



OPEN ACCESS

EDITED BY

Zefeng Yang,
Yangzhou University, China

REVIEWED BY

Yang-Jun Wen,
Nanjing Agricultural University, China
Mary-Francis LaPorte,
University of California, Davis, United States

*CORRESPONDENCE

Diego Jarquin,
✉ jhernandezjarqui@ufl.edu

RECEIVED 03 August 2024

ACCEPTED 05 May 2025

PUBLISHED 02 July 2025

CITATION

Evangelista JSPC, Dias KOdG, Pastina MM, Chaves S, Guimarães LJM, Hidalgo J, Garcia-Abadillo J, Persa R, Queiroz VAV, Silva DdD, Bhering LL and Jarquin D (2025) Optimizing the single-step model for predicting fumonisins resistance in maize hybrids accounting for the genotype-by-environment interaction.

Front. Genet. 16:1475452.

doi: 10.3389/fgene.2025.1475452

COPYRIGHT

© 2025 Evangelista, Dias, Pastina, Chaves, Guimarães, Hidalgo, Garcia-Abadillo, Persa, Queiroz, Silva, Bhering and Jarquin. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Optimizing the single-step model for predicting fumonisins resistance in maize hybrids accounting for the genotype-by-environment interaction

Jeniffer Santana Pinto Coelho Evangelista^{1,2},
Kaio Olimpo das Graças Dias², Maria Marta Pastina³,
Saulo Chaves⁴, Lauro José Moreira Guimarães³, Jorge Hidalgo⁵,
Julian Garcia-Abadillo^{1,6}, Reyna Persa¹,
Valéria Aparecida Vieira Queiroz³, Dagma Dionísia da Silva³,
Leonardo Lopes Bhering² and Diego Jarquin^{1*}

¹Agronomy Department, University of Florida, Gainesville, FL, United States, ²Departamento de Biologia Geral, Universidade Federal de Viçosa, Campus Universitário, Viçosa, Minas Gerais, Brazil, ³Embrapa Milho e Sorgo, Sete Lagoas, Minas Gerais, Brazil, ⁴Departamento de Genética, Escolha Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, São Paulo, Brazil, ⁵Department of Animal and Dairy Science, University of Georgia, Athens, GA, United States, ⁶Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM), Madrid, Spain

In Brazil, disease outbreaks in plant cultivars are common in tropical zones. For example, the fungus *Fusarium verticillioides* produces mycotoxins called fumonisins (FUMO) which are harmful to human and animal health. Besides the genetic component, the expression of this polygenic trait is regulated by interactions between genes and environmental factors (G × E). Genomic selection (GS) emerges as a promising approach to address the influence of multiple loci on resistance. We examined different manners to conduct the prediction of FUMO contamination using genomic and pedigree data, and combinations of these two via the single step model (**B**-matrix) which also offers the possibility of increasing training set sizes. This is the first study to apply the **B**-matrix approach for predicting FUMO in tropical maize breeding programs. Our research introduced a cross-validation approach to optimize the hyper-parameter w , which represents the fraction of total additive variance captured by the markers. We demonstrated the importance of selecting optimal w by environment in unbalanced datasets. A total of 13 predictive models considering General Combining Ability (GCA) and Specific Combining Ability (SCA) effects, resulted from five linear predictors and three different covariance structures including the single-step approach. Two cross-validation scenarios were considered to evaluate the model's proficiency: CV1 simulated the prediction of completely untested hybrids, where the individuals in the validation set had no phenotypic records in the training set; and CV2 simulated the prediction of partially tested hybrids, where individuals had been evaluated in some environments but not in the target environment. Results showed that using the **B**-matrix in the five tested linear models increased the

predictive ability compared to pedigree or genomic information. Under CV1, increasing training set sizes exhibit superior predictive accuracy. On the other hand, under CV2 the advantages of increasing the training set size are unclear and the improvements are due to better covariance structures. These insights can be applied to plant breeding programs where the GCA, SCA, and $G \times E$ interactions are of interest and pedigree information is accessible, but constraints related to genotyping costs for the entire population exist.

KEYWORDS

fumonisin resistance, genomic prediction, plant breeding, maize hybrid prediction, single-step model

1 Introduction

In tropical environments such as in Brazil, outbreaks of pests and diseases are more frequent and can exhibit significant variations between locations, years, and seasons within the same year, directly affecting crop productivity (Oliveira et al., 2014; Chaves et al., 2023). Among the most frequent diseases in corn crops, ear rot, caused by the fungus *Fusarium verticillioides*, is one of the most prevalent and economically significant diseases (Jorge et al., 2022). In addition to causing losses in productivity, the fungus produces mycotoxins. Fumonisin (FUMO) are particularly concerning due to their widespread occurrence and significant human and animal health impacts (Blacutt et al., 2018). FUMOs are a class of mycotoxins that mainly affect maize grains and can contaminate derived products, such as maize flour (Butoto et al., 2022). Therefore, it is of utmost importance that maize breeding programs consider plant resistance to FUMO contamination as a selection criterion (Lanubile et al., 2014; Holland et al., 2020).

However, phenotyping this trait can be expensive and time-consuming (Bush et al., 2004). Hence, the genomic prediction (GP) emerges as an important alternative strategy for evaluating maize cultivars for resistance to fumonisin contamination. Potentially, this approach could help to increase genetic gains by reducing the required time and associated costs in identifying the most promising materials (Heslot et al., 2015). The GP involves developing predictive models by integrating phenotypic and genomic information derived from single nucleotide polymorphisms (SNPs) - molecular markers. These models are then applied to estimate the genetic potential of individuals whose phenotypes have not been measured, based only on their marker profiles (Meuwissen et al., 2001). Genomic selection (GS), the breeding process that applies GP in breeding decisions, can be particularly useful in hybrid breeding. In GS models for hybrids, the effects of General Combining Ability (GCA) and Specific Combining Ability (SCA) have been widely used (Acosta-Pech et al., 2017; Jarquin et al., 2021; Fonseca et al., 2021; Zhang et al., 2022; Melchinger and Frisch, 2023). The GCA refers to the average performance of a parent in producing desirable traits in its progeny when crossed with different parents, while SCA reflects the specific interaction between two parents, indicating their compatibility and ability to produce superior progeny.

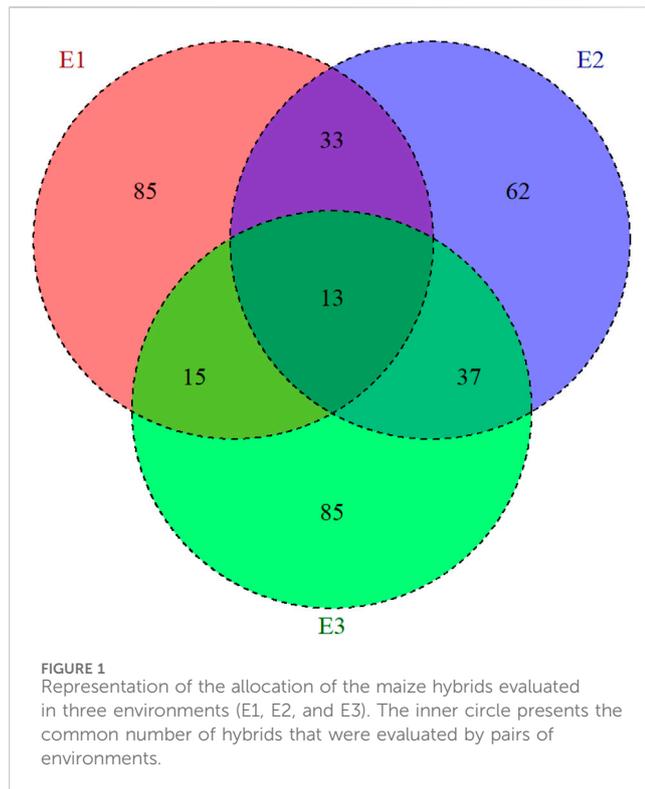
GS offers the opportunity to increase the selection intensity, expedite the breeding cycles, increase the genetic gains, and enable the efficient allocation of resources in breeding programs (Atanda et al., 2021; Beyene et al., 2021; Persa et al., 2021). These advantages

have led many breeding companies to incorporate GS into their programs, including efforts focused on fungal resistance traits. Recent research has demonstrated the potential of GS for improving resistance to fumonisin contamination in maize, highlighting its efficiency and cost-effectiveness compared to traditional phenotypic selection (Butoto et al., 2022). GS exploits the realized genomic relationships between genotypes based on the proportions of alleles they share through the genetic matrix (G ; VanRaden et al., 2008), whose entries describe the genomic similarities between pairs of individuals. This method provides a more accurate representation of genetic inheritance by accounting for Mendelian sampling, enabling the detection of genetic differences between individuals with identical expected similarities. In contrast, pedigree-based selection relies on a relationship matrix (A), constructed solely on the expected similarity between individuals (Hayes et al., 2009).

This does not mean that the pedigree information is expendable or lacks 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, G (the matrix of genomic relationships) can be enriched or complemented by A (the pedigree matrix) forming a single relationship matrix $B = w \times A + (1-w) \times G$ resulting in the well-known single step model (Misztal et al., 2009; Aguilar et al., 2010). In other words, the B -matrix combines genomic and pedigree information with a weighting factor/hyper-parameter (w). The w hyper-parameter represents the fraction of the total additive variance not captured by the markers (Velazco et al., 2019). To find the best way to build the B -matrix, different values for w are considered, and the value that returns the highest predictive ability in training sets is chosen for performing the prediction of individuals in the testing set (Velazco et al., 2019; Oliveira et al., 2020).

In addition, the single step approach allows combining the full set of genotypes (genotyped and non-genotyped but with pedigree information only) using a standard genomic selection method resulting in an increased training set size. The Genomic Best Linear Unbiased Predictor (GBLUP) linear predictor can be used to implement the single-step GBLUP, or ssGBLUP (Legarra et al., 2014) approach. Several 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).

In this study, we analyzed a dataset consisting of 373 single-cross tropical maize hybrids derived from 359 inbred lines. These hybrids were evaluated over 3 years, with each year's assessment taking place in the same location resulting in three different environments (year-



by-location combination). The objective was to evaluate the effectiveness of the single-step approach (referred to as the **B**-matrix) in maize breeding programs aimed at reducing FUMO levels in grains. Wherein the deployment of GP considered the following aims: *i*) optimization of the election of the w hyper-parameter to combine genomic information and pedigree data in prediction models via the **B**-matrix, where w represents the proportion of total additive genetic variance explained by markers; *ii*) comparing the predictive ability of different models based on main and interaction effects; SNPs matrix (**G**), pedigree information (**A**), and the hybrid matrix (**B**) via the single-step model, *iii*) evaluating the impacts on predictive ability by increasing training set sizes using phenotyped individuals with pedigree information only.

