Multi-Trait Regressor Stacking Increased Genomic Prediction Accuracy of Sorghum Grain Composition

Sirjan Sapkota [1,2,*], **J. Lucas Boatwright** [1,2], **Kathleen Jordan** [1], **Richard Boyles** [2,3] **and Stephen Kresovich** [1,2]

1   Advanced Plant Technology Program, Clemson University, Clemson, SC 29634, USA;
     jboatw2@clemson.edu (J.L.B.); kjorda7@clemson.edu (K.J.); skresov@clemson.edu (S.K.)
2   Department of Plant and Environmental Sciences, Clemson University, Clemson, SC 29634, USA;
     rboyles@clemson.edu
3   Pee Dee Research and Education Center, Clemson University, Florence, SC 29506, USA
*   Correspondence: ssapkot@clemson.edu

Abstract: Genomic prediction has enabled plant breeders to estimate breeding values of unobserved genotypes and environments. The use of genomic prediction will be extremely valuable for compositional traits for which phenotyping is labor-intensive and destructive for most accurate results. We studied the potential of Bayesian multi-output regressor stacking (BMORS) model in improving prediction performance over single trait single environment (STSE) models using a grain sorghum diversity panel (GSDP) and a biparental recombinant inbred lines (RILs) population. A total of five highly correlated grain composition traits—amylose, fat, gross energy, protein and starch, with genomic heritability ranging from 0.24 to 0.59 in the GSDP and 0.69 to 0.83 in the RILs were studied. Average prediction accuracies from the STSE model were within a range of 0.4 to 0.6 for all traits across both populations except amylose (0.25) in the GSDP. Prediction accuracy for BMORS increased by 41% and 32% on average over STSE in the GSDP and RILs, respectively. Prediction of whole environments by training with remaining environments in BMORS resulted in moderate to high prediction accuracy. Our results show regression stacking methods such as BMORS have potential to accurately predict unobserved individuals and environments, and implementation of such models can accelerate genetic gain.

Keywords: genomics; genomic selection; genomic prediction; marker-assisted selection; whole genome regression; grain quality; near infra-red spectroscopy; cereal crop; sorghum; multi-trait

1. Introduction

Cereal grains provide more than half of the total human caloric consumption globally and amount to over 80% in some of the poorest nations of the world [1]. Sorghum [*Sorghum bicolor* (L.) Moench], a drought-tolerant cereal crop, is a dietary staple for over half a billion people of semi-arid tropics which is inhabited by some of the most food insecure and malnourished populations [2]. In industrialized countries, such as the United States and Australia, grain sorghum is primarily grown for animal feed. But in recent years the uses of sorghum grain have expanded to baking, malting, brewing, and biofortification [3–5]. Therefore, genetic improvement of sorghum grain composition is crucial to mitigate the global malnutrition crisis, to increase efficiency of feed grains used in animal production, and to serve evolving niche markets for gluten-free grains.

In the last two decades, the use of genome-wide markers in prediction of genetic merit of individuals has revolutionized plant and animal breeding. Genomic prediction (GP) uses statistical models to estimate marker effects in a training population with phenotypic and genotypic data

which is then used to predict breeding values of individuals solely from genetic markers [6,7]. Training population size, genetic relatedness between individuals in training and testing population, marker density, span of linkage disequilibrium and genetic architecture of traits are some of the factors that can affect the predictive ability of the models [8–10]. Genomic prediction models are routinely studied and applied by breeding programs around the world in several crops. Novel statistical methods that are capable of incorporating pedigree, genomic, and environmental covariates into statistical-genetic prediction models have emerged as a result of extensive computational research [11].

