



Genomic Prediction of Grain Yield and Drought-Adaptation Capacity in Sorghum Is Enhanced by Multi-Trait Analysis

Julio G. Velazco^{1,2*}, David R. Jordan³, Emma S. Mace^{3,4}, Colleen H. Hunt^{3,4}, Marcos Malosetti² and Fred A. van Eeuwijk^{2*}

OPEN ACCESS

Edited by:

Rodomiro Ortiz,
Swedish University of Agricultural
Sciences, Sweden

Reviewed by:

Manje Gowda,
International Maize and Wheat
Improvement Center (CIMMYT),
Kenya
Tobias Würschum,
University of Hohenheim, Germany
Yusheng Zhao,
Leibniz-Institut für Pflanzengenetik
und Kulturpflanzenforschung (IPK),
Germany

*Correspondence:

Julio G. Velazco
velazco.julio@inta.gob.ar
Fred A. van Eeuwijk
fred.vaneeuwijk@wur.nl

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 10 May 2019

Accepted: 17 July 2019

Published: 31 July 2019

Citation:

Velazco JG, Jordan DR, Mace ES,
Hunt CH, Malosetti M and
van Eeuwijk FA (2019) Genomic
Prediction of Grain Yield
and Drought-Adaptation Capacity
in Sorghum Is Enhanced by Multi-Trait
Analysis. *Front. Plant Sci.* 10:997.
doi: 10.3389/fpls.2019.00997

¹ Department of Plant Breeding, EEA Pergamino, National Institute of Agricultural Technology (INTA), Pergamino, Argentina, ² Biometris – Mathematical and Statistical Methods, Wageningen University and Research, Wageningen, Netherlands, ³ Queensland Alliance for Agriculture and Food Innovation, Hermitage Research Facility, The University of Queensland, Warwick, QLD, Australia, ⁴ Department of Agriculture and Fisheries, Hermitage Research Facility, Queensland Government, Warwick, QLD, Australia

Grain yield and stay-green drought adaptation trait are important targets of selection in grain sorghum breeding for broad adaptation to a range of environments. Genomic prediction for these traits may be enhanced by joint multi-trait analysis. The objectives of this study were to assess the capacity of multi-trait models to improve genomic prediction of parental breeding values for grain yield and stay-green in sorghum by using information from correlated auxiliary traits, and to determine the combinations of traits that optimize predictive results in specific scenarios. The dataset included phenotypic performance of 2645 testcross hybrids across 26 environments as well as genomic and pedigree information on their female parental lines. The traits considered were grain yield (GY), stay-green (SG), plant height (PH), and flowering time (FT). We evaluated the improvement in predictive performance of multi-trait G-BLUP models relative to single-trait G-BLUP. The use of a blended kinship matrix exploiting pedigree and genomic information was also explored to optimize multi-trait predictions. Predictive ability for GY increased up to 16% when PH information on the training population was exploited through multi-trait genomic analysis. For SG prediction, full advantage from multi-trait G-BLUP was obtained only when GY information was also available on the predicted lines *per se*, with predictive ability improvements of up to 19%. Predictive ability, unbiasedness and accuracy of predictions from conventional multi-trait G-BLUP were further optimized by using a combined pedigree-genomic relationship matrix. Results of this study suggest that multi-trait genomic evaluation combining routinely measured traits may be used to improve prediction of crop productivity and drought adaptability in grain sorghum.

Keywords: genomic prediction, multi-trait analysis, sorghum, auxiliary trait, grain yield, stay-green, blended kinship matrix, BLUP

INTRODUCTION

Water scarcity in rain-fed cropping systems is a major challenge to a world of increasing food demand (UNCTAD, 2011). In this context, grain sorghum (*Sorghum bicolor* L. Moench) is a cereal crop that can play an important role for sustainable farming, as it is particularly resilient to stress conditions caused by drought and erratic rainfall (Paterson et al., 2009). This crop is a staple food in semi-arid regions of the world and is used for feed globally (Acquaah, 2012). Grain yield is the primary trait in sorghum, as it is a key measure of crop productivity and profitability of farmers. Another important trait is stay-green; a complex drought-adaptation mechanism associated with increased yield in environments where post-flowering drought occurs frequently (Borrell et al., 2000; Jordan et al., 2003). Accordingly, most efforts for increasing genetic progress in grain sorghum should concern both characteristics (Jordan et al., 2012; Borrell et al., 2014). Further improvement in productivity and drought adaptability requires knowledge-based selection strategies that efficiently exploit available phenotypic and genotypic information in sorghum breeding programs.

Selection of complex quantitative traits can be based on statistical methods that combine phenotypes and high-density marker data to predict genetic merit. This form of marker-assisted selection, known as genomic selection (Meuwissen et al., 2001), has been successfully implemented in animal and plant breeding (Meuwissen et al., 2016; Crossa et al., 2017). Several genomic prediction methods have been developed, including Bayesian regression (Gianola et al., 2009) and semiparametric approaches (de los Campos et al., 2010). An alternative method for genomic prediction within the linear mixed model framework is termed genomic best linear unbiased prediction or G-BLUP (VanRaden, 2008). This method is usually preferred in practice because it is simple to implement and computationally less demanding than competing procedures (Gianola et al., 2014). Moreover, G-BLUP is expected to perform similar to other models for prediction of complex agronomic traits such as grain yield (Heslot et al., 2012; Wimmer et al., 2013), which are typically affected by a large number of small-effect genes (Schön et al., 2004). Independent of the model used, one of the main features of genomic selection is that genetic merit can be predicted for selection candidates that have not yet been phenotyped. This application is particularly promising for reducing evaluation cost and generation interval in the sorghum breeding pipeline, where parental lines of commercial hybrids are currently selected on the base of extensive field progeny testing. Moreover, development of female lines as well as hybrid seed production is based on the use of cytoplasmic-genetic male sterility, which requires extra time and human resources. Despite the potential of genomic selection to increase rates of genetic gain in sorghum, studies on the application of genomic models are limited compared to other cereal crops, such as maize, wheat, and rice (Kulwal, 2016). A first genomic selection study in grain sorghum was reported by Hunt et al. (2018) for prediction of test-cross performance in individual trials. Velazco et al. (2019) investigated different genomic models including pedigree information for across-environment

prediction of parental breeding values in productivity and adaptability traits.

Most of genomic selection studies, including the ones mentioned above, have been based on separate analysis of individual traits. However, selection decisions in plant breeding programs typically rely on several measured characters. The joint analysis of multiple traits (MT) can increase the accuracy of genetic evaluations relative to single-trait (ST) analysis by exploiting information from correlated characters (Henderson and Quaas, 1976). The potential gain in accuracy depends on the strength of genetic and environmental correlations between traits. The benefit is expected to increase for lowly heritable traits, when analyzed together with strongly correlated traits of higher heritability (Thompson and Meyer, 1986). Additionally, MT models are able to reduce selection bias or culling bias introduced by contemporary or sequential selection on correlated traits, which are ignored by ST analysis (Mrode, 2005). The extension of MT analysis into the context of genomic prediction methods has been studied using real and simulated data (e.g., Calus and Veerkamp, 2011; Jia and Jannink, 2012). MT genomic models can be used to combine information from correlated traits and from relatives in an efficient way. When the breeder is interested in response to selection of a single target trait, other secondary or auxiliary traits can be incorporated in the prediction model to provide additional information on the primary character. Consequently, more phenotypes recorded on the reference population can be potentially exploited to assist predictions of the target trait in the testing population.

