Revista Brasileira de Milho e Sorgo
Brazilian Journal of Maize and Sorghum

ISSN 1980 - 6477
Journal homepage: www.abms.org.br/site/paginas

GENOMIC SELECTION AS A TOOL FOR MAIZE CULTIVARS DEVELOPMENT

Caillet Dornelles Marinho(1)(✉), Igor Ferreira Coelho(2,3), Marco Antônio Peixoto(2,3), Geraldo Afonso de Carvalho Junior(1) and Marcio Fernando Ribeiro de Resende Junior(3) (✉)

(1) Helix Sementes, Rio Claro, São Paulo
E-mail: caillet.marinho@agroceres.com,
g.acjunior@gmail.com

(2) Laboratório de Biometria, Universidade Federal de Viçosa
E-mail: igor.f.coelho@ufv.br,
marcopeixotom@gmail.com

(3) Horticultural Sciences Department, University of Florida, Gainesville
E-mail: mresende@ufl.edu

✉ Corresponding author

Abstract – The ability to predict genotypes that have not yet been tested is always a target of plant breeders. Over the last twenty years, many studies presented genomic selection (GS) as a tool contributing to this goal. Currently, many research papers have shown encouraging results in the application of GS. However, there are few examples of long-term, successful applications of GS in plant breeding programs. Furthermore, for breeders and researchers considering the application of GS, there are a series of important considerations on how to adapt a breeding program to maximize the benefit of GS, aiming to reduce the costs and maximize the genetic gains. Under this perspective, we present a review with a general view about applied GS in maize breeding, future perspectives of this technique, and an applied study case of a breeding program using GS. We attempt to provide a brief review of the literature with recent developments, as well as a discussion involving the number of markers required to deploy GS, the different statistical approaches to create GS models, the different ways to define training populations, and the incorporation of non-additive effects and genotype by environment interaction. We end with general recommendations and conclusions about some critical points about adopting GS in maize breeding.

Keywords: Predictive accuracy; training set; non-additive effects; genotype-by-environment interaction.

SELEÇÃO GENÔMICA COMO UMA FERRAMENTA PARA O DESENVOLVIMENTO DE CULTIVARES DE MILHO

Resumo - A capacidade de predição de genótipos futuros não testados em campo sempre foi um objetivo comum para melhoristas. Uma ferramenta que demonstra grande potencial nesse âmbito é a seleção genômica ampla (SGA). Embora os resultados apresentados na literatura representam fortes evidências em favor de incorporar a SGA aos programas de melhoramento de milho, exemplos do mundo real da implantação bem-sucedida ainda são escassos. Além disso, ainda existem questões sobre como converter ou adaptar um programa de melhoramento para maximizar o retorno do investimento na seleção genômica. Assim, apresentamos uma revisão com uma visão geral sobre a aplicação da seleção genômica em melhoramento de milho, um caso real de utilização em um programa de melhoramento privado e perspectivas futuras. Desta forma, exibimos uma breve revisão de literatura com desenvolvimentos recentes em SGA, além do número de marcas necessários para aplicar uma análise genômica de qualidade, métodos estatísticos adotados para criar modelos de SGA, maneiras diferentes de criar populações de treinamento, e adição de efeitos não-aditivos e de interação genótipos por ambientes nos modelos de predição genômicos. Ao final, destacamos alguns pontos importantes sobre a implementação de SGA em programas de melhoramento de milho.

Palavras-chave: Acurácia preditiva, População de treinamento; Efeitos não-aditivos; Interação genótipos por ambientes

How to cite

MARINHO, C. D.; COELHO, I. F.; PEIXOTO, M. A.; CARVALHO JUNIOR, G.; RESENDE JUNIOR, M. F. R. Genomic selection as a tool for maize cultivars development. **Revista Brasileira de Milho e Sorgo**, v. 21, e1285, 2022.

DOI: https://doi.org/10.18512/rbms2022v21e1285

Genomic-aided predictive methods, generally referred to as genomic selection or genomic prediction (hereafter GS), have constantly evolved since the early work of Meuwissen et al. (2001), who proposed the concept using simulations. The method's core idea lies in the possibility to predict future, unobserved phenotypes by combining high-density genotypic information and predictive modeling. The approach leads to higher genetic gains if molecular markers' prediction is more accurate than the conventional phenotypic prediction/selection. Furthermore, if the cost of genotyping is less than establishing field evaluation plots needed for phenotypic selection, the approach leads to cost savings. Finally, if GS enables faster identification of selection candidates, the approach reduces the breeding process time. Over the last 20 years, these premises have been demonstrated and validated for many different species and traits, making genomic selection a valuable toolkit in modern breeding.

Multiple studies have shown the potential of integrating genomic selection into maize breeding programs (SUPPLEMENTARY MATERIAL - Table 1). The literature has explored different prediction methods, different breeding designs, different traits, and many other ways in which genomic selection can complement traditional plant breeding systems (Crossa et al., 2017). The emergence of novel genotyping technologies has reduced genotyping turnaround time while reducing the cost of genotyping. The cheaper access to genomic information has favored the implementation of genomic selection (Galli et al., 2018). In parallel, the development of phenotyping technologies seeking precise phenotyping is increasing at an unprecedented rate. Such technologies bring extra power to modeling the link between traits and genomic information on crop improvement programs (Frietsch-Neto et al., 2021).

1. Genomic selection applied to maize breeding over the past 20 years.

Although these results encourage breeders to use GS in maize breeding programs, real-world examples of the successful deployment of genomic selection are still scarce publicly. Furthermore, questions still arise regarding how to convert or adapt a breeding program to maximize the return on the investment in genomic selection. With that in mind, this review aims to provide an overview of genomic selection applied to maize breeding. Thus, we summarize the history and results over the last twenty years since the proposed approach, the future perspectives and considerations of GS in maize breeding, and finally, we presented results from a private breeding program that successfully deployed GS.

1.1. The initial attempts to predict maize grain yield

Predicting future performance, directly or indirectly, is the goal of any plant breeder. In maize, the prediction of hybrid performance was first attempted in the 1930s, as Bernardo (2021) reviewed. Jenkins (1934) and Doxtator and Johson (1936) demonstrated the ability to predict double-cross hybrid performance based on the phenotypic performance of their single-cross parents (Jenkins, 1934; Doxtator & Johnson,

Table 1. Grain yield average of top twenty hybrids. The bold hybrids were selected based on genomic selection.

Hybrid	kg/ha†	Method
H01	**10484**	**Genomic selection**
H02	10294	Conventional approach
H03	**9908**	**Genomic selection**
H04	9776	Conventional approach
H05	9726	Conventional approach
H06	9674	Conventional approach
H07	9669	Conventional approach
H08	**9653**	**Genomic selection**
H09	**9562**	**Genomic selection**
H10	**9538**	**Genomic selection**
H11	**9505**	**Genomic selection**
H12	**9492**	**Genomic selection**
H13	**9487**	**Genomic selection**
H14	**9404**	**Genomic selection**
H15	**9384**	**Genomic selection**
H16	**9379**	**Genomic selection**
H17	9362	Conventional approach
H18	9356	Conventional approach
H19	**9350**	**Genomic selection**
H20	9332	Conventional approach

†: kilograms per hectare

1936). Since then, maize breeding and quantitative genetics have continued to evolve. Over the last decades, different procedures were proposed to utilize phenotypic data and experimental design to estimate and predict breeding values, genetic values, specific combining abilities, and general combining abilities (Griffing, 1962; Henderson, 1984). The first application of molecular markers to predict a complex trait was proposed by Bernardo (1994) before Meuwissen et al. (2001). Bernardo (1994) utilized a set of 220 RFLP (restriction fragment length polymorphism) loci to build a variance-covariance matrix that estimated genomic relationship. Using this matrix and a mixed model approach, the author presented the correlation among predicted and observed yield in single crosses equal to 0.80. The approach concept was very similar to what was later referred to as GBLUP (genomic best linear unbiased prediction), proposed by VanRaden

(2008) as an alternative GS method.