One of the main advantages of GP is that breeders can use phenotypic values from some lines in some environments to make predictions of new lines and environments. Genomic best linear unbiased prediction (GBLUP) proposed by VanRaden [12] is probably the most widely used genomic prediction model in both plant and animal breeding. Since then GBLUP model has been extended to include G × E interactions resulting in improved prediction accuracy of unobserved lines in environments [13,14]. Burgueño et al. [13] found an increase in prediction ability of unobserved wheat genotypes by about 20% in multi-environment GBLUP model compared to single environment model. Also an extension of the GBLUP model, Jarquín et al. [14] introduced a reaction norm model which introduces the main and interaction effects of markers and environmental covariates by using high-dimensional random variance-covariance structures of markers and environmental covariates. While most of the genomic prediction studies have been on individual traits, breeding programs use selection indices based on several traits to make breeding decisions. To address those challenges, expanded genomic prediction models that perform joint analysis of multiple traits have been studied using empirical and simulated data [15,16]. Subsequent improvement in prediction accuracy from multi-trait model over single-trait model depends on trait heritability and correlation between the traits involved [15,17].

Data generated in breeding programs span multiple environment and are recorded for multiple traits for each individual. While multi-environment models and multi-trait models are implemented separately, a single model to account for complexity of variance-covariance structure in a combined multi-trait multi-environment (MTME) model was lacking until Montesinos-López et al. [18] developed a Bayesian whole genome prediction model to incorporate and analyze multiple traits and multiple environments simultaneously. Montesinos-López et al. [18] also developed a computationally efficient Markov Chain Monte Carlo (MCMC) method that produces a full conditional distribution of the parameters leading to an exact Gibbs sampling for the posterior distribution. Another MTME model that employs a completely different method was proposed by Montesinos-López et al. [19]. This method, called the Bayesian multi-output regression stacking (BMORS), is a Bayesian version of multi-target regressor stacking (MTRS) originally proposed by Spyromitros-Xioufis et al. [20,21]. This method consists of training in two stages: first training multiple learning algorithms for the same dataset and then subsequently combining the predictions to obtain the final predictions.

Genomic prediction for grain quality traits has previously been reported in crops such as wheat [22–24], rye [25], maize [26], and soybean [27]. Hayes et al. [28] and Battenfield et al. [23] used near-infrared derived phenotypes in genomic prediction of protein content and end-use quality in wheat. Multi-trait genomic prediction models can simultaneously improve grain yield and protein content despite being negatively correlated [24,29]. In sorghum, grain macronutrients have shown to be inter-correlated among one another [30], which suggests the multi-trait models may increase predictive ability of individual grain quality traits. The ability to assess genetic merit of unobserved selection candidates across environments is promising for reducing evaluation cost and generation interval in the sorghum breeding pipeline where parental lines of commercial hybrids are currently selected on the basis of extensive progeny testing [31]. In order to extend capacities to performance index selection for multiple environments, we need to study and effectively implement MTME genomic prediction models in our breeding programs. In this study, we report the first implementation of genomic prediction for grain composition in sorghum, and the objective was to assess potential for improvement in prediction accuracy of multi-trait regressor stacking model over single trait model for five grain composition traits: amylose, fat, gross energy, protein and starch.

2. Results

2.1. Phenotypic Variation

A single calibration curve for near infra-red spectroscopy (NIRS) was used for the two populations studied. Table 1 outlines the summary statistics of NIRS predictions and phenotypic distribution and heritability of the grain composition traits. The cross validation accuracy (R^2) of the NIRS calibration curve was moderately high to high, except for fat which had a moderate R^2 value (0.41). We had a total of three environments (three years in one location) for the GSDP and four environments (two years in two locations) for the RILs. Traits were normally distributed except amylose in two 2014 environments in the RILs which had bimodal distribution (Figures S1 and S2). All traits showed significant variation in distribution across the environments, except starch in GSDP.

Table 1. **Summary statistics of near infra-red spectroscopy (NIRS) calibration and phenotypic distribution.** R^2 is the prediction accuracy and SECV is the standard error of cross validation for the NIRS calibration curve. Mean represents the phenotypic mean of the trait with its standard deviation (SD). h^2 is the estimate of genomic heritability.

| Trait | NIRS | | GSDP | | RILs | |
	R^2	SECV	Mean $\pm$ SD	h^2	Mean $\pm$ SD	h^2
Amylose	0.60	2.24	13.87 ± 2.98	0.24	11.49 ± 4.32	0.77
Fat	0.41	0.53	2.53 ± 0.57	0.54	3.07 ± 0.67	0.76
Gross energy	0.71	25.80	4108.33 ± 55.15	0.59	4124.56 ± 41.74	0.69
Protein	0.96	0.27	12.02 ± 1.45	0.39	11.43 ± 1.03	0.83
Starch	0.89	0.75	68.30 ± 2.44	0.44	68.37 ± 1.87	0.79