Genome-based MT analysis has been applied for breeding in hybrid crops of other major cereals like maize, rice and wheat (e.g., dos Santos et al., 2016; Wang et al., 2017; Schulthess et al., 2018). In sorghum, MT genomic prediction has been implemented only in biomass-type genotypes using a pre-breeding population (Fernandes et al., 2018). Here, we present a first study on the potential of exploiting trait associations for genomic prediction in advanced breeding material of grain sorghum. Our research is developed in the context of prediction for broad adaptation using testcross performance data across dryland sorghum production environments in Australia.

The aims of this study were to investigate if multi-trait analysis improves across-environment genomic predictions for grain yield and stay-green in sorghum, and to identify the combinations of traits that optimize results in different prediction scenarios. To attain these objectives, the optimal combination of traits was empirically determined for each scenario by evaluating the gain in prediction quality of alternative MT models relative to the ST model. In addition, we explored if the performance of best-predictive MT genomic models can be further enhanced by incorporating pedigree information.

MATERIALS AND METHODS

Data

The dataset used in this study is part of the sorghum breeding program of female parental lines conducted by the University of Queensland and the Department of Agriculture and Fisheries

in Queensland, Australia. The phenotypic records consisted of 26 testcross performance trials where a total of 646 female lines were tested in hybrid combinations across 12 locations between 2008 and 2014. Phenotypes of 2645 testcross hybrids were used to assess female lines performance across a target population of environments (TPE) covering the main sorghum cropping region of Australia. The series of trials belongs to an advanced stage of yield testing (AYT), where measurements are taken from relatively large plots. More details on field layout and structure of the dataset are given in Velazco et al. (2019). Four productivity and adaptability traits routinely measured in advanced testing were considered for this study: grain yield (GY), stay-green (SG), plant height (PH), and flowering time (FT). GY is the main target trait with direct economic value driving selection. SG is an integrated drought-adaptation trait that is expressed as delayed leaf senescence, which is a consequence of improved water balance in the plant under post-flowering drought stress (Borrell et al., 2014). This functional SG phenotype is also considered an important trait since it is associated with enhanced crop productivity in water-limited seasons (Jordan et al., 2003, 2012). Given that SG expression depends on the occurrence of terminal drought conditions, records of this trait were available at nine trials and for 603 lines in the present dataset. While PH and FT are mainly selected in earlier breeding stages to reduce extreme variation, these traits are considered in advanced testing to ensure appropriate agronomic type for commercial production (Jordan et al., 2011).

All the female lines were genotyped using an integrated DArT and genotyping-by-sequencing (GBS) methodology involving complexity reduction of the genomic DNA to remove repetitive sequences using methylation sensitive restriction enzymes prior to sequencing on Next Generation sequencing platforms (DArT)¹. The sequence data generated were then aligned to the most recent version (v3.1.1) of the sorghum reference genome sequence (Paterson et al., 2009; McCormick et al., 2018) to identify SNP (Single Nucleotide Polymorphism) markers. SNPs with minor allele frequency lower than 2.5% or more than 20% of missing values were discarded. Missing genotypes were imputed based on random sampling from marginal allele distributions using the synbreed package (Wimmer et al., 2012) in R (R Core Team, 2018). After quality filtering, 4781 evenly spaced SNPs remained for analysis.

Inbred parent lines were derived from pedigree breeding methods resulting in a highly structured breeding population. Pedigree data was available for the female lines and 499 ancestors tracing back 28 generations.

Single-Trait Analysis

Univariate analysis of each trait was performed within the REML-based mixed model framework using a stage-wise approach. In the first stage, adjusted testcross means were computed per trial after correcting for design factors and spatial field variation. For this, we used a novel flexible method for spatial analysis of individual trials based on P-splines

(see Velazco et al., 2017; Rodríguez-Álvarez et al., 2018b; for details). The specific model applied in the first stage is described in Velazco et al. (2019).

In the second stage, spatially adjusted testcross means from each trial were jointly analyzed to estimate line means across testers and environments. The model is as follows:

$$y_{ijk} = \mu + L_i + M_j + E_k + LM_{ij} + LE_{ik} + ME_{jk} + LME_{ijk}, \quad (1)$$

where y_{ijk} represents the best linear unbiased estimation (BLUE) of the i -th female line crossed with the j -th male tester in the k -th environment, which was fitted by a random line effect (L_i), a fixed male tester effect (M_j), a fixed environmental effect (E_k), and all possible interactions between these factors. Since line effects were considered random, all the interactions involving L_i were random, while ME_{jk} was fixed. Note that LME_{ijk} includes the error of genotype mean. All random effects were assumed independent homoscedastic and normally distributed with zero mean.

Due to shrinkage properties of BLUP (Robinson, 1991), random line effects are contracted toward the mean of the line population. Given that not all lines were crossed with the same number of testers or evaluated in the same number of trials, the amount of shrinkage is different for each BLUP of L_i . Moreover, using BLUP(L_i) as response variable in the genomic prediction model is problematic because it would result in double shrinkage of predicted breeding values. Therefore, to eliminate shrinkage in line BLUPs before the genomic prediction stage, we applied the deregression procedure of Garrick et al. (2009). This correction relies on individual reliabilities of BLUP(L_i), as obtained by inverting the coefficient matrix of the mixed model equation (Meyer, 1989).

The following G-BLUP model was used to predict parental breeding values of female lines from progeny performance:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}, \quad (2)$$

where the vector \mathbf{y} contains deregressed BLUPs of L_i derived from the second stage; $\mathbf{1}$ is a vector of ones with associated general mean μ ; \mathbf{Z} is a design matrix allocating deregressed BLUP(L_i) to genomic effects; \mathbf{g} is the vector of additive genomic effects with distribution $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$, where σ_g^2 is the additive genomic variance and \mathbf{G} is the genomic relationship matrix as computed with the first method of VanRaden (2008); and \mathbf{e} is the vector of residuals assuming $\mathbf{e} \sim N(0, \mathbf{D}\sigma_e^2)$, where σ_e^2 is the residual variance and \mathbf{D} is a diagonal weighting matrix accounting for heterogeneous residual variances due to differences in individual reliabilities of deregressed BLUP(L_i) (VanRaden, 2008; Garrick et al., 2009).

Given the variance components from the ST prediction model, the narrow-sense heritabilities (h^2) of line means were obtained as: $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$. Note that σ_e^2 comprises non-additive genetic effects and true errors associated with line mean estimates.

¹<http://www.diversityarrays.com>

Multi-Trait Analysis

For joint MT analysis models (1) and (2) were extended using multivariate mixed models. We present a general formulation for any combination among traits $T = (GY, SG, PH, FT)$. The multi-trait case of model (1) can be represented in vector notation as:

$$\mathbf{y}_{ijk} = \boldsymbol{\mu} + \mathbf{L}_i + \mathbf{M}_j + \mathbf{E}_k + \mathbf{LM}_{ij} + \mathbf{LE}_{ik} + \mathbf{ME}_{jk} + \mathbf{LME}_{ijk}, \quad (3)$$

where in this case \mathbf{y}_{ijk} is a vector collecting spatially adjusted BLUEs of multiple traits T from separate univariate analyses in stage one; and \mathbf{L}_i , \mathbf{LM}_{ij} , \mathbf{LE}_{ik} , \mathbf{LME}_{ijk} are the vectors of multi-trait random effects with respective assumed distributions $\mathbf{L}_i \sim \text{MVN}(\mathbf{0}, \mathbf{I}_L \otimes \boldsymbol{\Sigma}_L)$, $\mathbf{LM}_{ij} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_{LM} \otimes \boldsymbol{\Sigma}_{LM})$, $\mathbf{LE}_{ik} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_{LE} \otimes \boldsymbol{\Sigma}_{LE})$, $\mathbf{LME}_{ijk} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_{LME} \otimes \boldsymbol{\Sigma}_{LME})$, where $\boldsymbol{\Sigma}_q$, for $q = L, LM, LE, LME$, is a covariance matrix among traits and \otimes is the Kronecker product operator. For all random effects, the matrix $\boldsymbol{\Sigma}_q$ was modeled as unstructured, allowing for unequal variances across traits and specific covariances for each pair of traits.