1.2. What is Genomic Selection

For many decades, plant and animal breeders have sought to create approaches to predict quantitative traits using molecular markers. The primary motivation has been the potential increase in selection gain, especially for traits with low prediction accuracy or complex to evaluate using standard breeding methods. Initial approaches that use DNA information for breeding, generally named Marker Assisted Selection (MAS), relied initially on identifying polymorphisms on genes affecting the traits or markers linked adjacent to these genes. In 2001, Meuwissen et al. (2001) proposed a different method to perform MAS and predict quantitative traits, referred to as genomic selection. Contrary to traditional MAS, which aims to predict traits based on knowledge of their specific functional polymorphisms, in GS, the phenotype is modeled as a function of a large number of DNA markers distributed across the genome. In doing so, the authors hypothesized that every QTL affecting the trait would have its effect captured by at least one marker in linkage disequilibrium (LD) with it. Since it was proposed, new research demonstrated that when all the markers are modeled simultaneously, even markers that are not in LD with the QTL contribute to the trait prediction by capturing average genetic relationships among the individuals of the breeding population (Zhong et al., 2009; Daetwyler et al., 2013). From a statistical point of view, the main principle of GS lies in that all markers are modeled jointly (Meuwissen et al., 2001) to estimate genomic estimated breeding values (GEBV). In brief, the process develops and validates a model by including different genetic effects. The next step consists of training and validating this model with phenotypic data of a specific set of the dataset. Afterward, the model is used to predict tested and untested genotypes, preferable genotypes related to the ones used in training and validating sets during modeling (Crossa et al., 2017).

Some studies also have compared GS over phenotypic selection to check the importance of the genomic approach. A bi-parental population study encountered a higher gain by adopting GS instead of only phenotypic selection (PS). They compared two contrasting environments to assess drought tolerance. The drought stress environment raised the gain per cycle from 0.27 (PS) to 0.50 (GS) t/ha/cycle. Under optimal conditions, the gain increased from 0.34 to 0.55 t/ha/cycle when adopting GS (Vivek et al., 2017). Moro et al. (2019) investigated the superiority of GS over PS when using selection indices in maize. The authors concluded that there is a high potential of using selection indexes with genomic data. Selection in early stages permitted to increase the selection intensity and to fit two breeding cycles in a year. In 2014, a study compared GS and phenotypic BLUP using pedigree information and evaluated the influence of different breeding cycles on the predictive ability. The GS predictions were superior to the phenotypic BLUP in all scenarios and traits (Albrecht et al., 2014).

1.3. GS in maize and considerations studies in the last two decades.

Prediction methods

The basic framework for the GS in maize was implementing the linear mixed models. However, the application of GS typically requires mid- to high-density marker panels. This procedure creates a challenging situation where the number of parameters to be adjusted in the model is often more significant than the number of phenotypic observations collected, also denoted as $p>>n$. An approach to address these limitations is to model the markers as random effects and apply regularization, either in the form of shrinkage, dimensionality reduction methods, or a combination of both (De Los Campos et al., 2013; Knürr et al., 2013). One of these approaches, the ridge regression BLUP (RRBLUP), shrinks the estimates of the marker effects using a penalized least-square based on the L2 penalty to solve the regression problem. In addition, several variations of a hierarchical Bayesian linear regression were proposed, where the regularization is set up by imposing different priors (Gianola, 2013).

A good review of the different methods is provided by De Los Campos et al. (2013). After 20 years of method comparison in different species and traits, it is now well accepted that the RRBLUP has predictive performance similar to all other methods (Resende et al., 2012a). This conclusion was also achieved by a recent study in maize that compared parametric methods (RRBLUP, elastic net, least absolute shrinkage and selector operator, Bayes B, Bayes C, and reproducing kernel Hilbert Space - RKHS) and non-parametric (random forest and support vector machine) methods (Li et al., 2020). Our recommendation for breeders evaluating GS for the first time in their population is to assess the performance of three different methods: GBLUP/RR-BLUP, RKHS, and BayesB. The BayesB method was superior in a few specific cases where the genetic inheritance is oligogenic (Resende et al., 2012a; Almeida et al., 2016). RKHS was superior in instances in which non-additive effects were high. If no difference between these methods is perceived, the application of GBLUP/RR-BLUP is more accessible and less computationally intensive than other methods.

Recently, machine learning methods have also been applied to genomic selection in maize breeding (Montesinos-López et al., 2021). Different methods have been evaluated, including deep learning and multiple perceptron layers approaches (Rachmatia et al., 2017; Montesinos-López et al., 2018a, 2018b; Zingaretti et al., 2020). Nonetheless, it is worth noting that current machine learning implementations have not shown minor superior to conventional approaches. One hypothesis for these results is that the training datasets used to calibrate the models are relatively small. Future research is yet to demonstrate if higher prediction accuracies are possible when larger training datasets are available.

Training set

Genomic prediction models the relationship between breeding individuals and exploits the *linkage disequilibrium* between a marker and a region of the chromosome that controls a specific trait assessed (Heslot et al., 2015). However, to achieve effectiveness, three premises should be considered: (1) the use of markers with good coverage of the genome; (2) high-quality phenotyping in the training population; and (3) relationship between the individuals from the training population and the individuals from the predicted population.

Before building a prediction model, the training population needs to be phenotyped and genotyped. It is now well established that the model's prediction accuracy is highly affected by the training population used in the model calibration (Albrecht et al., 2011; Pszczola et al., 2012). Nowadays, establishing the training datasets (or calibration sets) that optimize the efficiency of GS is one of the critical considerations before deploying GS operationally (Rincent et al., 2012). The training set in maize is usually composed of inbred lines due to the easiness to genotype fully homozygous materials. However, deciding what forms the training set will need to balance the number of individuals in the training set, the relationship of the training set with the breeding population where GS will be applied, and the number of markers (or marker platform) utilized. All of these factors will affect the prediction accuracy of the model and the total investment required for GS.

The initial work to investigate the optimum training set focused on balancing training size and marker panel size (Maenhout et al., 2010; Rincent et al., 2012). The simulation provided evidence that incorporating disconnected individuals into the training set was not beneficial to the model. The work also attempted to maximize the prediction accuracy by using different training populations. The authors evaluated training populations that maximized genetic diversity, populations that minimized the mean of the prediction error variance of the contrast between non-phenotyped individuals and the mean of the panel, and populations that maximized the mean of the coefficient of determination of the contrast between non-phenotyped individuals and the mean of the panel. The optimization approaches tested superior to randomly selected training populations, highlighting the opportunity to increase prediction accuracy during the planning phase (deciding which individuals will get genotyped). Similar results with slightly different optimization schemes were later demonstrated by Akdemir et al. (2015). The importance of the relationship between the training set and the population where GS will be deployed was also shown with real maize data (Windhausen et al., 2012; Mastrodomenico et al., 2019). Both studies obtained low predictive accuracy due to the low genetic relationship between populations. In contrast, Li et al. (2021) evaluated the inclusion of related individuals into the training population and found a significant increase in the predictive accuracy.

Another important consideration when deciding the training population is how to leverage multi-year experiments to calibrate the GS model. This parameter is particularly relevant once GS has been partially implemented while phenotypic data is still collected on individuals genotyped. Hence, the breeder has the opportunity to decide how to combine the phenotypic data collected over multiple years into a single prediction model. An eight-year study developed in a maize breeding program concluded that increasing the training population size does not necessarily raise the predictive accuracy. The authors identified it as an obstacle to predicting specific cycles due to the introgression of new sources of diversity into the program. Thus, the best predictive performance was obtained using two previous cycles to predict the following (Dias et al., 2020) rather than utilizing all the available data. Wang et al. (2020) also studied the use of different cycles in the training population of the GS models and reached similar results. These studies stand out the importance of size and structure in a training population and highlight how this step is essential for GS success.

2. Optimization of genomic selection in maize breeding

2.1. Additive and non-additive effects in genomic selection

Total genetic effects can be partitioned into additive and non-additive, and the non-additive effect can be further partitioned into dominance end epistasis (Falconer & Mackay, 1996). Genomic selection models were initially modeled to predict the additive effect, which was accomplished by parameterizing the DNA markers with the additive dosage model (*e.g*., 0, 1, and 2). A natural expansion of these models was incorporating non-additive effects as genomic relationship matrices that capture non-additive effects (Vitezica et al., 2013; Muñoz et al., 2014) or direct marker estimates (Wellmann & Bennewitz, 2012).