The genomic heritabilities of all traits except gross energy were significantly higher ($p < 0.05$) in the RILs than in the GSDP (Table 1). Trait heritabilities were high in the RILs, with protein and gross energy having the highest and lowest heritabilities, respectively. In the GSDP, genomic heritability was moderately high for fat and gross energy, moderate for protein and starch, and low for amylose. The poor genomic heritability (0.24) of amylose in the GSDP was expected because only a very small proportion (1%) of accessions have low amylose as a result of *waxy* gene (Mendelian trait).

Figure 1 shows correlation between the adjusted phenotypic means for trait and environment combination. Starch was negatively correlated ($p < 0.001$) with all other traits in both populations except for amylose in the RILs. Fat, protein and gross energy were significantly positively ($p < 0.001$) correlated to each other across environments in both populations. The strongest positive correlation was between gross energy and fat, whereas the strongest negative correlations were found between starch gross energy and starch protein. Moderate (0.4) to high (0.73) positive correlation was observed between years for all traits except for amylose (r = 0.08) between 2014 and 2017 in GSDP (Figure 1). We conducted a principal component analysis (PCA) of correlation matrix for the traits in each environment. In both populations, the first component separated amylose and starch from the other three traits, whereas, the second component separated amylose from starch and gross energy from protein and fat (Figure S3). The first component explained 78.8% and 75.9% of variation, and second component explained 6.3% and 9.8% of variation in the GSDP and RILs, respectively. The third principal component in the RILs separated proteins from fat and explained about 7.6% of the variation.
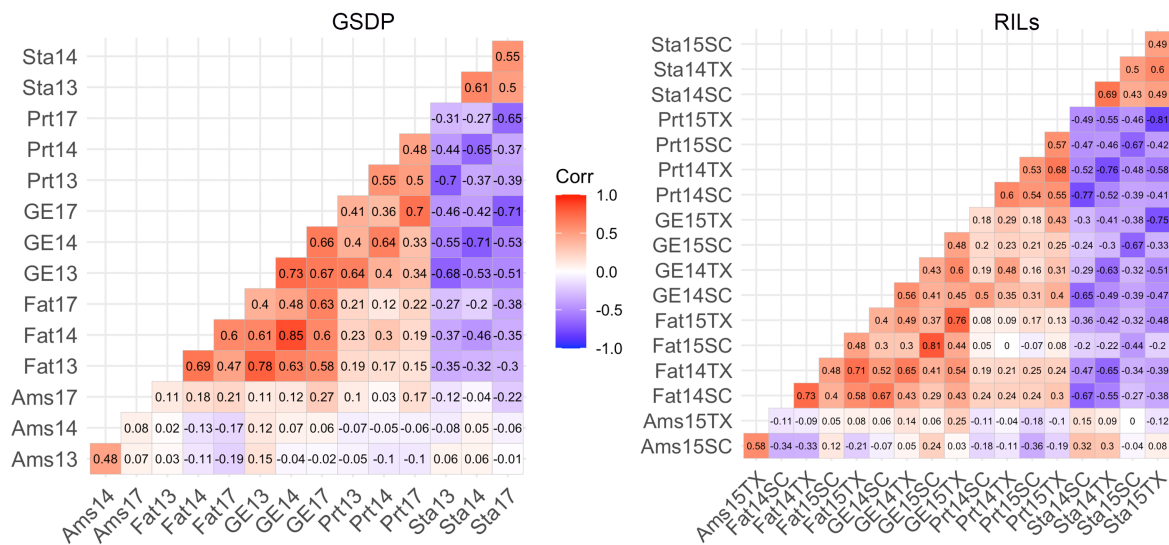
Figure 1. Correlation between traits across year and location combination for the two populations. Ams: amylose, GE: gross energy, Prt: protein, Sta: starch, SC: South Carolina, TX: Texas, and numbers in x and y-axes represent the year.

