Multi-Trait Genomic Prediction of Yield-Related Traits in US Soft Wheat under Variable Water Regimes

Jia Guo [1,†], Jahangir Khan [1,†], Sumit Pradhan [1], Dipendra Shahi [1], Naeem Khan [1],
Muhsin Avci [1], Jordan Mcbreen [1], Stephen Harrison [2], Gina Brown-Guedira [3],
Joseph Paul Murphy [4], Jerry Johnson [5], Mohamed Mergoum [5], Richanrd Esten Mason [6],
Amir M. H. Ibrahim [7], Russel Sutton [7], Carl Griffey [8] and Md Ali Babar [1,*]

[1] Department of Agronomy, University of Florida, Gainesville, FL 32611, USA; neojiaguo@gmail.com (J.G.);
   jkazrc@UFL.EDU (J.K.); sumitp@ufl.edu (S.P.); dshahi@ufl.edu (D.S.); naeemkhan@ufl.edu (N.K.);
   mavci@ufl.edu (M.A.); jcmcbreen@ufl.edu (J.M.)
[2] School of Plant Environment and Soil Sciences, Louisiana State University, Baton Rouge, LA 70803, USA;
   sharrison@agctr.lsu.edu
[3] USDA-ARS, North Carolina State University, Raleigh, NC 27607, USA; Gina.Brown-Guedira@ars.usda.gov
[4] Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27607, USA;
   paul_murphy@ncsu.edu
[5] Department of Crop and Soil Sciences, University of Georgia, Griffin, GA 32223, USA; jjohnso@uga.edu (J.J.);
   mmergoum@uga.edu (M.M.)
[6] Department of Crop Soil and Environmental Sciences, University of Arkansas, Fayetteville, AR 72701, USA;
   esten@uark.edu
[7] Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA;
   aibrahim@tamu.edu (A.M.H.I.); r-sutton@tamu.edu (R.S.)
[8] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA 24061, USA;
   cgriffey@exchange.vt.edu
* Correspondence: mababar@ufl.edu
† These authors contributed equally to this work.

Abstract: The performance of genomic prediction (GP) on genetically correlated traits can be improved through an interdependence multi-trait model under a multi-environment context. In this study, a panel of 237 soft facultative wheat (*Triticum aestivum* L.) lines was evaluated to compare single- and multi-trait models for predicting grain yield (GY), harvest index (HI), spike fertility (SF), and thousand grain weight (TGW). The panel was phenotyped in two locations and two years in Florida under drought and moderately drought stress conditions, while the genotyping was performed using 27,957 genotyping-by-sequencing (GBS) single nucleotide polymorphism (SNP) makers. Five predictive models including Multi-environment Genomic Best Linear Unbiased Predictor (MGBLUP), Bayesian Multi-trait Multi-environment (BMTME), Bayesian Multi-output Regressor Stacking (BMORS), Single-trait Multi-environment Deep Learning (SMDL), and Multi-trait Multi-environment Deep Learning (MMDL) were compared. Across environments, the multi-trait statistical model (BMTME) was superior to the multi-trait DL model for prediction accuracy in most scenarios, but the DL models were comparable to the statistical models for response to selection. The multi-trait model also showed 5 to 22% more genetic gain compared to the single-trait model across environment reflected by the response to selection. Overall, these results suggest that multi-trait genomic prediction can be an efficient strategy for economically important yield component related traits in soft wheat.

1. Introduction

From 2007 to 2050, farmers will need to increase the production of cereals by 60% to feed over 9.5 billion people in the world [1]. Meanwhile, this must be done under a continuously changing environment due to extreme weather conditions, land pressure, and increased energy use [2–5]. Hence, it is of paramount importance to renovate breeding technologies to increase food production while mitigating pressure on the environment. Genomic prediction (GP), originally proposed by Meuwissen et al. [6], is becoming widely used by plant breeders in recent years to advance breeding progress. The availability of high-throughput phenotyping and cost-effective genotyping technologies were the most important factors for the successful and effective implementation of GP in plant breeding [7,8]. With the help of improved statistical models, GP can be augmented to be more accurate and applicable in various scenarios in plant breeding such as multi-trait and multi-environment schemes.

Compared to traditional marker-assisted selection, GP does not require prior knowledge about a few, large-effect quantitative trait loci, since all genotypic markers are curated in training prediction models [6]. Essentially, the genomic estimated breeding value (GEBV) of individuals can be calculated using genome-wide molecular markers and phenotypic data. Then, a predictive model is constructed using a training set of individuals with known phenotypic and genotypic information. In the validation set of individuals, GEBV are calculated based on their genotypic information and the previously constructed model. Then, the accuracy of the predictive models is evaluated by cross-validation approaches within and among environment. Several empirical studies have shown that GP is effective in accelerating breeding cycles and improving genetic gains per unit of time in major crops [8–10].

A key component in GP is the choice and optimization of models that are used to estimate the marker effect. Statistical models with different capacities are required to handle the ever-growing magnitude of phenotypic and genotypic data. The number of predictor variables (p) is much larger than the number of observations (n) due to the improved availability of genotypic data compared to phenotypic. As such, penalized GP models such as ridge-regression best linear unbiased prediction (rrBLUP), least absolute shrinkage and selection operator (LASSO), and elastic net are employed to control the trade-offs between lack of fit and model complexity [6,11,12]. Bayesian methods are also used for parameterization in GP models [13–15]. As most of these models are univariate and focus on predicting one dependent variable at the time, a multivariate model incorporating associations among several dependent variables can improve the power of predictive models [16–18]. A multivariate model is also effective in predicting continuous variables closely associated with each other, which is a common situation for quantitative traits such as grain yield and nutrition content in cereal crops [16,19,20]. In large-scale plant breeding programs, multi-environmental trials add another layer of challenges in dissecting information from genotype × environment interaction (G×E). A three-way genomic model for evaluating the prediction accuracy of trait × genotype × environment could advance GP in modern plant breeding programs.

A few multivariate and/or multi-environment predictive models have been proposed for binary, ordinal, count, and continuous traits. Several studies applied a Bayesian-based model for multi-trait analysis and observed improved accuracies compared to single-trait analysis [16,19,21]. Burgueño et al. [22] and López et al. [23] extended the single-environment model to a multi-environment–multi-trait context and reported a significant improvement in GP model accuracy. In empirical field experiments, Montesinos-López et al. [24] and Guo et al. [25] observed that prediction models that incorporated hyperspectral data or other physiological traits (canopy temperature, membrane thermostability, chlorophyll content, stay green, and rate of senescence)

and spectrum/trait by environment interaction terms were more accurate than those that did not. Guo et al. [25], Crain et al. [26], and Krauss et al. [27] also reported improved prediction accuracies using a multi-environment model relative to a single-environment model. Two Bayesian-based mixed multi-trait models, Bayesian Multi-trait Multi-environment (BMTME) and Bayesian Multi-output Regressor Stacking (BMORS) models were proposed by Montesinos-Lopez et al. [28,29]. The BMTME model assesses the variance–covariance structure among trait, genotype, and environment, and it jointly predicts multiple traits evaluated in multiple environments [23]. The BMORS model first calculates genomic best linear unbiased predictions (GBLUP) for each trait and then corrects accuracy in a secondary model using the prediction of the first-stage GBLUP model [23,29]. An improved version of the BMTME model was proposed by Montesinos-Lopez et al. (Montesinos-López et al. 2019b), which was equipped with optimization algorithms for efficiently applying the software to real data. Deep learning (DL) algorithms have led to success in bioinformatics research due to its versatility and flexibility [30]. One of the DL algorithms, neural networks (NNs), has showed comparable prediction accuracy to statistical models for complex human and animal traits [31–33]. A few studies have reported the performance of DL algorithms in plant genomic prediction. Liu and Wang [34] indicated that NNs had higher prediction accuracies compared to Bayesian or ridge regression-based methods using a set of soybean data. Ma et al. [35] compared NNs to a GBLUP model in a large set of wheat data and reported a better performance for NNs in terms of higher phenotypic value in top selected individuals and lower sensitivity to outliers. Montesinos-Lopez et al. [36] proposed a multi-trait deep learning model and compared it to the GBLUP model using maize (*Zea mays* L.) and wheat data, which showed higher prediction accuracies for NNs when the genotypic by environmental (G × E) effect was ignored, while lower prediction accuracies were observed when G×E effect was involved. In addition to model selection, the genetic structure of the trait, marker density, sample size, and composition of training population (TP) and validation population (VP) are also important factors in GP accuracy [19,37–41].

Yield component traits such as harvest index (HI), spike fertility (SF: ratio of grain number per spike to chaff weight per spike), and thousand grain weight (TGW) play important roles in the determination of grain yield (GY) in wheat. Strong genetic correlations were observed among GY, HI, SF, and TGW under different environmental conditions [42–46]. Thousand grain weight is usually a highly inheritable trait that positively contributes to GY [47,48]. Increasing the grain number through maximizing the partitioning of assimilates (e.g., carbohydrate) to grain instead of a non-grain part of the spike is a noble and effective approach to increase grain yield [49–51]. Therefore, manipulation of the grain number and spike chaff dry weight is a potential avenue for yield increase in wheat, which is supported by several field studies [46,52–55]. In addition to the significant genetic correlation with the performance, SF demonstrated strong genetic variations in advanced breeding lines and early generation breeding populations in wheat [46,56,57]. Guo et al. [25] observed high prediction accuracy (0.3–0.5) for fertile spikelet number and spike length while observing low prediction accuracy (<0.2) for SF and spikelet density (ratio of spikelet number per spike to spike length) in spring wheat. In another study using two doubled haploid wheat populations, moderate to low accuracies were observed for grain number per spike in controlled (0.10–0.42) and osmotic stress (0.27–0.46) conditions [58]. Improvement of these yield component traits is an ideal solution for enhancing sink capacity in wheat. Currently, there is no information available on multi-trait genomic prediction for GY, HI, SF, and TGW in wheat. This approach could provide an accurate prediction for jointly improving grain yield-related traits in wheat. The proposed study will provide critical information in the development of wheat germplasm through optimized yield components traits using GP. Therefore, the objectives of this study were to (1) estimate the genetic correlations among GY, HI, SF, and TGW in a multi-environment scenario, (2) compare the prediction accuracies of single- against multi-trait models under a multi-environment context, and (3) estimate the response to selection (RS) of grain yield from single- and multi-trait models under a multi-environment context.

2. Materials and Methods

2.1. Site Description

The experiment was conducted over two growing seasons from 2016 to 2018 at Citra and Quincy, Florida (Table 1). Citra is characterized by sandy soil with loam at 20–80 inches with low water-holding capacity, whereas Quincy has well-drained loamy soil with higher water-holding capacity than Citra. Citra had moderate precipitation (212–447 mm) during 2016–2018, moderate humidity, and temperature rise above >30 °C multiple times during the grain-filling stages. Quincy received higher precipitation (582–625 mm), had high humidity, and experienced relatively fewer episodes of high temperatures (>30 °C) during the same period. Experiments were planted in between mid-November to the first week of December.

Table 1. Experimental site information including name of the sites, years evaluated, coordinates, and soil type.

Site	Year	Coordinates	Soil Type [1]
Citra	2016–2017	29°24′18″ N 82°10′22″ W	Well-drained sandy soil with loamy subsoil at 20–80 inches
	2017–2018	29°24′32″ N 82°10′46″ W	
Quincy	2016–2017	30°33′04″ N 84°35′51″ W	Well-drained loamy soils
	2017–2018	30°32′45″ N 84°35′46″ W	

[1] Source: Soil Map of Florida (EUDASM).