The multi-trait G-BLUP model was defined as:

$$\mathbf{y}_T = \mathbf{1}_T \boldsymbol{\mu}_T + \mathbf{Z}_T \mathbf{g}_T + \mathbf{e}_T,$$

where the subscript refers to traits T ; \mathbf{y}_T is now a multi-trait vector of deregressed BLUP(L_i), ordered as lines within traits, obtained from joint multivariate analysis using model (3); \mathbf{g}_T is the vector of multi-trait additive genomic effects with distribution $\mathbf{g}_T \sim \text{MVN}(\mathbf{0}, \mathbf{G} \otimes \boldsymbol{\Sigma}_g)$; and the multivariate residual effects were assumed $\mathbf{e}_T \sim \text{MVN}[\mathbf{0}, (\mathbf{I}_e \otimes \boldsymbol{\Sigma}_e) \mathbf{D}_T]$, where matrix \mathbf{D}_T has diagonal elements containing weights based on individual reliabilities of deregressed BLUP(L_i) for each trait, as obtained from the second stage of multi-trait analysis (model 3). The covariance matrices among traits for additive genomic effects ($\boldsymbol{\Sigma}_g$) and residuals effects ($\boldsymbol{\Sigma}_e$) were assumed unstructured.

We also considered multi-trait prediction models fitting a kinship matrix that combines pedigree and genomic information. For this, we extended the BLUP method advocated by Velazco et al. (2019) to the multi-trait context. The method, referred as K-BLUP, uses a combined kinship matrix formed as $\mathbf{K} = w\mathbf{A} + (1 - w)\mathbf{G}_s$, where \mathbf{A} is the numerator relationship matrix among lines computed from the full pedigree data and \mathbf{G}_s is a scaled matrix \mathbf{G} that is compatible with \mathbf{A} in reference to the base breeding population (Vitezica et al., 2011; Christensen, 2012). The weighting term w can be interpreted as the fraction of additive genetic variance that is not captured by SNPs and is explained by familial relationships. In this case, the vector \mathbf{g}_T has distribution $\text{MVN}(\mathbf{0}, \mathbf{K} \otimes \boldsymbol{\Sigma}_g)$, and collects additive genomic effects as well as residual polygenic effects. Under our approach, the optimal w is empirically determined in order to maximize predictive ability after evaluating a sequence of candidate weights within the range $0 < w < 1$. To assess the benefits of including additional pedigree information, the best-predictive multi-trait model using conventional G-BLUP (i.e., setting $w = 0$) was compared to the equivalent optimized multi-trait K-BLUP model.

Model Fitting

Spatial analyses in the first stage were performed with the R package SpATS (Rodríguez-Álvarez et al., 2018a), which is publicly available from CRAN². Uni- and multivariate models in the second and the genomic prediction stages were fitted by REML using the average information algorithm as implemented in ASReml-R (Butler et al., 2017). We used the correlation form parameterization of the unstructured covariance matrix available in ASReml-R to obtain direct estimates of trait correlations and corresponding standard errors.

Prediction Scenarios

To assess the value of multi-trait genomic analysis in sorghum, we considered the general predictive strategy where a single trait is the primary target of prediction (selection) and other auxiliary characters are used to potentially improve predictions of the target trait. Predictive performance of MT models was evaluated through fivefold cross-validation, where 80% of lines were randomly assigned to the training lines set (TLS) and the remaining 20% formed the validation lines set (VLS). We explored two potential scenarios for MT prediction (Figure 1). One scenario assuming that (new) lines in the VLS have not been evaluated in the field for any trait; therefore, phenotypes of auxiliary traits (and the target trait) were available only for the TLS (Aux@TL). A second MT prediction scenario assumes that lines of the VLS have been evaluated in the field, but for traits other than the target; therefore, phenotypes of auxiliary traits were available for both TLS and VLS (Aux@TL + VL). For comparison, we also considered ST prediction, where only records of the target trait in VLS are used for prediction.

The first MT prediction scenario was studied for two cases: when GY or SG were the traits of interest. However, the Aux@TL + VL scenario was considered only for SG as target trait. Given that one of the main purposes of the trial series used in this study is to obtain representative measurements of GY productivity, the fact that lines are field-tested for other traits,

²<https://cran.r-project.org/package=SpATS>

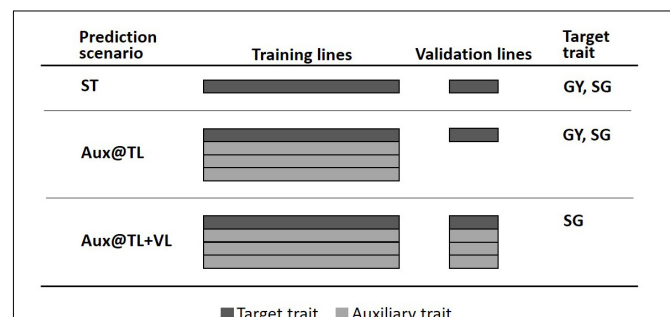


FIGURE 1 | Fivefold cross-validation schemes representing three prediction scenarios: ST, where only data of the target trait in the training lines set (TLS) are used for prediction in the validation lines set (VLS); Aux@TL, where data of auxiliary traits in TLS are included for prediction of the target trait in VLS; and Aux@TL + VL, where data of auxiliary traits in both TLS and VLS are included for prediction of the target trait in VLS.

but not for GY, would imply atypical circumstances seriously affecting GY during an experiment. Moreover, extremely reduced or aborted grain production affects normal canopy development; hence, data of traits such as PH or SG should be discarded or considered with extreme caution. For these reasons, the Aux@TL + VL scenario was not considered when GY was the target of prediction, as it would represent an unrealistic or exceptional situation in the context of an advanced stage of sorghum testing.

Validation of Genomic Prediction

In each prediction scenario, the performances of all possible MT models combining two, three, and the four traits were compared to that of the ST model, with the latter taken as benchmark. The response variables used for validation of genomic predictions were consistent with the models applied for analysis: \mathbf{y}_T and \mathbf{y} were used to validate $\tilde{\mathbf{g}}_T$ and $\tilde{\mathbf{g}}$, respectively. Model validation was evaluated by considering the predictive ability, unbiasedness, and accuracy of predictions in the VLS. Predictive ability was measured as the Pearson's correlation (r_{PA}) between $\tilde{\mathbf{g}}_T$ ($\tilde{\mathbf{g}}$) and \mathbf{y}_T (\mathbf{y}). Unbiasedness of genomic predictions was measured by the regression coefficient (b) of \mathbf{y}_T (\mathbf{y}) on $\tilde{\mathbf{g}}_T$ ($\tilde{\mathbf{g}}$), where $b = 1$ indicates an empirically unbiased predictor. Accuracy of predictions was assessed by computing the mean squared error of prediction (MSEP) from the linear regression. The evaluation of prediction models was based on average values over 20 replicates of each cross-validation scenario using the same random seed for all models.