2.2. Prediction Performance in Single and Multiple Environment

We first implemented GBLUP prediction model for single-trait single-environment (STSE). Prediction accuracies of the STSE model varied across environments in both populations (Figure 2). The environments 2014 in the GSDP and TX2014 in the RILs had highest average prediction accuracy but were not always the best predicted environment for all traits. While poorly predicted for amylose, the environments 2017 in the GSDP and TX2015 in the RILs had higher prediction accuracy for starch compared to all or most environments. Despite variation across environments and populations, the average prediction accuracies from the STSE were within the range of 0.4 to 0.6 for all traits except amylose (0.25) in the GSDP. The average prediction accuracy of the STSE model in the GSDP was positively correlated ($r = 0.86$) with the genomic heritability of the traits. In the RILs, there was a positive correlation ($r = 0.77$) between average prediction accuracy and genomic heritability for amylose, fat and gross energy, but the traits (protein and starch) with the highest heritabilities had relatively lower average prediction accuracies.
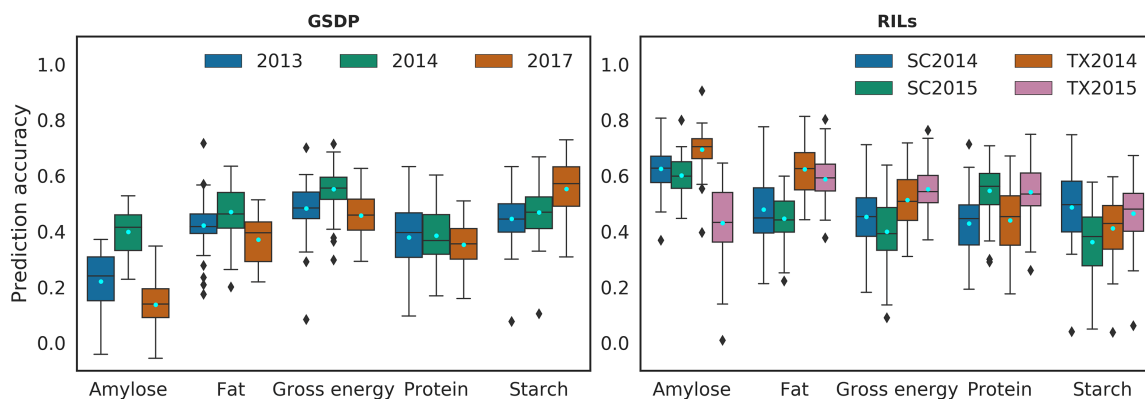

Figure 2. Prediction accuracy for single-trait single-environment (STSE) model. The y-axis shows prediction accuracy calculated as Pearson's correlation between observed values and predicted values of phenotypes. Legend represents the environment/years. SC: South Carolina, TX: Texas, GSDP: Grain sorghum diversity panel, RILs: recombinant inbred lines. Pale blue dots represent the mean of prediction accuracy.

We did not see substantial improvement in multi-environment (BME) model over the STSE prediction accuracies (Figure S4). In the GSDP, the multi-environment models resulted in a decline in average prediction accuracy compared to the STSE model for fat (21%), amylose (10%) and protein (4%), however, no significant change was observed for gross energy and starch (Table 2, Figure S5). The prediction accuracy in the RILs increased by an average of 3% in the BME compared to the STSE, however, the overall trend of prediction accuracy for traits and environments remained unchanged (Table 3, Figure S5). The environment SC2014 showed consistent increase in accuracy for BME over STSE model across all traits with about 10% increase for protein (Table 3). Amylose in TX2015 environment had the single greatest increase (12%) in average prediction accuracy in the BME among all trait-environment combinations for the RILs (Table 3).

Table 2. Percent change in mean prediction accuracy (r) over the single trait single environment (STSE) model in the diversity panel (GSDP). BME: Bayesian multi-environment, and BMORS: Bayesian multi-output regressor stacking. Values were rounded to the nearest whole number.