Maize breeding has traditionally benefitted from non-additive variance and designed the breeding pipeline to capitalize on hybrid vigor and specific combining abilities. Hence, GS models that predicted hybrid yield performance were more accurate when dominance was included in the model (Alves et al., 2019). Similar results were reported by Dias et al. (2020), where the predictive ability for grain yield increased by over 15% in the model that included additive and dominance effects (A + D). However, it is worth noting that A + D models did not increase the prediction accuracy of ear height and plant height (Alves et al., 2019), providing evidence of a more negligible contribution of non-additive effects to those traits. Implementing an RKHS model for hybrid prediction, Li et al. (2020) demonstrated that A + D models were more efficient when relatively high non-additive variance. However, for the eleven traits assessed in the study, improvement in prediction was observed only for grain yield. Ferrão et al. (2020) also showed an increase of 30% in the predictive accuracy by adopting the A + D model for grain yield, whereas no increment was observed for grain moisture. In summary, when GS is being explored for a new species, population or trait, the general recommendation is to evaluate the predictive ability of models that include A + D. Including dominance effects is advised in predictions of maize grain yield. However, for traits where prediction accuracies are equivalent to the additive models, the additive model would then be the easiest way to deploy GS, and no extra complexity is required in the model.

2.2. Optimization of genomic prediction by including G × E interaction effect

The genotype-by-environment (G × E) interaction is an effect well known to any breeder, increasing the complexity of predicting variety performance in maize breeding programs. Similarly, the G × E interaction also affects the predictive ability of genomic selection models because the "ground-truth" phenotype used to calibrate the model in a given environment will not have the same magnitude or ranking in another environment under G × E. Hence, the resulting GS model will not yield high predictive accuracies. This point was well demonstrated in clonal plant populations planted in different environments (Resende et al., 2012b). In maize, GS models that explicitly included G × E interaction as a random effect in a GBLUP model increased the predictive ability of hybrid performance (Acosta-Pech et al., 2017; Dias et al., 2018; Rogers et al., 2021). Another study modeling Zn concentration for maize production also indicated an improvement in the predictive models when they included the G × E interaction effect (Mageto et al., 2020).

Multi-environment GS models have also been evaluated jointly with multi-trait approaches. Multi-trait models are particularly relevant for traits with many missing data or lack of precision in the phenotyping. Using such an approach, Oliveira et al. (2020) observed the superiority of the multi-trait models and similar accuracies in a single environment and multi-environment models. However, while the overall estimates of prediction accuracy were similar, multi-environment GS models improved the prediction accuracy for new, untested hybrids. Genotype-by-environment models have also been evaluated together with the incorporation of dominance effects. In a joint GS analysis of test-cross hybrids that included non-additive effects and G × E interactions, dominance effects were relevant for grain yield, days to silking, and plant height. In the case of G × E interaction, the effect was critical only for grain yield (Ramstein et al., 2020). Ferrão et al. (2020) analyzed the inclusion of the G × E interaction effect and concluded that both dominance and G × E interaction effects contributed to the improvement of the genomic prediction. In addition, Rogers et al. (2021) presented evidence that the inclusion of dominance improved the prediction accuracy by 7-10% for grain yield compared with models accounting for only additive effects when the models accounted for the G × E interaction effect. These and other results show that G × E interaction should be included in the calibration and deployment of GS models.

2.3. Genomic selection applied to double haploid selection and assessment

Double haploid production is currently the basis of field corn breeding programs. In summary, the creation of haploids plants is induced, and haploid kernels are identified, typically by a color marker. The haploid plants have their chromosome doubled, and the resulting plant is a double haploid, diploid, inbred line (Geiger & Gordillo, 2009). The double haploid system is an excellent complement to GS (Messina et al., 2018) since the availability of genotypic data in inbred lines enables the creation of *in-silico* hybrids for every combination of crosses without the need to genotype each segregating individual independently. This process vastly reduces the overall cost of deploying GS. The second application of interest in the interface of DH production and GS is the use of prediction models to select the inducers with superior performance (adaptation to target environment and induction rate). Almeida et al. (2020) considered the haploid induction rate a trait and calibrated a GS model. Those authors observed high predictive abilities and indicated that GS could be successfully applied for inducer development, accelerating and increasing genetic gain in a costly and time-consuming trait to screen.

2.4. Genomic selection applied to public breeding programs

In general, modern public field corn breeding programs are heavily focused on pre-breeding efforts. While some programs have successfully developed inbred lines for niche markets or specific environments, a significant component of public breeding research revolves around identifying and characterizing germplasm with unique traits or adaptation profiles. Genomic selection can accelerate the targeted incorporation of landraces and other germplasm into an inbred/hybrid development pipeline, as demonstrated by simulation (Gorjanc et al., 2016) and using data from diverse accessions (Yu et al., 2016, 2020). Furthermore, recent work investigated introducing new lines, from a public germplasm bank, into a private program germplasm bank with elite lines (Allier et al., 2020), using GS as the proposed approach. The authors found considerably high predictive abilities and concluded that GS could be helpful as a pre-breeding approach to introduce targeted genetic diversity and define the resources that can further improve elite lines. Hence, we note that in addition to increasing genetic gains in inbred development breeding pipelines, similar success is likely to be obtained in developing elite populations or pre-breeding efforts with exotic germplasm.

2.5. Insights from a private commercial breeding program

The application of GS can fit in commercial breeding programs of any size, and its deployment will depend on many factors, including the resources available. Herein we describe a case of GS implementation in a mid-size breeding program owned by Helix Sementes e Mudas Ltda (owner of the commercial brands Sementes Biomatrix e Santa Helena Sementes) in Brazil. Helix started incorporating GS into its breeding pipeline in 2015. The initial strategy was to run a pilot project to test the viability of implementing GS by focusing on target environments relevant to the second maize cropping season in Brazil, also known as “safrinha.” For that, 128 inbred lines were genotyped with approximately 616,000 SNP markers. At the same time, a training population was developed by characterizing 450 single-cross hybrids across five locations within the targeted mega-environment.

Grain yield prediction modeling was developed for all possible combinations among the 128 lines, including only additive effects. The first 274 hybrids with the highest yield predicted performances were selected for field testing in the next cycle. Additionally, 1360 hybrids developed using the traditional breeding pipeline were added to the trials. Hence, 1634 hybrids were tested in the pilot project, in which 17% came from genomic prediction and 83%

from the conventional approach (Figure 1). The field experiments were planted and evaluated at the same environments for developing the training population, except that one location was lost due to climate constraints. Grain yield was measured and adjusted for 13% moisture. Average grain yield was estimated, and a 1.5% selection threshold was applied, resulting in the list of selected hybrids below sorted by grain yield. Interestingly, the genomic prediction approach generated 12 out of the 20 hybrids; in other words, 60% of the chosen hybrids with the highest yield performance (Table 1). These results initially suggested a potential to incorporate GS into the conventional breeding pipeline adopted by the company.