2 Materials and methods

2.1 Phenotypic data

The analyzed datasets correspond to three maize trials established at the Brazilian Agricultural Research Corporation (EMBRAPA) Maize and Sorghum headquarters in Sete Lagoas city at the Minas Gerais state, Brazil (19°28'S, 44°15'W). These trials were conducted in three consecutive agronomical years (2014/2015, 2015/2016, and 2016/2017), and each agronomical year was considered as a different environment (E1 = 2014/2015 year; E2 = 2015/2016 year, and E3 = 2016/2017 year). Genotypes were planted in a lattice design, with two repetitions. In total 373 single-cross hybrids were evaluated: 146 in environment 1 (E1), 145 in E2, and 150 in E3. Furthermore, care was taken to allow connectivity

TABLE 1 Dataset summary of the availability of pedigree and genomic information for Hybrids, and groups of parents (P1, and P2).

Information	Hybrids	P1	P2
Pedigree	127	113	21
Pedigree and genomic	246	30	211
Total	373	143	232

between environments: 33 hybrids are common between E1 and E2, 15 between E1 and E3, 37 between E2 and E3, and 13 across the three environments (Figure 1).

All 373 single-cross hybrids were scored for fumonisin concentration in parts per million (ppm) (FUMO), and this information was missing for the parents of these hybrids. The determination of the FUMO trait was initially based on a 500 g sample of corn grains that required a meticulous quantification process. The quantification took place at the Laboratory of Food Safety at Embrapa Milho e Sorgo where a 10 g subsample was finely ground. Then fumonisins were extracted in a solution of 100 mL water/methanol mixture (20/80), and 5 g NaCl using a blender for 1 min. The resulting mixture was then filtered through Whatman paper, and an aliquot of 10 mL of the filtered extract was subsequently diluted in 40 mL of 0.1% phosphate Tween-20 solution (phosphate buffer). This solution was then filtered again using a 1.0 mm microfiber filter. Then a subsample of 10 mL of this solution was passed through the FumoniTest column, which was washed with 10 mL of phosphate buffer solution followed by a second flow of 10 mL of phosphate buffer. The contents of the column were eluted with 1.0 mL of methanol (HPLC grade), collected, and mixed with 1 mL of developer. Finally, the fumonisin concentration in the grain was quantified using the FumoniTest™ and the Fluorometer VICAM, following the manufacturer's protocols (Jorge et al., 2022).

2.2 Pedigree and genotypic data

Pedigree and genomic information were available for a total of 246 hybrids originated from 236 unique inbred lines (Table 1) belonging to two heterotic groups: 142 inbred lines from the Flint group, and 94 inbred lines from the Dent group (see Supplementary Figure S1 for more details about the heterotic groups). The inbred lines were divided into two groups of parents: P1, 30 inbred lines used as male; and P2, 211 inbred lines used as female. A few inbred lines (only 5) were used as both male or female; thus, these genotypes are found in both P1 and P2 groups.

For genotyping, the young leaves of 236 inbred lines were used for extracting genomic DNA via the CTAB method (Hexadecyltrimethylammonium bromide; Saghai-Marooof 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, United States) 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

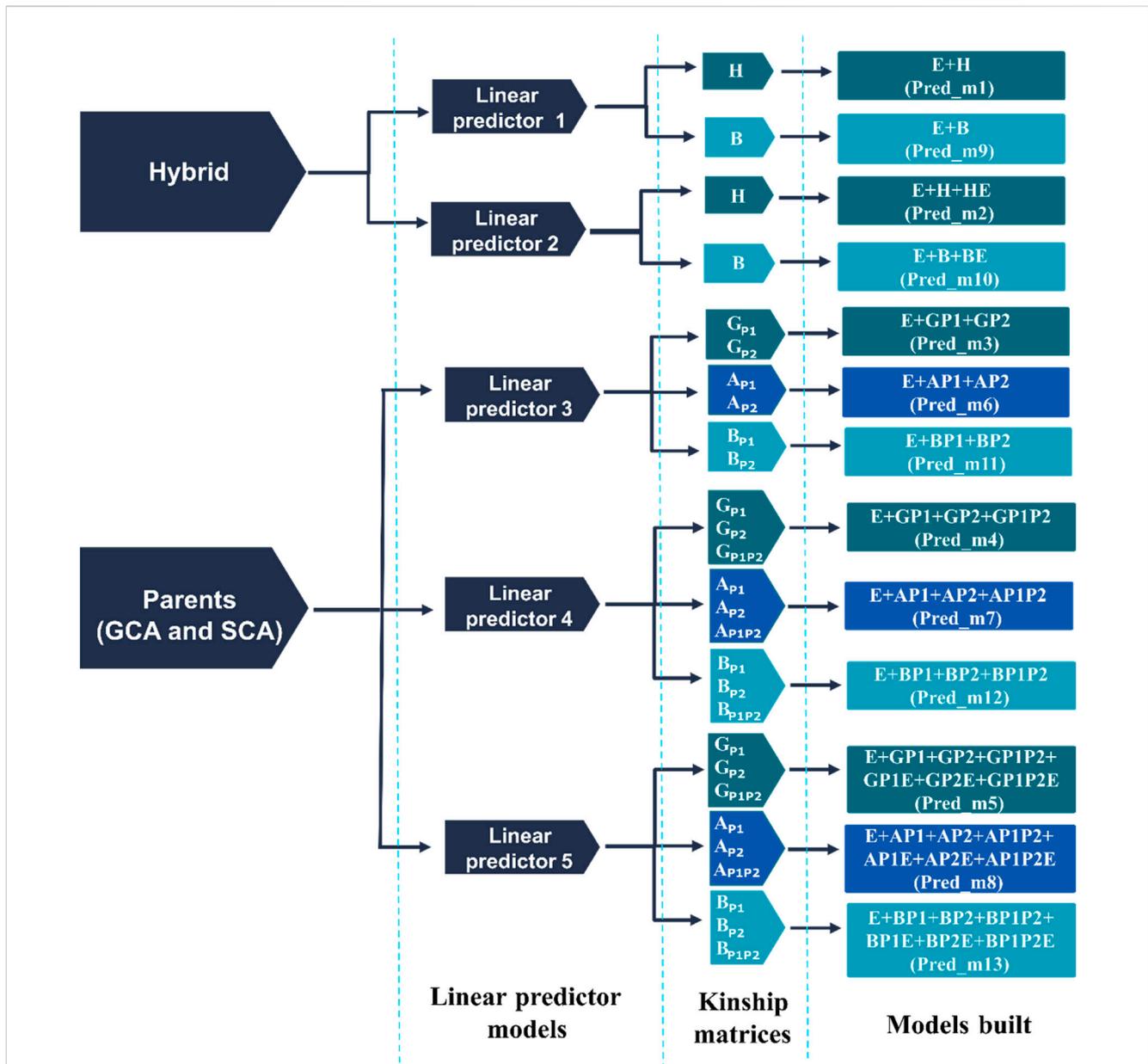


FIGURE 2 Prediction models derived from five linear predictors and four kinship matrices (H, G, A, and B). Where H is the genomic relationship matrix of the hybrids based on synthetic marker SNPs, G_{P1} is the genomic relationship matrix of the P1 group; G_{P2} is the genomic relationship matrix of the P2 group; $G_{P1P2} = G_{P1} \circ G_{P2}$, A_{P1} is the pedigree relationship matrix of the P1 group, A_{P2} is the pedigree relationship matrix of the P2 group, $A_{P1P2} = A_{P1} \circ A_{P2}$, 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. The first two linear predictors (1 and 2) were implemented to model the hybrid performance involving the H matrix alone or in combination with pedigree via the single-step B-matrix. The other three linear predictors (3–5) considered the parents' genomic and/or pedigree information via the general and specific combining ability (GCA and SCA) terms.

B73 reference genome (AGPv3). A total of 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) smaller 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 SNP markers of

the inbred lines were encoded as 0, 1, and 2, with 0 representing the allele with the minor frequency, one for the heterozygous, and two for the allele with the major frequency. The genotypes of the synthetic hybrids were built by combining the markers from the respective parental inbred lines using the expected value, which represents the mean allele dosage across parents for each marker.

In addition, there were also available phenotypic and pedigree information for an extra set of 127 hybrids totaling 373 when combined with the initial set of 246 hybrids that have both