Trait	2013		2014		2017	
	BME	BMORS	BME	BMORS	BME	BMORS
Amylose	−11	66	−5	11	−13	−13
Fat	−24	47	−12	47	−27	58
Gross energy	3	54	−2	40	1	57
Protein	−3	56	−1	55	−8	52
Starch	4	37	−2	38	1	17
Average	−6	52	−4	38	−9	34

Table 3. Percent change in mean prediction accuracy (r) over the single trait single environment (STSE) model in the recombinant inbred lines (RILs). BME: Bayesian multi-environment, BMORS: Bayesian multi-output regressor stacking, SC: South Carolina, and TX: Texas. Values were rounded to the nearest whole number.

Trait	SC2014		SC2015		TX2014		TX2015	
	BME	BMORS	BME	BMORS	BME	BMORS	BME	BMORS
Amylose	2	28	0	28	−1	25	12	43
Fat	5	33	1	15	2	18	2	20
Gross energy	7	28	−3	27	1	18	3	17
Protein	10	51	1	23	5	60	−4	33
Starch	5	36	−4	40	7	54	4	37
Average	6	35	−1	27	3	35	3	30

2.3. Bayesian Multi-Output Regression Stacking

We tested two different prediction schemes in the BMORS prediction model using the two functions *BMORS()* and *BMORS_Env()* as described in Montesinos-López et al. [19]. While the *BMORS()* function was used for a five-fold CV as described in the methods section, the *BMORS_Env()* was used to assess the prediction performance of whole environments while using the remaining environments as training. So in *BMORS_Env*, an environment was left out during training and correlation between the predicted values (obtained from training with remaining environments) and observed values for the test environment was measured as prediction accuracy for that environment in BMORS_Env model.

2.3.1. Five-Fold CV

The prediction accuracy from five-fold CV in BMORS increased by 41% and 32% on average over the STSE model in GSDP and RILs, respectively. Figure 3 shows the prediction accuracy of BMORS for each trait and environment combination across the two populations. While the percent change in

accuracy varied across environments, the BMORS model nonetheless had higher average prediction accuracy than the STSE and BME models for all traits (Figure S4). The increase in average accuracy in BMORS over STSE ranged from 11% (amylose, 2014) to 66% (amylose, 2013) in the GSDP with exception of amylose in 2017 (13% decrease), and 15% (fat, SC2015) to 60% (protein, TX2014) in the RILs (Tables 2 and 3). The increase in average prediction accuracy was higher (35%) for both locations in 2014 for the RILs, whereas, the year 2013 in the GSDP increased the most. Among the traits, protein (54%) had the greatest average increase in prediction accuracy in the GSDP, whereas in the RILs, protein and starch (42%) both showed the greatest increase.
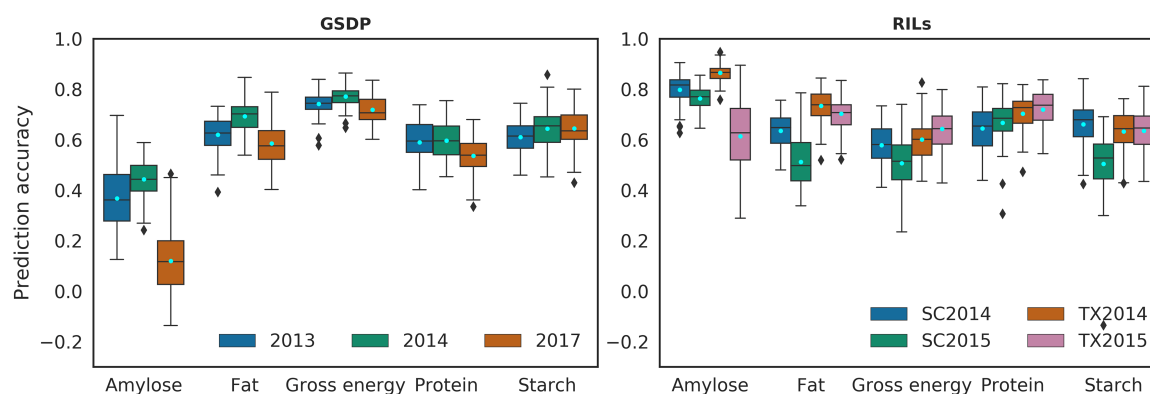

Figure 3. Prediction accuracy of Bayesian multi-output regressor stacking (BMORS) model using five-fold cross validation. Legend represents the years/environment. The y-axis shows prediction accuracy calculated as Pearson's correlation between observed values and predicted values of phenotypes. SC: South Carolina, TX: Texas, GSDP: Grain sorghum diversity panel, and RILs: Recombinant inbred lines. Pale blue dots show mean of the prediction accuracy.