Using the same methodology described before, we investigated the effects of adding kinship information into the modeling estimation and materials selection in the next cycle. A set of 52 hybrids from the training population was selected by predicting the most productive hybrids in grain yield using GBLUP (Figure 2 and Table 2). In parallel, an independent set of 138 combinations were selected based solely on BLUP estimations (no kinship information). These groups totalized 190 hybrids planted and evaluated for grain yield in the next cycle at the same sites described previously, including two more locations, totalizing seven sites. The phenotypic average grain yield across all locations showed that 13 out of the 20 best performing hybrids came from the set of hybrids derived from the GBLUP approach. This result

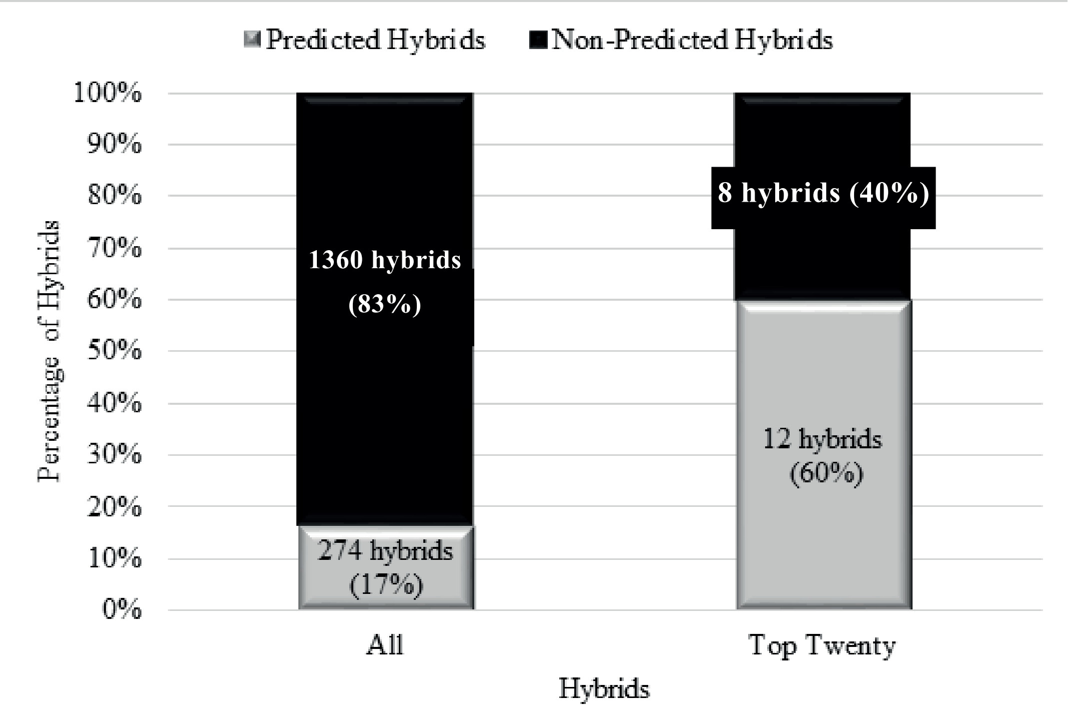

Figure 1. Total number of tested and selected hybrids (predicted x non-predicted).

Table 2. Average grain yield of top twenty hybrids. The bold hybrids were selected based on genomic selection.

Hybrid	kg/ha †	Method
H01	**9939**	**GBLUP**
H02	9833	BLUP
H03	9789	BLUP
H04	9778	BLUP
H05	**9721**	**GBLUP**
H06	**9718**	**GBLUP**
H07	**9534**	**GBLUP**
H08	9502	BLUP
H09	**9500**	**GBLUP**
H10	9483	BLUP
H11	**9452**	**GBLUP**
H12	**9445**	**GBLUP**
H13	**9430**	**GBLUP**
H14	**9428**	**GBLUP**
H15	**9408**	**GBLUP**
H16	**9405**	**GBLUP**
H17	**9375**	**GBLUP**
H18	9328	BLUP
H19	**9316**	**GBLUP**
H20	9311	BLUP

†: kilograms per hectare. BLUP: Best Linear Unbiased Prediction. GBLUP: Best Linear Unbiased Prediction.

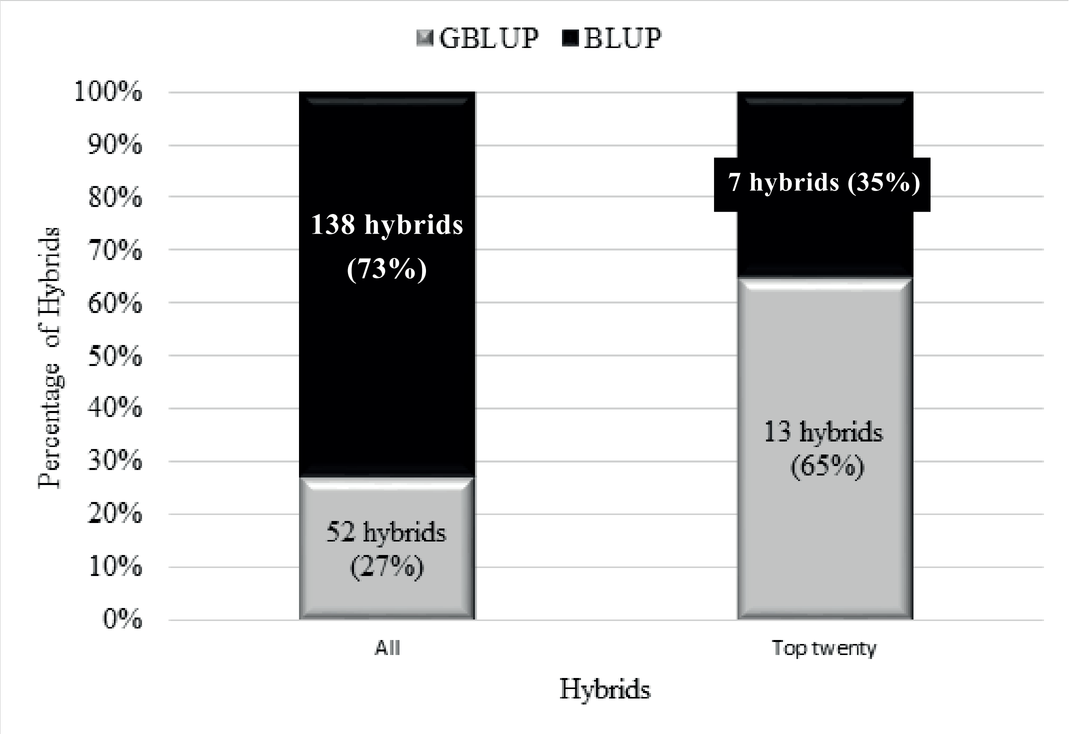

Figure 2. Total number of tested and selected hybrids. BLUP: Best Linear Unbiased Prediction; GBLUP: Genomic Best Linear Unbiased Prediction.

demonstrated that adding kinship information improved the prediction and selection efficiency.

Lastly, a hypothesis on whether increasing the number of molecular markers (SNPs) in the calibration modeling phase improves the predictive ability was tested for grain yield, grain moisture, plant height, and ear height. As demonstrated in Figure 3, there was no evident indication that increasing the numbers of markers altered the predictive ability for all traits evaluated. It is crucial to highlight that the minimum number of markers, in this case, was 10,000 markers increasing to up to 550,000 markers.

The results from this pilot project corroborated what the literature had suggested since the evolution of genomic selection in maize breeding. A good indication is that genomic prediction is a helpful tool in maize breeding. Unfortunately, our discussion and published data are limited due to companies' restrictions on sharing data and other relevant information. Nevertheless, GS was fully incorporated into the company's breeding pipeline since the development of this pilot project with some tweaks and improvements. Interestingly, three hybrids selected via genomic selection in this pilot study have recently reached the commercial phase with decent levels of adoption so far. This result reinforces GS's effectiveness as a tool to any breeding program.

3. Ongoing and future research developments

3.1. Integration of environmental covariates into genomic selection models

After 20 years since its inception, the field of GS is now relatively well established, and its potential has been demonstrated in many different crops and populations. Here we provide additional evidence in a commercial breeding program of how GS was utilized and selections moved forward until the commercial phase. In the following two sections, we provide a brief discussion of two recently emerging research fields proposed to further advance the application of genomic data into breeding programs. The first one uses environmental covariates to improve the predictive ability in untested environments. The use of genomic data combined with high-resolution environmental characterization (envirotyping) brings novel opportunities to i) utilize high-dimensional GS to predict phenotypes of non-tested genotypes in novel environments for which environmental characterization is available or predicted; ii) integrate genomics, environmental and phenotypic variables measured in distinct stages of crop development, and physiological understanding of crop growth processes to model a broad spectrum of interactions between genotypes and environments (Technow et al., 2015; Cooper et al., 2016); iii) integrate models using imagery technology associated with high-throughput phenotyping and genomic information in early testing generations to accelerate genetic gain.

