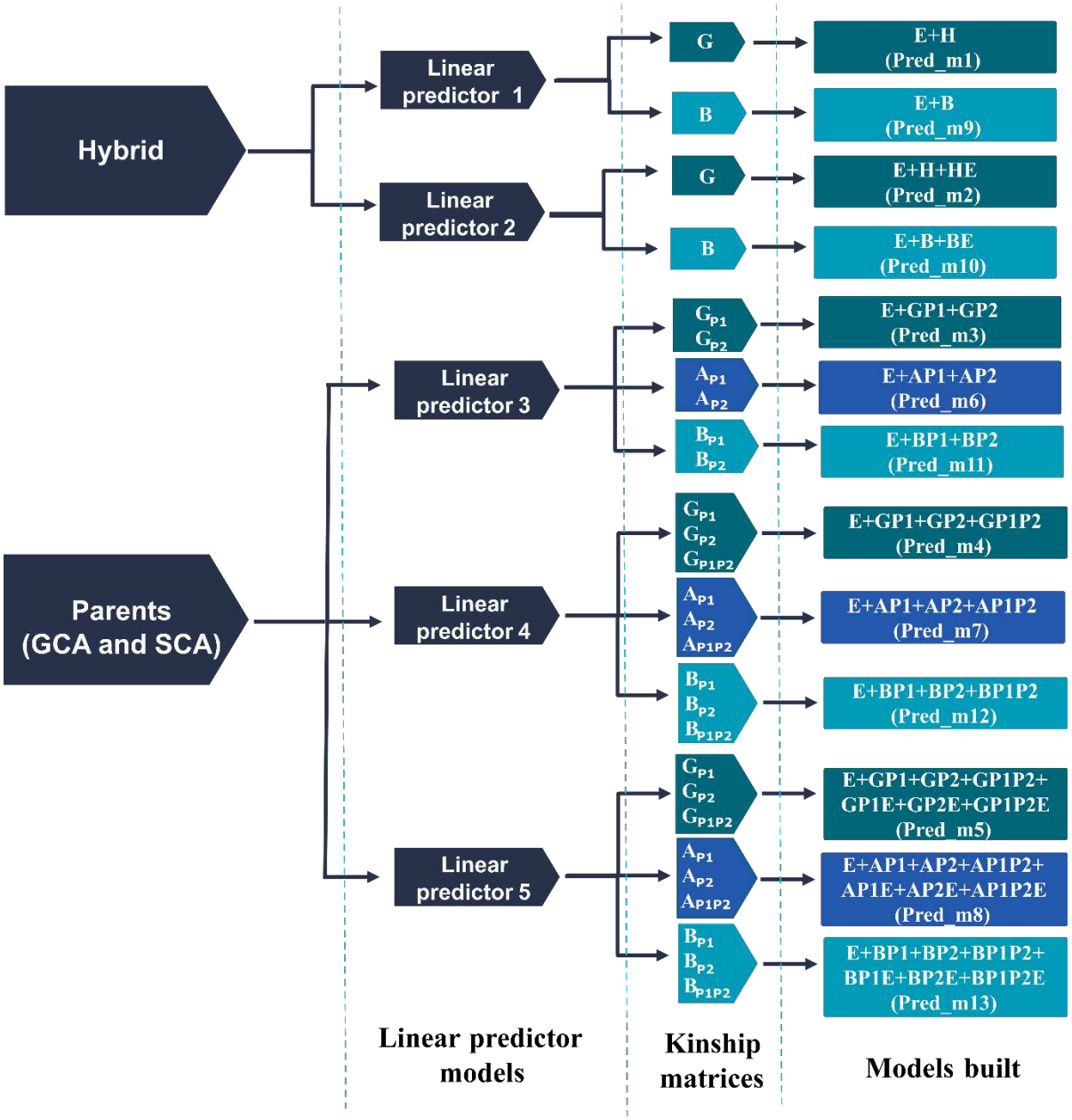


Figure 2. Training models built from five linear predictor models and three kinship matrices (G, A, and, B). The first two linear predictors (1, and 2) were implemented to model the hybrid performance. The other three linear predictors (3 to 5) considered the genomic and/or the pedigree information of the parents via the general and specific combining ability (GCA and SCA) terms. Where G is the genomic relationship matrix of the hybrid, G_{P1} is the genomic relationship matrix of the P1 group; G_{P2} is the genomic relationship matrix of the P2 group; G_{P1P2} is the $G_{P1}{}^{\circ}\, G_{P2}$, A_{P1} is the pedigree relationship matrix of the P1 group, A_{P2} is the pedigree relationship matrix of the P2 group, A_{P1P2} is the $A_{P1}{}^{\circ}\, A_{P2}$, B is the single-step relationship matrix of the hybrid; B_{P1} is the single-step relationship matrix of the P1 group, B_{P2} is the single-step relationship matrix of the P2 group; B_{P1P2} is the $B_{P1}{}^{\circ}\, B_{P2}$, and; $^{\circ}$ is the Hadamard or Shur product.

5.5 CROSS-VALIDATION AND B MATRIX OPTIMIZATION

For evaluating the predictive ability two cross-validation scenarios were implemented: 1) CV2, which consists of predicting the performance of hybrids that have been tested in some environments but not in others (incomplete field trials), and; 2) CV1, which consists of predicting hybrids that have not been tested at any of the observed environments. Furthermore, we used two models of cross-validation represented in **Figure 3**. For the models that used G and A matrices (eight models), the traditional fivefold cross-validation scheme was implemented (**Figure 3A**). This consists of a random five-fold (represented in **Figure 3** by the hexagon), where four folds were used for model training and the remaining fold was used as testing set for evaluating the predictive ability of the different models. In this case, around 20% of the hybrids were predicted using the 80% of the observed phenotypes for each one of the five replicates.

For the models including the B matrix (five models M9-M13), in principle, a slightly different cross-validation was implemented and it is shown in **Figure 3B**. The same five fold partition was considered; however, an optimization procedure to select *w* was implemented. For this three-fold were used for training the population, and the fourth fold was used to find the optimum *w* value that returns the highest correlation between predicted and observed values for the fourth fold. The last fold was used for testing the predictive ability of these models (**Figure 3B**). Thus, around 60% of the phenotypes was used for model training, 20% to find the optimum value for *w*, and 20% to test the model predictive ability given the optimum value for *w* computed with the fourth fold.

A sequence of 21 values for *w* from 0 to 1 (with increments of 0.05) were considered for constructing the $B = wA^* + (1 - w)G^*$ matrices (A^* can be pedigree matrix of the hybrid, P1 or P2, and, G^* can be genomic matrix of the hybrid, P1 or P2). Using the fourth fold, after the model was trained with the three folds predicting the fourth one, the correlation between predicted and observed values for the fourth fold was computed and the *w* value that returned the highest correlation was chosen to be used in the B matrix for predicting the fifth fold. The correlation between predicted and observed values was computed on a trial basis (within environments), thus the choice of the *w* value can be determined using two approaches: *1*) select the value of *w* optimum across environments, and; *2*) select the value *w* optimum for each environment. For the first case, the weighted average correlation across the three environments was computed following Tiezzi et al. (2019) (see equation 8 for more details), then the *w* value returning the highest average correlation was used for predicting the fifth fold. For the second

case, the *w* value that returned the highest correlation between predicted and observed values for each environment was selected to predict the fifth fold, thus up to three different *w* values were used for predicting the fifth fold, one for each one of the environments.

This procedure was repeated 10 times for each one of the cross-validation scenarios (CV1 and CV2) and applied to both data types (PGds and GOds). The statistical analyses were performed in R statistical package version 4.2.1 (R core team 2022), using the package BGLR, version 1.1.0 (Pérez and De Los Campos 2014).