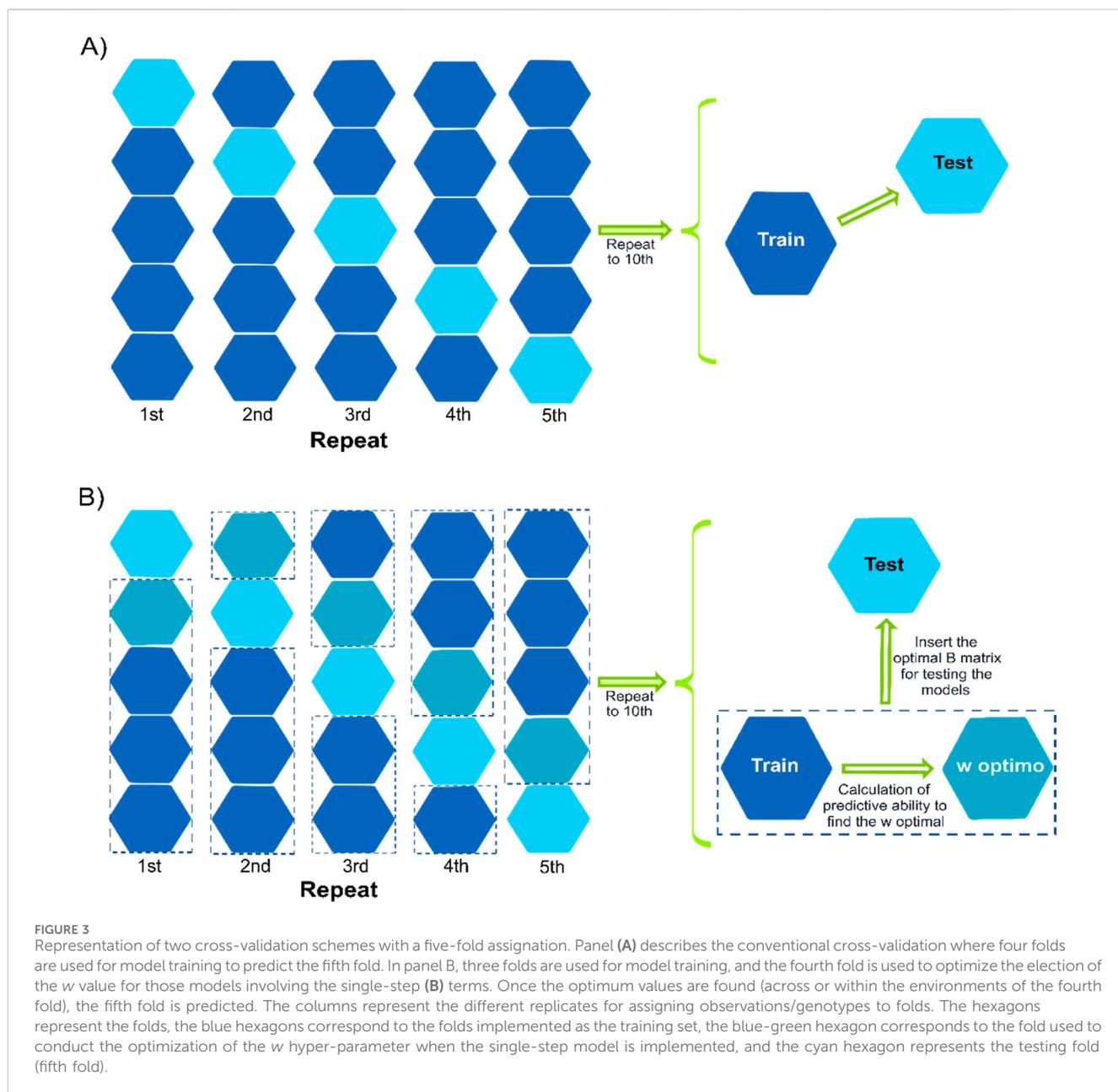


FIGURE 3

Representation of two cross-validation schemes with a five-fold assignment. Panel (A) describes the conventional cross-validation where four folds are used for model training to predict the fifth fold. In panel B, three folds are used for model training, and the fourth fold is used to optimize the election of the w value for those models involving the single-step (B) terms. Once the optimum values are found (across or within the environments of the fourth fold), the fifth fold is predicted. The columns represent the different replicates for assigning observations/genotypes to folds. The hexagons represent the folds, the blue hexagons correspond to the folds implemented as the training set, the blue-green hexagon corresponds to the fold used to conduct the optimization of the w hyper-parameter when the single-step model is implemented, and the cyan hexagon represents the testing fold (fifth fold).

pedigree and genomic data (Table 1). This set of 127 hybrids was originated from 131 inbred lines belonging to three heterotic groups: 74 inbred lines from the Flint group, 49 inbred lines from the Dent group, and four inbred lines from the C group. The C group is intermediate to Flint and Dent, representing lines of several origins. It is well known there is a good combining ability between genotypes from the C group crossed with Dent or Flint testers (groups) (Silva et al., 2020). There was also information available regarding the pedigree of the ancestors of the inbred lines. Hence it was possible compute the pedigree-based relationship matrix for the hybrids (A_H) and the corresponding matrices for the group of parents P1 (A_{P1}), and P2 (A_{P2}). It is worth noting that over 3 years, in the same experiment, there were hybrids whose parents were genotyped (246 hybrids) and hybrids whose parents were not genotyped (127 hybrids), due to limited resources of the breeding program.

2.3 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), which utilizes genealogical information to estimate genetic relatedness. This method is widely used in animal and plant breeding to quantify the genetic similarity among individuals based on their shared ancestry. The R package AGHmatrix (Amadeu et al., 2016) was used to build the additive relationship matrices for P1 (A_{P1}), P2 (A_{P2}), and the hybrids (A_H). The genomic data was used to compute the realized genomic relationship matrix (or G matrix) for P1 (G_{P1}), P2 (G_{P2}), and the hybrids (G_H) following VanRaden et al. (2008). VanRaden's method estimates genetic relatedness based on actual genetic marker data, providing a more precise measure of genetic similarity by

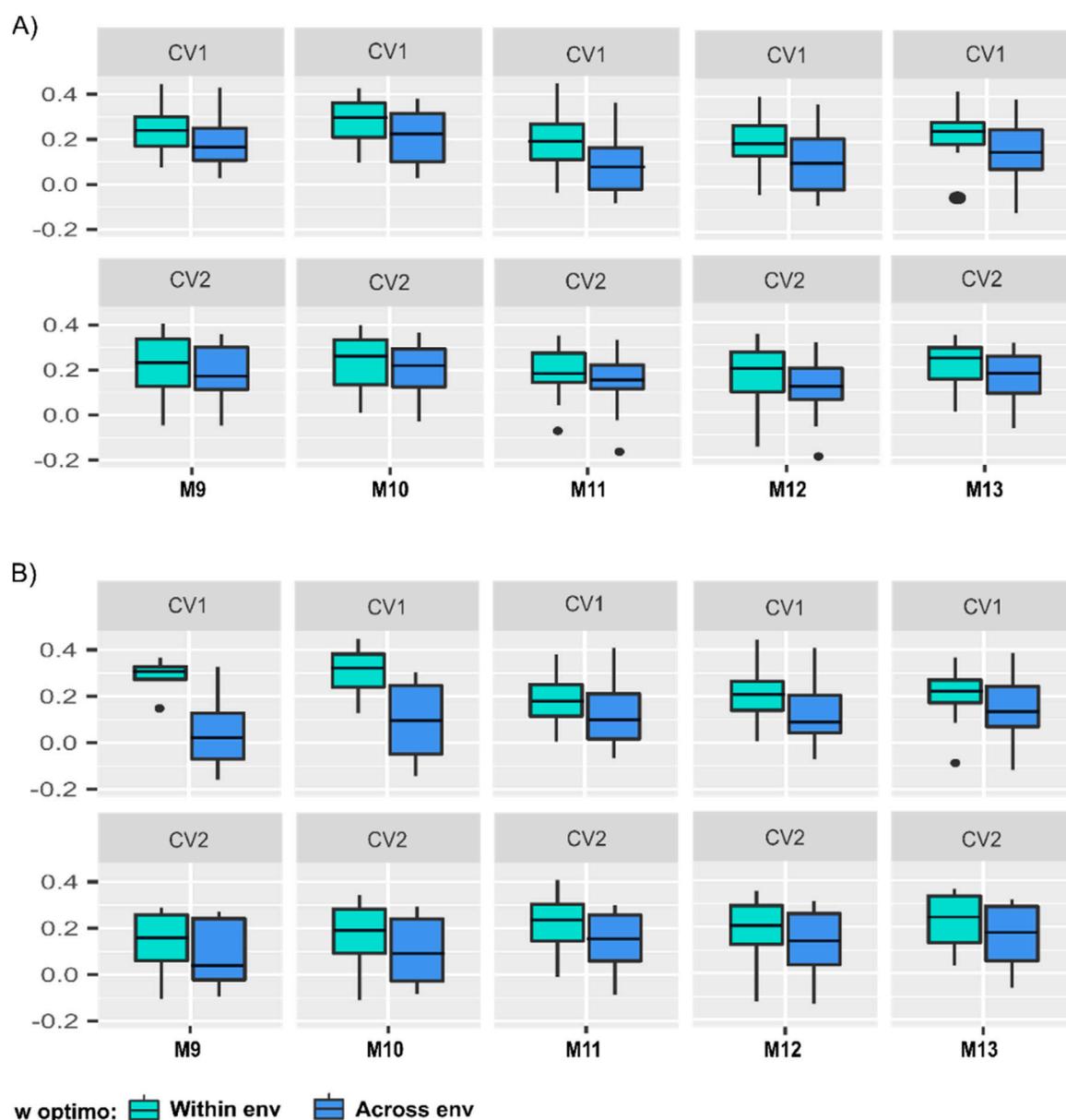


FIGURE 4
 Boxplot of the predictive abilities of five linear predictor models in two prediction scenarios (CV1 and CV2), using two approaches to find the optimum w parameter of the **B**-matrix for models M9-M13. In blue color are indicated the results from selecting the w optimum value across environments, while in green, the results from selecting the within environments w value for the calibration sets: **(A)** the GOds (genomic-only dataset), and **(B)** the PGds (pedigree and genomic dataset). The numbers on the vertical y-axis are the models' predictive abilities, and the horizontal x-axis represents models M9-M13.

considering the genetic variation at specific loci across the genome. In addition, *a*) to having a better characterization of the matrix of genomic relationships, and *b*) to increasing the training set size by including non-genotyped inbred lines, *c*) the single-step relationship matrices were built for P1, P2, and the hybrid components to include information of the ancestry of the parents as in [Khanna et al. \(2022\)](#).

In general, the matrix that combines genomic and pedigree data is commonly named the **H** matrix ([Velazco et al., 2019](#)); however, since this research involves the prediction of hybrids, the use of **H** was reserved for denoting the genomic relationship matrix computed with the synthetic matrix of marker SNPs (mean of

the allele dosage of the parents at each marker position). For this reason, the resulting matrix from the single-step procedure is named **B**-matrix. For example, the E + H model represents a linear predictor that includes the main effect of the environments and the main effect of the synthetic hybrid markers obtained as the mean across the marker information of the two parents involved in a specific crossing. This change was intended for an easier understanding for the reader. On the other hand, generically, the **B** (single step) matrix was built following [Aguilar et al. \(2010\)](#):

$$\mathbf{B} = w \times \mathbf{A} + (1 - w) \times \mathbf{G} \quad (1)$$



where **A** is the pedigree relationship matrix, **G** is the genomic relationship matrix, and *w* is the weighting factor (i.e., the fraction of total additive variance not addressed by the markers; Velazco et al., 2019). Here, the **B**-matrix in Equation 1 can be computed considering different sources of information. For example, combining the **H** and the **A** matrices, or the **G_{P1}** and the **A_{P1}** matrices, the **G_{P2}** and the **A_{P2}** matrices, and their corresponding interactions with environments, etc. Further details about all the different combinations and model terms considered in this study are below.