Envirotyping describes, in detail, environmental factors that affect the development and productivity of crops in multi-environment trials and groups of breeding target environments (Cooper et al., 2016; Xu, 2016). On the other hand, growth models or crop growth models use quantitative descriptors of eco-physiological process, soil chemistry, agricultural climatology, crop management, and areas related to predicting crop growth, crop development, and crop yield of plant species in a given environment (Hodson & White, 2010). Because of that, they can explain the impact of G x E interation and certain types of non-additive effects on the expression of a phenotype.

Genomic selection models grounded on reaction norms using G x E interation and environmental covariables (Jarquín et al., 2014) can increase the prediction accuracy by 7 to 20% depending on the species and trait studied (Crossa et al., 2017; Pérez-Rodríguez et al., 2017). Also, it allows the prediction of G x E interation in non-tested environments. Similarly, by adding plant stress covariables extracted from growth models using a large group of environmental variables, Heslot et al. (2014) demonstrated that the prediction accuracy of yield performance of wheat lines at non-tested environments increased by 11%, on average. In maize, a study by Technow et al. (2015) has shown a considerable increase in the capacity of predicting grain yield when using crop growth models and environmental variables simultaneously in the prediction process. For instance, the integration of growth models and GS in comparison with a GBLUP model increased (on average) the predictive accuracy by 32% when the estimated predictions were performed in the same environment and up to 500% when predictions were performed for non-tested sites. Although these results are exciting, extensive research is still needed to evaluate the behavior of such integrative GS models in different traits, genetic backgrounds, agroecosystems, and breeding pipelines. We expect future developments to validate further and improve these models, enabling the selection of high-performance genotypes that have not yet been tested in novel target environments.

3.2. Using simulation to optimize breeding programs accounting for GS

The field of computer simulation is a second field that has observed a resurgence in research interest, and that can significantly benefit the application of GS into specific breeding pipelines. In general, the size of maize breeding programs has increased in the last decades, and an extraordinary number of genotypes have been tested, requiring labor, field, and time resources. As the cost of these endeavors also increases, the interest and demand in precisely optimizing and maximizing genetic gain also increase. Because of that, a field that has re-emerged in the last few years is the use of computer simulation as an inexpensive, flexible, rapid, and reliable alternative to optimize breeding programs (Batista et al., 2021).

Also known as *in silico* experiments, simulation studies can design new breeding designs, calculate the return on the investment, or investigate how breeder’s choices affect the breeder’s equation (Gaynor et al., 2017). The advantages of this approach are that it can evaluate

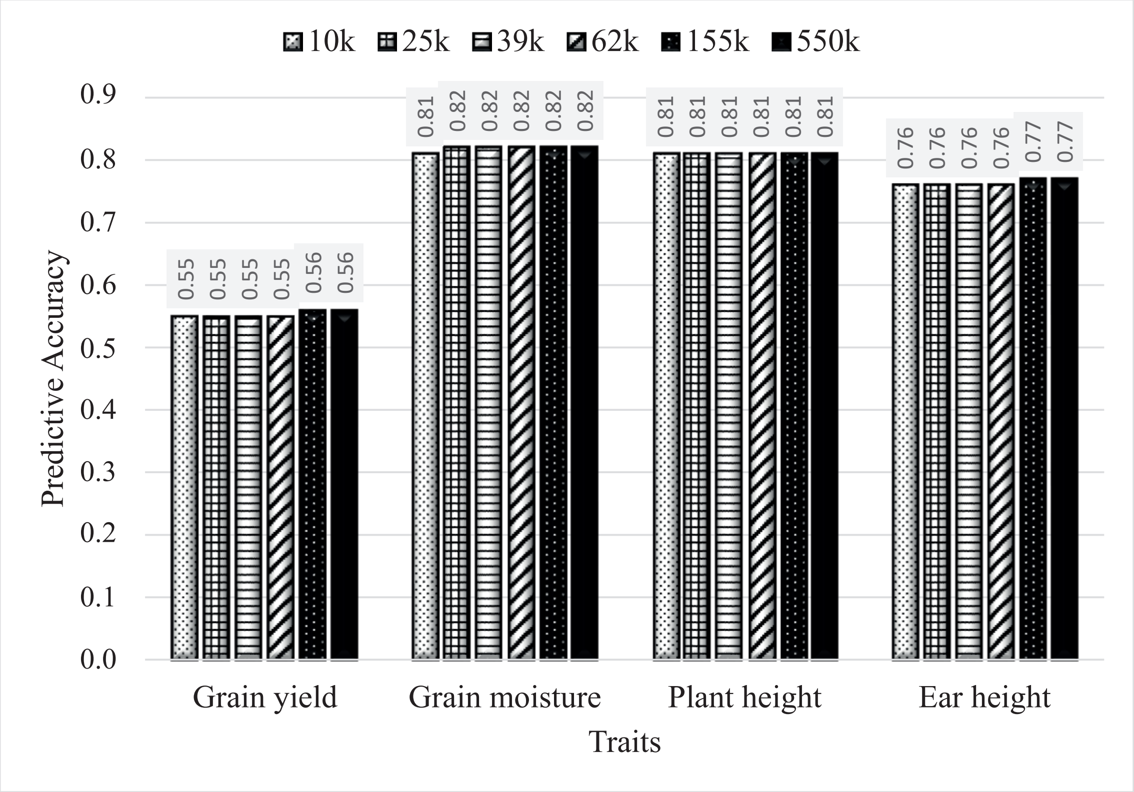

Figure 3. Predictive accuracy versus the number of makers for the traits grain yield, grain moisture, plant height, and ear height.

larger sets of these studied factors (*e.g.*, population size, replications, environments) (Andorf et al., 2019), as well as optimize the breeding program design by testing the genotypes` matting/crosses, and their outputs (Faux et al., 2016). The long-term performance of a specific strategy (pedigree or genomic relationship information) adopted in the early stages of the breeding pipeline was estimated using this approach (Gorjanc et al., 2018; Cowling et al., 2020). However, the number of questions that can be asked and the different parameters in a breeding program that can be optimized are extensive. Hence, we expect that more research in this field will continue over the next few years, and outcomes from computer simulation will likely help shape changes in breeding programs that optimize the process.

Conclusions

In conclusion, the field of genomic selection in the last twenty years has significantly expanded. GS as a tool is now well established, validated, and has been commercially deployed in many breeding programs of maize and other plant and animal species. In this review, our goal was to demonstrate preliminary results of such an initiative in a commercial breeding program and highlight key factors that need to be considered when applying GS. We also expect the field of predictive breeding to continue to evolve and expand as decision tools that complement field evaluations.

Acknowledgments

We appreciate Helix Sementes e Mudas Ltda for supporting our topic discussion with relevant results of its maize breeding program. Also, we appreciate the financial support of the Brazilian Government offered by the National Council for Scientific and Technological Development (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) - Finance Code 001, and the National Institute of Food and Agriculture (SCRI 2018-51181-28419 and AFRI 2019-05410 to MFRR)

References

ACOSTA-PECH, R.; CROSSA, J.; DE LOS CAMPOS, G.; TEYSSÈDRE, S.; CLAUSTRES, B.; PÉREZ-ELIZALDE, S.; PÉREZ-RODRÍGUEZ, P. Genomic models with genotype × environment interaction for predicting hybrid performance: an application in maize hybrids. **Theoretical and Applied Genetics**, v. 130, n. 7, p. 1431-1440, 2017. DOI: https://doi.org/10.1007/s00122-017-2898-0.

AKDEMIR, D.; SANCHEZ, J. I.; JANNINK, J. Optimization of genomic selection training populations with a genetic algorithm. **Genetics Selection Evolution**, v. 47, article 38, 2015. DOI: https://doi.org/10.1186/s12711-015-0116-6.

ALBRECHT, T.; AUINGER, H.-J.; WIMMER, V.; OGUTU, J. O.; KNAAK, C.; OUZUNOVA, M.; PIEPHO, H.-P.; SCHÖN, C.-C. Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. **Theoretical and Applied Genetics**, v. 127, n. 6, p. 1375-1386, 2014. DOI: https://doi.org/10.1007/s00122-014-2305-z.