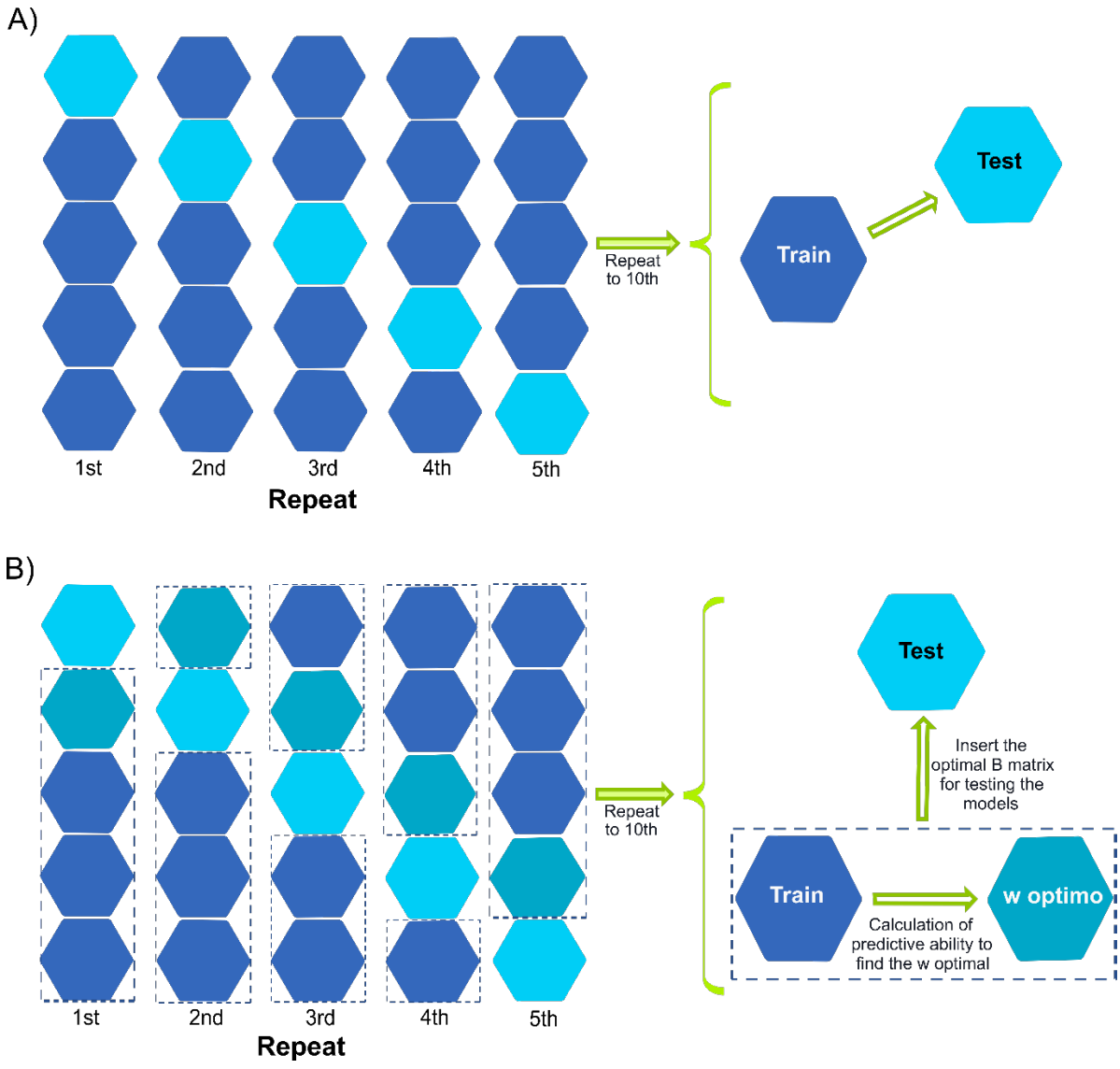


Figure 3. Representation of two cross-validation schemes with a five-fold assignation: (A) traditional where four-folds are used for model training to predict the fifth-fold, and (B) where three folds are used for model training and the fourth fold is implemented to conduct the optimization of the *w* value for those models involving the single-step matrix (B). Once the optimum values is chosen (across or within the environments of the fourth fold), the fifth-fold

is predicted. The columns represent the repetitions number, the hexagons represent the folds, the blue hexagons correspond to the folds implemented as training set, the blue-green hexagon correspond to the fold used to conduct of the *w* hyper-parameter (optimization) when the single-step model is implemented, and the cyan hexagon is the testing fold (fifth fold).

5.6 PREDICTIVE ABILITY WITHIN AND ACROSS ENVIRONMENTS

The predictive ability was assessed in trail basis computing the Pearson's correlation between predicted and observed values within the same environment. The average correlation across environments was calculated by accounting for uncertainty and the sample size of the environments (Tiezzi et al. 2017).

$$r_{\phi} = \frac{\sum_{j=1}^{J} \frac{r_j}{V(r_j)}}{\sum_{j=1}^{J} \frac{1}{V(r_j)}} \qquad (8)$$

where r_j is the Pearson's correlation between predicted and observed values at the j[th] environment, $V(r_j) = \frac{1-{r_j}^2}{n_j-2}$ corresponds to the sampling variance and n_j is the number of j observations at the j[th] environment. Furthermore, the mean squared prediction error (MSPE) across-environment was also computed.

$$MSPE = \frac{\sum_{j=1}^{J}\sum_{i=1}^{I}(y_{ij}-\hat{y}_{ij})}{Total\ phenotypes} \qquad (9)$$

where y_{ij} is phenotype observed of hybrid i^{th} at the j^{th} environment, and, $\hat{y}_{ij}$ is the corresponding predicted value.

5.7 SELECTION OF THE TOP HYBRIDS

For selecting the top hybrids, the complete dataset was used, and the following base model was considered:

$$y_{ijkl} = \mu + E_j + r_{jk} + b_{l(jk)} + g_i + gE_{ij} + \varepsilon_{ijkl} \qquad (10)$$

where y_{ijkl} is the phenotypic value of the i^{th} hybrid within environment j^{th}, in the block l^{th} of replicate k^{th}, μ is the general mean, E_j is the fixed effect of environment j, r_{jk} is the fixed effect of replication k within environment j, $b_{l(jk)}$ is the random effect of the block l^{th} within replication k in the environment j, with $b_{l(k)} \sim N(0, \sigma_b^2)$, and σ_b^2 as the block variance, g_i is

the random effect of hybrid i^{th}, with $g_i \sim N(0, \sigma_g^2)$, and σ_g^2 as the genetic variance, gE_{ij} is the random effect of the hybrid-by-environment interaction, with $gE_i \sim N(0, \sigma_{ge}^2)$, and σ_{ge}^2 as the hybrid-by-environment interaction variance, and, ε_{ijkl} is the residual random effect with $\varepsilon_{ikl} \sim N(0, \sigma_\varepsilon^2)$, with σ_ε^2 as the residual variance. Another important difference is in the covariance matrices of the random effects. Here, we considered the following relation:

$$\text{var}\left(\begin{bmatrix}\mathbf{g}\\ \mathbf{b}\\ \varepsilon\end{bmatrix}\right) = \begin{bmatrix}\boldsymbol{\Sigma}_\mathbf{g}\otimes\mathbf{B}\ \textbf{or}\ \mathbf{I}_i & 0 & 0\\ 0 & \sigma_\text{b}^2\mathbf{I}_{ijk} & 0\\ 0 & 0 & \boldsymbol{\Sigma}_\mathbf{r}\otimes\mathbf{I}_n\end{bmatrix} \qquad (11)$$

where $\boldsymbol{\Sigma}_\mathbf{g}$ and $\boldsymbol{\Sigma}_\mathbf{r}$ are square covariance matrices (3x3) of $\mathbf{g}$ and ε vectors, respectively. Here, we assume for $\text{var}(\mathbf{b})$ a diagonal structure ($\sigma_\text{b}^2\mathbf{I}_{ijk}$) and tested several structures for $\boldsymbol{\Sigma}_\mathbf{g}$ and $\boldsymbol{\Sigma}_\mathbf{r}$ (**Table 1**). We tested these structures in different models, considering the B matrix optimized or not adding a relationship matrix in the models (**Table 1**). We tested these structures in different models, considering the B matrix optimized or not adding a relationship matrix in the models (**Table 1**), note that, in this table, we built six different models (M1 to M6) using different covariance structures, models different that we used for the genomic selection. As our focus on this part is the marginal BLUP's prediction, we built the B matrix of the hybrids, using the w value optimum estimated across environments. For to find the w optimum across-environment, we applied the same strategy proposed in **Figure 3B** in our base model (eq. 10). We selected the best-fit model using the Akaike Information Criterion (AIC) (Akaike, 1974), and prediction error variance (PEV) that was obtained from the diagonal elements of the inverted matrix of coefficients. The statistical analyses were performed on R environment, version 4.2.1 (R Core Team 2022), using the package ASReml-R, version 4.1 (Butler et al. 2018).

Table 1. Covariance structures of the $\boldsymbol{\Sigma_g}$ and $\boldsymbol{\Sigma_r}$ matrices in different models (M1 to M4) with and without the **B** matrix.

B matrix	Name	Variance structure	Acronym	M1		M2		M3		M4		M5		M6	
				G	R	G	R	G	R	G	R	G	R	G	R
Without (P)	Identity variance	$\mathbf{I}_n\sigma_\varepsilon^2$	IDV		●				●				●		
	Compound symmetry	$(\sigma_g^2\mathbf{J} + \sigma_{ge}^2\mathbf{I_j}) \otimes \mathbf{I}_i$	CS	●		●									
	Diagonal	$\oplus_1^j \sigma_{\varepsilon/g}^2\mathbf{I}_j$	DIAG				●	●		●	●				●
	Heterogeneous compound symmetry	$\{\sqrt{\mathbf{D}}[\mathbf{I_j} + \rho(\mathbf{J} - \mathbf{I_j})\sqrt{\mathbf{D}}\} \otimes \mathbf{I}_i$	CSH									●		●	
With (B)	Identity variance	$\mathbf{I}_n\sigma_\varepsilon^2$	IDV		●				●				●		
	Compound symmetry	$(\sigma_g^2\mathbf{J} + \sigma_{ge}^2\mathbf{I_j}) \otimes \mathbf{I}_i$	CS	●		●									
	Diagonal	$\oplus_1^j \sigma_{\varepsilon/g}^2\mathbf{I}_j$	DIAG				●	●		●	●				●
	Heterogeneous compound symmetry	$\{\sqrt{\mathbf{D}}[\mathbf{I_j}\rho(\mathbf{J} - \mathbf{I_j})\sqrt{\mathbf{D}}\} \otimes \mathbf{I}_i$	CSH									●		●	

σ_ε^2, σ_g^2 and σ_{ge}^2 : variance components of residual, genetic and genotype-by-environment interaction effects; $\sigma_{\varepsilon/g}^2$: residual or genetic variance components; B: B matrix; **I** : identity matrix whose dimension depends on the effect it refers; ρ: genetic correlation coefficient between environments; e: number of environments; n is the number of observations; ° is the Hadamard or Shur product; ⊕ is the direct sum.

Generalized heritability (H^2) for FUMO was calculated according to Cullis et al. (2006):

$$H^2 = 1 - \frac{\bar{\Delta}_g}{2\sigma_g^2} \qquad (12)$$

where $\bar{\Delta}_g$ is the mean prediction error variance of BLUPs pairwise.