2.4 Prediction models

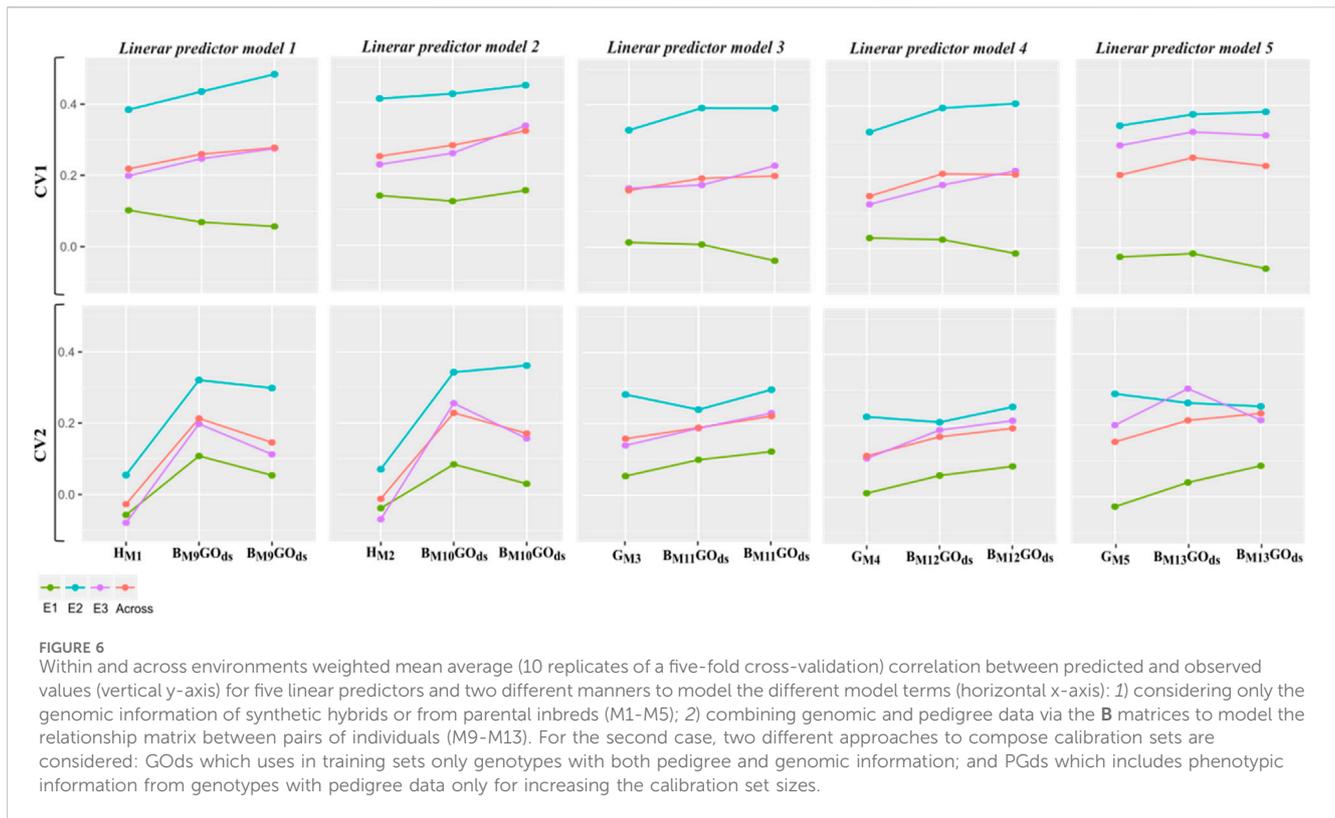
A two-stage approach was considered to implement the prediction models. In the first stage, within-environment adjusted means (best linear unbiased estimation–BLUE) were obtained with the following model:

$$y_{ikl} = \mu + h_i + r_k + b_{l(k)} + \epsilon_{ikl} \tag{2}$$

where *y_{ikl}* is the phenotypic value of the *i*th hybrid within the *l*th block of the *k*th replicate; μ is the constant, *h_i* is the fixed effect of *i*th hybrid

(*i* = 1, 2, ..., 373), *r_k* is the fixed effect of replication *k* (*k* = 1, 2), *b_{l(k)}* is the random effect of the *l*th block within the *k*th replicate (*l* = 1, 2, ..., 16), with *b_{l(k)}* ~ *N*(0, σ_b^2), where σ_b^2 is the block variance, and ϵ_{ikl} is the residual random effect with ϵ_{ikl} ~ *N*(0, σ_ϵ^2), and σ_ϵ^2 is the residual variance. A descriptive statistical analysis of the dataset was performed to assess the variance components and FUMO heritability for each environment and the results are presented in Supplementary Table S1. The analysis of the variance components employed a similar model to Equation 2 with the genetic effect of the hybrids considered as random such that *h_i* ~ *N*(0, σ_h^2), where σ_h^2 represents the corresponding variance component. 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 different linear predictors were considered to model the trait response, and three different ways to account for relationships between hybrids using genomic (**H**, **G_{P1}**, **G_{P2}**), pedigree (**A_{P1}**, **A_{P2}**, **A_{P1P2}**), and the single-step (**B**, **B_{P1}**, **B_{P2}**, **B_{P1P2}**) information. Thirteen different prediction models resulted from combining the five linear predictors and the three different ways to introduce the relationships between hybrids (Figure 2;



Supplementary Table S2). In principle, the first two linear predictors were implemented to model the hybrid performance directly, considering *i*) the main effects of the synthetic markers and *ii*) their corresponding interactions with environments. This model was also implemented to model the trait response using the pedigree matrix **A** of the hybrids and their corresponding interaction with environments. Hence, four different models can be constructed by combining the two linear predictors and the two ways to directly model the hybrids (M1, M2, M9, and M10 in Figure 2; Supplementary Table S2).

The other three linear predictors (3, 4, and 5) were considered to indirectly model the hybrids via the general and specific combining ability (GCA and SCA) terms. The GCA (P1+P2) term attempts to model the average effect of the parents involved in the crosses, while the SCA (P1 × P2) model term corresponds to the interaction of these effects. Similarly to the previous models, different sources of information were used to model the effects of the parents involved in the crosses: 1) molecular marker information (**G**_{P1}, **G**_{P2}, and **G**_{P1P2}), 2) pedigree data (**A**_{P1}, **A**_{P2}, and **A**_{P1P2}), and 3) the combination of these sources of information using the single-step method: **B**_{P1} = $w \times \mathbf{A}_{P1} + (1 - w) \times \mathbf{G}_{P1}$, **B**_{P2} = $w \times \mathbf{A}_{P2} + (1 - w) \times \mathbf{G}_{P2}$, and **B**_{P1P2} = $w \times \mathbf{A}_{P1P2} + (1 - w) \times \mathbf{G}_{P1P2}$. Further details are provided below to describe the election of the optimum value of the *w* hyper-parameter. In addition, the interactions of these terms with environments were also considered resulting in models M3, M4, M5, M6, M7, M9, M11, M12, and M13 (Figure 2; Supplementary Table S2).

Below, the linear predictors (1–5) and their corresponding components (Figure 2; Supplementary Table S2) are described.

2.4.1 Linear predictor 1

Consider that \bar{y}_{ij} represents the phenotypic response (FUMO) of the *i*th hybrid observed in the *j*th environment; depending on the different sources of information it can be modeled as follows:

$$\bar{y}_{ij} = \mu + E_j + k_i + \varepsilon_{ij} \tag{3}$$

where μ is the general mean (constant effect across genotypes and environments), E_j is the fixed effect of *j*th environment, k_i models the random effect of the *i*th hybrid, where $\mathbf{k} = \{k_i\} \sim \text{MVN}(\mathbf{0}, \mathbf{K}\sigma_k^2)$. The k_i can be modeled considering different sources of information. For example, **K** could represent the covariance structure computed with the synthetic SNPs markers **H**, or the **B**-matrix of the hybrids combining **A** and **H** matrices as $w \times \mathbf{A} + (1 - w) \times \mathbf{H}$, with σ_k^2 being the additive variance, and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, where σ_ε^2 is the residual variance. Thus, the possible models of this linear predictor described in Equation 3 are M1, and M9 (Figure 2).

2.4.2 Linear predictor 2

This model is similar to linear predictor 1, but it also includes the interaction effect between hybrids and environments ($G \times E$) 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 kE_{ij} , this linear predictor can be described as:

$$\bar{y}_{ij} = \mu + E_j + k_i + kE_{ij} + \varepsilon_{ij} \tag{4}$$

where kE_{ij} is the random effect of the hybrid-by-environment interaction such that $\mathbf{kE} = \{kE_{ij}\} \sim \text{MVN}[\mathbf{0}, (\mathbf{Z}_g \mathbf{K} \mathbf{Z}'_g) \circ (\mathbf{Z}_E \mathbf{Z}'_E) \sigma_{kE}^2]$, \mathbf{Z}_g and \mathbf{Z}_E are the incidence matrices that connect phenotypes with genotypes and environments, respectively, σ_{kE}^2 represents the variance component of the

hybrid-by-environment interaction, and \circ is the Hadamard or Shur product (the cell-by-cell product) between two matrices. Similar to the previous model in Equation 4, different sources of information (\mathbf{H} , or \mathbf{B} -matrix) can be used to model the $G \times E$ interaction term. In this case, the resulting models of this linear predictor are M2 and M10 (Figure 2).

2.4.3 Linear predictor 3

This linear predictor was built by modeling the general combining ability (GCA) of the male and female parents involved in the crosses of the hybrids. Consider the i th hybrid originated from crossing parent 1 (P1) and parent 2 (P2), with k_{P1i} and k_{P2i} representing their corresponding effects. The following linear predictor was implemented to modeling the performance of the i th hybrid in the j th environment via the GCA of the inbred lines:

$$\bar{y}_{ij} = \mu + E_j + k_{P1i} + k_{P2i} + \varepsilon_{ij} \quad (5)$$

where $\mathbf{k}_{P1} = \{k_{P1i}\} \sim \text{MVN}(\mathbf{0}, \mathbf{K}_{P1}\sigma_{P1}^2)$, with \mathbf{K}_{P1} representing the kinship matrix of the P1 parent, σ_{P1}^2 being the associated variance component; $\mathbf{k}_{P2} = \{k_{P2i}\} \sim \text{MVN}(\mathbf{0}, \mathbf{K}_{P2}\sigma_{P2}^2)$, \mathbf{k}_{P2} representing the kinship matrix of the P2 parent, and σ_{P2}^2 being the associated variance component. Similar to the previous models in Equation 5, different sources of information can be used to model the relationship matrices for parents P1 (\mathbf{K}_{P1}) and P2 (\mathbf{K}_{P2}). For example, these can be modeled using the genomic information of the parental inbreds (\mathbf{G}_{P1} , and \mathbf{G}_{P2}), the pedigree information of the parental inbreds (\mathbf{A}_{P1} , and \mathbf{A}_{P2} ; as in Khanna et al., 2022) or the combination of these components via the \mathbf{B} matrices (\mathbf{B}_{P1} , and \mathbf{B}_{P2}). The resulting models of this linear predictor described in Equation 6 are M3, M6, and M11 (Figure 2).