ALBRECHT, T.; WIMMER, V.; AUINGER, H.-J.; ERBE, M.; KNAAK, C.; OUZUNOVA, M.; SIMIANER, H.; SCHÖN, C.-C. Genome-based prediction of testcross values in maize. **Theoretical and Applied Genetics**, v. 123, n. 2, article 339, 2011. DOI: https://doi.org/10.1007/s00122-011-1587-7.

ALLIER, A.; TEYSSÈDRE, S.; LEHERMEIER, C.; CHARCOSSET, A.; MOREAU, L. Genomic prediction with a maize collaborative panel: identification of genetic resources to enrich elite breeding programs. **Theoretical and Applied Genetics**, v. 133, n. 1, p. 201-215, 2020. DOI: https://doi.org/10.1007/s00122-019-03451-9.

ALMEIDA, J. E. de; GUIMARÃES, J.; SILVA, F. E.; RESENDE, M.; MUÑOZ, P.; KIRST, M.; RESENDE, M. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. **Heredity**, v. 117, p. 33-41, 2016. DOI: https://doi.org/10.1038/hdy.2016.23.

ALMEIDA, V. C.; TRENTIN, H. U.; FREI, U. K.; LÜBBERSTEDT, T. Genomic prediction of maternal haploid induction rate in maize. **Plant Genome**, v. 13, n. 1, e20014, 2020. DOI: https://doi.org/10.1002/tpg2.20014.

ALVES, F. C.; GRANATO, I. S. C.; GALLI, G.; LYRA, D. H.; FRITSCHE-NETO, R.; DE LOS CAMPOS, G. Bayesian analysis and prediction of hybrid performance. **Plant Methods**, v. 15, n. 1, article 14, 2019. DOI: https://doi.org/10.1186/s13007-019-0388-x.

ANDORF, C.; BEAVIS, W. D.; HUFFORD, M.; SMITH, S.; SUZA, W. P.; WANG, K.; WOODHOUSE, M.; YU, J.; LÜBBERSTEDT, T. Technological advances in maize breeding: past, present and future. **Theoretical and Applied Genetics**, v. 132, n. 3, p. 817-849, 2019. DOI: https://doi.org/10.1007/s00122-019-03306-3.

BATISTA, L. G.; GAYNOR, R. C.; MARGARIDO, G. R. A.; BYRNE, T.; AMER, P.; GORJANC, G.; HICKEY, J. M. Long-term comparison between index selection and optimal independent culling in plant breeding programs with genomic prediction. **PLoS ONE**, v. 16, n. 5, e0235554, 2021. DOI: https://doi.org/10.1371/journal.pone.0235554.

BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science**, v. 34, n. 1, p. 20-25, 1994. DOI: https://doi.org/10.2135/cropsci1994.0011183X003400010003x.

BERNARDO, R. Predictive breeding in maize during the last 90 years. **Crop Science**, v. 61, n. 5, p. 2872-2881, 2021. DOI: https://doi.org/10.1002/csc2.20529.

COOPER, M.; TECHNOW, F.; MESSINA, C.; GHO, C.; TOTIR, L. R. Use of crop growth models with whole-genome prediction: application to a maize multienvironment trial. **Crop Science**, v. 56, n. 5, p. 2141-2156, 2016. DOI: https://doi.org/10.2135/cropsci2015.08.0512

COWLING, W. A.; GAYNOR, R. C.; ANTOLÍN, R.; GORJANC, G.; EDWARDS, S. M.; POWELL, O.; HICKEY, J. M. *In silico* simulation of future hybrid performance to evaluate heterotic pool formation in a self-pollinating crop. **Scientific Reports**, v. 10, n. 1, article 4037, 2020. DOI: https://doi.org/10.1038/s41598-020-61031-0.

CROSSA, J.; PÉREZ-RODRÍGUEZ, P.; CUEVAS, J.; MONTESINOS-LÓPEZ, O.; JARQUÍN, D.; DE LOS CAMPOS, G.; BURGUEÑO, J.; GONZÁLEZ-CAMACHO, J. M.; PÉREZ-ELIZALDE, S.; BEYENE, Y.; DREISIGACKER, S.; SINGH, R.; ZHANG, X.; GOWDA, M.; ROORKIWAL, M.; RUTKOSKI, J.; VARSHNEY, R. K. Genomic selection in plant breeding: methods, models, and perspectives. **Trends in Plant Science**, v. 22, n. 11, p. 961-975, 2017. DOI: https://doi.org/10.1016/j.tplants.2017.08.011.

DAETWYLER, H. D.; CALUS, M. P. L.; PONG-WONG, R.; DE LOS CAMPOS, G.; HICKEY, J. M. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, v. 193, n. 2, p. 347-365, 2013. DOI: https://doi.org/10.1534/genetics.112.147983.

DE LOS CAMPOS, G.; VAZQUEZ, A. I.; FERNANDO, R.; KLIMENTIDIS, Y. C.; SORENSEN, D. Prediction of complex human traits using the genomic best linear unbiased predictor. **PLoS Genetics**, v. 9, n. 7, e1003608, 2013. DOI: https://doi.org/10.1371/journal.pgen.1003608.

DIAS, K. O. G.; GEZAN, S. A.; GUIMARÃES, C. T.; NAZARIAN, A.; SILVA, L. C.; PARENTONI, S. N.; GUIMARÃES, P. E.; ANONI, C. O.; PÁDUA, J. M. V.; PINTO, M. O.; NODA, R. W.; RIBEIRO, C. A. G.; MAGALHÃES, J. V.; GARCIA, A. A. F.; SOUZA, J. C.; GUIMARÃES, L. J. M.; PASTINA, M. M. Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. **Heredity**, v. 121, n. 1, p. 24-37, 2018. DOI: https://doi.org/10.1038/s41437-018-0053-6.

DIAS, K. O. G.; PIEPHO, H. P.; GUIMARÃES, L. J. M.; GUIMARÃES, P. E. O.; PARENTONI, S. N.; PINTO, M. O.; NODA, R. W.; MAGALHÃES, J. V.; GUIMARÃES, C. T.; GARCIA, A. A. F.; PASTINA, M. M. Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. **Theoretical and Applied Genetics**, v. 133, n. 2, p. 443-455, 2020. DOI: https://doi.org/10.1007/s00122-019-03475-1.

DOXTATOR, C. W.; JOHNSON, I. J. Prediction of double cross yields in corn. **Agronomy Journal**, v. 28, n. 6, p. 460-462, 1936. DOI: https://doi.org/10.2134/agronj1936.00021962002800060008x.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4. ed. Boston: Pearson, 1996. 280 p.

FAUX, A.; GORJANC, G.; GAYNOR, R. C.; BATTAGIN, M.; EDWARDS, S. M.; WILSON, D. L.; HEARNE, S. J.; GONEN, S.; HICKEY, J. M. AlphaSim: software for breeding program simulation. **The Plant Genome**, v. 9, n. 3, plantgenome2016-02, 2016. DOI: https://doi.org/10.3835/plantgenome2016.02.0013.

FERRÃO, L. F. V.; MARINHO, C. D.; MUÑOZ, P. R.; RESENDE, M. F. R. Improvement of predictive ability in maize hybrids by including dominance effects and marker × environment models. **Crop Science**, v. 60, n. 2, p. 666-677, 2020. DOI: http://dx.doi.org/10.1002/csc2.20096.

GALLI, G.; LYRA, D. H.; ALVES, F. C.; GRANATO, I. S. C.; SOUSA, M. B.; FRITSCHE-NETO, R. Impact of phenotypic correction method and missing phenotypic data on genomic prediction of maize hybrids. **Crop Science**, v. 58, n. 4, p. 1481-1491, 2018. DOI: https://doi.org/10.2135/cropsci2017.07.0459.

GAYNOR, R. C.; GORJANC, G.; BENTLEY, A. R.; OBER, E. S.; HOWELL, P.; JACKSON, R.; MACKAY, I. J.; HICKEY, J. M. A two-part strategy for using genomic selection to develop inbred lines. **Crop Science**, v. 57, n. 5, p. 2372-2386, 2017. DOI: https://doi.org/10.2135/cropsci2016.09.0742.