2.2. Plant Genetic Material and Experimental Design

The genetic material used for the study consisted of 240 (237 + 3 checks) facultative soft wheat genotypes selected from the Gulf Atlantic Wheat Nursery (GAWN). The genotypes were developed by public wheat breeding programs (North Carolina State University, Texas A&M, Louisiana State University, University of Georgia, University of Arkansas, and Virginia Tech) targeting for the south and southeastern regions of the USA. The panel is referred to as GAWN panel. The genotypes in the panel generally require a short duration of cold treatment to satisfy the vernalization for flower induction. The panel was evaluated at two locations in Florida: Citra and Quincy during 2016–2018 (two years). To induce terminal drought stress at Citra, irrigation was stopped 2 weeks before anthesis (GS60) until maturity. Contrary to that, 1–2 supplemental irrigations were applied when needed in Quincy. Citra is considered to be a drought-stressed environment, and Quincy is moderately drought stressed. All trials were planted in six-row plots (3 m length × 1.5 m width) using a seeding rate of 100 kg h^{-1}. The GAWN panel was planted in an incomplete block augmented design with repeated checks (AGS2000, PI 656845; SS8641, PI 674197; Jamestown, PI 653731) in each block with 237 unreplicated new entries [59]. Three repeated checks are widely adapted and cultivated throughout the southeastern US. To control foliar and glume diseases, fungicides were sprayed three times. Herbicides were sprayed to control weeds as required. Fertilizers were applied through irrigation for best management practices for proper growth and yield.

2.3. Traits Measurement

Phenotypic data for days to heading (DTH), grain yield (GY), harvest index (HI), spike fertility (SF), and thousand grain weight (TGW) were collected. Days to heading (GS 59) were collected using the Zadoks scale [60]. GY was measured by harvesting all six rows using a small plot harvester and was calculated by dividing the total grain weight by plot area, adjusted to 12% moisture, and converted to t ha^{-1}. To measure SF, ten random spikes were sampled from the field at physiological maturity, dried for 72 h at 60 °C, and threshed by a single head thresher to determine chaff weight (the non-grain part of a spike), which was calculated as the difference between total spike dry weight and spike grain weight. Spike fertility (SF) was calculated as a ratio of grains m^{-2} to spike chaff weight m^{-2} [44]. Grain m^{-2} was obtained by using seeds from SF sample (grains per spike) multiplied to the number of

spikes m^{-2}. To get spike number m^{-2}, we harvested tillers at maturity from 0.5 m^2 middle two rows and counted and later converted to m^2. The HI was calculated as the ratio of grain weight m^{-2} to total dry matter m^{-2}. TGW was measured by weighing 1000 grains counted through a seed counter (Seedburo Equipment Co., Chicago, IL, USA).

2.4. Phenotypic Data Analysis

The best linear unbiased estimates (BLUEs) and standard errors were calculated for DTH, GY, HI, SF, and TGW using the following equation assuming genotype as a fixed effect and environment and block as random effects:

$$Y'_{ijk} = \mu + Gg_j + E_i + B_{i(k)} + GgE_{ji} + e_{ijk} \tag{1}$$

where Y_{ijk} is the observed value; μ was the general mean; Gg_j is the genotypic effect (j = 1 to 223); E_i is the environment effect (i = 1 to 4, corresponding to Citra 2017, Citra 2018, Quincy 2017, and Quincy 2018); $B_{j(k)}$ is the block effect (k = 1 to 12; $N[0, \sigma B2]$) nested within the ith environment; GgE_{ji} is the jth genotype by i^{th} environment interaction effect; and e_{ijk} is the random error ($N[0, \sigma e2]$). Block and environmental effects and error are commonly modeled to follow independent normal distributions [61]. To evaluate the influence of phenology, DTH was included as an additional fixed effect in model (1) for all following analyses. The broad sense heritability (H^2) from each environment was calculated using the following formula, $H^2 = (\sigma^2{}_G)/(\sigma^2{}_G + \sigma^2{}_e)$, where $\sigma^2{}_G$ and $\sigma^2{}_e$ were variances due to genotype and error, respectively. Genotype and block were considered as random effects. In order to estimate variance values, we used the following model below:

$$Y'_{ij} = \mu + Gg_j + B_k + e_{jk}. \tag{2}$$

Pearson correlation analyses among four phenotypic traits were also calculated.

2.5. Genotypic Data Analysis

Fresh green seedling leaf tissues of each line were used to get genomic DNA through the LGC Genomics Oktopure robotic extraction platform along with Sbeadex magnetic microplate reagent kits. Genotyping by sequencing (GBS) was performed using Illumina HiSeq 2500 after double digestion of genomic DNA with *Pstl* and *Msel* restriction enzymes [38]. SNP calling was carried out using TASSEL-GBS v5.2.49 [62,63]. The Illumina platform generated short reads were aligned using Burrows-Wheeler Aligner v0.7.17-r1188 to the Chinese Spring IWGSC RefSeq v1.0 wheat reference sequence [64]. Pre and post imputation filtering were used for retaining biallelic SNPs and removing SNP with missing data >50%, with minor allele frequencies <5%. We also remove genotypes with >85% missing data. Then, missing data were imputed using Beagle 5.1, and later, data was re-filtered to remove SNPs with minor allele frequency (MAF) <5% or heterozygous call frequency of <10%. A Fisher's exact test was used to test if the SNP alleles at each site were independent in a population of inbred lines, as described by Poland et al. [65]. The SNPs were assumed to be allelic in the population if the null hypothesis of independence for the two alleles was rejected (α = 0.001). This procedure typically lowers heterozygous calls due to sequencing errors, genome duplications, and homologous sequences on different genomes [38,65,66]. In the final genomic dataset, a total of 27,957 SNPs remained.

2.6. Prediction Models

Three statistical models including the Multi-environment Genomic Best Linear Unbiased Predictor (MGBLUP), Bayesian Multi-trait Multi-environment (BMTME) model, and Bayesian Multi-output Regressor Stacking (BMORS), and two deep learning (DL) models including Single-trait Multi-environment Deep Learning (SMDL) and Multi-trait Multi-environment Deep Learning (MMDL) were compared for predicting GY, HI, SF, and TGW.

2.6.1. Multi-Environment Genomic Best Linear Unbiased Predictor (MGBLUP) Model

According to Montesinos-López et al. [24,67], a brief summary of three statistical models are presented in the following sections. A univariate linear mixed model is often used for accounting for effects of environment and environment × genotype interaction:

$$Y_{ij} = E_i + G_j + GE_{ij} + \varepsilon_{ij} \tag{3}$$

where Y_{ij} is the best linear unbiased estimate (BLUE) of predicted trait for j^{th} genotype in i^{th} environment; E_i is the environment effect (i = 1 to 4, corresponding to Citra 2017, Citra 2018, Quincy 2017, and Quincy 2018); G_j is the genetic main effect (j = 1 to 223); the genetic main effect is assumed as a joint distribution of genotype effect with a multivariate normal distribution $\mathbf{G} = \left(G_1, \ldots, G_{j*}\right)^T \sim MN\left(0, \sigma_G^2\mathbf{G}\right)$, where σ_G^2 denotes the genomic variance and $\mathbf{G}$ represents the genomic relationship matrix; the $\mathbf{G}$ matrices were calculated as $\mathbf{G} = \frac{XX'}{p}$, where X is a matrix of the centered and standardized SNP marker matrix and p is the number of SNP markers; $\mathbf{GE}_{ji}$ is the jth genotype by i^{th} environment interaction effect; the term $\mathbf{GE}_{ji}$ was assumed to have a multivariate normal distribution, that is $\mathbf{GE}_{ji} = \left(GE_{11}, \ldots, GE_{ji}\right)^T \sim MN\left(0, \left(Z_gGZ_g^T\right)\#\left(Z_EG_E^T\right)\sigma_{GE}^2\right)$ where Z_g and Z_E are incidence matrices for the vector of genomics and environment effects, and σ_{GE}^2 is the variance component for $\mathbf{GE}_{ji}$; ε_{ij} is a random residual associated with the jth line in the ith environment distributed as $N(0, \sigma^2)$ where σ^2 is the residual variance.

2.6.2. Bayesian Multi-Trait Multi-Environment (BMTME) Model

For the BMTME model, a matrix-variate normal distribution is assumed denoted as $M \sim NM_{n\times p}(\mathbf{H}, \Omega, \Sigma)$. The $(np \times 1)$ random vector $vec(\mathbf{M})$ is distributed as multivariate normal as $N_{np}(vec(\mathbf{H}), \Sigma \otimes \Omega)$; $\mathbf{H}$ is a $n \times p$ location matrix, Σ is a $p \times p$ first covariance matrix, and Ω is a $n \times n$ second covariance matrix. n is the number of genotypes, and p is the number of SNPs; $vec(.)$ and $\otimes$ are the standard vector operator and Kronecker product, respectively. Therefore, a BMTME model is defined as follows:

$$Y = X\beta + Z_1b_1 + Z_2b_2 + E \tag{4}$$

where Y is the vector of multivariate responses of $n \times L$, with L being the number of predicted traits and $n = J \times I$, where J denotes the jth genotype and I denotes the ith environment, X is a vector of $n \times I$, β is of order $I \times L$; Z_1 is of order of $n \times J$, b_1 is of order $J \times L$ and represents the genotype × trait interaction; Z_2 is a vector of order $n \times IJ$, b_2 is a vector of order $IJ \times L$ and represents the genotype × environment × trait interaction. Vector b_1 is assumed under a matrix-variate normal distribution as $NM_{J\times L}\left(0, \mathbf{G}', \Sigma_t\right)$, where $\mathbf{G}'$ denotes the genomic relationship matrix; the $\mathbf{G}$ matrices were calculated as $\mathbf{G}' = \frac{WW'}{p}$, where W is a matrix of the centered and standardized SNP marker matrix of order $J \times p$, and p is the number of SNP markers; and Σ_t is a unstructured genetic covariance matrix of traits of order $L \times L$, b_2 is assumed under a matrix-variate normal distribution as $NM_{JI\times L}\left(0, \Sigma_E \otimes \mathbf{G}', \Sigma_t\right)$, where Σ_E is an unstructured covariance matrix of order $I \times I$ and E is the matrix of residuals of order $n \times L$ with $E \sim NM_{n\times L}(0, I_n, R_e)$, where R_e is the unstructured residual covariance matrix of traits of order $L \times L$. Genetic correlations between phenotypic traits and environments were calculated as $r_{G(a,b)} = \frac{\sigma_{G(a,b)}}{\sqrt{\sigma_{G(a)}^2 \sigma_{G(b)}^2}}$, where $\sigma_{G(a,b)}$ is the covariance of traits a and b; $\sigma_{G(a)}^2$ is the genotypic variance of trait a; and $\sigma_{G(b)}^2$ is the genotypic variance of trait b.

2.6.3. Bayesian Multi-Output Regressor Stacking (BMORS) Model

The BMORS model is a two-stage predictive model originally proposed by Spyromitros-Xioufis et al. [68,69]. In the first stage, single-trait GBLUP models are established for each trait according to model (3).

In the second training stage, the information of a single-trait model is implemented in a new meta-model as follows:

$$y_{ij} = \beta_1 \hat{Z}_{1ij} + \beta_2 \hat{Z}_{2ij} + \ldots + \beta_L \hat{Z}_{Lij} + e_{ij} \tag{5}$$

where $\hat{Z}_{Lij}$ represents the scaled predictions of each trait obtained from the single-trait MGBLUP model in the first stage analysis, and β_L is the β coefficients for each prediction. Each prediction was scaled by subtracting its mean $(\hat{\mu}_{Lij})$ and dividing by its standard deviation $(\hat{\sigma}_{Lij})$, that is, $\hat{Z}_{Lij} = (\hat{y}_{Lij} - \hat{\mu}_{Lij})\hat{\sigma}_{Lij}^{-1}$. The BMORS model is an expansion of the multi-label classification method exploiting dependencies between target variables (e.g., multiple phenotypic traits in GP) in order to improve prediction accuracy [69–71]. This method captures correlations between phenotypic traits by appropriate choices of covariance functions such as the weighted regressors used in the proposed model.