RESULTS

The narrow-sense heritability estimates using ST G-BLUP are given in **Table 1** along with the additive genetic and residual correlations between traits estimated by MT G-BLUP. Heritability estimates ranged from 0.36 to 0.76, with GY presenting the lowest h^2 . The heritability of SG was higher than that of GY, but lower than those of PH and FT. Additive genetic correlations were significant only for GY with SG and PH, while FT was statistically uncorrelated with the other traits. Significant residual correlations were estimated between GY and the rest of the traits as well as between PH and FT.

Figure 2 shows measures of predictive performance from MT G-BLUP when additional information from auxiliary traits in

the TLS was used to predict GY or SG (Aux@TL scenario). MT models that included PH data alone or combined with other auxiliary traits increased predictive ability for GY, compared with the ST model. The highest predictive ability represented a 16% increase and it was achieved by combining GY, SG and PH information. Regression coefficients (b) for all models were below 1 for GY, which indicates overestimation of genomic predictions. The average empirical bias of GY predictions from the ST model was slightly reduced only by the MT model including PH alone as auxiliary trait. The relative decrease (%) in MSEP by MT models was consistent with their increases in predictive ability; MT G-BLUP exploiting PH information had a greater impact on improving accuracy of genomic prediction for GY. When SG was the target of prediction, ST G-BLUP outperformed MT G-BLUP models in predictive ability. Even though prediction models incorporating GY data in the TLS reduced the absolute deviation of b from 1, these models tended to over-predict additive genetic values for SG ($b < 1$). The inclusion of auxiliary traits in TLS produced higher MSEP relative to using only SG data for prediction.

Different results were found for SG predictions when MT G-BLUP models used auxiliary trait records on both TLS and VLS (**Figure 3**). Under this prediction scenario, all MT models including GY as auxiliary trait improved predictive abilities for SG. Relative increases of up to 19% in predictive ability were observed for MT models adding combined GY-PH or GY-PH-FT information from TLS and VLS. Moreover, predictions from these models tended to be less biased than those from ST G-BLUP and over-prediction was eliminated. Finally, MSEP for SG was consistently reduced when GY information on VLS was incorporated by MT G-BLUP models.

After evaluating predictive performance of G-BLUP method, we explored the use of a combined pedigree-genomic matrix **K** for optimization of multi-trait genomic prediction. **Table 2** presents results from the best-predictive MT G-BLUP models and from the optimized MT K-BLUP models for the scenarios where MT analysis outperformed ST analysis (Aux@TL for GY and Aux@TL + VL for SG). In these prediction scenarios, the optimal use of combined trait information, from a predictive perspective, resulted from exploiting genetic correlations among GY, SG, and PH (**Figures 2, 3**). When GY was the target trait, validation results of MT G-BLUP were improved by MT K-BLUP (with **K** using $w = 0.25$) for all evaluation criteria to a marginal extent. Under the prediction scenario for SG, bias and MSEP from multi-trait models were slightly higher when **K** (using $w = 0.30$) was used instead of **G**. However, an additional 10% gain in predictive ability was obtained by MT K-BLUP in this scenario.

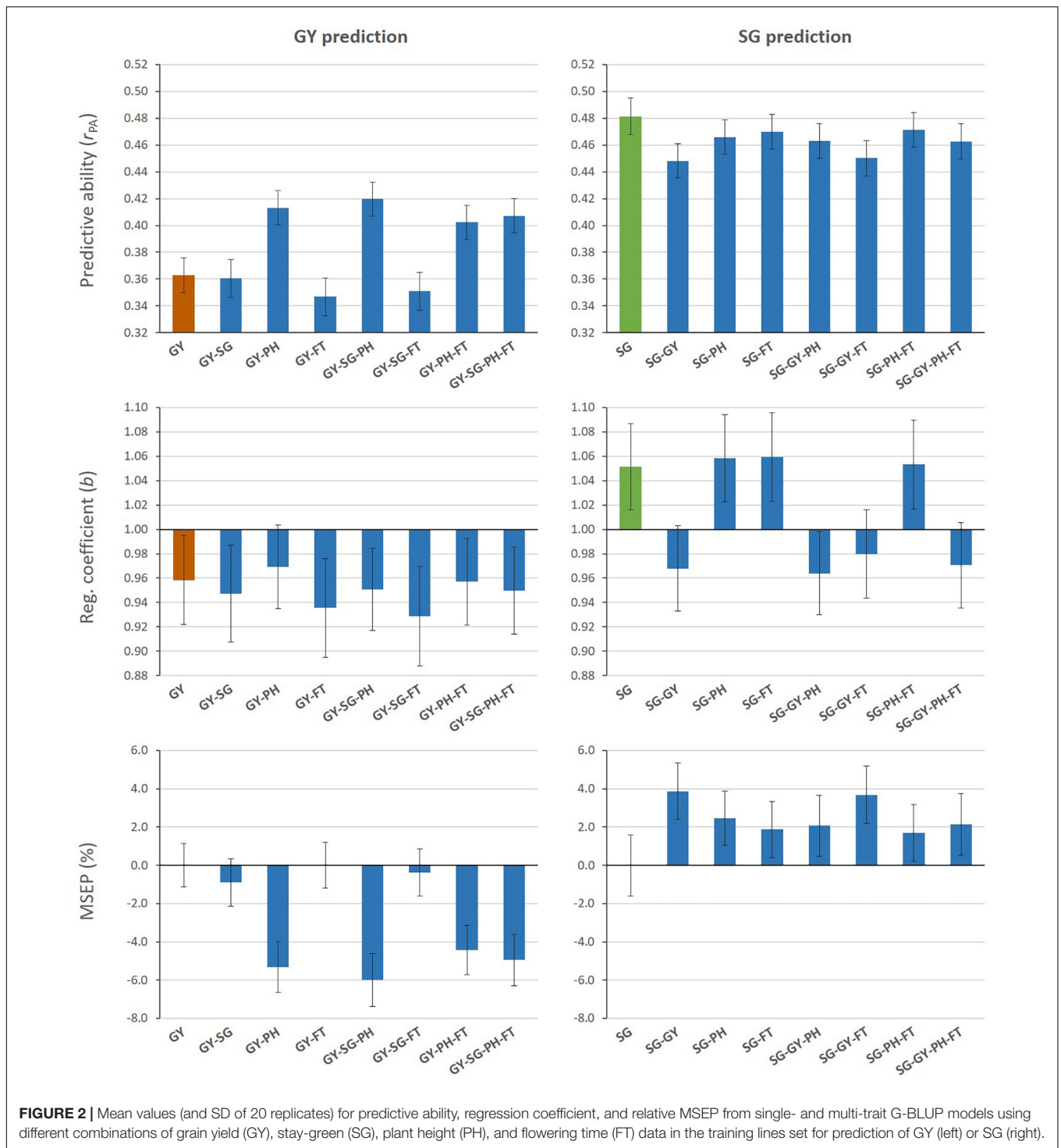
DISCUSSION

This study intended to establish the value of using multi-trait analysis to improve across-environment genomic predictions for grain yield and stay-green in sorghum by exploiting information from auxiliary traits. Our approach to predict is consistent with selection for broad adaptation, where the set of trials is considered to be representative of the TPE. To determine the

TABLE 1 | Heritabilities (diagonal; in parentheses), additive genetic (above diagonal), and residual (below diagonal) correlations^a for grain yield (GY), stay-green (SG), plant height (PH), and flowering time (FT).