GEIGER, H. H.; GORDILLO, G. A. Doubled haploids in hybrid maize breeding. **Maydica**, v. 54, n. 4, p. 485-499, 2009.

GIANOLA, D. Priors in whole-genome regression: the bayesian alphabet returns. **Genetics**, v. 194, n. 3, p. 573-596, 2013. DOI: https://doi.org/10.1534/genetics.113.151753.

GORJANC, G.; GAYNOR, R. C.; HICKEY, J. M. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. **Theoretical and Applied Genetics**, v. 131, n. 9, p. 1953-1966, 2018. DOI: https://doi.org/10.1007/s00122-018-3125-3.

GORJANC, G.; JENKO, J.; HEARNE, S. J.; HICKEY, J. M. Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. **BMC Genomics**, v. 17, n. 1, p. 1-15, 2016. DOI: https://doi.org/10.1186/s12864-015-2345-z.

GRIFFING, B. Prediction formulae for general combining ability selection methods utilizing one or two random-mating populations. **Australian Journal of Biological Sciences**, v. 15, n. 4, p. 650-665, 1962. DOI: https://doi.org/10.1071/BI9620650.

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

HESLOT, N.; AKDEMIR, D.; SORRELLS, M. E.; JANNINK, J.-L. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. **Theoretical and Applied Genetics**, v. 127, n. 2, p. 463-480, 2014. DOI: https://doi.org/10.1007/s00122-013-2231-5.

HESLOT, N.; JANNINK, J.-L.; SORRELLS, M. E. Perspectives for genomic selection applications and research in plants. **Crop Science**, v. 55, n. 1, p. 1-12, 2015. DOI: https://doi.org/10.2135/cropsci2014.03.0249.

HODSON, D.; WHITE, J. GIS and crop simulation modelling applications in climate change research. In: REYNOLDS, M. (ed.). **Climate change and crop production**. Wallingford: CABI Publishing, 2010. p. 245-262.

JARQUÍN, D.; CROSSA, J.; LACAZE, X.; DU CHEYRON, P.; DAUCOURT, J.; LORGEOU, J.; PIRAUX, F.; GUERREIRO, L.; PÉREZ, P.; CALUS, M.; BURGUEÑO, J.; DE LOS CAMPOS, G. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. **Theoretical and Applied Genetics**, v. 127, n. 3, p. 595-607, 2014. DOI: https://doi.org/10.1007/s00122-013-2243-1.

JENKINS, M. T. Methods of estimating the performance of double crosses in corn. **Agronomy Journal**, v. 26, n. 3, p. 199-204, 1934. DOI: https://doi.org/10.2134/agronj1934.00021962002600030004x.

KNÜRR, T.; LÄÄRÄ, E.; SILLANPÄÄ, M. J. Impact of prior specifications in a shrinkage-inducing Bayesian model for quantitative trait mapping and genomic prediction. **Genetics Selection Evolution**, v. 45, article 24, 2013. DOI: https://doi.org/10.1186/1297-9686-45-24.

LI, D.; XU, Z.; GU, R.; WANG, P.; XU, J.; DU, D.; FU, J.; WANG, J.; ZHANG, H.; WANG, G. Genomic prediction across structured hybrid populations and environments in maize. **Plants**, v. 10, n. 6, article 1174, 2021. DOI: https://doi.org/10.3390/plants10061174.

LI, G.; DONG, Y.; ZHAO, Y.; TIAN, X.; WÜRSCHUM, T.; XUE, J.; CHEN, S.; REIF, J.

C.; XU, S.; LIU, W. Genome-wide prediction in a hybrid maize population adapted to Northwest China. **Crop Journal**, v. 8, n. 5, p. 830-842, 2020. DOI: https://doi.org/10.1016/j.cj.2020.04.006.

MAENHOUT, S.; DE BAETS, B.; HAESAERT, G. Graph-based data selection for the construction of genomic prediction models. **Genetics**, v. 185, n. 4, p. 1463-1475, 2010. DOI: https://doi.org/10.1534/genetics.110.116426.

MAGETO, E. K.; CROSSA, J.; PEREZ-RODRIGUEZ, P.; DHLIWAYO, T.; PALACIOS-ROJAS, N.; LEE, M.; GUO, R.; VICENTE, F. S.; ZHANG, X.; HINDU, V. Genomic prediction with genotype by environment interaction analysis for kernel zinc concentration in tropical maize germplasm. **G3: Genes, Genomes, Genetics**, v. 10, n. 8, p. 2629-2639, 2020. DOI: https://doi.org/10.1534/g3.120.401172.

MASTRODOMENICO, A. T.; BOHN, M. O.; LIPKA, A. E.; BELOW, F. E. Genomic selection using maize ex-plant variety protection germplasm for the prediction of nitrogen-use traits. **Crop Science**, v. 59, n. 1, p. 212-220, 2019. DOI: https://doi.org/10.2135/cropsci2018.06.0398.

MESSINA, C. D.; TECHNOW, F.; TANG, T.; TOTIR, R.; GHO, C.; COOPER, M. Leveraging biological insight and environmental variation to improve phenotypic prediction: integrating crop growth models (CGM) with whole genome prediction (WGP). **European Journal of Agronomy**, v. 100, p. 151-162, 2018. DOI: https://doi.org/10.1016/j.eja.2018.01.007.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-1829, 2001. DOI: https://doi.org/10.1093/genetics/157.4.1819.

MONTESINOS-LÓPEZ, A.; MONTESINOS-LÓPEZ, O. A.; GIANOLA, D.; CROSSA, J.; HERNÁNDEZ-SUÁREZ, C. M. Multi-environment genomic prediction of plant traits using deep learners with dense architecture. **G3: Genes, Genomes, Genetics**, v. 8, n. 12, p. 3813-3828, 2018a. DOI: https://doi.org/10.1534/g3.118.200740.

MONTESINOS-LÓPEZ, O. A.; MONTESINOS-LÓPEZ, A.; CROSSA, J.; GIANOLA, D.; HERNÁNDEZ-SUÁREZ, C. M.; MARTÍN-VALLEJO, J. Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. **G3: Genes, Genomes, Genetics**, v. 8, n. 12, p. 3829-3840, 2018b. DOI: https://doi.org/10.1534/g3.118.200728.

MONTESINOS-LÓPEZ, O. A.; MONTESINOS-LÓPEZ, A.; HERNANDEZ-SUAREZ, C. M.; BARRÓN-LÓPEZ, J. A.; CROSSA, J. Deep-learning power and perspectives for genomic selection. **The Plant Genome**, v. 14, n. 3, e20122, 2021. DOI: https://doi.org/10.1002/tpg2.20122.

MÔRO, G. V.; SANTOS, M. F.; SOUZA JÚNIOR, C. L. Comparison of genome-wide and phenotypic selection indices in maize. **Euphytica**, v. 215, n. 4, article 76, 2019. DOI: https://doi.org/10.1007/s10681-019-2401-x.

MUÑOZ, P. R.; RESENDE, M. F. R.; GEZAN, S. A.; RESENDE, M. D. V.; DE LOS CAMPOS,

G.; KIRST, M.; HUBER, D.; PETER, G. F. Unraveling additive from nonadditive effects using genomic relationship matrices. **Genetics**, v. 198, n. 4, p. 1759-1768, 2014. DOI: https://doi.org/10.1534/genetics.114.171322.

OLIVEIRA, A. A.; RESENDE, M. F. R.; FERRÃO, L. F. V.; AMADEU, R. R.; GUIMARÃES, L. J. M.; GUIMARÃES, C. T.; PASTINA, M. M.; MARGARIDO, G. R. A. Genomic prediction applied to multiple traits and environments in second season maize hybrids. **Heredity**, v. 125, p. 60-72, 29, 2020. DOI: https://doi.org/10.1038/s41437-020-0321-0._

PÉREZ-RODRÍGUEZ, P.; CROSSA, J.; RUTKOSKI, J.; POLAND, J.; SINGH, R.; LEGARRA, A.; AUTRIQUE, E.; DE LOS CAMPO, G.; BURGUENO, J.; DREISIGACKER, S. Single-step genomic and pedigree genotype × environment interaction models for predicting wheat lines in international environments. **The Plant Genome**, v. 10, n. 2, p. 1-15, 2017. DOI: https://doi.org/10.3835/plantgenome2016.09.0089.