2.6.4. Deep Learning (DL) Models

Single-trait Multi-environment Deep Learning (SMDL) and Multi-trait Multi-environment Deep Learning (MMDL) models delineated by Montesinos-López et al. [36] were also included in prediction analyses. In brief, a densely connected neural network consisting of an input layer, multiple output layers, and multiple hidden layers between them was constructed. Then, the input variables (e.g., SNPs) were fed into the neural network and transformed by the neurons on each hidden layer with geometric non-linear functions. The final output layer is a vector of numbers (e.g., phenotypic values), or a matrix of multiple variables (e.g., multi-trait phenotypic values) predicted by the neural network. The MMDL model has multiple output neurons instead of one neuron in the SMDL model. The success of implementing DL models relies on a fine-tuning process which is involved with selecting hyperparameters including the number of neurons, number of epochs, number of layers, type of regularization, and type of action function. Based on previous studies using similar types of data [35,36,72], we included three hidden layers and used the rectified linear activation unit as an activation function and the dropout type (25% dropout rate) as the regularization method. For our study, a second-order response surface search method with a full factorial design was implemented to find the optimal combination of number of neurons and epochs for our dataset. We evaluated numbers of neurons from 5 to 70 with an increment of 5 and numbers of epochs from 10 to 80 with an increment of 10. A quadratic plateau non-linear model was used to locate the optimal number of neurons for each level of number of epochs.

2.7. Model Evaluation

All five predictive models were evaluated using a five-fold cross-validation (CV) approach for their prediction accuracies. Under this CV, the dataset was partitioned into five subgroups of equal size; four of the five subgroups (i.e., the training population) were used to fit each prediction model, while the remaining subgroup (i.e., the validation population) was used to assess the correlation between the observed and predicted trait values. This process was repeated five times, with each subgroup being used as the prediction set once. A stratification method was employed to evaluate the influence of population structure on prediction accuracies for all three models. Briefly, the population was split into 10 clusters based on the discriminant analysis of principal components (*DAPC*) [73] clustering approach using all 27,957 SNPs, so that a similar number of lines belonging to the same cluster were present in either the validation or training population. We also used a random partitioning method without considering the underlying population structure in the panel. For DL models, the response surface search optimization was performed before CV, and an optimal combination of number of neurons and epochs was used to compare the results with the other three models. Prediction accuracies were calculated as $r_{GY} = r_p / \sqrt{H^2}$, where r_p is the mean predictive correlations across five folds. In addition, the prediction accuracy of the BMORS model was evaluated across four environments in which the dataset from each environment was predicted by the dataset from the other

three environments. The model is denoted as BMORS. Finally, both the BMTME and BMOR models were implemented with 15,000 iterations, of which 10,000 were used as burn-in to fit the models.

The standard error of prediction accuracy for each environment and each model was calculated based on SE $_{GYP} = \sigma_{r_p} / \sqrt{fH^2}$, where σ_{r_p} is the standard deviation of the predictive correlation; f is the number of folds (five in this case). Response to selection (RTS) was calculated using the formula $R = H^2S$ [74], where H^2 is the heritability for grain yield and S is the selection differential (in unit of kg ha^{-1}). To be specific, all 237 lines were ordered according to their GEBV calculated from each model in each environment. Then, the top 10% lines were chosen as the selected population (i.e., selection intensity of 10%). The selection differential was calculated as the difference of grain yield between the means of selected lines and the whole population: $S = \hat{\mu}_S - \hat{\mu}_P$, where μ_S is the mean yield of 10% selected lines based on GEBV and μ_P is the mean yield of the population. The response to selection for all three models at each environment were computed with and without correction for DTH. The mean of RTS was calculated for each environment and each model across five folds. The standard error of RTS was calculated based on SE $_{GYRTS} = \sigma_{RTS} / \sqrt{f}$, where σ_{RTS} is the standard deviation of the RTS; and f is the number of folds (five in this case).

2.8. Software Implementation

Phenotypic data analysis, including BLUPs and heritability calculation, and correlation analyses, were performed using R (R Development Core Team 2018). Basic models (1–2) were fit with the "lme4" package [75]. Genetic correlations between phenotypic traits were calculated using the "BMTME" package [67]. Prediction models (4) and (5) were fit with the package "BGLR" and "BMTME", respectively [67,76]. Two DL models were evaluated using "Keras" and "tensorflow" packages [77,78]. The *DAPC* analysis was performed using an "adegenet" package [73]. The response surface search was conducted with "rsm" package [79]. Cross-validation and prediction accuracy calculation were conducted using customized codes in R.

2.9. Data Availability

All data generated or analyzed during this study are available in the supplemental files, including phenotypic data in "multi-trait GS phenotypic data.csv" and genotypic data in "multi-trait GS genotypic data.txt".

3. Results

3.1. Descriptive Statistics

A description of GY, HI, SF, and TGW phenotypic traits is presented in Table 2. Phenotypic BLUEs and heritability values varied significantly among four environments. For Quincy, a generally lower temperature environment compared to Citra showed the highest BLUEs of GY, HI, and TGW (5.3 t ha^{-1}, 42.7%, and 40.9 g, respectively) in 2017, compared to other three environments. For Citra, a hotter and drier environment compared to Quincy had the lowest BLUEs of GY, HI, and SF (2.0 t ha^{-1}, 30.5%, and 63.9 grains/g of chaff weight, respectively) in 2018. Citra 2018 had the lowest value for TGW (34.1 g) and the highest value for SF (98.3 g). In general, Citra 2017 and Citra 2018 showed higher broad-sense heritability values than Quincy 2017 and Quincy 2018 for all four traits. For GY, Citra 2018 had the highest heritability (0.80), while Quincy 2018 had the lowest value (0.24). For HI, Citra 2017 had the highest heritability (0.78), while Quincy 2018 had the lowest value (0.26). Quincy 2017 showed the lowest heritability for SF (0.22), and Citra 2018 had the highest value (0.68). Quincy 2018 showed the lowest heritability for TGW (0.44), and Citra 2018 had the highest value (0.87). In general, a higher coefficient of variation was shown in Citra 2017 and Quincy 2017 compared to that in Citra 2018 and Quincy 2018 for all four traits.

Table 2. Description of grain yield (GY) (ton ha^{-1}), harvest index (HI) (%), spike fertility (SF) (grains/g of chaff weight), and thousand grain weight (TGW) (g) phenotypic traits† evaluated at Citra, FL and Quincy, FL in 2017 and 2018. The best linear unbiased estimates (BLUEs), standard error (SE), heritability (H2), coefficient of variation (CV), maximum and minimum value were calculated for each trait in four environments.

	Trait	BLUE	SE	H^2	CV	Min	Max
Citra 2017	GY	2.0	0.1	0.71	28.3	0.3	4.5
	HI	30.5	0.8	0.78	17.8	16	52
	SF	63.9	1.7	0.38	27.2	12	142
	TGW	34.7	0.4	0.48	10.8	24	48
Citra 2018	GY	3.8	0.1	0.80	11.5	1.0	7.0
	HI	37.4	0.4	0.74	6.7	20	48
	SF	98.3	1.2	0.68	9.5	62	161
	TGW	34.1	0.4	0.87	5.1	19	46
Quincy 2017	GY	3.3	0.1	0.36	16.6	1.5	5.6
	HI	34.3	0.4	0.43	12.6	20	47
	SF	83.2	1.7	0.22	25.1	34	148
	TGW	39.4	0.3	0.58	7.4	26	50
Quincy 2018	GY	5.3	0.1	0.24	18.4	2.1	8.8
	HI	42.7	0.3	0.26	9.8	28	54
	SF	94.6	1.4	0.32	15.6	52	158
	TGW	40.9	0.4	0.44	9.7	30	54

Genetic correlations among four traits and four environments are presented in Tables 3 and 4, respectively. The highest positive genetic correlation among traits was found between GY and HI (0.67). Relatively low genetic correlations were found between GY and SF (0.17), GY and TGW (0.18), and HI and SF (0.17). The HI and TGW had the lowest positive genetic correlation (0.10). The SF and TGW showed a negative genetic correlation (−0.32). Correlations between environments were generally low and ranged from 0.16 to 0.24. The highest and lowest correlations were found between 2017 Quincy and 2018 Quincy (0.24), and 2018 Quincy and 2018 Citra (0.16), respectively.

Table 3. Estimates of averaged genetic correlation (above diagonal) and Pearson correlation of phenotypic values (below diagonal) among GY (ton ha^{-1}), HI (%), SF (grains/g of chaff weight), and TGW (g) across four environments.

	GY	HI	SF	TGW
GY		0.67	0.17	0.18
HI	0.76		0.17	0.10
SF	0.36	0.30		−0.32
TGW	0.33	0.24	−0.23	

Table 4. Estimates of genetic correlation among four environments averaged over four traits at Citra, FL and Quincy, FL in 2017 and 2018.

|	Quincy 2017	Citra 2017	Quincy 2018	Citra 2018
Quincy 2017		0.24	0.26	0.19
Citra 2017			0.19	0.17
Quincy 2018				0.16

3.2. Prediction Accuracy

Population structure was determined by using the *DAPC* algorithm, and the panel was clustered into 10 groups (Figure 1). Each subgroup consisted of 14 to 38 lines, which were then randomly assigned to five different folds for cross-validation analysis. This process is considered as a stratification of both training and validation populations.
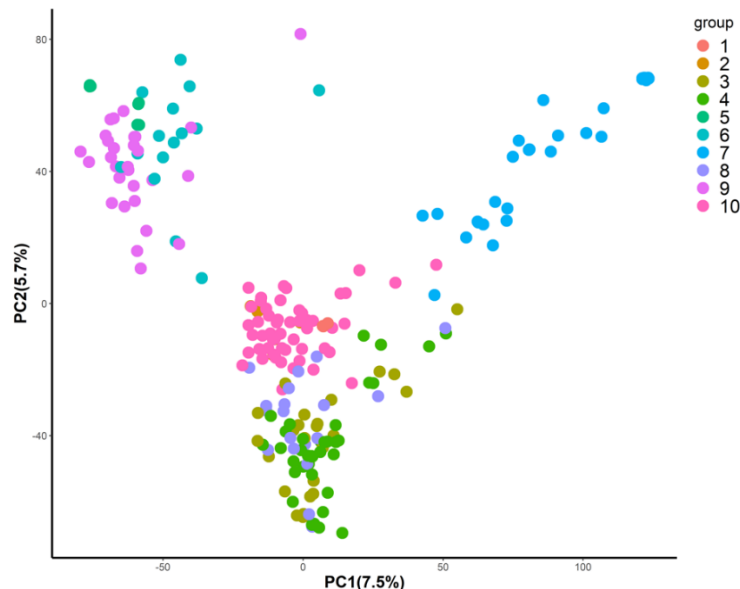

Figure 1. Stratification of genomic prediction panel inferred from discriminant analysis of principal components (DAPC) using 27,957 SNPs data. The first two principal components (account for variance of 7.5% and 5.7%, respectively) are used to represent each line in the genomic prediction (GP) panel. Each line was colored based on the posterior of probability assigned to 10 genetic groups inferred from DAPC.