The concordances of the top hybrids ranked between models were calculated using the Kappa coefficient (K) (Cohen 1960). Similar to correlation coefficients, this coefficient can range from -1 to +1, where 0 represents the no concordance of ranking between models, and 1 represents perfect concordance (McHugh 2012), given by:

$$K = \frac{A-C}{D-C} \qquad (13)$$

where: *A* is the number of matching selected hybrids, *C* is the number of selected hybrids due to chance ($C = bD$, where b is the selection percentage) and *D* is the number of selected hybrids (where D is equal to 20, 40, or 59 hybrids). The Kappa coefficient result can be interpreted as: values ≤ 0 indicating no concordances between models, 0.01 to 0.20 as none to slight concordances, 0.21 to 0.40 as fair concordance, 0.41 to 0.60 as moderate concordance, 0.61 to 0.80 as substantial concordance, and 0.81 to 1.00 as an almost perfect concordance (McHugh 2012).

The selection gain in percentage (SG%), from 20, 40, or 59 top hybrids ranked, were obtained using the following expression:

$$\text{SG\%} = \left(\frac{\bar{g}_s - \bar{g}_o}{\bar{y}}\right) * 100 \qquad (14)$$

where $\bar{g}_s$ is the mean BLUP of the selected hybrids and $\bar{g}_o$ is the mean BLUP of the evaluated population, and $\bar{y}$ is the phenotypic mean. The SG% was calculated considering a selection percentage of 5% (20 hybrids), 10% (40 hybrids), and 15% (59 hybrids).

6 RESULTS

Our results and discussion were divided into the intermediate and final stages of a maize breeding program for FUMO.

6.1 MAIZE BREEDING PIPELINE: INTERMEDIATE STAGES

6.1.1 Optimization of the B matrix

Based on the two approaches (within and across environments) used to find the optimal B matrix, the predictive ability improved for all the tested models for both prediction scenarios (CV1 and CV2) and both datasets (GOds and PGds) (**Figure 4**) when selecting *w* within environments. in GOds the average predictive ability ranged from 0.11 to 0.21 in CV1 and, 0.09 to 0.18 in CV2 selecting the *w* value across environments (**Figure 4A**). However, when used the value optimum *w* value nested in the environment the average predictive ability increased, ranging from 0.18 to 0.26 in CV1 and 0.14 to 0.21 in CV2 (**Figure 4A**). The model that showed the best predictive ability, for the two approaches used and prediction scenarios in this dataset was E+B+BE (**Figure 4A**).

For PGds using the *w* optimum within environments the average predictive ability ranged from 0.20 to 0.32 and 0.12 to 0.21 for CV1 and CV2, respectively (**Figure 4B**). These estimates were higher than those found when optimizing *w* across environments, to both prediction scenarios, and these ranged from 0.05 to 0.15 and 0.02 to 0.17 for CV1 and CV2 respectively (**Figure 4B**). The model that showed the highest predictive ability for CV1 using *w* optimum across environments was E+BP1+BP2+BP1P2+BP1E+BP2E+BP1P2E. Nonetheless, when used *w* optimal nested in environments the highest predictive ability was shown with E+B+BE model (**Figure 4B**). These results indicate that the manner for computing the optimized B matrix influences the election of the best prediction model. For CV2, the model that showed the best predictive ability, for the two approaches, within and across environments, was E+BP1+BP2+BP1P2+BP1E+BP2E+BP1P2E. Due to the results presented here, we used the w optimum value nested within environments in the linear predictor models that used the B matrix in the results showed below.

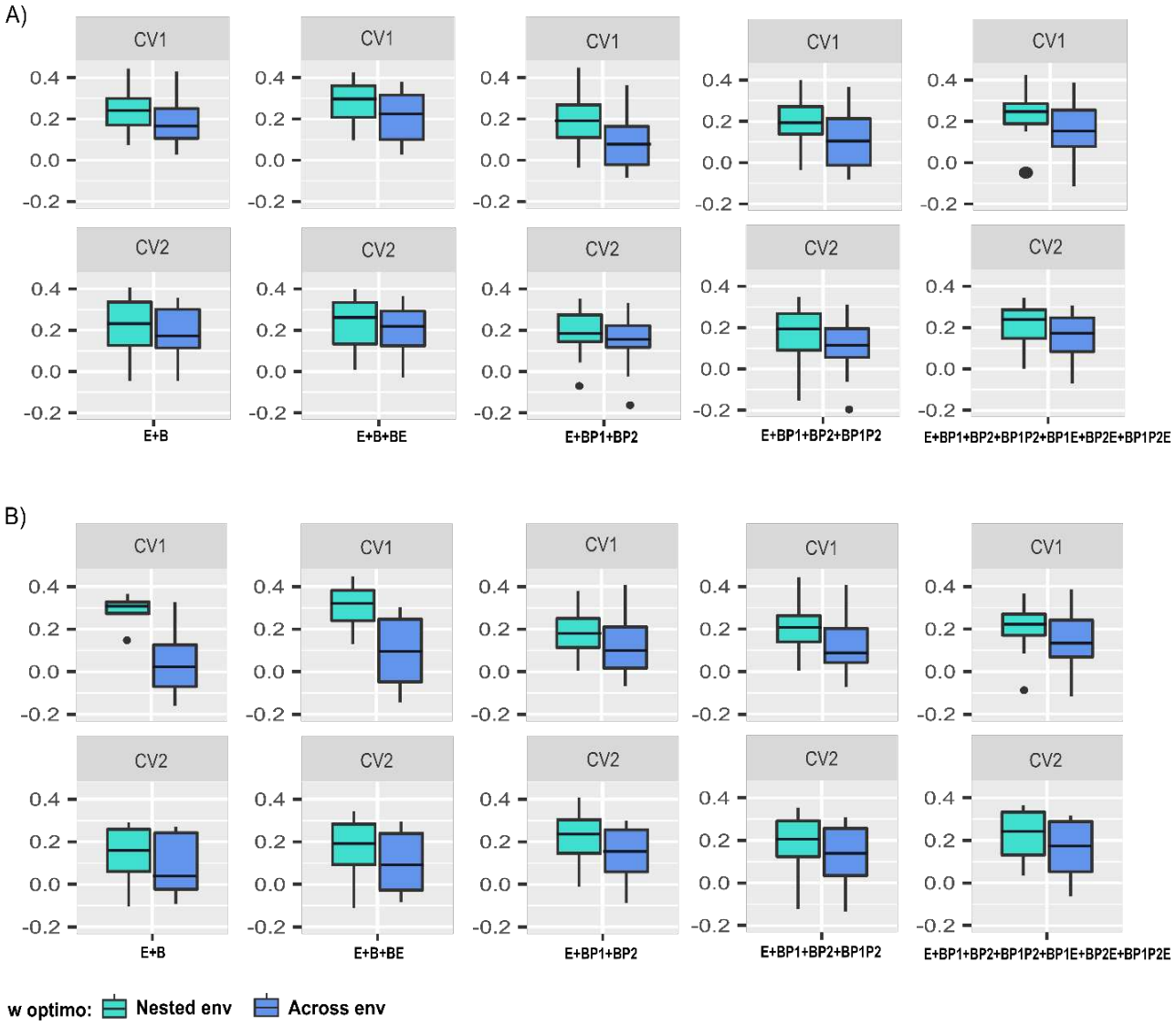


Figure 4. Boxplot of the predictive abilities of five linear predictor models in two prediction scenarios (CV1 and CV2), using two approaches to find the optimum B matrix. In blue color are indicated the results from selecting the *w* optimum value across environments, while in green the results from selecting the *w* optimum value nested in environment for: A) the GOds (genomic-only dataset), and; B) the PGds (pedigree and genomic dataset).

6.1.2 Prediction models

Predictive abilities varied considerably through environments for both prediction scenarios (CV1 and CV2) and both datasets (**Figure 5**). Overall, for all linear models, there was an improvement in the correlation across environments and in the MSPE when the B matrix was inserted (**Figure 5**). Note that, independently of whether we used all individuals (PGds) or not (GOds) in the training population, the use of the B matrix improved the predictive capacity

of the linear predictor models when compared to these using the G matrix (**Figure 6**). In general, the models with the best ability predictive, for each environment and across environments, considered the interaction between the genetic effect and the environmental effect (**Figure 5**).

CV1 Scheme

In the scenario that predict untested hybrids in evaluated environments, the predictive ability across the environments between the prediction models to GOds ranged from 0.11 to 0.28 (**Figure 5**). In this dataset, the best prediction model for E1 was E+H+HE (r = 0.139). In the E2 and E3, the models with the best predictive abilities were E+B (r = 0.436) and, E+BP1+BP2+BP1P2+BP1E+BP2E+BP1P2E (r = 0.325), respectively (**Figure 5**). However, the E+B+BE showed the best predictive ability across environments with a 0.280 correlation, and low MSPE (3.399). In the PGds, the predictive ability across the environments ranged from 0.050 to 0.321 (**Figure 5**). The models with the B matrix showed the best predictive ability for all environments. For PGds, the model with the best predictive ability across the environments also was the E+B+BE with a correlation of 0.321, and MSPE of 3.364 (**Figure 5**). Note that, if we compare the model E+B+BE, using GOds and PGds, we can observe an improvement in the predictive ability across the environments when using the dataset with all hybrids evaluated.

CV2 Scheme

In the scenery that predicts the performance of incomplete environments, the predictive abilities were lower than the CV1 scheme (Figura 3). For GOds, the predictive ability across the environments ranged from -0.027 to 0.229 (**Figure 5**). For all environments, the model with the B matrix showed the best predictive ability, however, different prediction models were considered for each environment (**Figure 5**). The E+B+BE showed a higher correlation across the environments, with 0.229, and a low MSPE (3.447) (**Figure 5**). In the PGds, the predictive ability across the environments ranged from -0.042 to 0.223. The models with the higher correlation for E1 and E2 were E+B (r = 0.107) and, E+B+BE (r = 0.343), respectively. For the E3, E+AP1+AP2+AP1P2+AP1E+AP2E+AP1P2E showed a higher correlation, with 0.298 (**Figure 5**). However, for prediction across the environments, the E+BP1+BP2+BP1P2+BP1E+BP2E+BP1P2E showed a higher correlation (r = 0.233). In this prediction scenario also observed that the E+BP1+BP2+BP1P2+BP1E+BP2E+BP1P2E model in GOds, improved the predictive ability across the environments when inserting more individuals in the training model, using the PGds. Note that, for this model, the correlation increased by 8.6% when used PGds (**Figure 5**).