PSZCZOLA, M.; STRABEL, T.; MULDER, H. A.; CALUS, M. P. L. Reliability of direct genomic values for animals with different relationships within and to the reference population. **Journal of Dairy Science**, v. 95, n. 1, p. 389-400, 2012. DOI: https://doi.org/10.3168/jds.2011-4338.

RACHMATIA, H.; KUSUMA, W. A.; HASIBUAN, L. S. Prediction of maize phenotype based on whole-genome single nucleotide polymorphisms using deep belief networks. **Journal of Physics: Conference Series**, v. 835, n. 1, 012003, 2017. DOI: http://dx.doi.org/10.1088/1742-6596/835/1/012003.

RAMSTEIN, G. P.; LARSSON, S. J.; COOK, J. P.; EDWARDS, J. W.; ERSOZ, E. S.; FLINT-GARCIA, S.; GARDNER, C. A.; HOLLAND, J. B.; LORENZ, A. J.; MCMULLEN, M. D.; MILLARD, M. J.; ROCHEFORD, T. R.; TUINSTRA, M. R.; BRADBURY, P. J.; BUCKLER, E. S.; CINTA ROMAY, M. Dominance effects and functional enrichments improve prediction of agronomic traits in hybrid maize. **Genetics**, v. 215, n. 1, p. 215-230, 2020. DOI: https://doi.org/10.1534/genetics.120.303025.

RESENDE, M. F. R.; MUÑOZ, P.; ACOSTA, J. J.; PETER, G. F.; DAVIS, J. M.; GRATTAPAGLIA, D.; RESENDE, V.; KIRST, M.; KIRST, M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. **New Phytologist**, v. 193, n. 3, p. 617-624, 2012a. DOI: https://doi.org/10.1111/j.1469-8137.2011.03895.x._

RESENDE, M. F. R.; MUÑOZ, P.; RESENDE, M. D. V.; GARRICK, D. J.; FERNANDO, R. L.; DAVIS, J. M.; JOKELA, E. J.; MARTIN, T. A.; PETER, G. F.; KIRST, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). **Genetics**, v. 190, n. 4, p. 1503-1510, 2012b. DOI: https://doi.org/10.1534/genetics.111.137026.

RINCENT, R.; LALOË, D.; NICOLAS, S.; ALTMANN, T.; BRUNEL, D.; REVILLA, P.; RODRÍGUEZ, V. M.; MORENO-GONZALEZ, J.; MELCHINGER, A.; BAUER, E.; SCHOEN,

C.-C.; MEYER, N.; GIAUFFRET, C.; BAULAND, C.; JAMIN, P.; LABORDE, J.; MONOD, H.; FLAMENT, P.; CHARCOSSET, A.; MOREAU, L. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). **Genetics**, v. 192, n. 2, p. 715-728, 2012. DOI: https://doi.org/10.1534/genetics.112.141473.

ROGERS, A. R.; DUNNE, J. C.; ROMAY, C.; BOHN, M.; BUCKLER, E. S.; CIAMPITTI, I. A.; EDWARDS, J.; ERTL, D.; FLINT-GARCIA, S.; GORE, M. A.; GRAHAM, C.; HIRSCH, C. N.; HOOD, E.; HOOKER, D. C.; KNOLL, J.; LEE, E. C.; LORENZ, A.; LYNCH, J. P.; MCKAY, J. et al. The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. **G3: Genes, Genomes, Genetics**, v. 11, n. 2, jkaa050, 2021. DOI: https://doi.org/10.1093/g3journal/jkaa050.

TECHNOW, F.; MESSINA, C. D.; TOTIR, L. R.; COOPER, M. Integrating crop growth models with whole genome prediction through approximate bayesian computation. **PLoS ONE**, v. 10, n. 6, e0130855, 2015. DOI: https://doi.org/10.1371/journal.pone.0130855.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414-4423, 2008. DOI: https://doi.org/10.3168/jds.2007-0980.

VITEZICA, Z. G.; VARONA, L.; LEGARRA, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. **Genetics**, v. 195, n. 4, p. 1223-1230, 2013. DOI: https://doi.org/10.1534/genetics.113.155176.

VIVEK, B. S.; KRISHNA, G. K.; VENGADESSAN, V.; BABU, R.; ZAIDI, P. H.; KHA, L. Q.; MANDAL, S. S.; GRUDLOYMA, P.; TAKALKAR, S.; KROTHAPALLI, K.; SINGH, I. S.; OCAMPO, E. T. M.; XINGMING, F.; BURGUEÑO, J.; AZRAI, M.; SINGH, R. P.; CROSSA, J. Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. **The Plant Genome**, v. 10, n. 1, p. 1-8, 2017. DOI: https://doi.org/10.3835/plantgenome2016.07.0070.

WANG, N.; WANG, H.; ZHANG, A.; LIU, Y.; YU, D.; HAO, Z.; ILUT, D.; GLAUBITZ, J. C.; GAO, Y.; JONES, E. Genomic prediction across years in a maize doubled haploid breeding program to accelerate early-stage testcross testing. **Theoretical and Applied Genetics**, v. 133, n. 10, p. 2869-2879, 2020. DOI: https://doi.org/10.1007/s00122-020-03638-5.

WELLMANN, R.; BENNEWITZ, J. Bayesian models with dominance effects for genomic evaluation of quantitative traits. **Genetics Research**, v. 94, n. 1, p. 21-37, 2012. DOI: https://doi.org/10.1017/S0016672312000018.

WINDHAUSEN, V. S.; ATLIN, G. N.; HICKEY, J. M.; CROSSA, J.; JANNINK, J. L.; SORRELLS, M. E.; RAMAN, B.; CAIRNS, J. E.; TAREKEGNE, A.; SEMAGN, K.; BEYENE, Y.; GRUDLOYMA, P.; TECHNOW, F.; RIEDELSHEIMER, C.; MELCHINGER,

A. E. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. **G3: Genes, Genomes, Genetics**, v. 2, n. 11, p. 1427-1436, 2012. DOI: https://doi.org/10.1534/g3.112.003699.

XU, Y. Envirotyping for deciphering environmental impacts on crop plants. **Theoretical and Applied Genetics**, v. 129, n. 4, p. 653-673, 2016. DOI: https://doi.org/10.1007/s00122-016-2691-5.

YU, X.; LEIBOFF, S.; LI, X.; GUO, T.; RONNING, N.; ZHANG, X.; MUEHLBAUER, G. J.; TIMMERMANS, M. C. P.; SCHNABLE, P. S.; SCANLON, M. J.; YU, J. Genomic prediction of maize microphenotypes provides insights for optimizing selection and mining diversity. **Plant Biotechnology Journal**, v. 18, n. 12, p. 2456-2465, 2020. DOI: https://doi.org/10.1111/pbi.13420.

YU, X.; LI, X.; GUO, T.; ZHU, C.; WU, Y.; MITCHELL, S. E.; ROOZEBOOM, K. L.; WANG, D.; WANG, M. L.; PEDERSON, G. A.; TESSO, T. T.; SCHNABLE, P. S.; BERNARDO, R.; YU, J. Genomic prediction contributing to a promising global strategy to turbocharge gene banks. **Nature Plants**, v. 2, n. 10, article 16150, 2016. DOI: https://doi.org/10.1038/nplants.2016.150.

ZHONG, S.; DEKKERS, J. C. M.; FERNANDO, R. L.; JANNINK, J.-L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, n. 1, p. 355-364, 2009. DOI: https://doi.org/10.1534/genetics.108.098277.

ZINGARETTI, L. M.; GEZAN, S. A.; FERRÃO, L. F. V.; OSORIO, L. F.; MONFORT, A.; MUÑOZ, P. R.; WHITAKER, V. M.; PÉREZ-ENCISO, M. Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species. **Frontiers in Plant Science**, v. 11, article 25, 2020. DOI: https://doi.org/10.3389/fpls.2020.00025.