For DL models, optimal epoch and neuron combinations were identified based on the results of response surface research (Supplement Figure S1). Then, the prediction accuracy of each trait was calculated based on the identified optimal epoch and neuron combination. When populations were not stratified or randomly sampled (noted as "un-stratified"), prediction accuracies ranged between −0.23 and 0.59 for GY, 0.07 and 0.55 for HI, 0.13 and 0.78 for SF, 0.20 and 0.88 for TGW among four environments and three models (Figure 2). Although predictive correlations of all models for Quincy 2018 were not significantly different from zero ($p > 0.05$), the low heritability of GY in this environment contributed to the negative predication accuracies in general. Overall, statistical models including MGBLUP and BMTME showed higher prediction accuracies than DL models (SMDL and MMDL). The BMOR model showed the highest prediction accuracies in the majority of the cases except for SF in Citra 2017. However, the differences between statistical models and DL models were minimal in some environments and traits. For example, DL models were comparable to statistical models for HI across environments. For SF, two DL models showed higher prediction accuracies compared to two statistical models in Quincy 2017 and Quincy 2018. When comparing results from four environments, Citra 2017 showed high prediction accuracies for all four traits. For GY, Quincy 2018 had lower prediction accuracies compared to other environments. Citra 2018 had lower prediction accuracies for SF and TGW compared to other environments. For HI, the prediction accuracies varied between models and environments.
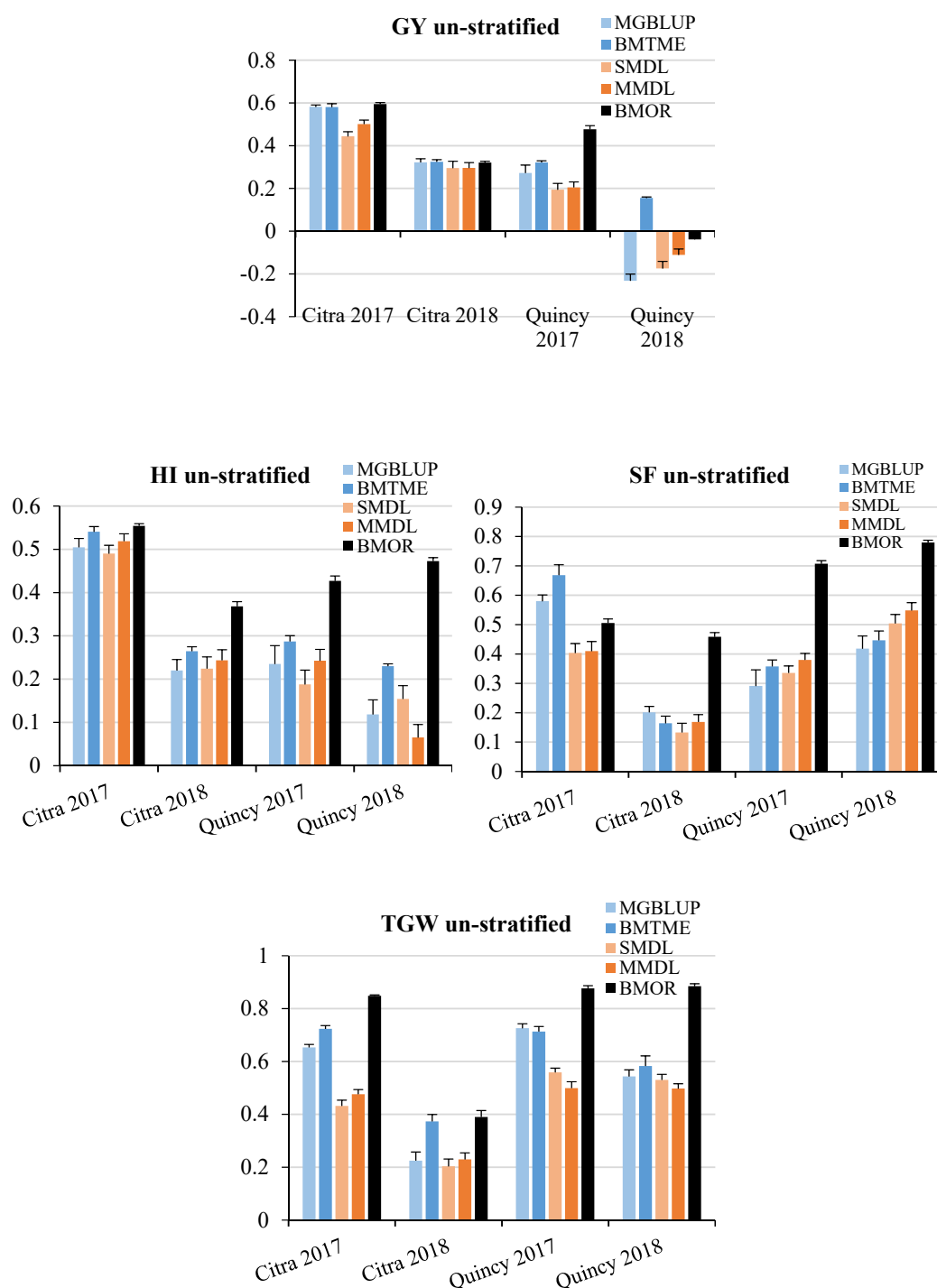
Figure 2. Prediction accuracies for GY, HI, SF, and TGW without population stratified. GY, grain yield; HI, harvest index; SF, spike fertility; TGW, thousand grain weight. Mean Pearson's correlations and standard errors for each environment were presented for each trait. Statistical models were colored in light blue and blue while DL models were colored in light orange and orange. The Bayesian Multi-output Regressor (BMOR) model was colored in black.

When populations were stratified (noted as "stratified"), prediction accuracies ranged between −0.22 and 0.62 for GY, 0.03 and 0.55 for HI, 0.16 and 0.83 for SF, and 0.21 and 0.85 for TGW among four environments and three models (Figure 3). A similar pattern was observed between the stratified and un-stratified strategy for prediction accuracy across environments (Figures 2 and 3).
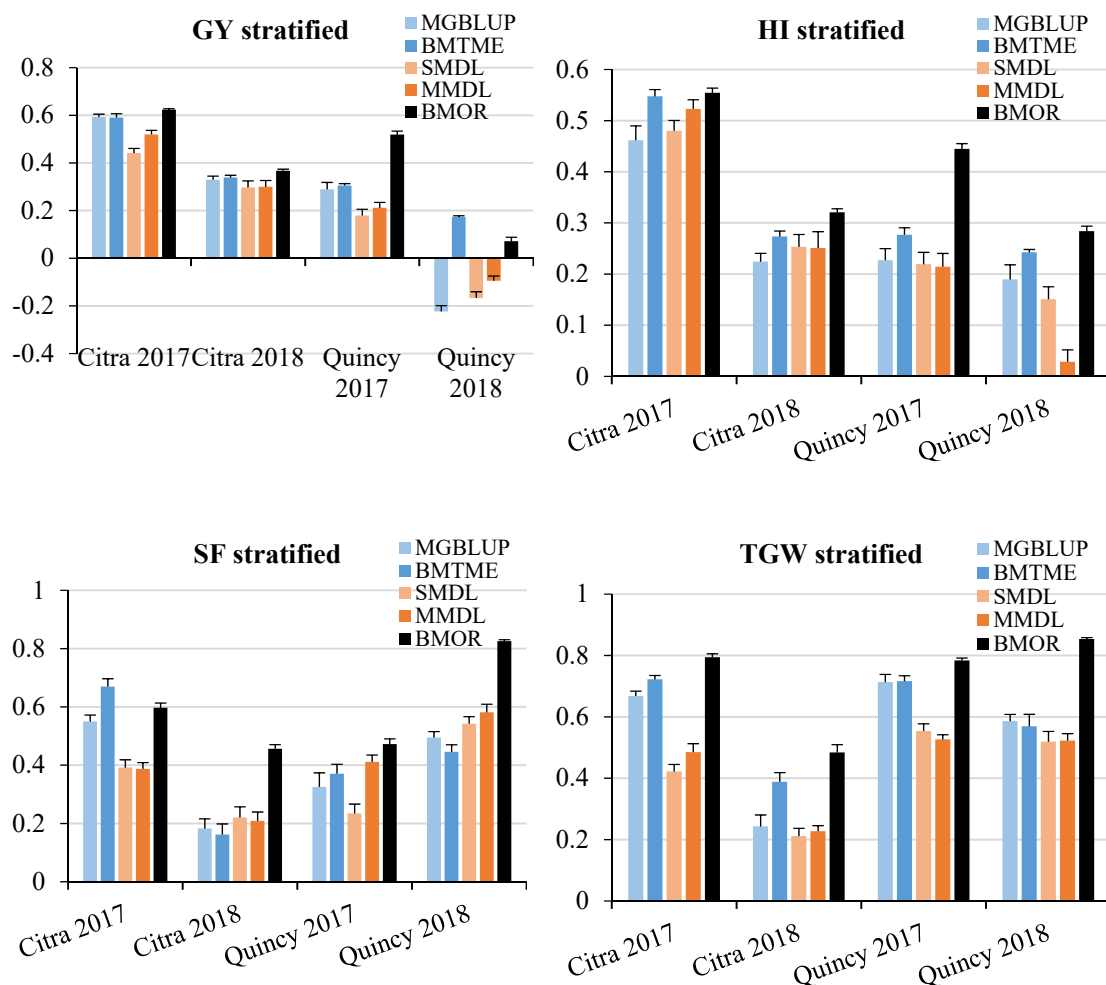
Figure 3. Prediction accuracies for GY, HI, SF, and TGW with population stratified. GY, grain yield; HI, harvest index; SF, spike fertility; TGW, thousand grain weight. Mean Pearson's correlations and standard errors for each environment were presented for each trait. Statistical models were colored in light blue and blue while DL models were colored in light orange and orange. The BMOR model is colored in black.

For MGBLUP and BMOR models, the averaged prediction accuracies across environments were higher for TGW, which was followed by SF, HI, and GY in order (Figure 4). For the BMTME model, GY had higher prediction accuracy than HI. Prediction accuracies were not significantly affected by the stratification of populations. The SMDL and MMDL models followed the same pattern and had lower prediction accuracies than statistical models when comparing the averaged values. However, the multi-traits models including BMTME and MMDL showed higher prediction accuracies than their counter-part single-trait models for all four traits. When prediction accuracies were averaged for each model, the BMOR model showed the highest prediction accuracy followed by BMTME, MGBLUP, MMDL, and SMDL in order (Figure 5).

We also applied the BMOR model to predict whole environments using the remaining environments as training datasets (Figure 6). Prediction accuracies ranged between 0.31 and 0.59 for GY, 0.14 and 0.54 for HI, 0.35 and 0.82 for SF, 0.54 and 0.91 for TGW.