2.4.4 Linear predictor 4

This linear predictor is similar to the linear predictor 3; however, in addition to the GCA of the inbred lines it also includes the specific combining ability (SCA) of the parents involved in the corresponding crossing. The SCA was modeled as the interaction effect between the pair of parents according to Acosta-Pech et al. (2017). Combining the assumptions from the linear predictor 3 with the SCA term, the resulting linear predictor is:

$$\bar{y}_{ij} = \mu + E_j + k_{P1i} + k_{P2i} + k_{P1P2i} + \varepsilon_{ij} \quad (6)$$

where $\mathbf{k}_{P1P2} = \{k_{P1P2i}\} \sim \text{MVN}(\mathbf{0}, \mathbf{K}_{P1P2}\sigma_{P1P2}^2)$, with $\mathbf{K}_{P1P2} = \mathbf{K}_{P1} \circ \mathbf{K}_{P2}$, and \mathbf{K}_{P1} and \mathbf{K}_{P2} being as previously described in the linear predictor 3, and σ_{P1P2}^2 is the associated variance component with this interaction term. This interaction term can be modeled using the genomic information of the parental inbreds, the pedigree information of the parental inbreds or the combination of these via the \mathbf{B} matrices. The resulting models of this linear predictor are M4, M7, and M12 (Figure 2).

2.4.5 Linear predictor 5

One of the disadvantages of the previous linear predictor is that it returns the same estimated predicted value for each genotype across environments. In addition, many of the hybrids share a common parent (either P1 or P2) but observed in other environments. In these cases, it is possible borrow information between hybrids sharing a common parent across environments. Acosta-Pech et al. (2017) and Jarquin et al. (2021) showed that

inclusion of the GCA and SCA components significantly increase maize hybrid prediction. The resulting linear predictor is an extension of the fourth linear predictor that also includes the interaction between the components of the GCA ($\mathbf{k}_{P1}, \mathbf{k}_{P2}$) and SCA (\mathbf{k}_{P1P2}) terms with the environment via the reaction norm model. The resulting linear predictor is as follows:

$$\bar{y}_{ij} = \mu + E_j + k_{P1i} + k_{P2i} + k_{P1P2i} + kE_{P1i} + kE_{P2i} + kE_{P1P2i} + \varepsilon_{ij} \quad (7)$$

where $\mathbf{kE}_{P1} = \{kE_{P1i}\} \sim \text{MVN}[\mathbf{0}, (\mathbf{Z}_{P1}\mathbf{K}_{P1}\mathbf{Z}'_{P1}) \circ (\mathbf{Z}_E\mathbf{Z}'_E)\sigma_{P1e}^2]$, $\mathbf{kE}_{P2} = \{kE_{P2i}\} \sim \text{MVN}[\mathbf{0}, (\mathbf{Z}_{P2}\mathbf{K}_{P2}\mathbf{Z}'_{P2}) \circ (\mathbf{Z}_E\mathbf{Z}'_E)\sigma_{P2e}^2]$, $\mathbf{kE}_{P1P2} = \{kE_{P1P2i}\} \sim \text{MVN}[\mathbf{0}, (\mathbf{Z}_{P1}\mathbf{K}_{P1}\mathbf{Z}'_{P1}) \circ (\mathbf{Z}_{P2}\mathbf{K}_{P2}\mathbf{Z}'_{P2}) \circ (\mathbf{Z}_E\mathbf{Z}'_E)\sigma_{P1P2e}^2]$, \mathbf{Z}_{P1} , and \mathbf{Z}_{P2} are the incidence matrices that connect phenotypes with inbred parents P1 and P2; σ_{P1e}^2 , σ_{P2e}^2 and σ_{P1P2e}^2 are the corresponding variance components of the interaction between the GCA (P1 and P2) and SCA (P1 \times P2) terms with the environments E . Since different sources of information can be used to model the covariance structures \mathbf{K}_{P1} , and \mathbf{K}_{P2} , the resulting models of the linear predictor described in Equation 7 are M5, M8, and M13 (Figure 2).

2.5 Cross-validation and \mathbf{B} -matrix optimization

For evaluating the model's predictive ability, two cross-validation scenarios were implemented: 1) CV1, which consists of predicting hybrids that have not been tested at any of the observed environments, and 2) CV2, which consists of predicting the performance of hybrids that have been tested in some environments but not in others (incomplete field trials). Furthermore, two different ways of composing training sets represented in Figure 3 were considered. The traditional five-fold cross-validation scheme was implemented for the models that only use \mathbf{H} , \mathbf{G} , or the \mathbf{A} matrices (models M1-M8; 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 a testing set for evaluating the predictive ability. In this case, around 20% of the hybrids were predicted using 80% of the observed phenotypes. Ten replicates were considered to randomly assign phenotypes (CV2)/genotypes (CV1) to folds.

A slightly modified cross-validation approach was employed for the models incorporating the \mathbf{B} -matrix (five models M9-M13), as depicted in Figure 3B. While maintaining the same five-fold partition, an optimization procedure was introduced to select the optimal w value to combining both covariance structures. In this setup, three folds were utilized for training the model, and the fourth fold was used to determine the optimum w value that yields the highest correlation between predicted and observed values. Regarding the fourth fold, a sequence of 21 values ranging from 0 to 1 (with increments of 0.05) for w was considered for constructing $\mathbf{B} = w \times \mathbf{A} + (1 - w) \times \mathbf{G}$ matrices. After fitting the models using three folds, the correlation between predicted and observed values on the fourth fold was computed. The w value that returned the highest correlation was chosen to construct the \mathbf{B} -matrix for predicting the fifth fold (testing set). Subsequently,

the last fold was employed to assess the predictive performance of the models (Figure 3B). This approach allocated approximately 60% of the phenotypic data for model training, 20% for w optimization, and another 20% for evaluating the model's predictive ability given the optimized w value obtained at the fourth 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) selecting the optimum value of w across environments and 2) selecting the optimum value of w for each environment. For the first case, the weighted average correlation across the three environments was computed following Tiezzi et al. (2017) (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 for predicting the remaining fifth fold. Therefore, up to three different w values could result from selecting the optimum w value for each environment. This procedure was repeated 10 times for each one of the cross-validation scenarios (CV1 and CV2). 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).

2.6 Training set composition

The single-step model has shown improvements in predictive ability in plant breeding implementations (Legarra et al., 2009; Cappa et al., 2019; Oliveira et al., 2020). These improvements are attributed to the fact that *i*) combining genomic and pedigree information potentially results in a better matrix describing relationships between pairs of individuals, and *ii*) the ability to increase the training set size for model calibration (Oliveira et al., 2020). This study considered two scenarios for composing training sets to assess the factors contributing to improving predictive ability when considering the single-step model (B). The first scenario, named genomic only dataset (GOds), consists of only individuals with genomic and pedigree information (246 hybrids). While the second scenario, named pedigree and genomic dataset (PGds), increases the training set size by including phenotypic information of individuals with pedigree data (373 hybrids). Some of these have both genomic and pedigree information. For the two training set composition scenarios, in principle, the same training set partitions described in Figure 3 were considered; however, for the second scenario, the training set was augmented with the phenotypic information of non-genotyped inbreds for whose pedigree information was available to connect with genotyped inbreds.

2.7 Within and across environments predictive ability

The predictive ability was assessed on a trial basis by computing 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) as follows in Equation 8:

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

where r_j is the Pearson's correlation between the predicted and the 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 corresponding number of observations. Furthermore, the across-environments mean squared prediction error (MSPE) was also computed as described in Equation 9.

$$MSPE = \frac{\sum_{j=1}^J \sum_{i=1}^I (y_{ij} - \hat{y}_{ij})^2}{Total\ phenotypes} \quad (9)$$

where y_{ij} is the phenotypic value of the i^{th} hybrid at the j^{th} environment, and \hat{y}_{ij} is the corresponding predicted value using the above described 13 models (M1-M13).

2.8 Summary of the prediction strategies

Two cross-validation scenarios (CV1 and CV2) were considered to assess the predictive ability of the prediction models. A total of 13 prediction models (M1-M13) were derived from the five linear predictors and the three different sources of information to modeling the different model terms. For models M1-M8, the training set was composed of four folds, while the fifth fold corresponds to the testing set. On the other hand, the models M9-M13 involve single-step parameterization. Thus, these require an optimization process to select the value of the w hyper-parameter that returns the highest predictive ability using three folds for model calibration and the fourth fold to evaluate different values for w (21 values ranging from zero to one in steps of 0.05). In both groups of models (M1-M8, and M9-M13), the same partitions were considered for a direct comparison.

In addition, for all models using only genotypes with genomic data (GOds) for composing the training sets, an alternative was considered by augmenting the four folds with phenotypic information of inbreds with pedigree data (PGds) only. Thus, in this case, it is possible to evaluate whether the improvements in predictive ability are because *i*) a better relationship matrix is computed when combining genomic and pedigree data, or *ii*) direct benefits of increasing the training set by including phenotypic information of non-genotyped inbreds, or *iii*) interaction of both events occurring at the same time.

3 Results

3.1 Optimization of the B-matrix

Two different approaches were considered to find the optimum value of the w hyper-parameter used to blend/combine the A and the genomic (H or G) matrices according to the different linear predictors (1–5), the different ways for fitting the model terms (models M9-M13), and the different manners to compose

calibration sets (i.e., GOds or PGds). The first case consists of optimizing w to return the highest weighted average correlation between predicted and observed values across environments, while the second approach focuses on finding the optimum w value for each environment such that up to three different values can be obtained. Figure 4 presents the weighted average (10 replicates) correlation across the three environments for both cross-validation schemes (CV1 and CV2), five prediction models M9-M13, two different ways to compose training sets (GOds and PGds), and both methods to find the optimum value of the w hyper-parameter (across and within environments).

As expected, when comparing these two approaches, for both prediction scenarios (CV1 and CV2) and both datasets (GOds and PGds), the predictive ability always improved when selecting w within environments for models M9-M13. Regarding the training set compositions, considering only information of genotyped individuals (GOds), under the CV1 cross-validation scheme, the average predictive ability ranged from 0.11 to 0.21, while between 0.09 and 0.18 for CV2 when performing the selection of the w value across environments (Figure 4A). On the other hand, when the w value was optimized within environments, the corresponding average predictive abilities increased, ranging from 0.18 to 0.26 (CV1) and 0.14 to 0.21 in (CV2; Figure 4A). The model that showed the highest predictive ability, for the two approaches (within and across environments), considering the optimization of the w hyper-parameter in both prediction scenarios (CV1 and CV2) and considering training sets of only genotyped individuals (GOds) was M10 (Figure 4A).

Augmenting the training set considering phenotypic information of individuals with pedigree data (PGds) when using the within environments w optimum value, the average predictive ability ranged from 0.20 to 0.32 and 0.12 to 0.21 for CV1 and CV2, respectively (Figure 4B). The model that showed the highest predictive ability for CV1 using w optimum across environments was M13. Nonetheless, when using the within environments w optimal, the highest predictive ability was shown with M10 (Figure 4B). These results indicate that the manner of computing the optimized **B**-matrix influences the election of the best prediction model. For CV2, the model that showed the best predictive ability, for both approaches (GOds, and PGds), within and across environments optimization, was M13. Due to the results presented, the w optimum value computed within environments was used on the models considering the **B** matrices (M9-M13) for comparing the predictive ability of the 13 models (M1-M13) for GOds and PGds training sets (see Figure 4).