Trait	GY	SG	PH	FT
GY	(0.36)	0.52	0.66	-0.01
SG	0.36	(0.50)	0.03	-0.02
PH	0.71	0.04	(0.76)	-0.07
FT	-0.48	0.04	-0.29	(0.65)

^aBoldfaced correlations are significant based on approximate 95% confidence interval (Holland, 2006).



efficiency of this predictive method, multi-trait BLUP models were evaluated in terms of prediction quality measures for two cases: when the target trait and auxiliary traits were recorded only on relatives of the predicted lines; and when records of auxiliary traits were also available for the predicted lines *per se*. Multi-trait BLUP can be seen as a generalized (linear) selection index, with the additional abilities to account for unbalanced information

from any set of relatives while properly adjusting for fixed effects in the data (Lynch and Walsh, 1998; Mrode, 2005). In our index, and from mixed model theory, data of several traits were optimally weighted based on the multi-trait genetic and residual covariance matrices in order to maximize accuracy of predicted genetic merit by BLUP estimation. An important decision is which auxiliary traits to include in a multi-trait model. In this

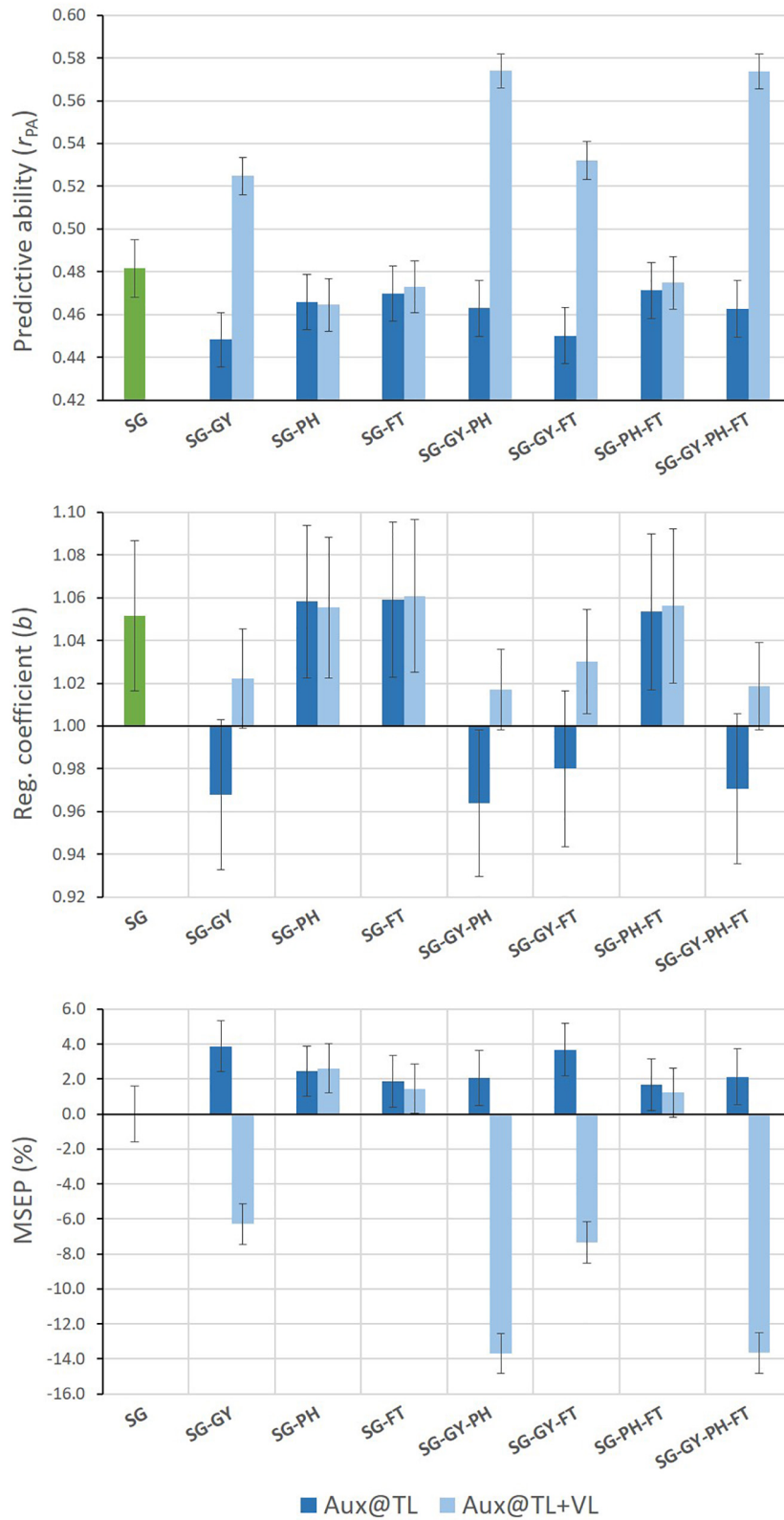


FIGURE 3 | Mean values (and SD of 20 replicates) for predictive ability, regression coefficient, and relative MSEP for SG predictions from single- and multi-trait G-BLUP models using different combinations of grain yield (GY), stay-green (SG), plant height (PH), and flowering time (FT) data only in the training lines set (Aux@TL), and in both the training and validation lines sets (Aux@TL + VL).

TABLE 2 | Predictive abilities (r_{PA}), regression coefficients (Bias), and relative MSEP from single-trait G-BLUP and K-BLUP, and from the best-predictive multi-trait G-BLUP and K-BLUP models for GY prediction using auxiliary traits data on training lines (Aux@TL) and for SG prediction using auxiliary traits data on both training and validation lines (Aux@TL + VL).

Prediction scenario	Quality criterion	Single-trait		Multi-trait ^a	
		G-BLUP	K-BLUP	G-BLUP	K-BLUP ^b
GY: (Aux@TL)	r_{PA}	0.363 (0.013)	0.373 (0.014)	0.420 (0.013)	0.429 (0.015)
	Bias (b)	0.958 (0.037)	0.965 (0.037)	0.951 (0.034)	0.967 (0.039)
	MSEP (%)	0 (1.1)	-0.7 (1.1)	-6.0 (1.3)	-6.6 (1.6)
SG: (Aux@TL + VL)	r_{PA}	0.482 (0.013)	0.508 (0.015)	0.574 (0.008)	0.630 (0.009)
	Bias (b)	1.052 (0.035)	1.074 (0.036)	1.017 (0.019)	1.038 (0.022)
	MSEP (%)	0 (1.6)	2.7 (1.6)	-13.7 (1.2)	-12.0 (1.2)

The best value for each evaluation criterion is boldfaced. ^aMulti-trait models combining GY, SG, and PH information, ^bUsing $K = wA + (1 - w)G_s$ with optimal weights $w = 0.25$ and $w = 0.30$ for GY and SG prediction, respectively.

article, we examined all possible combinations of available traits, and the optimal multi-trait BLUP model for each prediction scenario was empirically determined on the base of predictive outcomes. The present study considered the general situation where the breeding objective is to improve a single character with economic value (GY or SG), while other traits would contribute to that genetic goal. This is equivalent to assigning a relative economic weight of 1 to the target trait and 0 weights to auxiliary traits in a selection index. It is worth mentioning that, in a real breeding program, SG is unlikely to be considered as an independent target of selection since its economic value is linked to its positive effect on GY and reduced lodging under stressing conditions. Our research could be easily extended to the case of simultaneous improvement for GY and SG by constructing an index with non-zero economic weights for both traits. This would give predictive results that are intermediate of those presented in our research, with variations depending on the relative economic weights assigned (not shown). Given that defining the appropriate relative economic value of GY and SG is beyond the scope of the present study, here we presented the evaluation of multi-trait prediction for a target trait at a time without losing generality of results.