3.3. Response to Selection

Response to selection (RTS) was compared in the same fashion as prediction accuracy for each model × environment combination. When populations were not stratified, RTS ranged from −0.05 to

0.5 ton ha^{-1} for GY, 0.09 to 4.94% for HI, 0.45 to 3.90 grains g^{-1} of chaff weight for SF, and 0.99 to 2.02 g for TGW among four environments and three models (Figure 7). In general, statistical models had higher RTS than DL models with exceptions of GY in Citra 2018, HI in Citra 2017, and SF in Citra 2017, Quincy 2017, and Quincy 2018. For all five models, the highest and lowest RTS for GY and HI was found in Citra 2017 and Quincy 2018, respectively. For SF, the highest and lowest RTS showed in Citra 2018 and Quincy 2017. For TGW, the highest and lowest RTS values were found in Quincy 2017 and Quincy 2018.
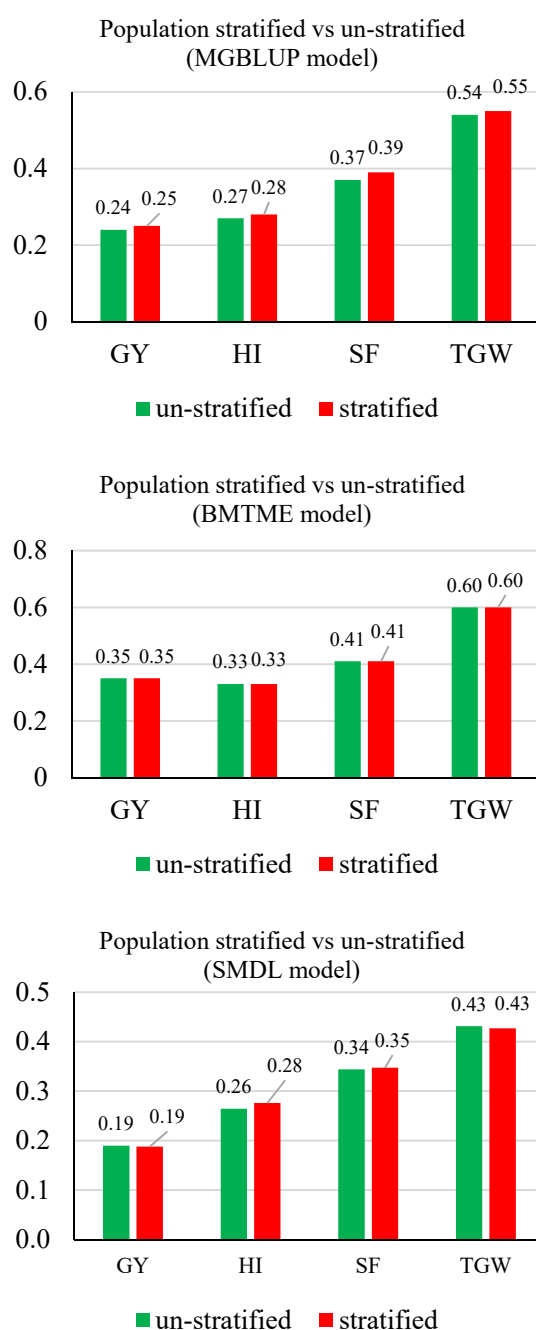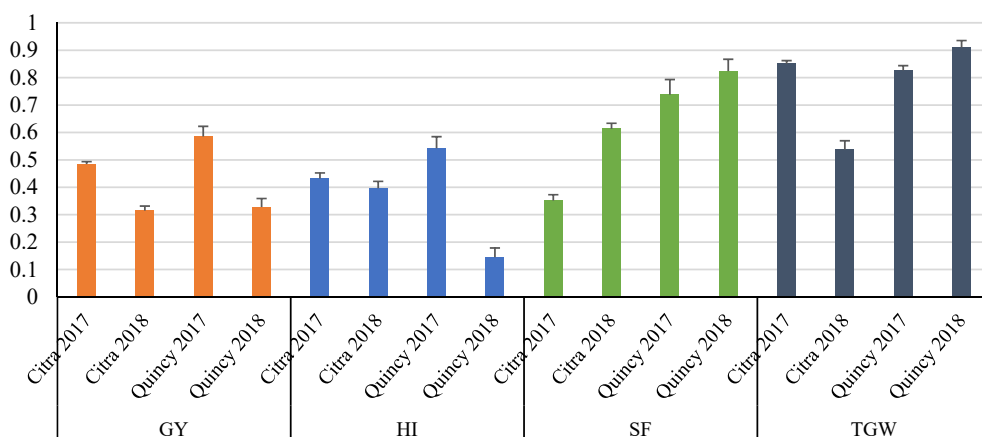

Figure 4. *Cont.*
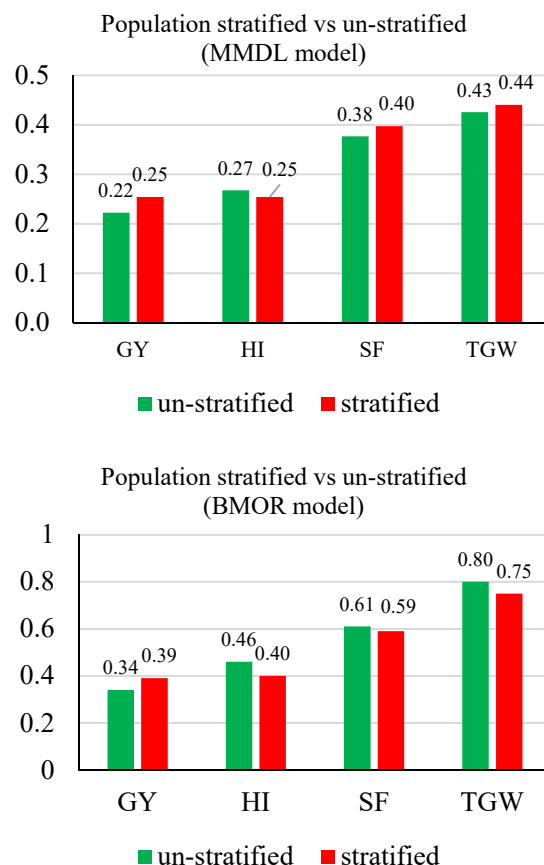
Figure 4. Average prediction accuracies for GY, HI, SF, and TGW with/without population stratified for five models. GY, grain yield; HI, harvest index; SF, spike fertility; TGW, thousand grain weight. Mean Pearson's correlations for each trait were presented and labeled. The stratified scheme was colored in green, and the un-stratified scheme was colored in red.
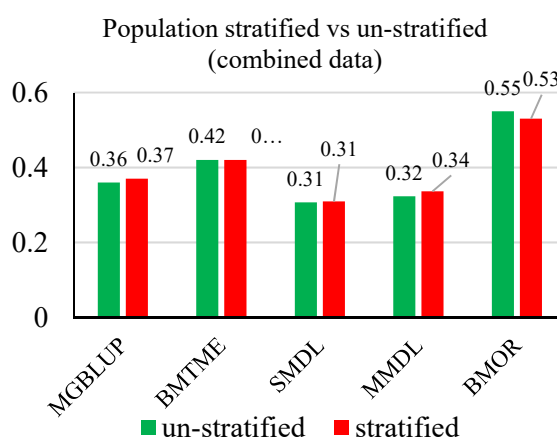

Figure 5. Average prediction accuracies for combined data with/without population stratified for five models Mean Pearson's correlations for each trait were presented and labeled. The stratified scheme was colored in green, and the un-stratified scheme was colored in red.

When populations were stratified, a similar pattern was observed for RTS compared to an un-stratified strategy. Response to selection ranged between −0.04 and 0.48 ton ha^{-1} for GY, 0.13 and 5.53% for HI, 0.26 and 4.46 for SF grains g^{-1} of chaff weight, and 1.01 g and 2.70 g for TGW among four environments and five models (Figure 8). However, the BMOR model showed significantly higher RTS for SF and TGW in Citra 2018 and Quincy 2018 compared to the other two models.

Figure 6. Prediction accuracies for GY (t ha⁻¹), HI (%), SF (grains/g chaff weight), and TGW (g) across environments using the BMOR model. GY, grain yield; HI, harvest index; SF, spike fertility; TGW, thousand grain weight. Mean Pearson's correlations and standard error for each environment were presented for each trait. The results of GY, HI, SF, and TGW trait were colored in orange, blue, light green, and blue-gray.
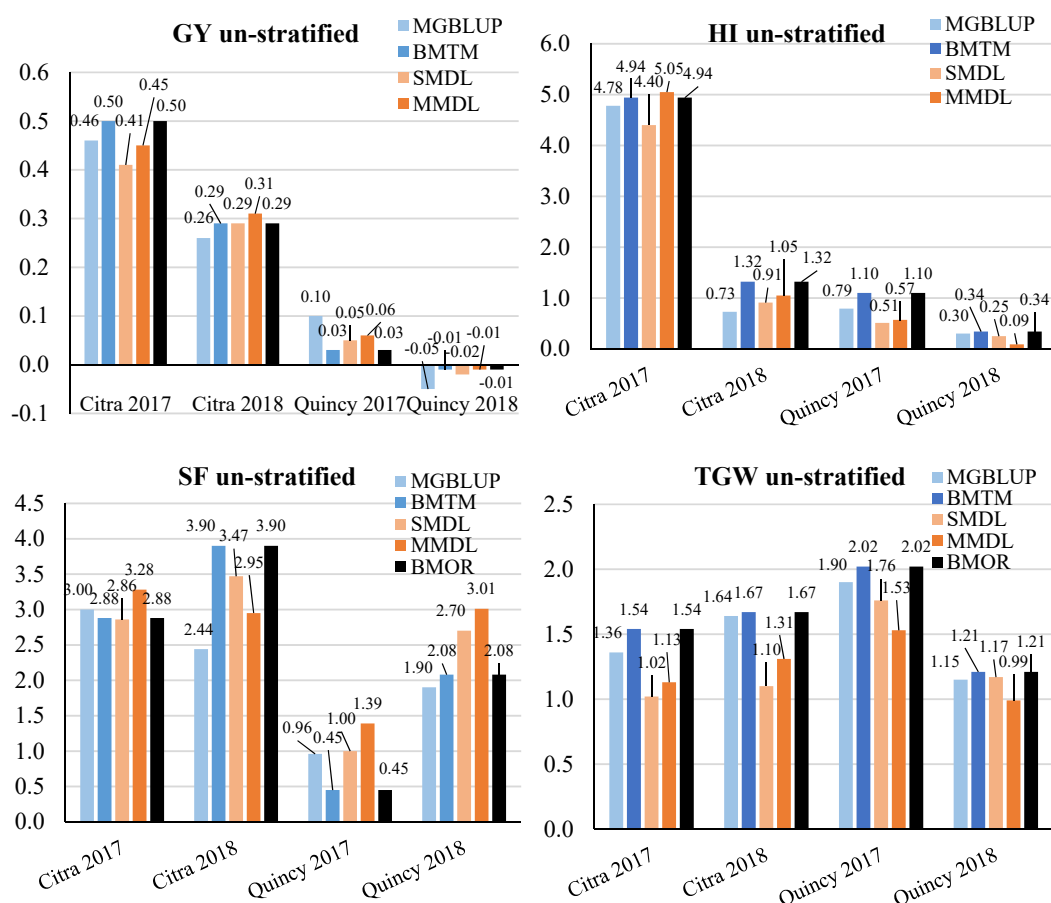
Figure 7. Response to selection for GY (t ha⁻¹), HI (%), SF (grains/g chaff weight), and TGW (g) without population stratified. GY, grain yield; HI, harvest index; SF, spike fertility; TGW, thousand grain weight. Response to selection for each environment was presented and labeled for each trait. Statistical models were colored in light blue and blue, while DL models were colored in light orange and orange. The BMOR model was colored in black.
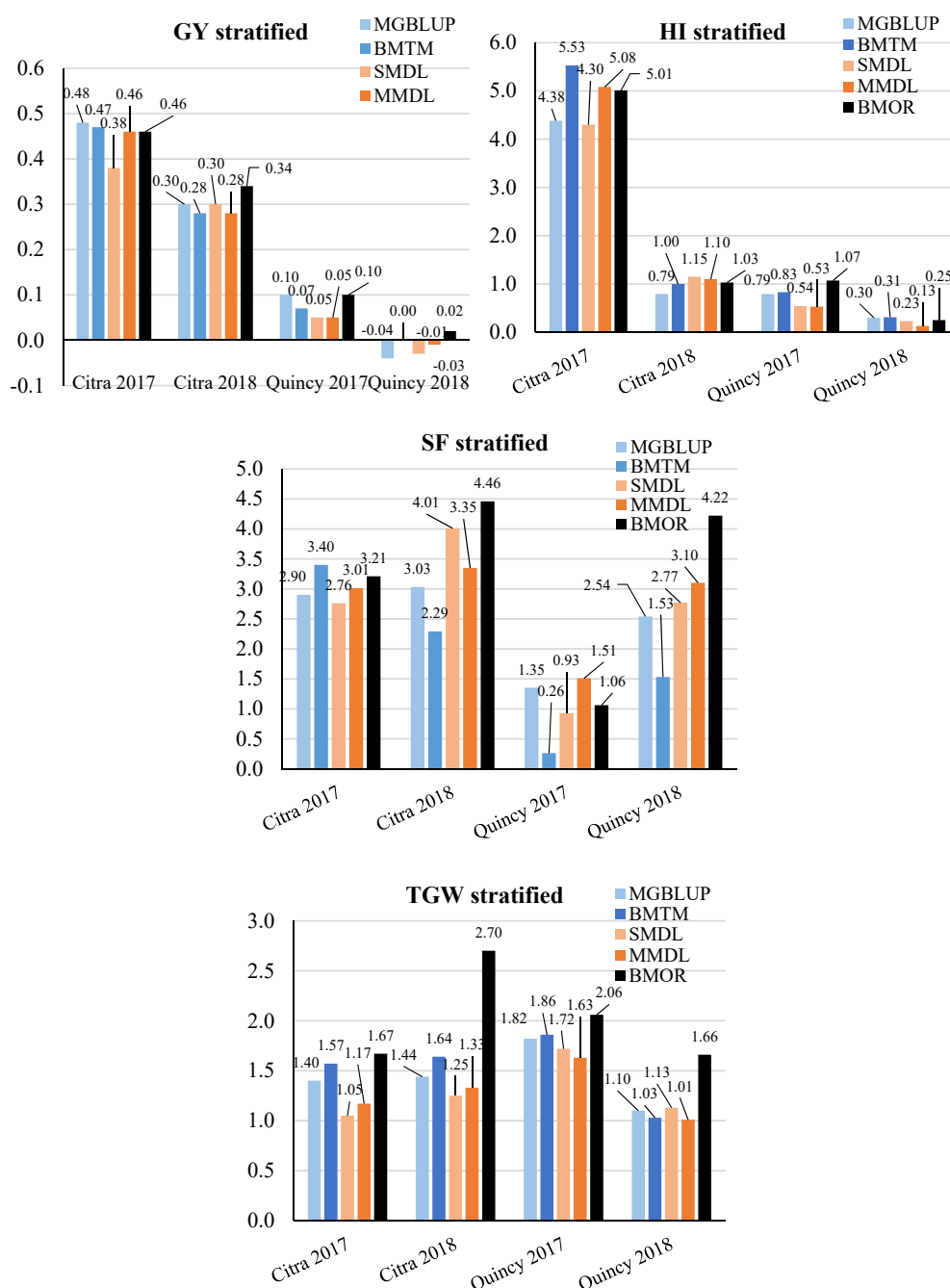
Figure 8. Response to selection for GY (ton ha^{-1}), HI (%), SF (grains/g chaff weight), and TGW (g) with population stratified. GY, grain yield; HI, harvest index; SF, spike fertility; TGW, thousand grain weight. Response to selection for each environment was presented and labeled for each trait. Statistical models were colored in light blue and blue, while the DL models were colored in light orange and orange. The BMOR model was colored in black.