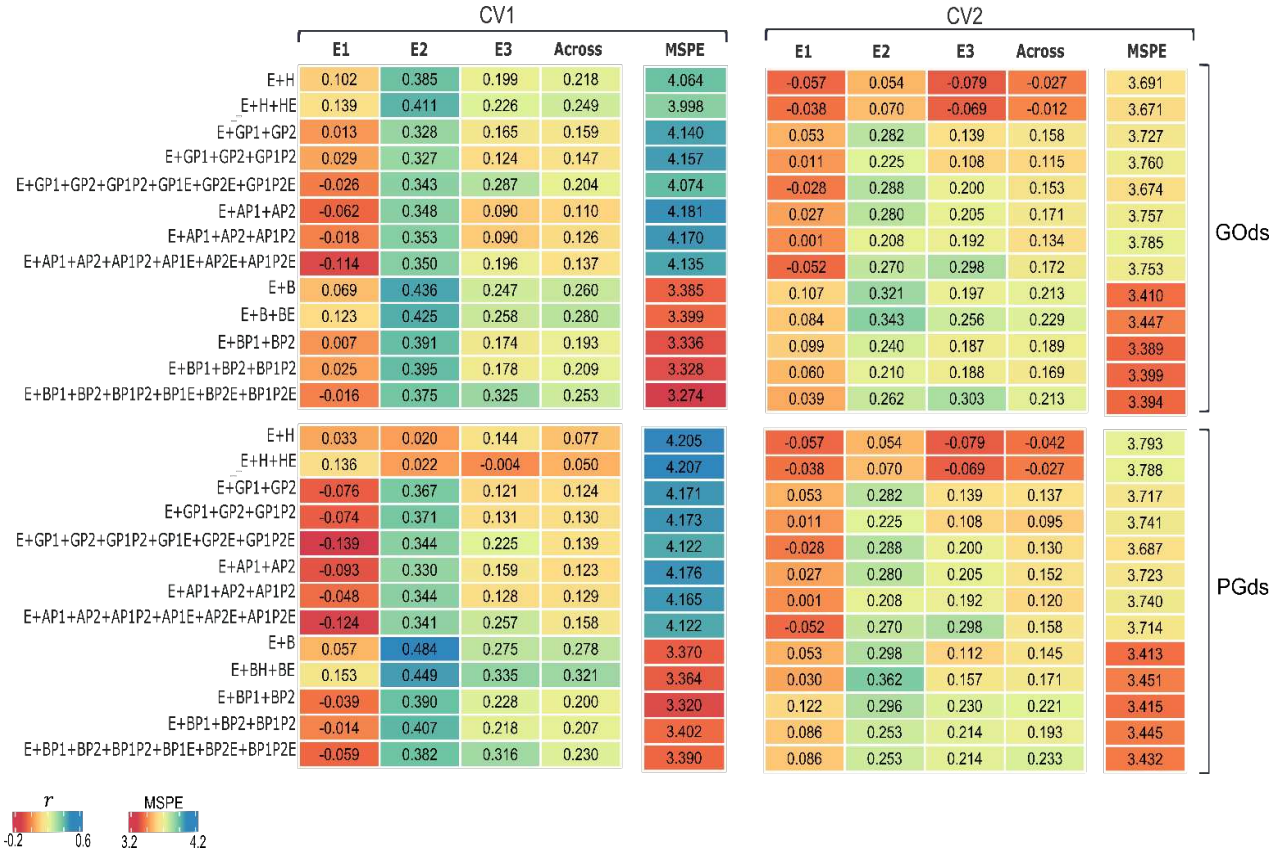


Figure 5. Heatmap of the predictive abilities of 13 models in two prediction scenarios (CV1 and CV2) to predict concentration of fumonisins (FUMO). Where r is the correlation between predicted and observed, for each environment (E1 to E3) and across environments; MSPE is the mean squared prediction error across environments; GOds is the genomic-only dataset and; PGds is the pedigree and genomic dataset.

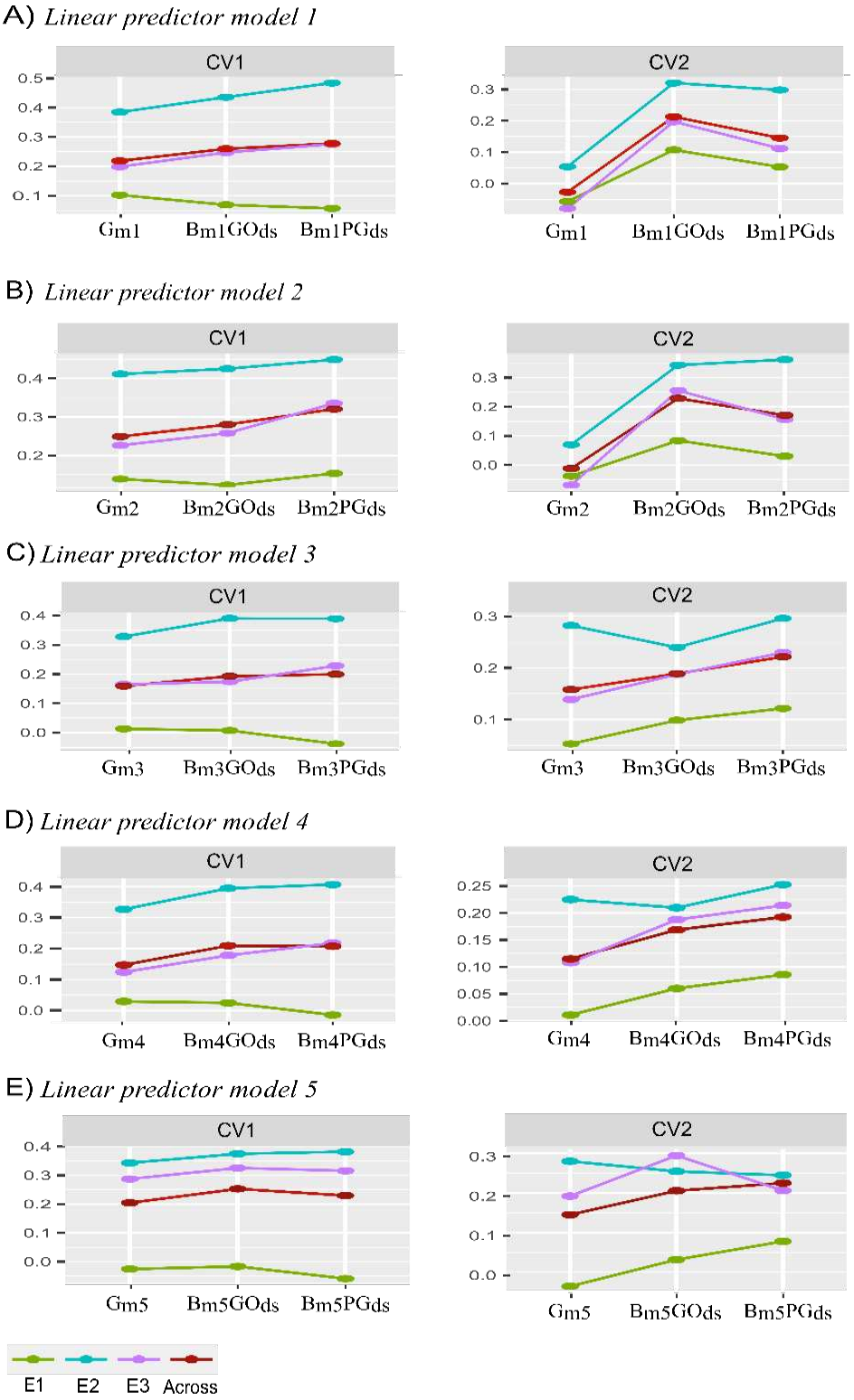


Figure 6. Correlation between predicted and observed value for the five linear predictor models tested using the G matrix, B matrix built only with individuals that have genomics information (GOds), and, B matrix built with all individuals evaluates (PGds). G_{m1}, G_{m2}, G_{m3}, G_{m4}, and, G_{m5}: genomic matrices used in the linear predictor models 1, 2, 3, 4, and 5, respectively; $B_{m1}GO_{ds}$, $B_{m2}GO_{ds}$, $B_{m3}GO_{ds}$, $B_{m4}GO_{ds}$, and, $B_{m5}GO_{ds}$: B matrices used in the linear predictor

models 1, 2, 3, 4, and 5, respectively, to GOds; and, $B_{m1}PG_{ds}$, $B_{m2}PG_{ds}$, $B_{m3}PG_{ds}$, $B_{m4}PG_{ds}$, and, $B_{m5}PG_{ds}$: B matrices used in the linear predictor models 1, 2, 3, 4, and 5, respectively, to PGds.

6.2 MAIZE BREEDING PIPELINE: FINAL STAGES

6.2.1 Selection of the top hybrid

For the selection of the most accurate model for top hybrids ranking, six models with different covariance structures were tested (**Table 2**). In **Table 2**, we added P if the model (M) does not consider the relative's information using the B optimized; and B otherwise. There was an average reduction of 30% to 40% in the PEV when we inserted the B matrix into the models (**Table 2**), indicative of higher accuracy in the genetics values prediction. In general, for all models tested, with or without the B matrix, was observed lowest values of AIC when using the diagonal structure (DIAG) for the residual effect, consequently, this structure is best for the fit of this effect (**Table 2**). Conversely, the addition of the relative information did not induce major changes in the AIC values (**Table 2**). In fact, when the relative information was put into the models there was a took longer to reach convergence, consequently, there was an increase in the log-likelihood (LogL), which provided the increase of the AIC for all the models that inserted relative information (**Table 2**). Considering the AIC low and PEV lowest as the best-fit model selection criteria, the BM4 was considered the fit model for ranking the top hybrids to the FUMO trait.