Multi-Trait Genomic Prediction for Grain Yield

Genetic correlations between traits estimated in this research are consistent with previous results showing strong associations of GY with SG, PH, and FT in sorghum hybrids (Jordan et al., 2003). In our case, additive parental effects of GY were only correlated with additive effect of SG and PH, whereas associations between GY and FT were basically explained by residual effects. In addition, the estimated heritability for GY in our study was lower than for the rest of the traits. Based on these findings, and according to multivariate BLUP and selection index theories, prediction of breeding values for GY should mainly benefit by utilizing additional information from SG and PH, while introducing an uncorrelated trait, such as FT, would reduce prediction efficiency due to incorporation of sampling error. These expectations were corroborated by our predictive results from cross-validation when records of auxiliary

traits were available for the training lines (Figure 2). Multi-trait models combining GY-SG-PH information produced the highest improvements in predictive ability for GY, representing up to 18% increase relative to ST G-BLUP (Table 2). These results are somewhat more promising than those previously reported for GY prediction in wheat (Rutkoski et al., 2016; Sun et al., 2017), maize (Lyra et al., 2017), and rye (Schulthess et al., 2016), which found null increases in predictive ability under equivalent prediction scenarios. The differences in response from using MT analysis are highly dependent of the experimental data and the genetic target of prediction. In our research, the inclusion of PH information in multi-trait models appeared to be essential for improving predictive ability, unbiasedness and accuracy of GY predictions. This may be attributed to the fact that PH presented the strongest additive genetic correlation and the largest difference in heritability with GY. The key role of these two factors for benefiting from multi-trait prediction of low-heritability traits has been demonstrated by simulation studies in plant and animal breeding contexts (Jia and Jannink, 2012; Guo et al., 2014). In addition, the present study assumes that the genetic target of selection is across-environment breeding value, which depends on overall heritabilities and trait correlations across the TPE. It should be considered, however, that the levels of heritability and correlations are variable among experiments due to the influence genotype-by-environment interaction, inducing variability of predictive results in specific trials.

Multi-Trait Genomic Prediction for Stay-Green

Our validation results were different for SG prediction when only MT information on the training set was used. As expected from considering heritabilities and additive genetic trait correlations, predictive performance of the ST model was generally better than those of MT models. This may be explained because SG was genetically correlated only with GY, but this trait had the lowest heritability (Table 1). Under these circumstances, information from PH and FT could not be borrowed to predict SG, while GY data would be genetically less informative than the target trait itself. On the other hand, predictive results favored multi-trait genomic models when GY information was available for both the

training lines and the predicted lines *per se*, which was reflected by increases in predictive ability and reductions in bias and MSE of SG predictions (Figure 3). The comparison of both prediction scenarios for SG allowed an empirical assessment of the value of two different sources of information from correlated traits. When auxiliary traits are only available for the training set, the estimates of covariance parameters among traits in Σ_g , which are then used for prediction, contain information from the training lines exclusively. Alternatively, if records of auxiliary traits are also available for the validation set, better parameter estimates in Σ_g can be obtained for prediction of the target trait. This is not only because more data is used for parameter estimation, but mainly because the extra information contained in estimated trait correlations is directly sourced from the predicted lines. Our results for SG suggest that, when this extra information is exploited, the strength of correlations among traits is more important than the relative levels of trait heritability. This would explain why SG predictions were improved by multi-trait G-BLUP using GY data, even when this trait had lower heritability than the target trait. Moreover, predictive performance of the MT model combining SG, GY, and PH indicates that, despite SG and PH were uncorrelated, additional information from PH could be transmitted via GY to further enhance predictions of SG (Figure 3). Similar benefits of exploiting information from secondary traits in the validation set have been reported in crops for other traits, but not for SG (Rutkoski et al., 2016; Fernandes et al., 2018; Lado et al., 2018). Even though the link between stay-green trait and yield stability has been demonstrated in sorghum, maize and wheat (Christopher et al., 2008; Jordan et al., 2012; Gregersen et al., 2013), the predictive use of SG-GY association in cereals had not been studied yet in the context of genomic selection.

Combining Pedigree and Genomic Information for Multi-Trait Prediction

This study explored if additional genealogical information from pedigree could further improve multi-trait genomic prediction. Merging pedigree and marker-based relationships has been shown to be beneficial for ST prediction in animals (e.g., Rodríguez-Ramilo et al., 2014; Ilska et al., 2017) and more recently in plant breeding (Velazco et al., 2019). In the context of multi-trait analysis, models relying exclusively on SNP information may give distorted estimates of genetic variances and correlations between traits, mainly due to incomplete linkage disequilibrium (LD) of markers with causal loci (de los Campos et al., 2015; Gianola et al., 2015). Pedigree information can be used to account for residual polygenic effects not traced by SNPs, capturing LD patterns between loci that are due to common ancestral identity. Our study showed that better predictions of GY and SG were obtained by multi-trait models using a weighted combination of **A** and **G** instead of **G** alone (Table 2). Multi-trait K-BLUP produced the highest increase in predictive ability (of about 30%) for the SG prediction scenario, where trait correlations seemed to play a more important role in model performance. This result suggests that the **K** matrix optimized estimates of genetic correlation between traits, from a predictive perspective. The same approach has been applied

to infer trait correlations from multi-trait prediction models in chicken (Momen et al., 2017).

Here, we used a different optimal weight w to construct **K** according to the trait of interest for prediction ($w = 0.25$ and $w = 0.30$ to predict GY and SG, respectively). This is in line with the idea that genetic (genomic) similarities between relatives are actually trait-specific (Fernando and Gianola, 2018), which contrasts with the assumption of a common relationship matrix used in standard multi-trait G-BLUP. A more elaborate multi-trait G-BLUP model using trait-dependent weights has been recently proposed by Karaman et al. (2018) as a computationally less demanding alternative to multi-trait Bayesian methods. In this BLUP model, weights used to compute **G** are derived from posterior trait-specific (co)variance estimates of SNP effects, which are obtained from a previous Bayesian analysis. The authors found, based on simulations, that the benefits of their weighting method were generally not significant for the low-heritability trait that more closely fitted an infinitesimal model (using 500 QTLs). Alternatively, our multi-trait K-BLUP uses a more straightforward approach that assumes a common weight across the genome, while still relaxing the assumptions of conventional multi-trait G-BLUP by allowing the similarity matrix to vary across traits in order to optimize prediction of the target. Given the complexity of traits and the highly structured breeding population used for this study, genomic predictions are likely to rely more on familial relationships and less on information from specific SNPs in LD with QTL (Habier et al., 2007). In this context, the BLUP-based multi-trait models applied here are expected to perform well compared to more refined alternatives (Jia and Jannink, 2012; Haile et al., 2018).

Implications of Multi-Trait Genomic Prediction for Grain Sorghum Breeding

While genotyping costs are being constantly reduced, the efficient use of phenotypic data becomes more relevant for plant breeding programs, since field phenotyping is still costly and labor-intensive. This work has shown for the first time how phenotypes of several traits routinely measured in grain sorghum breeding can be efficiently utilized through multi-trait analysis to assist genome-based selection of a target trait. We demonstrated empirically that genomic prediction of parental breeding values for GY benefits mainly from using PH as auxiliary trait. This trait is particularly promising in practice, since PH phenotype can be potentially measured in all trials, as opposed to SG, which can only be phenotyped under specific environmental conditions. In addition, high throughput field phenotyping technologies are expected to increase accuracy of PH measurement in sorghum (Wang et al., 2018), which could favor GY selection indirectly through multi-trait prediction. The potential of exploiting genetic association between GY and PH predictively brings a new perspective regarding selection strategies in advanced sorghum testing. Our results imply that, despite strong selection for appropriate height is generally imposed in early stages of breeding, multi-trait evaluation could be implemented for increasing the capacity to use remaining genetic variation in PH when selecting for GY. That is, selection

based on multi-trait models is able to directly use the variation in GY that is not associated with PH, and can simultaneously exploit GY variability indirectly through the available variation in PH (within the acceptable range).