For the average RTS of GY and HI across environments, they were not significantly different between un-stratified and stratified strategy (Figure 9). The highest average RTS for GY was found using the BMOR model with a stratified strategy (0.23 ton ha^{-1}) (Figure 9). The highest average RTS for HI was found using the BMTME (1.93%) and BMOR model (1.93%) with an un-stratified strategy (Figure 9). For SF, the highest and lowest average RTS values were found using the BMOR model with a stratified strategy (3.74 grains/g of chaff weight) and the BMTME model with a stratified strategy (1.87 grains/g of chaff weight), respectively (Figure 9). For TGW, the highest and lowest average RTS

values were found using the BMOR model with a stratified strategy (2.02 g) and the MMDL model
with a stratified strategy (1.24 g), respectively (Figure 9). In general, the BMOR model showed the
highest RTS followed by BMTME and MGBLUP in order. However, the differences of RTS among three
models were smaller in magnitude when comparing to prediction accuracy. Notably, the DL models
only showed higher RTS than statistical models for SF. The multi-trait models had higher RTS than
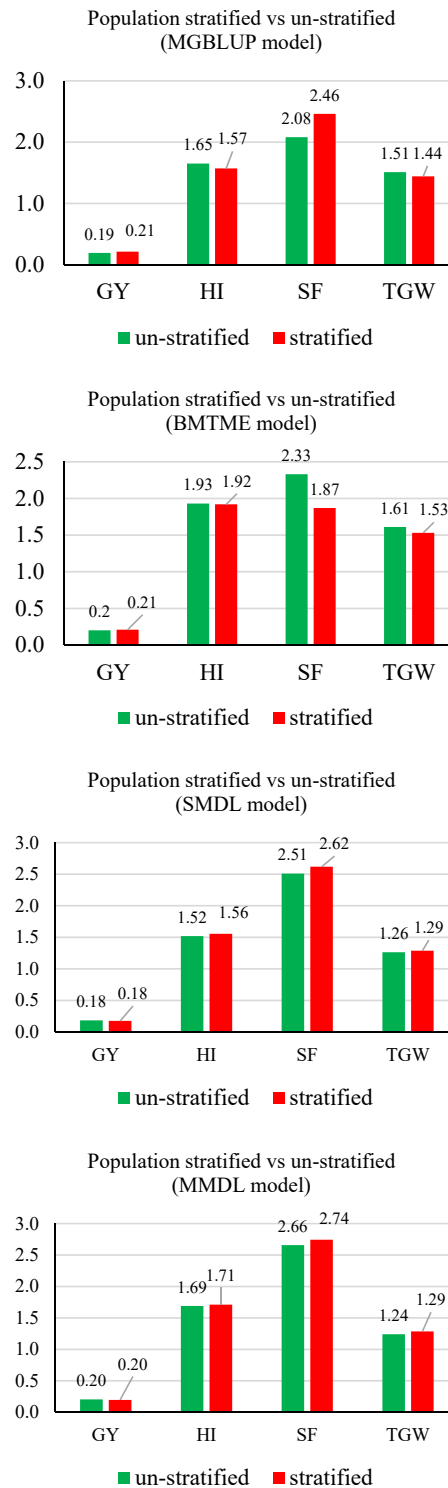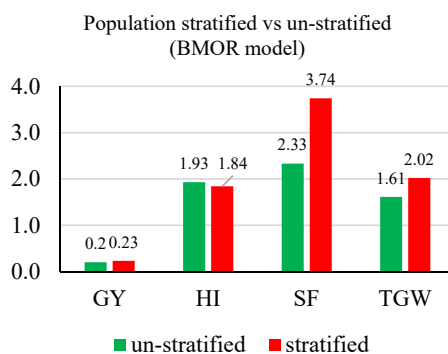single-trait models.

Figure 9. *Cont.*

Figure 9. Average response to selection for GY (ton ha^{-1}), HI (%), SF (grains/g chaff weight), and TGW (g) with/without population stratified for five models. GY, grain yield; HI, harvest index; SF, spike fertility; TGW, thousand grain weight. Mean response to selection for each trait were presented and labeled. Stratified scheme was colored in green and un-stratified scheme was colored in red.

When applying the BMOR model to predict RTS across environments, the RTS ranged from 0.26 to 0.69 ton ha^{-1} for GY, 0.45 to 5.22% for HI, 6.52 to 13.33 grains/g of chaff weight for SF, and 3.03 to 4.36 g for TGW (Figure 10).
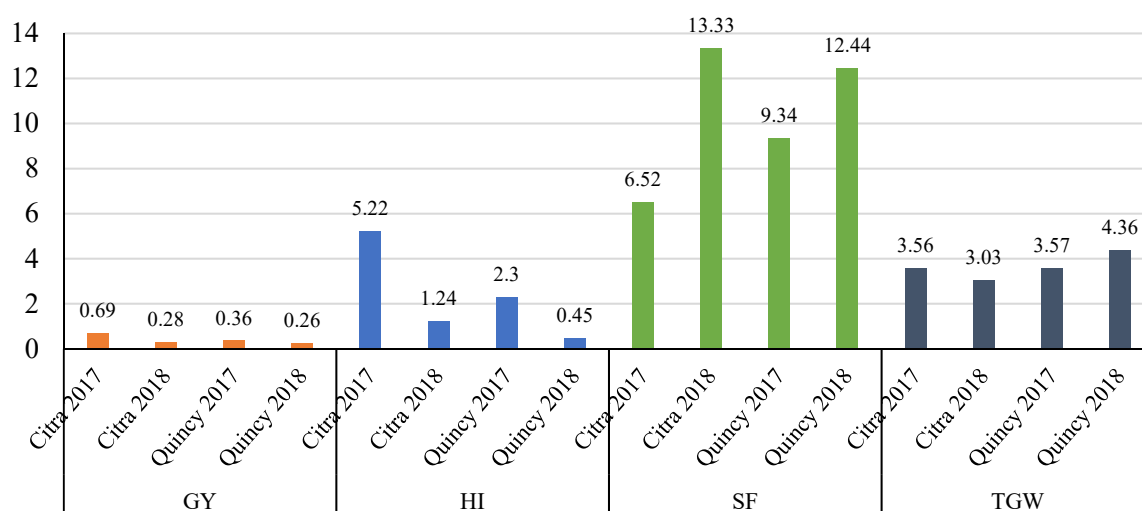
Figure 10. Response to selection for GY (ton ha^{-1}), HI (%), SF (grains/g chaff weight), and TGW (g) across environments using the BMOR model. GY, grain yield; HI, harvest index; SF, spike fertility; TGW, thousand grain weight. Mean response to selection for each environment were presented for each trait. The results of GY, HI, SF, and TGW trait were colored in orange, blue, light green, and blue-gray.

4. Discussion

In plant breeding programs, plant breeders usually perform selection for the improvement of different traits that raise the economic value of plants. When performing selection for an environment, breeders generally apply selection for several traits simultaneously associated with the most important economic traits [74]. For example, when a small grain breeder selects for GY, he also selects indirectly for other yield components, such as grain number, TGW, HI, or different physiological traits, such canopy temperature or NDVI, which are associated with grain yield. Grain number in wheat is a product of spike dry weight and grain number per unit of spike chaff weight, which is known as SF [80]. It is a major potential component of grain number m^{-2}. Evidence have supported the idea of manipulating SF in breeding programs to increase sink strength and ultimately increase yield potential [46,50,53,81,82]. A strong association between SF and GY, HI, and grain number has been reported in wheat [46]. These studies suggest that the increases in SF would be related to a greater

partitioning of photo-assimilates to increase GY and HI in wheat. SF has moderate heritability and is difficult and expensive to estimate as it requires spike count m^{-2}, spike harvest and threshing, and separation of grain and chaff from the spike. Due to its difficulty and the cost of estimation and moderate heritability, as well as correlation with GY, HI, and grain number, SF is a perfect candidate for a multi-trait genomic selection approach to increase the predictive accuracy of GY, HI, and other associated traits [16]. Currently, plant breeding programs are mostly practicing targeted single-trait GP approaches, not considering a full exploitation of genetic information (linkage and pleiotropic effect) from correlated traits. The joint prediction of multiple traits through Multi-Trait Genomic Prediction (MTGP) approach is designed to benefit from genetic correlation between traits and the indirect selection of a target trait with relatively low heritability that genetically correlated with other high-heritability traits [19,21]. Thus, the joint multi-trait model obtained higher prediction accuracy than single-trait methods, especially for a low-heritability trait. One of the major limitations of using a multi-trait model is correlations between traits that are in practice undesirable for plant breeders. In our present study, we used four traits, and all are positively correlated with each other except for the association between SF and TGW. Thus, generally, the prediction accuracy for the different traits in our study was higher for multi-trait than single-trait genomic prediction, which is different from a previously reported study in US soft wheat [83], but it is in agreement with the results reported in European rye by Schulthess et al. [19]. Additionally, the testing environments in the present study were stressed by drought and heat, which usually makes the phenotyping of complex traits more complicated through adding environmental uncertainty. The inclusion of genotype × environment interaction in the multi-trait model improved the prediction accuracy for the joint prediction of multiple traits. The increased accuracy and RTS using a multi-trait multi-environment model for stressed environments certainly demonstrates the effectiveness of the model when the right correlated traits are included.