The BM4 had the residual and genetic covariance matrices structured in a diagonal form (DIAG), meaning that the model that provides more reliable results considers heterogeneous variances and the absence of covariance's between environments, this model also had the PEV value lowest. Coincidently, this fact would also be true if we did not consider relationship genetic information, i.e., between the models without B matrix, PM4 had better results. In the following, we will compare four models: PM1, BM1, PM4, and BM4. These correspond to the baseline model with and without the B matrix (BM1 and PM1), and the model with the most appropriate covariance structures, with and without the B matrix, respectively (BM4 and PM4).

Table 2. Values of Akaike (AIC) and prediction error variance (PEV) of the six models without (P) and with the genetic relationship information (B) for evaluating fumonisin concentration in grains (FUMO) in tropical maize.

Relationship Information	Model	Matrix		AIC	PEV
		G	R		
Without	PM1	CS	IDV	2314.85	0.58
	PM2	CS	DIAG	2280.51	0.58
	PM3	DIAG	IDV	2314.70	0.54
	PM4	DIAG	DIAG	2281.00	0.53
	PM5	CSH	IDV	2316.63	0.58
	PM6	CSH	DIAG	2283.99	0.57
With	BM1	CS	IDV	2386.93	0.41
	BM2	CS	DIAG	2357.90	0.33
	BM3	DIAG	IDV	2388.32	0.31
	BM4†	DIAG	DIAG	2355.10	0.31
	BM5	CSH	IDV	2390.36	0.34
	BM6	CSH	DIAG	2356.91	0.34

†: Best-fit model selection.

In **Table 3**, note that when we used the across-environments model (PM1 and BM1) the genotype-by-environment interaction (GEI) captured most of the phenotypic variation, compared to the genetic main effects (Supplementary Figure 1). Consequently, the estimates of the generalized heritability of these models tended to be zero. Conversely, when using the M4 models (PM4 and BM4), the genetic effects were nested in each environment and there was not an explicit GEI effect (**Table 3**). In other words, in this case, nested in the environment models provide better visualization of the genetic variance and genetic parameters, which was almost imperceptible using M1. Moreover, high estimates were found for generalized heritability when using PM4 and HM4, thus showing a favorable scenario for the selection of the tropical hybrid. However, when inserting genetic relationship information, we observed an increase of 13% in the estimate of this parameter (**Table 3**).

Table 3. Variance components (VC) and genetic parameters (GP) estimated by models M1 (PM1 and BM1) and M4 (PM4 and BM4) models using only phenotypic (P) data and using genetic relationship information (B) for fumonisin in grains (FUMO) in maize tropical.

Model	VC/GP		Relationship Information	
			Without (P)	With (B)
M1 (Across)	**σ^2g**		0.12	0.16
	σ^2ge		1.99	2.20
	σ^2b		0.01	0.04
	σ^2 ε		2.60	3.64
	H^2		-	-
	SG	**5%**	9.63	26.00
		10%	10.74	21.16
		15%	10.62	17.89
M4 (Nested)	**σ^2g**	**E1**	2.32	4.44
		E2	2.06	2.03
		E3	1.94	1.67
	σ^2b		0.06	0.06
	σ^2 ε	**E1**	1.34	1.82
		E2	2.75	4.00
		E3	3.64	4.72
	H^2		0.75	0.85
	SG	**5%**	11.46	27.10
		10%	11.41	22.76
		15%	10.45	19.72

σ^2g: genotypic variance; σ^2ge: genotype x environment interaction variance; σ^2 ε: residual variance; E1: environment 1; E2: environment 2; E3: environment 3; H^2: generalized heritability; SG: selection gain in percentage of 5% (20 hybrids), 10% (40 hybrids), and 15% (59 hybrids).

In the selection of the top hybrids, considering the top 5%, 10%, and 15% in the ranking, there was not a 100% coincidence between the selected hybrids of the compared four models (**Figure 7B**, **7C**, and **7D**). The coincidences were lower between P and B models, i.e. models without and with the relative's information (**Figure 7B**, **7C**, and **7D**). Note that in **Figure 7A**, 20% of hybrids ranked to the base model (PM1) were not ranked by any other model. Likewise, the PM4, BM1, and BM4 models also selected hybrids that were not ranked by any other model (**Figure 7A**). Consequently, the difference in the ranking between these models provided

different selection gains (**Table 3**). Note that, the selection gains doubled when the B matrix was inserted in the models (**Table 3** and Supplementary Figure 2). Furthermore, the best-fit model (BM4) showed the highest gain considering the 5%, 10%, and 15% of the top hybrids selected. Showing that by using B matrix plus the proper modeling of the covariance structure the breeder could reach better results in the selection of the top hybrid.

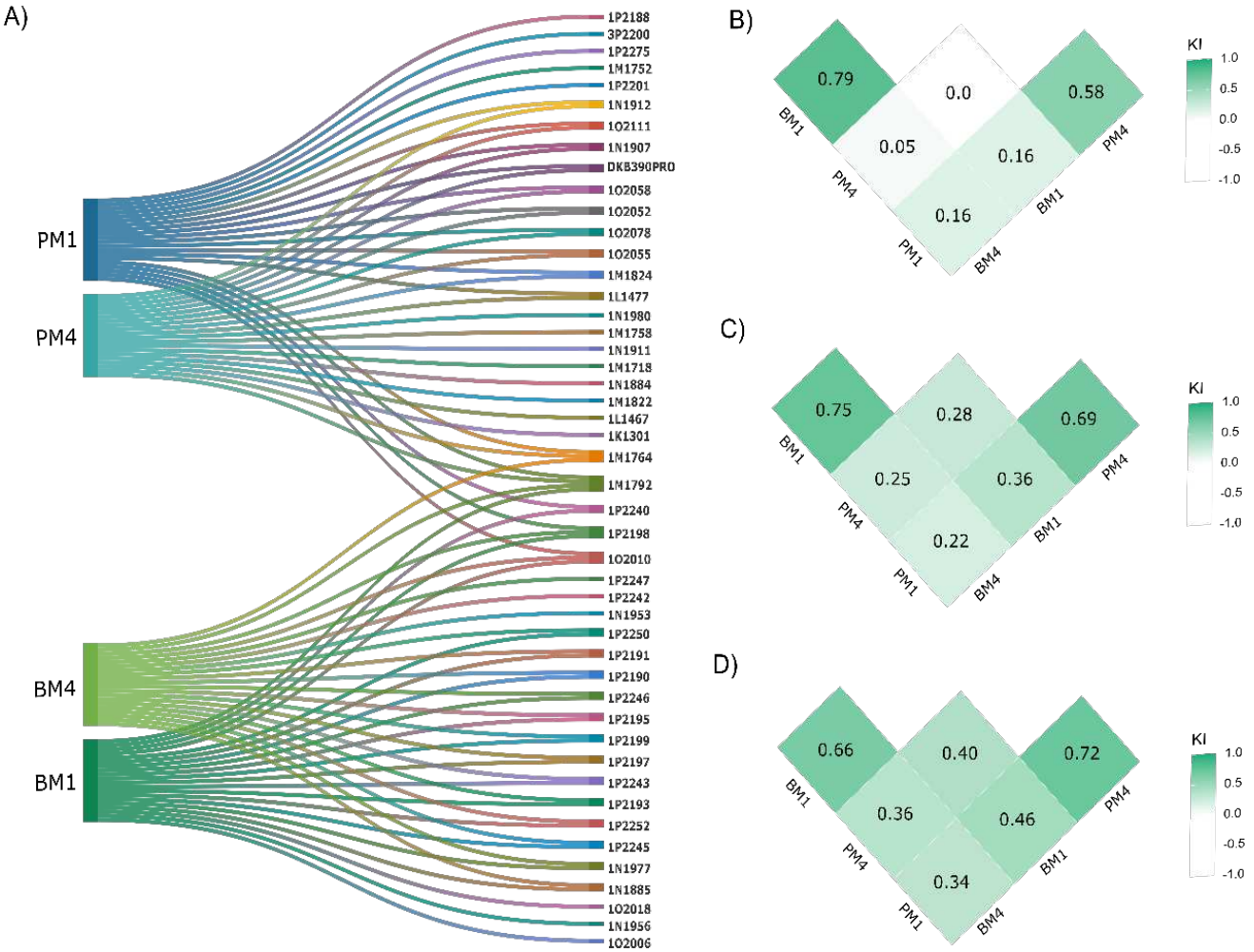


Figure 7. A) Ranking of 20 top hybrids (5% selected) by PM1, PM4, BM1, and BM4 models; and, Kappa Coefficient (KC) to PM1, PM4, BM1, and BM4 models considering: B) 5% top hybrids ranked; C) 10% top hybrids ranked; D) 15% top hybrids ranked.

7 DISCUSSION

7.1 MAIZE BREEDING PIPELINE: INTERMEDIATE STAGES

7.1.1 B matrix optimization

The single-step approach (B matrix) combines genomic and pedigree information using a weighting factor (w). The w represents the fraction of the total additive variance that is not captured by the markers (Velazco et al. 2019). To find the best way to build the B matrix, researchers assume different values for w, and the value that reaches the best predictive ability is chosen for the building of this matrix (Velazco et al. 2019; de Oliveira et al. 2020). Here, we proposed to find w through the fourth fold of the cross-validation to build the B matrix optima, before of used the fifth fold to find the best prediction model. Using this approach, we obtain to optimize the B matrix reaching satisfactory results. This suggests that using the fourth fold for B matrix optimization can be a strategy to guide the researcher to find the w optimal.