3.2 Prediction models

For an easy visualization, Figure 5 presents the heatmap of the within and across-environments predictive ability and the across-environments mean squared prediction error (MSPE) of the 13 models (M1-M13) for the two predicting scenarios (CV1 and CV2) and the two manners to compose training sets (GOds and PGds). As mentioned before, for models M9-M13, only the results derived from the within environments optimization are considered.

The predictive abilities varied considerably through environments for both prediction scenarios (CV1 and CV2) and

both datasets used for model training (GOds, and PGds). Overall, there was an improvement in the across-environments weighted predictive ability and a reduction of the MSPE when the **B**-matrix was considered (higher correlations and reduced MSPE) in models M9-M13. Note that despite using phenotypic information from genotyped individuals (GOds) or combining it with also individuals with only pedigree data for increasing training set size (PGds), the models including the **B**-matrix improved the predictive ability compared with those that only used the genomic information (**H** or **G** matrices) in M1-M5 models. In general, the models presenting the best results, within and across environments, are those that considered the interaction between the genetic (**H**, **G**, **A**, or **B**) and the environmental component (**E**).

3.2.1 CV1 scheme

In the scenario of predicting untested hybrids in evaluated environments (CV1), the across-environments predictive ability of the prediction models using GOds data to compose training sets ranged from 0.11 to 0.28 (Figure 5). In this case, the best prediction models for E1, E2, and E3 were M2 (0.139), M9 (0.436), and M13 (0.325), respectively. However, across environments, the M10 model (**E** + **B** + **BE**) showed the highest predictive ability (0.280) and a reduced MSPE (3.399) compared with most of the models. When the training sets were augmented with individuals with only pedigree data (PGds), the across-environments predictive ability ranged from 0.050 to 0.321 (Figure 5), with the models including the different **B** matrices showing the best results across environments. For PGds, the model with the best predictive ability across the environments was also M10 returning a correlation of 0.321 and a MSPE of 3.364. Note that a relative improvement in the predictive ability of around 15% (i.e., 0.280 vs. 0.321) was accomplished when the training set size was increased also including phenotyped individuals (PGds) with model M10.

3.2.2 CV2 scheme

Contrary to what was initially expected (based on similar studies), in the scenario that predicts the performance of already observed hybrids in some environments but not in others, the predictive abilities were lower than those from the CV1 scheme (Figure 5), which is a more complex prediction problem. Thus, for this particular dataset, it is better to ignore the available information from the target hybrids observed in environments different from those of interest and assume these as totally unobserved hybrids across environments (CV1 scheme).

For the GOds manner of composing training sets, the across-environments predictive ability ranged from -0.027 to 0.229 . For all environments, the models including the different **B** matrices showed the highest predictive ability; however, different prediction models showed the best results for each environment. The M10 showed the highest across-environments correlation (0.229) and a low MSPE (3.447). Regarding the PGds training sets, the predictive ability across the different environments ranged from -0.042 to 0.223 . The corresponding models with the highest correlations for E1, E2, and E3 were M9 (0.107), M10 (0.343), and M8 (0.298). However, across environments, model M13 showed the highest correlation (0.233). In this prediction scenario, it was also observed that the across-environments results of the M13 model trained with GOds (0.213) dataset were improved when increasing the individuals in the

training set using the PGDs set (0.233). Note that, for this model, the correlation increased by 9.4% when using the PGDs. In contrast, model M10 did not improve the predictive ability adding phenotypes of non-genotyped individuals (i.e., from Gods to PGOs), showing a reduction of ~33.9% (0.229 vs. 0.171).

3.3 Impacts of combining genomic information and pedigree data

In the previous section, across environments, mixed results were shown with the increasing of the training set by adding those individuals with pedigree information only. For CV1, the best results were observed when the training set was augmented (0.321) via the PGDs data set in comparison with the GOs set (0.280). Recalling that in both cases, for the most promising models (M10-M13), the genomic information and the pedigree data are combined via the **B** matrices (after an optimization process). However, the only difference is that for the GOs the training set is composed for only those individuals that have both genomic and pedigree data, while for the PGDs the previous data set is augmented by adding the phenotypic information from individuals with only pedigree information.

For CV2, an opposite trend was observed, with GOs returning better results (0.229) than PGDs (0.171) when considering the most promising model (M10) from the previous cross-validation scenario. However, in this case, for PGDs M13 returned the highest correlation (0.233). In any case, for M11-M13 models, slightly improvements were observed when using PGDs vs. GOs. These differences might not be significant; thus, no clear advantages were observed with the increasing training set size as for the CV1 case. Figure 6 depicts a direct comparison of the effects in predictive ability *i*) from combining genomic and pedigree information (GOs), *ii*) with also an increased training set size (PGDs) with respect to the models based on only a single data type (**H**, **G** or **A** used to describe similarities between pairs of individuals).

More specifically, in Figure 6, the results of the within and across environments predictive ability of the five linear predictors are depicted for single **H** (M1-M2), **G** (M3-M5) and combined with **A** through **B** (M9-M13) covariance structures considering phenotypic information of only individuals with genomic and pedigree data (GOs), and adding information from individuals with only pedigree data (PGDs).

Under the CV1 scheme, for the linear predictor 1, on average moderate improvements are observed by combining the matrices **H** and **A** (especially for E2) compared with only using the **H** matrix to model relationships between pairs of individuals (M9 vs. M1). Also, in average, slight improvements were observed when the training set size was increased with individuals with only pedigree data (PGDs) compared to the case where only the phenotypic information of individuals with both genomic and pedigree data (GOs) was used for model calibration.

A similar pattern was observed for the second linear predictor, where the interaction with environmental factors was included using the same covariance structures (**H** vs. **B**) contrasting models M2 and M10. In addition, the linear predictors 3-5 showed similar patterns. In all the cases, very low correlation values were observed for E1, suggesting this environment's high influence contributing to the $G \times E$ interaction.

Regarding the CV2, more pronounced improvements in predictive ability were found by combining **H** and **A** matrices for linear predictors 1 and 2. However, these decreased when the information from non-genotyped was also included in the calibration sets, except for E2, where it practically remained the same. For the linear predictors 3 and 4, some improvements were observed with the PGDs compared to GOs, while with the most complex linear predictor (5), mixed results were found depending on the environment.

4 Discussion

4.1 B-matrix optimization

This study introduces an approach for optimizing the hyper-parameter w using a novel method. To find the optimum w value, a cross-validation strategy using four of the original five folds was implemented (three folds for model fitting and one-fold to optimize w for 21 different values evenly spaced between 0 and 1 in steps of 0.05). Then the phenotypic information of the fifth fold was predicted considering the obtained w optimum value. This strategy offers a robust method for selecting the hyper-parameter w , returning satisfactory results regarding the accuracy of genomic predictions, which is particularly crucial for complex traits such as FUMO. This suggests that using the fourth fold for **B**-matrix optimization can be a good strategy to guide the election of the w hyper-parameter.

Available studies have shown that the election of the w optimum value varies according to the different characteristics being evaluated (Liu et al., 2011; Ashraf et al., 2016; Velazco et al., 2019). Nonetheless, our study presents a significant contribution by demonstrating the importance of selecting the optimal w by environment in unbalanced datasets, an approach not yet utilized in previous research. Until now, the studies have focused on selecting the w optimal value across environments (i.e., a common optimum value for all the environments). However, individuals have different genetic responses when changing environmental conditions, with non-static variation (e.g., 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).

Our results reveal that, particularly in situations with unbalanced datasets, the difference in gene expressions of individuals between environments can significantly influence the election of the w optimal weighting factor. This finding has important implications for multi-environment and multi-trait studies, suggesting that researchers should consider not only the selection of w by trait, as is customary, but also by environment. The change in the value of w with unbalanced data occurs due to the alteration in the relative contribution of genomic and pedigree information to the genetic structure, which may necessitate a reevaluation of the hyper-parameter w to ensure that the modeling of genetic relationships is accurate and appropriate for the specific data in question. Consequently, our study demonstrates that the best strategy for building the **B**-matrix should be optimizing the w hyper-parameter within in environments, which is particularly relevant for crop improvement in tropical environments where $G \times E$ interactions are pronounced, and datasets are often unbalanced.

4.2 Prediction models

Over the decades, breeders have sought strategies to reduce the FUMO content in grain (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 (USFDA, 2021) recommends a maximum of 2–4 mg kg⁻¹ of contamination of FUMO in corn products. In Brazil, the contamination limit is 2 mg kg⁻¹ of these toxins in grains (Agência Nacional de Vigilância, 2011). Exceeding these limits prevents the exportation and national marketing of maize batches, leading to substantial economic losses for producers. Therefore, in addition to high grain yield performance, developing cultivars resistant to FUMO is a primary objective. However, phenotyping this trait is laborious and expensive (Bush et al., 2004). Consequently, implementing genomic selection becomes an important alternative strategy for developing fumonisin-resistant genotypes. Comparing genomic to traditional phenotypic selection for FUMO Butoto et al. (2022) found that both methods performed similarly. Nevertheless, the authors highlight that genomic selection has the potential to be more efficient than phenotypic selection due to the employment of cheaper and faster genotyping methods.

To our knowledge, no studies in the literature have combined genomic and pedigree information to construct relationship covariance structures for predicting FUMO. Our results demonstrate that the use of the **B**-matrix improved predictive ability in all tested linear predictors for both prediction scenarios and datasets (GPDs and PODs) compared to conventional GS models. This improvement in predictive ability is crucial for complex traits like fumonisin resistance, where accurate predictions can significantly impact breeding program efficiency. As observed, the use of a single-step approach was superior to the use of only the **H** matrix, regardless of whether the training set size was increased with un-genotyped individuals. Since some markers may not be in linkage disequilibrium with QTLs, when combining the **A** matrix and **H** or **G** matrix in a relationship the pedigree information may have contributed to capturing associations between causative alleles due to common ancestral identity, improving predictions models (Velazco et al., 2019). This finding has important practical implications, especially in situations where genotyping costs are a constraint. The **B**-matrix approach allows increase the size of the training set by including un-genotyped individuals through pedigree information, potentially leading to more robust and accurate predictions.