Based on findings from this study, when SG is the target of prediction, most advantage from multi-trait genomic analysis can be obtained only if predicted lines have been phenotyped for GY. This may be beneficial when the interest is to predict the genetic aptitude for drought-adaptation in lines that have been field-tested for GY performance but could not yet experience post-flowering water stress. Such scenario is compatible with selection for broad adaptation across the TPE, which is generally the best approach to deal with the largely unpredictable occurrence of drought in the Australian sorghum region (Chapman et al., 2000). Expression of functional SG can also be a consequence of reduced sink demand relative to source, due to low grain production at plant level (Henzell and Gillieron, 1973; Duncan et al., 1981). For this reason, Borrell et al. (2014) pointed out that simultaneous selection for SG and GY should be applied in sorghum breeding programs to correct for functional SG that is actually driven by low sink demand. This correction can be automatically performed by joint analysis of both traits using multi-trait genomic models. For instance, in our case, selection of parental lines for broad adaptation would be based on a predicted genetic score that is optimally derived by combining three sources of information: direct information from own GY breeding value in well-watered environments, and indirect information from SG and GY breeding value of relatives in water-limited environments.

CONCLUSION

This study demonstrates, based on extensive breeding data, that there is potential to improve genome-based predictions of grain yield and stay-green traits in grain sorghum by using multi-trait genomic analysis. Results suggest that better predictive abilities and accuracies for GY prediction are obtained when PH information on the training lines is included in multi-trait

genomic models. When SG is the target, the quality of predictions is likely to improve only if GY performance data is available for both the training and the predicted lines *per se* since, in this case, direct information from SG-GY genetic correlations is exploited predictively through multi-trait analysis. This article also shows, for the first time in plant breeding, how a similarity matrix using trait-specific combinations of pedigree and marker-based relatedness can further enhance multi-trait genomic prediction. Collectively, our results imply useful properties of multi-trait BLUP to evaluate alternative prediction schemes for genetic improvement of crop productivity and drought adaptability in grain sorghum. Given that the traits considered in this study are commonly measured in cereal breeding programs, findings presented here can be also relevant for practical implementation in other major crops.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

JV, DJ, MM, and FvE designed the research. JV performed the statistical analyses and wrote the manuscript. DJ and FvE edited the manuscript. DJ coordinated the experiments and data collection. CH and EM processed and prepared the dataset. All authors read and approved the final version of the manuscript.

FUNDING

The data used in this research was generated with funding from the Grains Research and Development Corporation (GRDC), Australia. JV received financial support from the National Institute of Agricultural Technology (INTA), Argentina, Res. DN 1126/13.

REFERENCES

- Acquaah, G. (ed.) (2012). "Breeding Sorghum". In *Principles of Plant Genetics and Breeding, Second Edition* (Chichester: John Wiley & Sons, Ltd.).
- Borrell, A. K., Hammer, G. L., and Henzell, R. G. (2000). Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. *Crop Sci.* 40, 1037–1048. doi: 10.2135/cropsci2000.4041026x
- Borrell, A. K., Mullet, B. R., George-Jaeggli, B., van Oosterom, E. J., Hammer, G. L., Klein, P. E., et al. (2014). Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. *J. Exp. Bot.* 65, 6251–6263. doi: 10.1093/jxb/eru232
- Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. G., and Thompson, R. (2017). *ASReml-R Reference Manual Version 4*. Hemel Hempstead: VSN International Ltd.
- Calus, M. P., and Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genet. Sel. Evol.* 43, 26. doi: 10.1186/1297-9686-43-26
- Chapman, S. C., Cooper, M., Hammer, G. L., and Butler, D. G. (2000). Genotype by environment interactions affecting grain sorghum. II Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Aust. J. Agric. Res.* 51, 209–221. doi: 10.1071/AR99021
- Christensen, O. F. (2012). Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet. Sel. Evol.* 44, 37. doi: 10.1186/1297-9686-44-37
- Christopher, J. T., Manschadi, A. M., Hammer, G. L., and Borrell, A. K. (2008). Developmental and physiological traits associated with high yield and stay-green phenotype in wheat. *Aust. J. Agric. Res.* 59, 354–364. doi: 10.1071/AR07193
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquin, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011
- de los Campos, G., Gianola, D., Rosa, G. J., Weigel, K. A., and Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet. Res.* 92, 295–308. doi: 10.1017/S0016672310000285