Our study exploited both single- and multi-trait and multi-environment models to predict yield and yield component traits including GY, HI, SF, and TGW using a diversity soft wheat panel. The results are in favor of the multi-trait and multi-environment statistical model (BMTME) for prediction accuracy and response to selection of all four traits when comparing to the single-trait and multi-environment model (MGBLUP), single-trait and multi-environment deep learning model, and multi-trait and multi-environment deep learning model. This result is in concordance to previous studies that reported that multi-trait and multi-environment GP models could be implemented to increase the prediction accuracy and RTS for low-heritability traits correlated with higher-heritability traits [16,19,83–85]. Jia and Jannink [16] also indicated that a multi-trait model is more effective when the genetic correlation is moderate between these traits. For prediction accuracy, traits with lower heritability such as GY showed more benefit compared to high heritability traits such as TGW using the BMTME model (46% and 11% increase, respectively). In regard to RTS, the multi-trait statistical model also showed 5 to 22% more genetic gain compared to a single-trait model across the environment from the current study. However, the benefit of the multi-trait model for RTS was varied among traits and less relevant to their heritability values based on this study compared to prediction accuracy. The deep learning models showed comparable performance to statistical models, especially for RTS. The multi-trait DL model also performed better than a single-trait DL model in most of the scenarios. Although the prediction accuracy was lower for DL models comparing to statistical models, DL models were less time consuming when computing predicted values for our dataset (23 min on average for DL models and 436 min on average for statistical models). It is also believed that a high dimensional and large dataset could benefit DL models significantly in genomic prediction [36,86]. However, it is important to recognize that the performance of DL models is highly dependent on the SNP set and phenotype. A deep learning model must be curated and calibrated specifically for traits with complex genetic structure [32].

The use of a stratified cross-validation scheme with all five models did not increase the prediction accuracy compared with using an un-stratified cross-validation scheme in the present study.

One possible reason is that the alleles of representative quantitative trait loci (QTL) associated with target traits are commonly shared between training and validation populations in both stratified and un-stratified schemes. Ward et al. [83] also found that using un-related training and validation population schemes did not affect the predictive ability compared with using a related cross-validation scheme. Rutkoski et al. [84] also reported that using a multi-trait model including secondary traits had no influence on prediction accuracy if secondary trait phenotypes were not replicated in the validation test.

Although the BMOR model showed higher prediction accuracy and comparable RTS to that of the BMTME model, it does not estimate the covariances between traits and environments because it implements univariate analysis at both stages [29]. However, it is more computationally efficient than the MGBLUP and BMTME models (436 min on average) in terms of the computational resource and running time for the model to converge (42 min on average). Thus, it is advantageous when investigators are exploring the performance of genomic prediction in some preliminary studies. Therefore, we implemented the BMOR model to predict yield and yield component traits among environments. Heslot et al. [87] pointed out that GP results could be largely affected by an interaction between un-selected trait and environment being tested for selected traits, especially when selection were guided toward the un-selected traits such as stress tolerance traits with a large QTL effect. In our study, this is reflected by the inconsistent prediction accuracy and RTS when the BMOR model is applied in two cross-validation schemes and prediction among environments. For example, our soft wheat lines generally showed varying degrees of heat stress tolerance and were evaluated in Citra, FL where heat stress was common during the anthesis and grain-filling stages. The phenotype of target traits such as SF and TGW could be masked by the stress tolerance characteristics of each line. Based on our study, SF is the most affected trait, as the prediction accuracy and RTS for SF showed the largest difference between the two environments compared to other traits.

5. Conclusions

The study demonstrates that the multi-trait model has in general higher predictive accuracy than the single-trait model under a multiple-environmental analysis and has the capacity to predict the performance of genotypes for different test environments. It is useful for plant breeding scenarios where several economically important traits are inter-correlated. The findings of the present study could be potentially applied in plant breeding to achieve more cycles of selection by unit of time for multiple traits, to assess accurately genotype performance due to the low number of testing environments or due to a lack of replication, and to predict the performance of genotypes for stressed environments with low heritability. The analysis also showed that statistical models were superior to DL models for the studied traits, but DL models were comparable to statistical models in many cases. In conclusion, our study showed that for our population and traits of interest, multi-trait and multi-environment models can be exploited to achieve generally higher increases in prediction accuracy and RTS in several focal traits.

Author Contributions: Conceptualization, M.A.B., J.G. and J.K.; methodology, J.K., M.A.B. and J.G.; software, J.G.; formal analysis, J.G. and J.K.; investigation, J.K., S.P., D.S., N.K., M.A. and J.M.; resources, M.A.B.; data curation, J.G. and J.K.; writing—original draft preparation, J.G. and J.K.; writing—review and editing, M.A.B., S.H., G.B.-G., J.P.M., J.J., M.M., R.E.M., A.M.H.I., R.S. and C.G.; supervision, M.A.B.; funding acquisition, M.A.B. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Mann, J.; Cummings, J.; Englyst, H.; Key, T.; Liu, S.; Riccardi, G.; Summerbell, C.; Uauy, R.; Van Dam, R.; Venn, B. FAO/WHO scientific update on carbohydrates in human nutrition: Conclusions. *Eur. J. Clin. Nutr.* **2007**, *61*, S132–S137. [CrossRef] [PubMed]
2. Blum, A. *Plant Breeding for Water-Limited Environments*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010; ISBN 1-4419-7491-1.
3. Ceccarelli, S.; Grando, S.; Maatougui, M.; Michael, M.; Slash, M.; Haghparast, R.; Rahmanian, M.; Taheri, A.; Al-Yassin, A.; Benbelkacem, A. Plant breeding and climate changes. *J. Agric. Sci.* **2010**, *148*, 627–637. [CrossRef]
4. Tester, M.; Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **2010**, *327*, 818–822. [CrossRef] [PubMed]
5. Mu, J.E.; Sleeter, B.M.; Abatzoglou, J.T.; Antle, J.M. Climate impacts on agricultural land use in the USA: The role of socio-economic scenarios. *Clim. Chang.* **2017**, *144*, 329–345. [CrossRef]
6. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819.
7. Marroni, F.; Pinosio, S.; Morgante, M. The quest for rare variants: Pooled multiplexed next generation sequencing in plants. *Front. Plant Sci.* **2012**, *3*, 133. [CrossRef]
8. Bhat, J.A.; Ali, S.; Salgotra, R.K.; Mir, Z.A.; Dutta, S.; Jadon, V.; Tyagi, A.; Mushtaq, M.; Jain, N.; Singh, P.K. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.* **2016**, *7*, 221. [CrossRef]
9. Eathington, S.R.; Crosbie, T.M.; Edwards, M.D.; Reiter, R.S.; Bull, J.K. Molecular markers in a commercial breeding program. *Crop Sci.* **2007**, *47*, S154–S163. [CrossRef]
10. Battenfield, S.D.; Guzmán, C.; Gaynor, R.C.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; Fritz, A.K.; Poland, J.A. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* **2016**, *9*, 1–12. [CrossRef]
11. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
12. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; CRC Press: Boca Raton, FL, USA, 2015; ISBN 1-4987-1217-7.
13. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform.* **2011**, *12*, 186. [CrossRef] [PubMed]
14. Gianola, D. Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* **2013**, *194*, 573–596. [CrossRef] [PubMed]
15. Pérez-Rodríguez, P.; Crossa, J.; Rutkoski, J.; Poland, J.; Singh, R.; Legarra, A.; Autrique, E.; de Los Campos, G.; Burgueño, J.; Dreisigacker, S. Single-step genomic and pedigree genotype × environment interaction models for predicting wheat lines in international environments. *Plant Genome* **2017**, *10*, 1–15. [CrossRef]
16. Jia, Y.; Jannink, J.-L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* **2012**, *192*, 1513–1522. [CrossRef] [PubMed]
17. Lyra, D.H.; de Freitas Mendonça, L.; Galli, G.; Alves, F.C.; Granato, Í.S.C.; Fritsche-Neto, R. Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. *Mol. Breed.* **2017**, *37*, 80. [CrossRef]
18. Fernandes, S.B.; Dias, K.O.; Ferreira, D.F.; Brown, P.J. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theor. Appl. Genet.* **2018**, *131*, 747–755. [CrossRef]
19. Schulthess, A.W.; Wang, Y.; Miedaner, T.; Wilde, P.; Reif, J.C.; Zhao, Y. Multiple-trait-and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor. Appl. Genet.* **2016**, *129*, 273–287. [CrossRef] [PubMed]
20. Hayes, B.; Panozzo, J.; Walker, C.; Choy, A.; Kant, S.; Wong, D.; Tibbits, J.; Daetwyler, H.; Rochfort, S.; Hayden, M. Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor. Appl. Genet.* **2017**, *130*, 2505–2519. [CrossRef]

21. Jiang, J.; Zhang, Q.; Ma, L.; Li, J.; Wang, Z.; Liu, J. Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity* **2015**, *115*, 29–36. [CrossRef] [PubMed]

22. Burgueño, J.; Crossa, J.; Cotes, J.M.; Vicente, F.S.; Das, B. Prediction assessment of linear mixed models for multienvironment trials. *Crop Sci.* **2011**, *51*, 944–954. [CrossRef]

23. Montesinos-López, O.A.; Montesinos-López, A.; Pérez-Rodríguez, P.; de los Campos, G.; Eskridge, K.; Crossa, J. Threshold models for genome-enabled prediction of ordinal categorical traits in plant breeding. *G3 Genes Genomes Genet.* **2015**, *5*, 291–300. [CrossRef] [PubMed]

24. Montesinos-López, A.; Montesinos-López, O.A.; Cuevas, J.; Mata-López, W.A.; Burgueño, J.; Mondal, S.; Huerta, J.; Singh, R.; Autrique, E.; González-Pérez, L. Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. *Plant Methods* **2017**, *13*, 62. [CrossRef]

25. Guo, J.; Pradhan, S.; Shahi, D.; Khan, J.; Mcbreen, J.; Bai, G.; Murphy, J.P.; Babar, M.A. Increased prediction Accuracy Using combined Genomic information and physiological traits in A Soft Wheat panel evaluated in Multi-environments. *Sci. Rep.* **2020**, *10*, 1–12. [CrossRef]

26. Crain, J.; Mondal, S.; Rutkoski, J.; Singh, R.P.; Poland, J. Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome* **2018**, *11*, 1–14. [CrossRef]

27. Krause, M.R.; González-Pérez, L.; Crossa, J.; Pérez-Rodríguez, P.; Montesinos-López, O.; Singh, R.P.; Dreisigacker, S.; Poland, J.; Rutkoski, J.; Sorrells, M. Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes Genomes Genet.* **2019**, *9*, 1231–1247. [CrossRef] [PubMed]

28. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Toledo, F.H.; Pérez-Hernández, O.; Eskridge, K.M.; Rutkoski, J. A genomic Bayesian multi-trait and multi-environment model. *G3 Genes Genomes Genet.* **2016**, *6*, 2725–2744. [CrossRef] [PubMed]

29. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Cuevas, J.; Montesinos-López, J.C.; Gutiérrez, Z.S.; Lillemo, M.; Philomin, J.; Singh, R. A Bayesian genomic multi-output regressor stacking model for predicting multi-trait multi-environment plant breeding data. *G3 Genes Genomes Genet.* **2019**, *9*, 3381–3393. [CrossRef]

30. Min, S.; Lee, B.; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **2017**, *18*, 851–869. [CrossRef] [PubMed]

31. Kelley, D.R.; Snoek, J.; Rinn, J.L. Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* **2016**, *26*, 990–999. [CrossRef]

32. Bellot, P.; de los Campos, G.; Pérez-Enciso, M. Can deep learning improve genomic prediction of complex human traits? *Genetics* **2018**, *210*, 809–819. [CrossRef] [PubMed]

33. Abdollahi-Arpanahi, R.; Gianola, D.; Peñagaricano, F. Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genet. Sel. Evol.* **2020**, *52*, 1–15. [CrossRef]

34. Liu, Y.; Wang, D. *Application of Deep Learning in Genomic Selection*; IEEE: Piscataway, NJ, USA, 2017; p. 2280.

35. Ma, W.; Qiu, Z.; Song, J.; Cheng, Q.; Ma, C. DeepGS: Predicting phenotypes from genotypes using Deep Learning. *bioRxiv* **2017**. [CrossRef]

36. 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 Genet.* **2018**, *8*, 3829–3840. [CrossRef] [PubMed]

37. Jannink, J.-L.; Lorenz, A.J.; Iwata, H. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genom.* **2010**, *9*, 166–177. [CrossRef]

38. Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **2012**, *7*, e32253. [CrossRef]

39. 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. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3 Genes Genomes Genet.* **2012**, *2*, 1427–1436. [CrossRef]

40. Isidro, J.; Jannink, J.-L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M.E. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* **2015**, *128*, 145–158. [CrossRef]

41. Arruda, M.; Lipka, A.E.; Brown, P.J.; Krill, A.; Thurber, C.; Brown-Guedira, G.; Dong, Y.; Foresman, B.; Kolb, F.L. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol. Breed.* **2016**, *36*, 84. [CrossRef]

42. Shearman, V.; Sylvester-Bradley, R.; Scott, R.; Foulkes, M. Physiological processes associated with wheat yield progress in the UK. *Crop Sci.* **2005**, *45*, 175–185.