Studies showed that the w optimal varies when different characteristics are evaluated (Liu et al. 2011; Ashraf et al. 2016; Velazco et al. 2019). Nonetheless, it was never discussed under the behavior change of this factor under the environmental influence. Studies have focused on selecting the w-value optimal across-environments so far, i.e., an optimal value in common for all environments. However, individuals have different genetic responses when changing environmental conditions, with non-static variation (eg: temperature, precipitation, humidity) being the main source of genotype-environment interaction for maize (Cullis et al. 2000; Kleinknecht et al. 2013; Dias et al. 2020; Krause et al. 2022). In this paper, results shown that the difference in the gene expressions of the individuals between environments can influence the optimal weighting factor. Indicating that the best strategy for the B matrix optimization is the choice of w optimo nested in the environment.

7.1.2 Prediction models

Over the decades, breeders have searched for strategies to reduce the fumonisin content in grains (Duvick 2001; Eller et al. 2008; Pádua et al. 2016; Santiago et al. 2020; Butoto et al. 2022). The USFDA Center for Food Safety and Applied Nutrition, (2001) recommends a maximum of 2 to 4 mg kg^{-1} of contamination of FUMO in corn products. In Brazil, the contamination limit is 2 mg g^{-1} of these toxins in grains (Agência Nacional de Vigilância 2011).

Limit that can prevent the export and national marketing of maize batches, consequently causing large losses to producers. Therefore, it is paramount to develop cultivars resistant to FUMO beyond high grain yield. However, the phenotyping of this trait is laborious and expensive (Bush et al. 2004). Hence, genomic selection is an important strategy for breeding this trait. Comparing genomic selection and phenotypic selection traditional for FUMO, Butoto et al., (2022), noted that both selections performed similarly. Nonetheless, the authors emphasize that genomic selection has the potential to be more efficient than phenotypic selection, due to cheaper and faster genotyping methods, when compared to the phenotyping of this trait.

Until the moment, there is no research using the combination of genomic and pedigree information in a relationship matrix for the genomic prediction of the FUMO trait. In addition, research with different maize traits (e.g. grain yield, number of ears, anthesis date) shows the efficiency of genomic prediction when using the B matrix (Crossa et al. 2013; de Oliveira et al. 2020). Our results shown that the use of the B matrix improved the predictive ability in all tested linear models, for both prediction scenarios and both datasets. Besides that, the use of a single-step approach was superior to the G matrix, regardless of whether or not we included more individuals in the training population. As some markers may not be in linkage disequilibrium with QTLs when combining the A matrix and G matrix in a relationship matrix, the pedigree information may have contributed to capturing associations between causative alleles due to common ancestral identity, improving predictions models (Velazco et al. 2019). Moreover, all linear predictor models tested reached the lowest MSPEs when using the B matrix, result is consistent with those obtained by Velazco et al., (2019). The comparison of prediction models through the MSPE minimization has been recommended due to this statistic considers both the precision and bias of the models (Vitezica et al. 2011; González-Recio et al. 2014).

Mainly in animal breeding, the use of matrix B in genomic prediction has been widely discussed (Martini et al. 2018; Macedo et al. 2020; Mäntysaari et al. 2020; Masuda et al. 2021). However, in the plant breeding, this methodology has not yet become popular (de Oliveira et al. 2020). Our results shown the advantage of using the single-step approach in the intermediate stage of a breeding program, in two different contexts. The first context is when the program does not have all individuals genotyped (PGds), for example, due to new materials inserted in the pipeline. The B matrix enables the construction of the relationship matrix incorporating all individuals, increasing the selection intensity and providing an improvement in the predictive capacity of the models. As a strategy, de Oliveira et al. (2020) used the B matrix in multi-trait

multi-environment genomic prediction models, due to the lack of genotypic information for some maize hybrids evaluated.

The second context consists of companies that obtain all individuals genotyped and with pedigree information (GOds). Our results show that by combining both information, the predictive ability of the models increases, helping to discard the hybrids most susceptible to FUMO for the advancement of the next stages of the program. In this second context, other researchers have also observed that combining genomic and pedigree information optimizes genomic prediction for complex traits (Crossa et al. 2013; Basnet et al. 2019; Velazco et al. 2019). Breeding programs also are interested in the development of superior cultivars that get a favorable response for a diverse of environmental conditions (Jarquin et al. 2021). Our results was consistent with those obtained by several researchers that observed better predictive ability in models that consider that genetic and environmental interaction effect (Jarquin et al. 2014, 2021; Lado et al. 2016; Basnet et al. 2019; Khanna et al. 2022). Showing the importance of this effect for accurate prediction of complex traits.

7.2 MAIZE BREEDING PIPELINE: FINAL STAGES

7.2.1 Selection of the top hybrid

For the launch of a cultivar on the market, each country obtains specific laws to certify the agronomic value of the candidate material to be released. For this finality, the Ministry of Agriculture, Livestock, and Supply, (MAP) Brazilian requires the cultivation and use value trials (VCU's). One of the VCU's requirements is that the candidate cultivar has at least three years of field evaluation. For an accurate selection of the superior genotypes that will be advanced in the VCU's trials, it is extremely important to choose the appropriate model for the accurate BLUPs prediction. An efficient strategy to increase the reliability of genetic selection is the use of modeling the covariance structure of the genetic and non-genetic effects of the statistical model (Rocha et al. 2019; Melo et al. 2020; Chaves et al. 2022). Another strategy that reduces BLUPs prediction biases is the inclusion of a relationship matrix in the models (Piepho et al. 2008). Therefore, we used both approaches with the main aim of auxiliary the breeder to select the target cultivar of the final stages with high accuracy. The covariance structures modeling already have been used in research aimed at corn breeding, with a focus on the fumonisin reduction (Pádua et al. 2016). However, so far, a relationship matrix and covariance structures modeling has not been used with a focus on BLUPs prediction for this trait.

In the context of choosing the best-fit model, that is, one that provides more accurate parameter estimates, the AIC is widely used (Faveri et al. 2015; Melo et al. 2020; Chaves et al. 2022). However, when the aim is the BLUPs prediction, it is appropriate to use other statistics to support the best model choice. The PEV is considered a measure of the accuracy of a statistic model's prediction (Adebola et al. 2020). Therefore, we consider both statistics (AIC and, PEV) to select the best-fit model. The diagonal covariance structure proved to be more suitable for modeling both tested effects (residual and genetic), regardless of whether we inserted the B matrix in the models. The structure assumes heterogeneity of variances, showing the environmental difference and the imbalance of the genotypes evaluated over the years.

In addition, results showed that the use of the covariance structures modeling can influence the top hybrids ranking when comparing models PM1 and PM4. Result is consistent with those obtained in other researches (Souza et al. 2021; Chaves et al. 2022). However, with the inclusion of the B matrix plus the choice of the best model, the accuracy of BLUPs prediction increased considerably, allowing the ranking of the best hybrids to be highly reliable. In fact, not using these two combined strategies can induce the breeder to choose inferior genetic materials and discard superior materials, as shown by the Kappa coefficient. Consequently, this result reflects in the selection gain of the trait. We can observe smaller selection gains when using models PM1, PM4, and BM1. In this sense, the use of the B matrix combined with the modeling of genetic and non-genetic effects is an efficient strategy to help the breeder to select with more precision the top hybrids that will go to the next stages of the VCU's trials. Due to the complexity of the evaluated trait, and the excellent results achieved, this strategy has the potential to be extended to other complex maize traits, aiming at the selection of the final stages breeding program.

8 CONCLUSION

The best way to find the optimal B matrix is to select w within environments to increase the predictive ability of the prediction models. Using the single-step approach we can improve the predictive abilities of the linear predictor models for the FUMO trait, consequently, predict the hybrids untested in the intermediate stages of the program with better precision. The B matrix plus the selection of the best-fit model increases selective accuracy, consequently, the genetic gains, thus contributing to the breeder’s decision of the final stages of the breeding program.

9 REFERENCES

Adebola FB, Fasoranbaku OA, Kupolusi JA (2020) On prediction error variance to determining optimal design for two variable quadratic logistic model. Cogent Math Stat 7:1853888. https://doi.org/10.1080/25742558.2020.1853888.

Agência Nacional de Vigilância (2011) Dispõe sobre os limites máximos tolerados (LMT) para micotoxinas em alimentos (Resolução RDC no 7, de 18 de fevereiro de 2011). Diário Of [da] República Fed do Bras.

Aguilar I, Misztal I, Johnson DL, et al (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci 93:743–752.

Akaike H (1974) A new look at the statistical model identification. In: Selected Papers of Hirotugu Akaike. Springer, pp 215–222.

Ashraf B, Edriss V, Akdemir D, et al (2016) Genomic prediction using phenotypes from pedigreed lines with no marker data. Crop Sci 56:957–964.

Atanda SA, Olsen M, Burgueño J, et al (2021) Maximizing efficiency of genomic selection in CIMMYT's tropical maize breeding program. Theor Appl Genet 134:279–294.

Basnet BR, Crossa J, Dreisigacker S, et al (2019) Hybrid Wheat Prediction Using Genomic, Pedigree, and Environmental Covariables Interaction Models. Plant Genome 12:180051. https://doi.org/10.3835/plantgenome2018.07.0051.

Benin G, Storck L, Marchioro VS, et al (2013) Improving the precision of genotype selection in wheat performance trials. Crop Breed Appl Biotechnol 13:233–239. https://doi.org/10.1590/s1984-70332013000400003.

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20–25.

Beyene Y, Gowda M, Pérez-Rodríguez P, et al (2021) Application of genomic selection at the early stage of breeding pipeline in tropical maize. Front Plant Sci 12:685488.

Bhat JA, Ali S, Salgotra RK, et al (2016) Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding. 7:1–11. https://doi.org/10.3389/fgene.2016.00221.