The observed improvements in prediction accuracy using the **B**-matrix align with findings from previous studies by Imai et al. (2019) and Cappa et al. (2019). These authors found that in terms of prediction accuracy, the **B**-matrix matched or surpassed the use of only the **G** matrix or the **A** matrix only. Additionally, over the years, several other authors have reported enhanced prediction accuracy with the use of the **B**-matrix in plant and animal species (Ashraf et al., 2016; Gao et al., 2012; Koivula et al., 2012; Lourenco et al., 2015; Pérez-Rodríguez et al., 2017). Furthermore, the single-step method allows the utilization of larger phenotypic datasets compared to the GBLUP method, as demonstrated herein for maize. Moreover, all the tested models corresponding to the different linear predictors reached the lowest MSPEs when using the **B**-matrix in comparison to those using only one data type to

model similarities between pairs of individuals (**H**, **G**, or **A**). These results are consistent with those obtained by Velazco et al. (2019). The selection of prediction models based on minimizing MSPE has been recommended because 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 the **B**-matrix in genomic prediction has been widely discussed (Martini et al., 2018; Macedo et al., 2020; Mäntysaari et al., 2020; Masuda et al., 2021). However, this methodology has not yet become very popular in plant breeding (Oliveira et al., 2020). Our results demonstrate 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. In this case, the **B**-matrix enables the construction of the relationship matrix incorporating all individuals improving the predictive capacity of the models. The simultaneous use of genotyped and un-genotyped individuals relies on projecting genomic relationships to un-genotyped individuals based on the conditional distribution of breeding values for un-genotyped and genotyped individuals (Legarra et al., 2009). The genomic relationships can improve the pedigree relationships, while the un-genotyped individuals provide more phenotypic information. 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 evaluated maize hybrids. Likewise, Cappa et al. (2019) noted that the utilization of the single-step method resulted in elevated prediction accuracies and reduced bias of the genetic component of unobserved (non-phenotyped) but genotyped individuals compared to the standard GBLUP by using additional phenotypic information from non-genotyped individuals.

In our study, including phenotypes from un-genotyped individuals increased the predictive ability in most but not all the tested scenarios. In agreement with our findings, in a study with a broiler population, Hidalgo et al. (2021) found that the two most recent years of pedigree, phenotypic, and genomic data were sufficient to maintain prediction accuracies in selection candidates (i.e., the last generation of individuals), adding phenotypes of un-genotyped individuals from previous years did not increase the accuracy. Lourenco et al. (2014) stated that distant ancestors have minor contributions, explaining the null or marginal increase in predictive ability, and sometimes their inclusion can deteriorate predictive ability.

The second context consists of companies with genomic and pedigree information for all individuals (GODs). Our results showed that by combining both sources of information for modeling covariance structures, the predictive ability of the models increases, helping to prevent (discard) of advancing to the next stages of the program the most susceptible hybrids to FUMO. The increased predictive ability can be explained because the markers do not capture all the genetic variance; blending the genomic relationship matrix with a portion of the pedigree relationship matrix (with an optimal value w) implicitly fits a residual polygenic effect in the statistical model via the modified genomic relationship matrix. Although the genomic relationship matrix accounts for most of the genetic variation, its combination with the pedigree relationship matrix increases the captured genetic

variance. Furthermore, the combined use of pedigree and genomic information in the **B**-matrix can help mitigate the limitations of using the **G** matrix alone, such as incomplete linkage disequilibrium between markers and causal variants, especially in cases where the trait is influenced by rare or less common genetic variants, or imperfect genomic data. By leveraging the complementary information from both sources, the **B**-matrix provides a more comprehensive and accurate representation of the genetic relatedness among individuals, ultimately enhancing the predictive ability of the models. 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).

Our study employed two cross-validation schemes, CV1 and CV2, to evaluate the predictive ability of various models. The CV1 scheme, which predicts untested hybrids in evaluated environments, generally showed higher predictive abilities compared to the CV2 scheme. This finding was contrary to our initial expectations based on similar studies in the literature (Jarquin et al., 2021; Khanna et al., 2022; Persa et al., 2021). One of the factors that could contribute to this unexpected outcome is the unbalanced nature of our dataset that could potentially impact the effectiveness of the CV2 approach, which relies on information from other environments. These findings underscore the need for further research to understand the factors influencing the relative performance of CV1 and CV2 schemes in different contexts, particularly for traits showing strong environmental influences. Finally, the breeding programs are also interested in developing superior cultivars that respond favorably to diverse environmental conditions (Jarquin et al., 2021). Our results were consistent with those obtained by several researchers that observed better predictive ability in models that consider genetic and environmental interaction effects (Jarquin et al., 2014; 2021; Lado et al., 2016; Basnet et al., 2019; Khanna et al., 2022), further emphasizing the importance of accounting for $G \times E$ interactions for reaching improved prediction accuracies of complex traits in tropical maize breeding programs.

5 Conclusion

The findings of this study highlight the importance of combining genomic and pedigree data. Particularly, optimizing the election of the w hyper-parameter to construct the **B**-matrix when dealing with diverse environments and unbalanced datasets. The most convenient optimization resulted when it was implemented within environments compared to across environments. This strategy improves the predictive ability of the models used for making predictions. The **B**-matrix was shown to enhance the predictive ability of the tested linear models for different prediction scenarios and datasets, compared to the **G** and **A** matrices. Hence, the single-step approach helps improve the selection accuracy of FUMO trait.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

JE: Data curation, Formal Analysis, Investigation, Methodology, Writing – original draft, Writing - review and editing. KD: Investigation, Resources, Supervision, Writing - review and editing. MP: Resources, Supervision, Writing - review and editing. SC: Supervision, Writing - review and editing. LG: Resources, Supervision, Writing - review and editing. JH: Conceptualization, Methodology, Validation, Writing - review and editing, JG-A: Investigation, Supervision, Writing - review and editing. RP: Conceptualization, Supervision, Writing - review and editing, VQ: Resources, Supervision, Writing - review and editing. DS: Resources, Supervision, Writing - review and editing, LB: Investigation, Resources, Supervision, Writing - review and editing. DJ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Writing - review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), and Embrapa (Brazilian Agricultural Research Corporation). This study was also financed in part by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil - Finance Code 001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1475452/full#supplementary-material>

References

- Acosta-Pech, R., Crossa, J., de los Campos, G., Teysse, S., Claustres, B., Pérez-Elizalde, S., et al. (2017). Genomic models with genotype × environment interaction for predicting hybrid performance: an application in maize hybrids. *Theor. Appl. Genet.* 130, 1431–1440. doi:10.1007/s00122-017-2898-0
- 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.* Available online at: https://bvsms.saude.gov.br/bvs/saudelegis/anvisa/2011/res0007_18_02_2011_rep.html.
- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (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 (2), 743–752. doi:10.3168/jds.2009-2730
- Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A., Resende Jr, M. F., and Muñoz, P. R. (2016). AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. *plant genome* 9 (3). doi:10.3835/plantgenome2016.01.0009
- Ashraf, B., Edriss, V., Akdemir, D., Autrique, E., Bonnett, D., Crossa, J., et al. (2016). Genomic prediction using phenotypes from pedigreed lines with no marker data. *Crop Sci.* 56 (3), 957–964. doi:10.2135/cropsci2015.02.0111
- Atanda, S. A., Olsen, M., Burgueño, J., Crossa, J., Dzidzieny, D., Beyene, Y., et al. (2021). Maximizing efficiency of genomic selection in CIMMYT's tropical maize breeding program. *Theor. Appl. Genet.* 134 (1), 279–294. doi:10.1007/s00122-020-03696-9
- Basnet, B. R., Crossa, J., Dreisigacker, S., Pérez-Rodríguez, P., Manes, Y., Singh, R. P., et al. (2019). Hybrid wheat prediction using genomic, pedigree, and environmental covariables interaction models. *Plant Genome* 12 (1), 180051. doi:10.3835/plantgenome2018.07.0051
- Beyene, Y., Gowda, M., Pérez-Rodríguez, P., Olsen, M., Robbins, K. R., Burgueño, J., et al. (2021). Application of genomic selection at the early stage of breeding pipeline in tropical maize. *Front. Plant Sci.* 12, 685488. doi:10.3389/fpls.2021.685488
- Blacutt, A. A., Gold, S. E., Voss, K. A., Gao, M., and Glenn, A. E. (2018). *Fusarium verticillioides*: advancements in understanding the toxicity, virulence, and niche adaptations of a model mycotoxigenic pathogen of maize. *Phytopathology* 108 (3), 312–326. doi:10.1094/PHYTO-06-17-0203-RVW
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23 (19), 2633–2635. doi:10.1093/bioinformatics/btm308
- Browning, B. L., and Browning, S. R. (2016). Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98 (1), 116–126. doi:10.1016/j.ajhg.2015.11.020
- Bush, B. J., Carson, M. L., Cubeta, M. A., Hagler, W. M., and Payne, G. A. (2004). Infection and fumonisin production by *Fusarium verticillioides* in developing maize kernels. *Phytopathology* 94 (1), 88–93. doi:10.1094/PHYTO.2004.94.1.88
- Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., and Thompson, R. (2018). *ASReml-R reference manual*. Hemel Hempstead, United Kingdom: VSN International Ltd. Version 4.
- Butoto, E. N., Brewer, J. C., and Holland, J. B. (2022). Empirical comparison of genomic and phenotypic selection for resistance to *Fusarium* ear rot and fumonisin contamination in maize. *Theor. Appl. Genet.* 2022 1 (0123456789), 2799–2816. doi:10.1007/S00122-022-04150-8
- Callister, A. N., Bradshaw, B. P., Elms, S., Gillies, R. A., Sasse, J. M., and Brawner, J. T. (2021). Single-step genomic BLUP enables joint analysis of disconnected breeding programs: an example with *Eucalyptus globulus* Labill. *G3* 11 (10), jkab253. doi:10.1093/g3journal/jkab253
- Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Garcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Sci.* 284, 9–15. doi:10.1016/j.plantsci.2019.03.017
- Chaves, S. F., Evangelista, J. S., Trindade, R. S., Dias, L., Guimaraes, A., Alves, R. S., et al. (2023). Employing factor analytic tools for selecting high-performance and stable tropical maize hybrids. *Crop Sci.* 63 (3), 1114–1125. doi:10.1002/csc2.20911
- Crossa, J., Beyene, Y., Semagn, K., Pérez, P., Hickey, J. M., Chen, C., et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3 Genes, Genomes, Genet.* 3 (11), 1903–1926. doi:10.1534/g3.113.008227
- Cullis, B. R., Smith, A., Hunt, C., and Gilmour, A. (2000). An examination of the efficiency of Australian crop variety evaluation programmes. *J. Agric. Sci.* 135 (3), 213–222. doi:10.1017/S0021859699008163
- Dias, K. O. G., Piepho, H. P., Guimarães, L. J. M., Guimarães, P. E. O., Parentoni, S. N., Pinto, M. O., et al. (2020). Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. *Theor. Appl. Genet.* 133 (2), 443–455. doi:10.1007/s00122-019-03475-1
- Duvick, J. (2001). Prospects for reducing fumonisin contamination of Maize through genetic modification. *Environ. Health Perspect.* 109 (Suppl. 2), 337–342. doi:10.1289/ehp.01109s237
- Eller, M. S., Holland, J. B., and Payne, G. A. (2008). Breeding for improved resistance to fumonisin contamination in maize. *Toxin Rev.* 27 (3–4), 371–389. doi:10.1080/15569540802450326
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS one* 6 (5), e19379. doi:10.1371/journal.pone.0019379
- Fonseca, J. M., Klein, P. E., Crossa, J., Pacheco, A., Perez-Rodriguez, Ramasamy, P., et al. (2021). Assessing combining abilities, genomic data, and genotype × environment interactions to predict hybrid grain sorghum performance. *Plant Genome* 14 (3), e20127. doi:10.1002/tpg2.20127
- Gao, H., Christensen, O. F., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S., et al. (2012). Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet. Sel. Evol.* 44, 8. doi:10.1186/1297-9686-44-8
- González-Recio, O., Rosa, G. J. M., and Gianola, D. (2014). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest. Sci.* 166, 217–231. doi:10.1016/j.livsci.2014.05.036
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. *J. dairy Sci.* 92 (2), 433–443. doi:10.3168/jds.2008-1646
- Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69–83. doi:10.2307/2529339
- Heslot, N., Jannink, J. L., and Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Sci.* 55 (1), 1–12. doi:10.2135/cropsci2014.03.0249
- Hidalgo, J., Lourenco, D., Tsuruta, S., Masuda, Y., Breen, V., Hawken, R., et al. (2021). Investigating the persistence of accuracy of genomic predictions over time in broilers. *J. Animal Sci.* 99 (9), skab239–10. doi:10.1093/jas/skab239
- Holland, J. B., Marino, T. P., Manching, H. C., and Wissner, R. J. (2020). Genomic prediction for resistance to *Fusarium* ear rot and fumonisin contamination in maize. *Crop Sci.* 60 (4), 1863–1875. doi:10.1002/csc2.20163
- Imai, A., Kuniga, T., Yoshioka, T., Nonaka, K., Mitani, N., Fukamachi, H., et al. (2019). Single-step genomic prediction of fruit-quality traits using phenotypic records of non-genotyped relatives in citrus. *PLoS One* 14 (8), e0221880. doi:10.1371/journal.pone.0221880
- Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127 (3), 595–607. doi:10.1007/s00122-013-2243-1
- Jarquín, D., de Leon, N., Romay, C., Bohn, M., Buckler, E. S., Ciampitti, I., 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, 592769. doi:10.3389/fgene.2020.592769
- Jorge, K., Guimarães, C. T., Tinoco, S. M. D. S., Bernardino, K. D. C., Trindade, R. D. S., Queiroz, V. A. V., et al. (2022). A genome-wide association study investigating fumonisin contamination in a panel of tropical maize elite lines. *Euphytica*, 1–12. doi:10.1007/s10681-022-03082-0
- Khanna, A., Anumalla, M., Castolos, M., Bhosale, S., Jarquín, D., and Hussain, W. (2022a). Optimizing predictions in International Rice Research Institute's rice drought breeding program by leveraging 17 years of historical data and pedigree information. *Front. Plant Sci.* doi:10.3389/fpls.2022.983818
- Kleinknecht, K., Möhring, J., Singh, K. P., Zaidi, P. H., Atlin, G. N., and Piepho, H. (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 (4), 1384–1391. doi:10.2135/cropsci2013.02.0073
- Koivuola, M., Strandén, I., Su, G., and Mäntysaari, E. A. (2012). Different methods to calculate genomic predictions—comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). *J. dairy Sci.* 95 (7), 4065–4073. doi:10.3168/jds.2011-4874
- Krause, M. D., Dias, K. O. G., Singh, A. K., and Beavis, W. D. (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, P. G., Quincke, M., Silva, P., and Gutiérrez, L. (2016). Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding program. *Crop Sci.* 56 (5), 2165–2179. doi:10.2135/cropsci2015.04.0207
- Lanubile, A., Maschietto, V., and Marocco, A. (2014). Breeding maize for resistance to mycotoxins. *Mycotoxin Reduct. Grain Chains*, 37–58. doi:10.1002/9781118832790.ch4
- Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi:10.3168/jds.2009-2061