2.3.2. Prediction of Whole Environment

Predicting a whole environment using the BMORS model usually yielded higher accuracy than the mean prediction accuracy from the STSE or BME model where only portions of each environment was tested instead of whole environment as in *BMORS_Env* model (Figures 2 and 4, Figure S5, Table 4). This shows that BMORS model can be reliably used in predicting unobserved environment with the same accuracy as from STSE or BME models from training within the environments. The distribution of prediction accuracy across trait and environment combination were, however, similar to the results from the STSE model. In the GSDP, little variation in prediction accuracies was observed across environments for gross energy, starch and protein, whereas, amylose and fat showed greater variability in prediction accuracy between environments. In the RILs, prediction accuracy for all traits except protein had high variability across the environments (Table 4).

In order to assess predictability by location or year in the RILs, we tested one location or year by training the BMORS model using the other location or year, respectively (Table 4). The Texas location had higher accuracy of prediction for fat (+0.11) and gross energy (+0.1) compared to South Carolina, but rest of the traits had similar prediction accuracy (difference < 0.02). Prediction accuracy of whole years varied across traits, amylose (+0.09) and fat(+0.04) were higher in 2014, protein was higher (+0.05) in 2015, and starch and gross energy were similar.
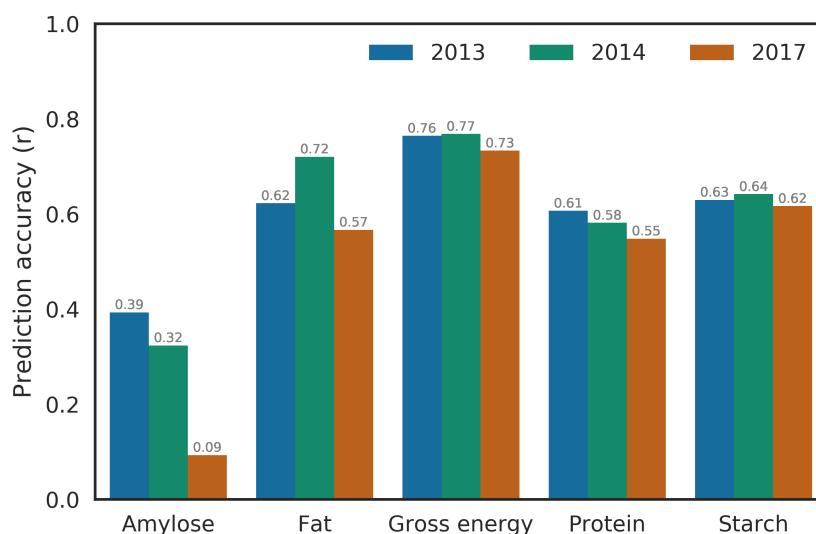
Figure 4. **Prediction accuracy of whole environment predicted using the Bayesian multi-output regressor stacking (BMORS_Env) in the diversity panel (GSDP).** The y-axis shows prediction accuracy calculated as Pearson's correlation between observed values and predicted values of phenotypes. Values on top of the bar represent the height of the bar.

Table 4. **Prediction accuracy of the test environments predicted using the Bayesian multi-output regressor stacking (BMORS_Env) in the recombinant inbred lines (RILs).** SC: South Carolina, TX, Texas. Prediction accuracy was calculated as Pearson's correlation between observed values and predicted values of phenotypes.