Blacutt AA, Gold SE, Voss KA, et al (2018) Fusarium verticillioides: Advancements in understanding the toxicity, virulence, and niche adaptations of a model mycotoxigenic pathogen of maize. Phytopathology 108:312–326.

Bush BJ, Carson ML, Cubeta MA, et al (2004) Infection and fumonisin production by Fusarium verticillioides in developing maize kernels. Phytopathology 94:88–93.

Butler DG, Cullis BR, Gilmour AR, et al (2018) ASReml-R reference manual. Version 4. VSN International Ltd.

Butoto EN, Brewer JC, Holland JB (2022) Empirical comparison of genomic and phenotypic selection for resistance to Fusarium ear rot and fumonisin contamination in maize. Theor Appl Genet 2022 1:1–18. https://doi.org/10.1007/S00122-022-04150-8.

Chaves SFS, Evangelista JSPC, Alves RS, et al (2022) Application of linear mixed models for multiple harvest/site trial analyses in perennial plant breeding. Tree Genet Genomes 18:1–12. https://doi.org/10.1007/s11295-022-01576-5.

Chaves SFS, Evangelista JSPC, Trindade RS, et al (2023) Employing factor analytic tools for selecting high-performance and stable tropical maize hybrids. Crop Sci 1–12. https://doi.org/10.1002/csc2.20911.

Chen J, Wen J, Tang Y, et al (2021) Research progress on fumonisin b1 contamination and toxicity: A review. Molecules 26:5238.

Cohen J (1960) A coefficient of agreement for nominal scales. Sage Publications Sage CA: Thousand Oaks, CA.

CONAB BNFSA (2022) Acompanhamento da Safra Brasileira. Bol da Safra 2021 9:60.

Crossa J (1990) Statistical analyses of multilocation trials. Adv Agron 44:55–85.

Crossa J, Beyene Y, Semagn K, et al (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes, Genomes, Genet 3:1903–1926. https://doi.org/10.1534/g3.113.008227.

Crossa J, Pérez-rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant Breeding : Methods , Models , and Perspectives. 22:961–975. https://doi.org/10.1016/j.tplants.2017.08.011.

Cullis BR, Smith A, Hunt C, Gilmour A (2000) An examination of the efficiency of Australian crop variety evaluation programmes. J Agric Sci 135:213–222. https://doi.org/10.1017/S0021859699008163.

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11:381–393.

da Silva KJ, Guimarães CT, Tinoco SM de S, et al (2022) A genome-wide association study investigating fumonisin contamination in a panel of tropical maize elite lines. Euphytica 218:130.

de Oliveira AA, Resende MFR, Ferrão LFV, et al (2020) Genomic prediction applied to multiple traits and environments in second season maize hybrids. Heredity (Edinb) 125:60–72. https://doi.org/10.1038/s41437-020-0321-0.

Dias KOG, Piepho HP, Guimarães LJM, et al (2020) Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. Theor Appl Genet 133:443–455. https://doi.org/10.1007/s00122-019-03475-1.

Duvick J (2001) Prospects for reducing fumonisin contamination of Maize through genetic modification. Environ Health Perspect 109:337–342. https://doi.org/10.1289/ehp.01109s2337.

Eller MS, Holland JB, Payne GA (2008) Breeding for improved resistance to fumonisin contamination in maize. Toxin Rev 27:371–389.

Evangelista JSPC, Peixoto MA, Coelho IF, et al (2023) Modeling covariance structures and optimizing Jatropha curcas breeding. Tree Genet Genomes 19:1–12. https://doi.org/10.1007/s11295-023-01596-9.

Faveri J De, Verbyla PA, Pitchford WS, et al (2015) Statistical methods for analysis of multi-harvest data from perennial pasture variety selection trials. Crop Pasture Sci 947–962. https://doi.org/http://dx.doi.org/10.1071/CP14312.

Gaikpa DS, Miedaner T (2019) Genomics-assisted breeding for ear rot resistances and reduced mycotoxin contamination in maize: methods, advances and prospects. Theor Appl Genet 132:2721–2739.

Gauch Jr HG, Zobel RW (1997) Identifying mega-environments and targeting genotypes. Crop Sci 37:311–326.

González-Recio O, Rosa GJM, Gianola D (2014) Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livest Sci 166:217–231.

Haug B, Messmer MM, Enjalbert J, et al (2021) Advances in breeding for mixed cropping–incomplete factorials and the producer/associate concept. Front Plant Sci 11:620400.

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 69–83.

Heslot N, Jannink J, Sorrells ME (2015) Perspectives for Genomic Selection Applications and Research in Plants. 1–12. https://doi.org/10.2135/cropsci2014.03.0249.

Holland JB, Marino TP, Manching HC, Wisser RJ (2020) Genomic prediction for resistance to Fusarium ear rot and fumonisin contamination in maize. Crop Sci 60:1863–1875. https://doi.org/10.1002/csc2.20163.

Jarquin D, Crossa J, Lacaze X, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607. https://doi.org/10.1007/s00122-013-2243-1.

Jarquin D, de Leon N, Romay C, et al (2021) Utility of Climatic Information via Combining Ability Models to Improve Genomic Prediction for Yield Within the Genomes to Fields Maize Project. Front Genet 11:1–11. https://doi.org/10.3389/fgene.2020.592769.

Jorge K, Teixeira C, Sylvia G, et al (2022) A genome - wide association study investigating fumonisin contamination in a panel of tropical maize elite lines. Euphytica 1–12. https://doi.org/10.1007/s10681-022-03082-0.

Kamle M, Mahato DK, Devi S, et al (2019) Fumonisins: Impact on agriculture, food, and human health and their management strategies. Toxins (Basel) 11:328.

Khanna A, Anumalla M, Catolos M, et al (2022) Optimizing predictions in IRRI's rice drought breeding program by leveraging 17 years of historical data and pedigree information. Front Plant Sci 13:. https://doi.org/10.3389/fpls.2022.983818.

Kleinknecht K, Möhring J, Singh KP, et al (2013) Comparison of the performance of best linear unbiased estimation and best linear unbiased prediction of genotype effects from zoned Indian maize data. Crop Sci 53:1384–1391. https://doi.org/10.2135/cropsci2013.02.0073.

Krause MD, Dias KOG, Singh AK, Beavis WD (2022) Using large soybean historical data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors.

Lado B, Barrios PG, Quincke M, et al (2016) Modeling genotype× environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Sci 56:2165–2179.

Lanubile A, Maschietto V, Marocco A (2014) Breeding maize for resistance to mycotoxins. Mycotoxin Reduct grain Chain 37–58.

Legarra A, Christensen OF, Aguilar I, Misztal I (2014) Single Step, a general approach for genomic selection. Livest Sci 166:54–65.

Linnaeus C (1799) Species plantarum. Impensis GC Nauk.

Liu Z, Seefried FR, Reinhardt F, et al (2011) Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet Sel Evol 43:1–9.

Macedo FL, Christensen OF, Astruc JM, et al (2020) Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. Genet Sel Evol 52:1–10. https://doi.org/10.1186/s12711-020-00567-1.

Mäntysaari EA, Koivula M, Strandén I (2020) Symposium review: Single-step genomic evaluations in dairy cattle. J Dairy Sci 103:5314–5326. https://doi.org/10.3168/jds.2019-17754.

Martini JWR, Schrauf MF, Garcia-Baccino CA, et al (2018) The effect of the H -1 scaling factors τ and ω on the structure of H in the single-step procedure. Genet Sel Evol 50:1–9. https://doi.org/10.1186/s12711-018-0386-x.

Masuda Y, Tsuruta S, Bermann M, et al (2021) Comparison of models for missing pedigree in single-step genomic prediction. J Anim Sci 99:1–10. https://doi.org/10.1093/jas/skab019.

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem medica 22:276–282.

Melo DT, Marc DS, Silva R, Anjos P (2020) Modeling (co) variance structures for genetic and non- genetic effects in the selection of common bean progenies. 9:. https://doi.org/10.1007/s10681-020-02607-9.

Mesterhazy A, Lemmens M, Reid LM (2012) Breeding for resistance to ear rots caused by Fusarium spp. in maize–a review. Plant Breed 131:1–19.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92:4648–4655.

Oakey H, Cullis B, Thompson R, et al (2016) Genomic selection in multi-environment crop trials. G3 Genes, Genomes, Genet 6:1313–1326.

Oliveira CM, Auad AM, Mendes SM, Frizzas MR (2014) Crop losses and the economic impact of insect pests on Brazilian agriculture. Crop Prot 56:50–54

Oliveira JP, Moreira Jr. WN, Duarte JB, et al (2003) Genotype-environment interaction in maize hybrids: an application of the AMMI model. Crop Breed Appl Biotechnol 3:185–192. https://doi.org/10.12702/1984-7033.v03n03a02.

Pádua JMV, Das Graças Dias KO, Pastina MM, et al (2016) A multi-environment trials diallel analysis provides insights on the inheritance of fumonisin contamination resistance in tropical maize. Euphytica 211:277–285. https://doi.org/10.1007/s10681-016-1722-2.

Paterniani E, CAMPOS MS (1999) Melhoramento do milho. Melhor espécies Cultiv Viçosa UFV 429–485.

Pereira FAC, De Carvalho SP, Rezende TT, et al (2018) Selection of coffea arabica L. Hybrids using mixed models with different structures of variance-covariance matrices. Coffee Sci 13:304–311. https://doi.org/10.25186/cs.v13i3.1444.

Pérez P, De Los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. Genetics 198:483–495. https://doi.org/10.1534/genetics.114.164442.

Persa R, Grondona M, Jarquin D (2021) Development of a Genomic Prediction Pipeline for Maintaining Comparable Sample Sizes in Training and Testing Sets across Prediction Schemes Accounting for the Genotype-by-Environment Interaction.

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. https://doi.org/10.1007/s10681-007-9449-8.