- de los Campos, G., Sorensen, D., and Gianola, D. (2015). Genomic heritability: what is it? *PLoS Genet.* 11:e1005048. doi: 10.1371/journal.pgen.1005048
- dos Santos, J. P. R., Vasconcellos, R. C. C., Pires, L. P. M., Balestre, M., and Von Pinho, R. G. (2016). Inclusion of dominance effects in the multivariate GBLUP model. *PLoS One* 11:e0152045. doi: 10.1371/journal.pone.0152045
- Duncan, R. R., Bockholt, A. J., and Miller, F. R. (1981). Descriptive comparison of senescent and nonsenescent *Sorghum* genotypes. *Agron. J.* 73, 849–853. doi: 10.2134/agronj1981.00021962007300050024x
- Fernandes, S. B., Dias, K. O. G., Ferreira, D. F., and Brown, P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass *Sorghum*. *Theor. Appl. Genet.* 131, 747–755. doi: 10.1007/s00122-017-3033-y
- Fernando, R. L., and Gianola, D. (2018). “Bayesian inference of genomic similarity among individuals from markers and phenotypes”, In *Proceedings of the World Congress on Genetics Applied to Livestock Production*, Auckland.
- Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41, 55. doi: 10.1186/1297-9686-41-55
- Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. L. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* 183, 347–363. doi: 10.1534/genetics
- Gianola, D., de los Campos, G., Toro, M. A., Naya, H., Schön, C.-C., and Sorensen, D. (2015). Do molecular markers inform about pleiotropy? *Genetics* 201, 23–29. doi: 10.1534/genetics.115.179978
- Gianola, D., Weigel, K. A., Krämer, N., Stella, A., and Schön, C.-C. (2014). Enhancing genome-enabled prediction by bagging genomic BLUP. *PLoS One* 9:e91693. doi: 10.1371/journal.pone.0091693
- Gregersen, P. L., Culetic, A., Boschian, L., and Krupinska, K. (2013). Plant senescence and crop productivity. *Plant. Mol. Biol.* 82, 603–622. doi: 10.1007/s11103-013-0013-8
- Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., and Su, G. (2014). Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 15:30. doi: 10.1186/1471-2156-15-30
- Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi: 10.1534/genetics.107.081190
- Haile, J. K., N’Diaye, A., Clarke, F., Clarke, J., Knox, R., Rutkoski, J., et al. (2018). Genomic selection for grain yield and quality traits in durum wheat. *Mol. Breed.* 38, 75. doi: 10.1007/s11032-018-0818-x
- Henderson, C. R., and Quaas, R. L. (1976). Multiple trait evaluation using relatives’ records. *J. Anim. Sci.* 43, 1188–1197. doi: 10.2527/jas1976.4361188x
- Henzell, R. G., and Gillieron, W. (1973). Effects of partial and complete panicle removal on the rate of death of some *Sorghum* bicolor genotypes under moisture stress. *Qld. J. Agric. Anim. Sci.* 30, 291–299.
- Heslot, N., Yang, H. P., Sorrells, M. E., and Jannink, J. L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Sci.* 52, 146–160. doi: 10.2135/cropsci2011.06.0297
- Holland, J. B. (2006). Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS proc MIXED. *Crop Sci.* 46, 642–654. doi: 10.2135/cropsci2005.0191
- Hunt, C. H., van Eeuwijk, F. A., Mace, E. S., Hayes, J. B., and Jordan, D. R. (2018). Development of genomic prediction in sorghum. *Crop Sci.* 58, 690–700. doi: 10.2135/cropsci2017.08.0469
- Ilska, J. J., Meuwissen, T. H. E., Kranis, A., and Woolliams, J. A. (2017). Use and optimization of different sources of information for genomic prediction. *Genet. Sel. Evol.* 49, 90. doi: 10.1186/s12711-017-0365-7
- Jia, Y., and Jannink, J. L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192, 1513–1522. doi: 10.1534/genetics.112.144246
- Jordan, D. R., Hunt, C. H., Cruickshank, A. W., Borrell, A. K., and Henzell, R. G. (2012). The relationship between the stay-green trait and grain yield in elite *Sorghum* hybrids grown in a range of environments. *Crop Sci.* 52, 1153–1161. doi: 10.2135/cropsci2011.06.0326
- Jordan, D. R., Mace, E. S., Cruickshank, A. W., Hunt, C. H., and Henzell, R. G. (2011). Exploring and exploiting genetic variation from unadapted *Sorghum* germplasm in a breeding program. *Crop Sci.* 51, 1444–1457. doi: 10.2135/cropsci2010.06.0326
- Jordan, D. R., Tao, Y., Godwin, I. D., Henzell, R. G., Cooper, M., and McIntyre, C. L. (2003). Prediction of hybrid performance in grain sorghum using RFLP markers. *Theor. Appl. Genet.* 106, 559–567. doi: 10.1007/s00122-002-1144-5
- Karaman, E., Lund, M. S., Anche, M. T., Janss, L., and Su, G. (2018). Genomic prediction using multi-trait weighted GBLUP accounting for heterogeneous variances and covariances across the genome. *G3 (Bethesda)* 8, 3549–3558. doi: 10.1534/g3.118.200673
- Kulwal, P. L. (2016). “Association mapping and genomic selection—where does sorghum stand?”, In *The Sorghum Genome. Compendium of Plant Genomes*, eds S. Rakshit and Y. H. Wang (Cham: Springer).
- Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., and Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor. Appl. Genet.* 131, 2719–2731. doi: 10.1007/s00122-018-3186-3
- Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.
- Lyra, D. H., Mendonça, L. F., Galli, G., Alves, F. C., Granato, I. S. C., and Fritschete-Neto, R. (2017). Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. *Mol. Breed.* 37, 80. doi: 10.1007/s11032-017-0681-1
- McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., et al. (2018). The *Sorghum* bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354. doi: 10.1111/tpj.13781
- 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, 1819–1829.
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2016). Genomic selection: a paradigm shift in animal breeding. *Anim. Front.* 6, 6–14. doi: 10.2527/af.2016-0002
- Meyer, K. (1989). Approximate accuracy of genetic evaluation under an individual animal model. *Livest. Prod. Sci.* 21, 87–100. doi: 10.1016/0301-6226(89)90041-9
- Momen, M., Mehrgardi, A. A., Sheikhy, A., Esmailizadeh, A., Fozzi, M. A., Kranis, A., et al. (2017). A predictive assessment of genetic correlations between traits in chickens using markers. *Genet. Sel. Evol.* 49, 16. doi: 10.1186/s12711-017-0290-9
- Mrode, R. A. (2005). *Linear Models for the Prediction of Animal Breeding Values*, 2nd Edn. Oxfordshire: CABI Publishing.
- Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The *Sorghum* bicolor genome and the diversification of grasses. *Nature* 457, 551–556. doi: 10.1038/nature07723
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* 6, 15–32. doi: 10.1214/ss/1177011926
- Rodríguez-Álvarez, M. X., Boer, M. P., Eilers, P. H. C., and van Eeuwijk, F. A. (2018a). *SpATS: Spatial Analysis of Field Trials With Splines. R Package Version 1.0–8*. <https://CRAN.R-project.org/package=SpATS> (accessed September, 2018).
- Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., and Eilers, P. H. C. (2018b). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spat. Stat.* 23, 52–71. doi: 10.1016/j.spasta.2017.10.003
- Rodríguez-Ramilo, S. T., García-Cortés, L. A., and González-Recio, O. (2014). Combining genomic and genealogical information in a reproducing kernel Hilbert spaces regression model for genome-enabled predictions in dairy cattle. *PLoS One* 9:e93424. doi: 10.1371/journal.pone.0093424
- Rutkoski, J., Poland, J., Mondal, S., Autrique, E., González Pérez, L., Crossa, J., et al. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 (Bethesda)* 6, 2799–2808. doi: 10.1534/g3.116.032888
- Schön, C.-C., Utz, H. F., Groh, S., Truberg, B., Openshaw, S., and Melchinger, A. E. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167, 485–498. doi: 10.1534/genetics.167.1.485
- Schulthess, A. W., Wang, Y., Miedaner, T., Wilde, P., Reif, J. C., and Zhao, Y. (2016). Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor. Appl. Genet.* 129, 273–287. doi: 10.1007/s00122-015-2626-6
- Schulthess, A. W., Zhao, Y., Longin, C. F. H., and Reif, J. C. (2018). Advantages and limitations of multiple-trait genomic prediction for *Fusarium* head blight

- severity in hybrid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 131, 685–701. doi: 10.1007/s00122-017-3029-7
- Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J., and Sorrells, M. E. (2017). Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome* 10. doi: 10.3835/plantgenome2016.11.0111
- Thompson, R., and Meyer, K. (1986). A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest. Prod. Sci.* 15, 299–313. doi: 10.1016/0301-6226(86)90071-0
- UNCTAD (2011). *Water for Food. Innovative Water Management Technologies for Food Security and Poverty Alleviation*. New York, NY: United Nations, 1–3.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- 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, 2055–2067. doi: 10.1007/s00122-019-03337-w
- Velazco, J. G., Rodríguez-Álvarez, M. X., Boer, M. P., Jordan, D. R., Eilers, P. H. C., Malosetti, M., et al. (2017). Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. *Theor. Appl. Genet.* 130, 1375–1392. doi: 10.1007/s00122-017-2894-4
- Vitezica, Z. G., Aguilar, I., Misztal, I., and Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genet. Res.* 93, 357–366. doi: 10.1017/S001667231100022X
- Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., et al. (2017). Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity (Edinb)* 118, 302–310. doi: 10.1038/hdy.2016.87
- Wang, X., Singh, D., Marla, S., Morris, G., and Poland, J. (2018). Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. *Plant Methods* 14, 53. doi: 10.1186/s13007-018-0324-5
- Wimmer, V., Albrecht, T., Auinger, H. J., and Schön, C.-C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28, 2086–2087. doi: 10.1093/bioinformatics/bts335
- Wimmer, V., Lehermeier, C., Albrecht, T., Auinger, H. J., Wang, Y., and Schön, C.-C. (2013). Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics* 195, 573–587. doi: 10.1534/genetics.113.150078
- Conflict of Interest Statement:** 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.
- Copyright © 2019 Velazco, Jordan, Mace, Hunt, Malosetti and van Eeuwijk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.