Trait	Year × Location				Location		Year	
	SC2014	SC2015	TX2014	TX2015	SC	TX	2014	2015
Amylose	0.79	0.80	0.88	0.60	0.76	0.74	0.74	0.65
Fat	0.69	0.49	0.78	0.74	0.60	0.71	0.64	0.60
Gross energy	0.56	0.49	0.62	0.66	0.48	0.58	0.56	0.56
Protein	0.65	0.66	0.66	0.70	0.59	0.58	0.61	0.66
Starch	0.64	0.52	0.68	0.60	0.55	0.56	0.56	0.55

3. Discussion

Phenotyping for grain compositional traits is—(1) challenging and labor-intensive, (2) destructive for most accurate results, and (3) only performed after plants reach physiological maturity and are harvested. The use of genomic prediction for compositional traits will be extremely valuable because it increases selection intensity and decreases generational interval by overcoming the phenotyping challenges. Moreover, these traits are complex and quantitatively inherited so will benefit from genomic prediction's ability to account for many small effect QTLs in estimating breeding values.

3.1. Trait Architecture and Prediction Accuracy

While the accuracy of NIRS calibration for traits in this study ranged from moderate to high, there was prediction error associated with NIRS prediction. However, it is unclear if and what effects NIRS prediction error had on genomic prediction. No direct correlation was observed between the genomic prediction accuracy and NIRS statistics for the traits studied. The trait with the lowest NIRS R^2, fat, was predicted as well as or better than starch, protein and gross energy, which had NIRS $R^2 > 0.7$. Despite varying strength of correlations between traits across the two populations studied,

the nature of relationship was similar for a given pair of traits, which is also in agreement with previous studies [30,32,33]. The strong negative relationship of starch and amylose to protein, fat and gross energy was further elucidated by the PCA analysis of phenotypic correlation matrix (Figure S3). Since starch, protein and fat were measured on a percent dry matter basis, the strong correlation between them is expected.

Genetic relatedness and trait architecture are known to affect the accuracy of genomic prediction [8,34]. The genetic relatedness between individuals and heritability of the traits were higher in the RILs than in GSDP (Figure S6, Table 1). Those factors could be contributing to higher average prediction accuracy in the RILs. However, the average prediction accuracies for gross energy and starch were comparable between GSDP and RILs (Figure S4). Prediction accuracy in the GSDP could have been boosted by greater genetic diversity despite lower genetic relatedness [35]. Heffner et al. [22] observed a prediction accuracy of 0.5–0.6 for wheat flour protein in two biparental populations. Guo et al. [26] reported prediction accuracies of 0.44 and 0.8 for protein and amylose in rice diversity panel. Similar results were observed in our STSE models for protein content (Figure 2). Whereas, lower prediction accuracy of amylose in our diversity panel is probably due to the lack of sufficient low-amylose lines with the *waxy* gene [30]. While genomic prediction study for starch, fat and gross energy has not been reported in sorghum, these traits are nutritionally one of the most important traits for any cereal grain. The moderate to high prediction accuracy observed suggests implementation of genomic selection can improve genetic gain for these grain quality traits.

3.2. Multi-Trait Regressor Stacking

One of the daunting tasks of genomic prediction is estimating the effects of unobserved individuals and environments. As multiple traits are analyzed across several environments, the ability to combine information from multiple traits and environments can be crucial in increasing accuracy of prediction [13,15,16]. When the correlations among traits are high, prediction accuracies of complex traits can be increased by using multivariate model that takes this correlation into account [15,18]. We fit a Bayesian multi-environment (BME) model (2) that takes the genotype × environment effects into consideration. In the GSDP, where environments were three years at the same location, the BME model showed a slight decline (7%) in average prediction accuracy which was mostly due to the two traits, amylose and fat (Table 2). The RILs showed slight increase (2–3%) in prediction accuracy of traits when averaged over the environments, but there was variability across the environments (Table 3).