R Core Team (2022) R: A language and environment for statistical computing.

Richard JL (2007) Some major mycotoxins and their mycotoxicoses—An overview. Int J Food Microbiol 119:3–10.

Rocha JR do AS de C, Nunes KV, Carneiro ALN, et al (2019) Selection of superior inbred progenies toward the common bean ideotype. Agron J 111:1181–1189. https://doi.org/10.2134/agronj2018.12.0761.

Rosa Junior OF, Dalcin MS, Nascimento VL, et al (2019) Fumonisin production by Fusarium verticillioides in maize genotypes cultivated in different environments. Toxins (Basel) 11:215.

Salla DA, Furlaneto F de PB, Cabello C, Kanthack RAD (2010) Estudo energético da produção de biocombustível a partir do milho. Ciência Rural 40:2017–2022.

Samayoa LF, Cao A, Santiago R, et al (2019) Genome-wide association analysis for fumonisin content in maize kernels. BMC Plant Biol 19:1–11.

Sandhu S, Dhillon BS (2021) Breeding plant type for adaptation to high plant density in tropical maize—A step towards productivity enhancement. Plant Breed 140:509–518.

Santiago R, Cao A, Malvar RA, Butrón A (2020) Genomics of maize resistance to fusarium ear rot and fumonisin contamination. Toxins (Basel) 12:1–16. https://doi.org/10.3390/toxins12070431.

Silva KJ, Guimarães CT, Guilhen JHS, et al (2020) High-density SNP-based genetic diversity and heterotic patterns of tropical maize breeding lines. Crop Sci 60:779–787. https://doi.org/10.1002/csc2.20018.

Smith A, Cullis B, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138–1147.

Smith AB, Cullis BR (2018) Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. Euphytica 214:143.

Souza VF de, Ribeiro PC de O, Vieira Junior IC, et al (2021) Exploring genotype× environment interaction in sweet sorghum under tropical environments. Agron J 113:3005–3018.

Sprague GF, Federer WT (1951) A Comparison of Variance Components in Corn Yield Trials: II. Error, Year x Variety, Location x Variety, and Variety Components 1 . Agron J 43:535–541. https://doi.org/10.2134/agronj1951.00021962004300110003x.

Tabery J (2008) R. A. Fisher, Lancelot Hogben, and the origin(s) of genotype-environment interaction. J Hist Biol 41:717–761. https://doi.org/10.1007/s10739-008-9155-y.

Tiezzi F, de los Campos G, Parker Gaddis KL, Maltecca C (2017) Genotype by environment (climate) interaction improves genomic prediction for production traits in US Holstein cattle. J Dairy Sci 100:2042–2056. https://doi.org/10.3168/jds.2016-11543.

Uate JV, Nuvunga JJ, da Silva CP, et al (2019) Genetic progress, adaptability and stability of maize cultivars for value of cultivation and use trials. Acta Sci - Agron 41:1–11. https://doi.org/10.4025/actasciagron.v41i1.42624.

Ukrainetz NK, Mansfield SD (2020) Prediction accuracy of single-step BLUP for growth and wood quality traits in the lodgepole pine breeding program in British Columbia. Tree Genet Genomes 16:1–13.

USFDA Center for Food Safety and Applied Nutrition (2001) Guidance for industry: fumonisin levels in human foods and animal feeds. US Food Drug Adm.

VanRaden PM, Van Tassell CP, Wiggans GR, et al (2008) Reliability of genomic predictions for North American dairy bulls. J Dairy Sci 91:305.

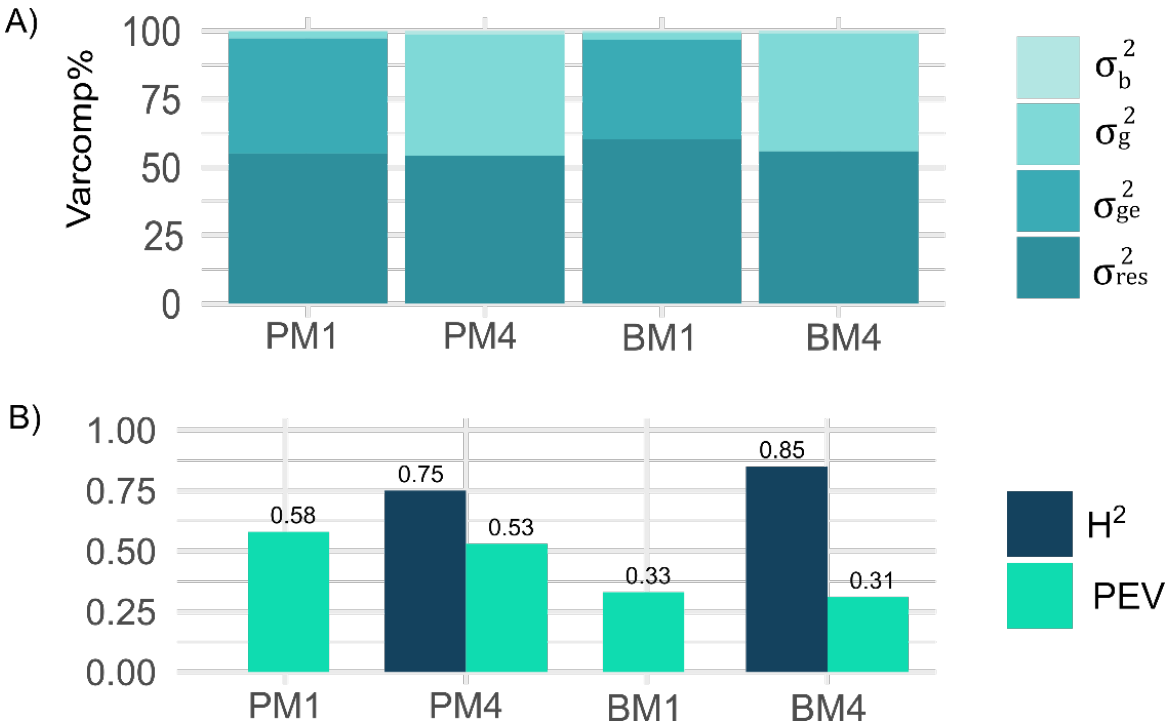
Velazco JG, Malosetti M, Hunt CH, et al (2019) Combining pedigree and genomic information to improve prediction quality: an example in sorghum. Theor Appl Genet 132:2055–2067. https://doi.org/10.1007/s00122-019-03337-w.

Vitezica Z-G, Aguilar I, Misztal I, Legarra A (2011) Bias in genomic predictions for populations under selection. Genet Res (Camb) 93:357–366.

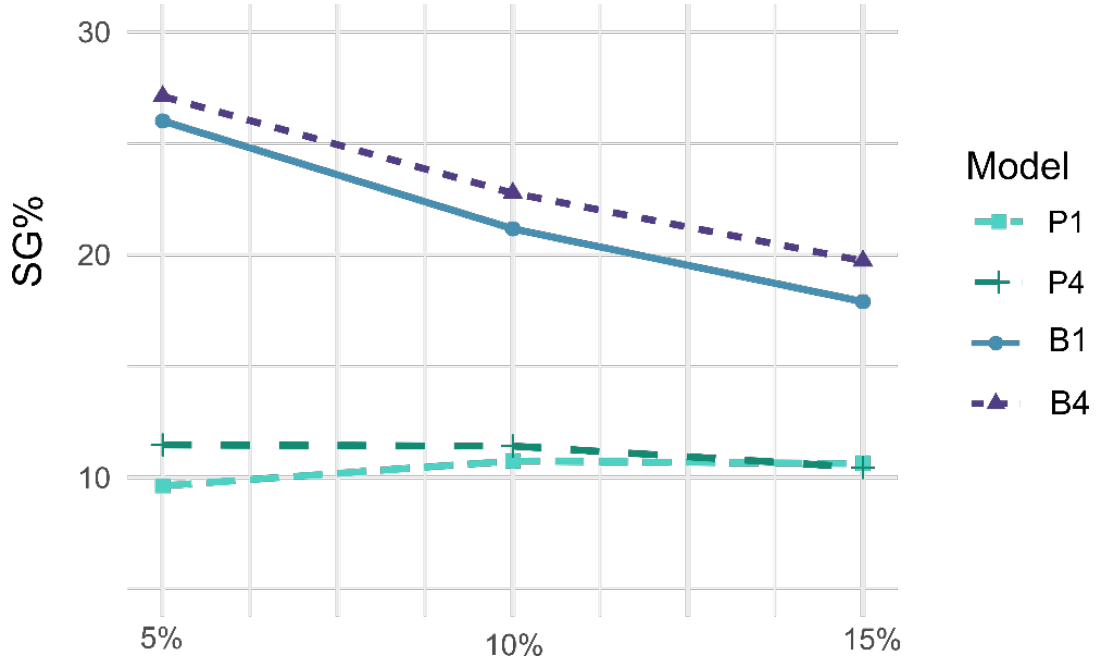
Zea L, Kandus M, Almorza D, et al (2010) Statistical models for evaluating the genotype-environment interaction in maize. 39–46.

Zhang M, Cui Y, Liu Y-H, et al (2020) Accurate prediction of maize grain yield using its contributing genes for gene-based breeding. Genomics 112:225–236.

10 SUPPLEMENT MATERIALS



Supplementary Figure 1. Variance components as percentage of the total variation (Var.comp%), accuracy and generalized heritability estimated using M1 models (PM1 and BM1) and M4 models (PM4 and BM4) for fumonisin concentration in grains (FUMO) in tropical maize.



Supplementary Figure 2. Selection gain in percentage (SG%) considering 5% (20 hybrids), 10% (40 hybrids), and, 15% (59 hybrids) of the top hybrids ranked by PM1, PM4, BM1, and BM4 models for fumonisin concentration in grains (FUMO).