- Legarra, A., Christensen, O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livest. Sci.* 166, 54–65. doi:10.1016/j.livsci.2014.04.029
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324
- Liu, Z., Seefried, F. R., Reinhardt, F., Rensing, S., Thaller, G., and Reents, R. (2011). Impacts of both residence population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet. Sel. Evol.* 43 (1), 19–9. doi:10.1186/1297-9686-43-19
- Lourenco, D. A. L., Misztal, I., Tsuruta, S., Aguilar, I., Lawlor, T. J., Forni, S., et al. (2014). Are evaluations of young animals benefiting from the past generations? *J. Dairy Sci.* 97 (6), 3930–3942. doi:10.3168/jds.2013-7769
- Lourenco, D. A. L., Tsuruta, S., Fragomeni, B. O., Masuda, Y., Aguilar, I., Legarra, A., et al. (2015). Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. *J. animal Sci.* 93 (6), 2653–2662. doi:10.2527/jas.2014-8836
- Macedo, F. L., Christensen, O. F., Astruc, J. M., Aguilar, I., Masuda, Y., and Legarra, A. (2020). Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genet. Sel. Evol.* 52 (1), 47–10. doi:10.1186/s12711-020-00567-1
- Mäntysaari, E. A., Koivula, M., and Strandén, I. (2020). Symposium review: single-step genomic evaluations in dairy cattle. *J. Dairy Sci.* 103 (6), 5314–5326. doi:10.3168/jds.2019-17754
- Martini, J. W. R., Schrauf, M. F., Garcia-Baccino, C. A., Pimentel, E. C. G., Munilla, S., Rogberg-Muñoz, A., 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), 16–19. doi:10.1186/s12711-018-0386-x
- Masuda, Y., Tsuruta, S., Bermann, M., Bradford, H. L., and Misztal, I. (2021). Comparison of models for missing pedigree in single-step genomic prediction. *J. Anim. Sci.* 99 (2), skab019–10. doi:10.1093/jas/skab019
- Melchinger, A. E., and Frisch, M. (2023). Genomic prediction in hybrid breeding: II. Reciprocal recurrent genomic selection with full-sib and half-sib families. *Theor. Appl. Genet.* 136 (9), 203. doi:10.1007/s00122-023-04446-3
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 (4), 1819–1829. doi:10.1093/genetics/157.4.1819
- Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92 (9), 4648–4655. doi:10.3168/jds.2009-2064
- Oliveira, A. A., Resende, M. F. R., Ferrão, L. F. V., Amadeu, R. R., Guimarães, L. J. M., Guimarães, C. T., et al. (2020). Genomic prediction applied to multiple traits and environments in second season maize hybrids. *Hered. (Edinb)* 125 (1–2), 60–72. doi:10.1038/s41437-020-0321-0
- Oliveira, C. M., Auad, A. M., Mendes, S. M., and Frizzas, M. R. (2014). Crop losses and the economic impact of insect pests on Brazilian agriculture. *Crop Prot.* 56, 50–54. doi:10.1016/j.cropro.2013.10.022
- Pádua, J. M. V., Das Graças Dias, K. O., Pastina, M. M., de Souza, J. C., Queiroz, V. A. V., da Costa, R. V., et al. (2016). A multi-environment trials diallel analysis provides insights on the inheritance of fumonisin contamination resistance in tropical maize. *Euphytica* 211 (3), 277–285. doi:10.1007/s10681-016-1722-2
- Pérez, P., and De Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198 (2), 483–495. doi:10.1534/genetics.114.164442
- Pérez-Rodríguez, P., Crossa, J., Rutkoski, J., Poland, J., Singh, R., Legarra, A., et al. (2017). Single-step genomic and pedigree genotype × environment interaction models for predicting wheat lines in international environments. *plant genome*. 10 (2). doi:10.3835/plantgenome2016.09.0089
- Persa, R., Grondona, M., and 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, 11, 932. doi:10.3390/agriculture11100932
- R Core Team (2022). R: a language and environment for statistical computing. Available online at: <https://www.r-project.org/>.
- Saghai-Marouf, M. A., Soliman, K. M., Jorgensen, R. A., and Allard, R. (1984). Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci.* 81 (24), 8014–8018. doi:10.1073/pnas.81.24.8014
- Santiago, R., Cao, A., Malvar, R. A., and Butrón, A. (2020). Genomics of maize resistance to fusarium ear rot and fumonisin contamination. *Toxins (Basel)*. 12 (7), 431. doi:10.3390/toxins12070431
- Silva, K. J., Guimarães, C. T., Guilhen, J. H. S., Guimarães, P. E. de O., Parentoni, S. N., Trindade, R. d. S., et al. (2020). High-density SNP-based genetic diversity and heterotic patterns of tropical maize breeding lines. *Crop Sci.* 60 (2), 779–787. doi:10.1002/csc2.20018
- Souza, V. F., Ribeiro, P. C. de O., Vieira Junior, I. C., Oliveira, I. C. M., Damasceno, C. M. B., Schaffert, R. E., et al. (2021). Exploring genotype environment interaction in sweet sorghum under tropical environments. *Agron. J.* 113 (4), 3005–3018. doi:10.1002/agj2.20696
- Tiezzi, F., de los Campos, G., Parker Gaddis, K. L., and Maltecca, C. (2017). Genotype by environment (climate) interaction improves genomic prediction for production traits in US Holstein cattle. *J. Dairy Sci.* 100 (3), 2042–2056. doi:10.3168/jds.2016-11543
- Ukrainetz, N. K., and Mansfield, S. D. (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, 64–13. doi:10.1007/s11295-020-01456-w
- USFDA (2021). Guidance for industry: fumonisin levels in human foods and animal feeds. *U. S. Food Drug Adm.*
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., et al. (2008). Reliability of genomic predictions for North American dairy bulls. *J. Dairy Sci.* 91 (Suppl. 1), 305. doi:10.3168/jds.2008-1514
- Velazco, J. G., Malosetti, M., Hunt, C. H., Mace, E. S., Jordan, D. R., and van Eeuwijk, F. A. (2019). Combining pedigree and genomic information to improve prediction quality: an example in sorghum. *Theor. Appl. Genet.* 132 (7), 2055–2067. doi:10.1007/s00122-019-03337-w
- Vitezica, Z.-G., Aguilar, I., Misztal, I., and Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genet. Res. (Camb)*. 93 (5), 357–366. doi:10.1017/S001667231100022X
- Zhang, A., Pérez-Rodríguez, P., San Vicente, F., Palacios-Rojas, N., Dhliwayo, T., Liu, Y., et al. (2022). Genomic prediction of the performance of hybrids and the combining abilities for line by tester trials in maize. *Crop J.* 10 (1), 109–116. doi:10.1016/j.cj.2021.04.007