43. Abbate, P.E.; López, J.R.; Brach, A.M.; Gutheim, F.; Gonzalez, F.; Kruk, B.; Serrago, R. Fertilidad de las espigas de trigo en ambientes sub-potenciales. In *Workshop Internacional: Ecofisiología Vegetal Aplicada al Estudio de la Determinación del Rendimiento y la Calidad de los Cultivos de Granos, Mar del Plata, Buenos Aires, Argentina, 6–7 September 2007*; Kruk, B., Serrago, R., Eds.; FAUBA: Buenos Aires, Argentina, 2007; pp. 2–3.

44. Abbate, P.E.; Pontaroli, A.C.; Lázaro, L.; Gutheim, F. A method of screening for spike fertility in wheat. *J. Agric. Sci.* **2013**, *151*, 322–330. [CrossRef]

45. Acreche, M.M.; Briceño-Félix, G.; Sánchez, J.A.M.; Slafer, G.A. Physiological bases of genetic gains in Mediterranean bread wheat yield in Spain. *Eur. J. Agron.* **2008**, *28*, 162–170. [CrossRef]

46. Pradhan, S.; Babar, M.A.; Robbins, K.; Bai, G.; Mason, R.E.; Khan, J.; Shahi, D.; Avci, M.; Guo, J.; Hossain, M.M. Understanding the Genetic Basis of Spike Fertility to Improve Grain Number, Harvest Index, and Grain Yield in Wheat Under High Temperature Stress Environments. *Front. Plant Sci.* **2019**, *10*, 1481. [CrossRef]

47. Botwright, T.L.; Condon, A.G.; Rebetzke, G.J.; Richards, R.A. Field evaluation of early vigour for genetic improvement of grain yield in wheat. *Aust. J. Agric. Res.* **2002**, *53*, 1137–1145. [CrossRef]

48. Kuchel, H.; Williams, K.J.; Langridge, P.; Eagles, H.A.; Jefferies, S.P. Genetic dissection of grain yield in bread wheat. I. QTL analysis. *Theor. Appl. Genet.* **2007**, *115*, 1029–1041. [CrossRef]

49. Reynolds, M.; Foulkes, M.J.; Slafer, G.A.; Berry, P.; Parry, M.A.; Snape, J.W.; Angus, W.J. Raising yield potential in wheat. *J. Exp. Bot.* **2009**, *60*, 1899–1918. [CrossRef]

50. Fischer, R. Wheat physiology: A review of recent developments. *Crop Pasture Sci.* **2011**, *62*, 95–114. [CrossRef]

51. Parry, M.; Slafer, G. Achieving yield gains in wheat. *Plant Cell Environ.* **2012**, *35*, 17991823Sears.

52. Gaju, O.; Reynolds, M.P.; Sparkes, D.L.; Foulkes, M.J. Relationships between Large-Spike Phenotype, Grain Number, and Yield Potential in Spring Wheat. *Crop Sci.* **2009**, *49*, 961–973. [CrossRef]

53. González, F.G.; Terrile, I.I.; Falcón, M.O. Spike Fertility and Duration of Stem Elongation as Promising Traits to Improve Potential Grain Number (and Yield): Variation in Modern Argentinean Wheats. *Crop Sci.* **2011**, *51*, 1693–1702. [CrossRef]

54. Rivera-Amado, C.; Trujillo-Negrellos, E.; Sylvester-Bradley, R.; Molero, G.; Sierra-Gonzalez, A.; Reynolds, M.; Foulkes, J. Achieving increases in spike growth, fruiting efficiency, and harvest index in high biomass wheat cultivars. In Proceedings of the 2nd International TRIGO (Wheat) Yield Potential; CIMMYT: Mexico City, Mexico, 2016; p. 70.

55. Molero, G.; Joynson, R.; Pinera-Chavez, F.J.; Gardiner, L.; Rivera-Amado, C.; Hall, A.; Reynolds, M.P. Elucidating the genetic basis of biomass accumulation and radiation use efficiency in spring wheat and its role in yield potential. *Plant Biotechnol. J.* **2019**, *17*, 1276–1288. [CrossRef]

56. Lopes, M.S.; El-Basyoni, I.; Baenziger, P.S.; Singh, S.; Royo, C.; Ozbek, K.; Aktas, H.; Ozer, E.; Ozdemir, F.; Manickavelu, A.; et al. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot.* **2015**, *66*, 3477–3486. [CrossRef]

57. Martino, D.L.; Abbate, P.E.; Cendoya, M.G.; Gutheim, F.; Mirabella, N.E.; Pontaroli, A.C. Wheat spike fertility: Inheritance and relationship with spike yield components in early generations. *Plant Breed.* **2015**, *134*, 264–270. [CrossRef]

58. Thavamanikumar, S.; Dolferus, R.; Thumma, B.R. Comparison of Genomic Selection Models to Predict Flowering Time and Spike Grain Number in Two Hexaploid Wheat Doubled Haploid Populations. *G3 Genes Genomes Genet.* **2015**, *5*, 1991. [CrossRef]

59. Federer, W.T.; Raghavarao, D. On augmented designs. *Biometrics* **1975**, *31*, 29–35. [CrossRef]

60. Zadoks, J.C.; Chang, T.T.; Konzak, C.F. A decimal code for the growth stages of cereals. *Weed Res.* **1974**, *14*, 415–421. [CrossRef]

61. Littell, R.C.; Milliken, G.A.; Stroup, W.W.; Wolfinger, R.D. SAS system for mixed models. *Technometrics* **1997**, *39*, 344.

62. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]

63. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [CrossRef]

64. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [CrossRef]

65. Poland, J.; Endelman, J.; Dawson, J.; Rutkoski, J.; Wu, S.; Manes, Y.; Dreisigacker, S.; Crossa, J.; Sánchez-Villeda, H.; Sorrells, M. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* **2012**, *5*, 103–113. [CrossRef]

66. Bansal, V.; Harismendy, O.; Tewhey, R.; Murray, S.S.; Schork, N.J.; Topol, E.J.; Frazer, K.A. Accurate detection and genotyping of SNPs utilizing population sequencing data. *Genome Res.* **2010**, *20*, 537–545. [CrossRef]

67. Montesinos-López, O.A.; Montesinos-López, A.; Luna-Vázquez, F.J.; Toledo, F.H.; Pérez-Rodríguez, P.; Lillemo, M.; Crossa, J. An R package for Bayesian analysis of multi-environment and multi-trait multi-environment data for genome-based prediction. *G3 Genes Genomes Genet.* **2019**, *9*, 1355–1369. [CrossRef]

68. Spyromitros-Xioufis, E.; Tsoumakas, G.; Groves, W.; Vlahavas, I. Multi-label classification methods for multi-target regression. *arXiv* **2012**, arXiv:12116581.

69. Spyromitros-Xioufis, E.; Tsoumakas, G.; Groves, W.; Vlahavas, I. Multi-target regression via input space expansion: Treating targets as inputs. *Mach. Learn.* **2016**, *104*, 55–98. [CrossRef]

70. Džeroski, S.; Demšar, D.; Grbović, J. Predicting chemical parameters of river water quality from bioindicator data. *Appl. Intell.* **2000**, *13*, 7–17. [CrossRef]

71. Kocev, D.; Džeroski, S.; White, M.D.; Newell, G.R.; Griffioen, P. Using single-and multi-target regression trees and ensembles to model a compound index of vegetation condition. *Ecol. Model.* **2009**, *220*, 1159–1168. [CrossRef]

72. Angermueller, C.; Lee, H.J.; Reik, W.; Stegle, O. DeepCpG: Accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biol.* **2017**, *18*, 67. [CrossRef]

73. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 94. [CrossRef]

74. Falconer, D.S.; Mackay, T.F.; Frankham, R. Introduction to quantitative genetics (4th edn). *Trends Genet.* **1996**, *12*, 280.

75. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *arXiv* **2014**, arXiv:14065823.

76. Pérez, P.; de Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [CrossRef]

77. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M. Tensorflow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.

78. Gulli, A.; Pal, S. *Deep Learning with KERAS*; Packt Publishing Ltd.: Birmingham, UK, 2017; ISBN 1-78712-903-9.

79. Lenth, R.V. Response-surface methods in R, using rsm. *J. Stat. Softw.* **2009**, *32*, 1–17. [CrossRef]

80. Banta, S.J. *Symposium on Potential Productivity of Field Crops under Different Environments*; International Rice Research Institute: Los Banos, Philippine, 1983; Volume 15, ISBN 971-10-4114-6.

81. Foulkes, M.J.; Slafer, G.A.; Davies, W.J.; Berry, P.M.; Sylvester-Bradley, R.; Martre, P.; Calderini, D.F.; Griffiths, S.; Reynolds, M.P. Raising yield potential of wheat. III. Optimizing partitioning to grain while maintaining lodging resistance. *J. Exp. Bot.* **2011**, *62*, 469–486. [CrossRef]

82. Slafer, G.A.; Elia, M.; Savin, R.; García, G.A.; Terrile, I.I.; Ferrante, A.; Miralles, D.J.; González, F.G. Fruiting efficiency: An alternative trait to further rise wheat yield. *Food Energy Secur.* **2015**, *4*, 92–109. [CrossRef]

83. Ward, B.P.; Brown-Guedira, G.; Tyagi, P.; Kolb, F.L.; Van Sanford, D.A.; Sneller, C.H.; Griffey, C.A. Multienvironment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. *Crop Sci.* **2019**, *59*, 491–507. [CrossRef]

84. Rutkoski, J.; Poland, J.; Mondal, S.; Autrique, E.; Pérez, L.G.; Crossa, J.; Reynolds, M.; Singh, R. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genet.* **2016**, *6*, 2799–2808. [CrossRef]

85. Montesinos-López, O.A.; Montesinos-López, A.; Montesinos-López, J.C.; Crossa, J.; Luna-Vázquez, F.J.; Salinas-Ruiz, J. A Bayesian Multiple-Trait and Multiple-Environment Model Using the Matrix Normal Distribution. *Phys. Methods Stimul. Plant Mushroom Dev.* **2018**, 19. [CrossRef]

86. Li, B.; Zhang, N.; Wang, Y.-G.; George, A.W.; Reverter, A.; Li, Y. Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Front. Genet.* **2018**, *9*, 237. [CrossRef]

87. Heslot, N.; Jannink, J.-L.; Sorrells, M.E. Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Sci.* **2013**, *53*, 921–933. [CrossRef]