We implemented two functions [*BMORS*() and *BMORS_Env*()] which are not only used to evaluate prediction accuracy but are also computationally efficient [19]. The BMORS model (3) performs two-stage training by stacking the multi-environment models from all the traits. The five-fold cross validation conducted for BMORS was similar to the CV1 strategy of Montesinos-López et al. [18]. The use of multi-trait models has been consistently shown to increase prediction accuracy over single-trait models across different crops and traits [15–17,36]. The multi-target regressor stacking increased average prediction accuracy by 41% and 32% in the GSDP and RILs, respectively, as compared to the STSE prediction accuracy. Average prediction accuracy of all traits improved in BMORS over STSE and BME across both the populations (Figure S4). Consistent improvement in accuracy of BMORS is a result of the ability to use not only correlation between traits but also between environment in the model training [18,19]. The ability to accurately estimate genetic merit of lines in unobserved environments is of tremendous value in plant breeding. Our results show potential of *BMORS_Env()* function for predicting the whole environment. Testing a whole environment by training BMORS model using all other environments resulted in higher prediction accuracy for that trait-environment combination than using STSE or BME model. Prediction accuracy of all environments were 0.5 or higher with exception of amylose in GSDP, the reason for which we have discussed above (Figure 4, Table 4).

3.3. Application for Breeding

Grain quality traits such as starch and protein content have been under selection since the inception of phenotypic selection in modern breeding practices. More recently, total energy supplement of grain has gained attention for increasing feed efficiency in animal production, and a need exists for increasing total calories for human nutrition in the wake of global malnutrition crisis. Despite high correlations among these traits, the genetic variation underlying starch, protein and fat can be decoupled. Boyles et al. [30] showed major and minor effect QTLs underlying the three traits are distributed across the genome and are segregating in biparental populations. However, in practice, selection would be conducted simultaneously for these traits using a selection index rather than for individual traits. Velazco et al. [31] observed an increase in predictive ability by using a multi-trait model for grain yield and stay green in sorghum, and argue that such an exercise would allow for using selection index for implementation of genomic selection for correlated traits. Increased prediction accuracy, improved selection index, and estimation of precise genetic, environmental and residual co-variances makes multi-trait multi-environment models preferable over univariate models [18]. The multi-trait regression stacking model we tested shows large scale improvement in model prediction and can be used in tandem with Bayesian multi-trait multi-environment (BMTME) model for parameter estimation and assessing prediction accuracy. The ability to estimate genetic effects and breeding values of unobserved environments will be of great advantage to predict performance in diverse environments and for implementation of selection theory.

4. Materials and Methods

4.1. Plant Material

4.1.1. Grain Sorghum Diversity Panel

A grain sorghum diversity panel (GSDP) of 389 diverse sorghum accessions was planted in randomized complete block design with two replications in 2013, 2014, and 2017 field seasons at the Clemson University Pee Dee Research and Education Center in Florence, SC. The GSDP included a total of 332 accessions from the original United States sorghum association panel (SAP) developed by Casa et al. [37]. The details on experimental field design and agronomic practices are described in Boyles et al. [38] and Sapkota et al. [35]. Briefly, the experiments were planted in a two row plots each 6.1 m long, separated by row spacing of 0.762 m with an approximate planting density of 130,000 plants ha^{-1}. Fields were irrigated only when signs of drought stress was seen across the field.

4.1.2. Recombinant Inbred Population

A biparental population of 191 recombinant inbred lines (RILs) segregating for grain quality traits was studied along with the GSDP. The parents of the RIL population were BTx642, a yellow-pericarp drought tolerant line, and BTxARG-1, a white pericarp waxy endosperm (low amylose) line. The population was planted in two replicated plots in randomized complete block design across two years (2014 and 2015) in Blackville, SC and College Station, TX. Field design and agronomic practices have previously been described in detail in Boyles et al. [30].

4.2. Phenotyping

The primary panicle of three plants selected from each plot were harvested at physiological maturity. The plants from beginning and end of the row were excluded to account for border effect. Panicles were air dried to a constant moisture (10–12%) and threshed. A 25 g subsample of cleaned and homogenized grain ground to 1-mm particle size with a CT 193 Cyclotec Sample Mill (FOSS North America) was used in near-infrared spectroscopy (NIRS) for compositional analysis.

Grain composition traits such as total fat, gross energy, crude protein, and starch content can be measured using NIRS. Previous studies have shown high NIRS predictability of the traits used in

feed analysis [39,40]. We used a DA 7250$^{\text{TM}}$ NIR analyzer (Perten Instruments). The ground sample was packed in a gradually rotating Teflon dish positioned under the instrument's light source and predicted phenotypic values was generated based on calibration curve for spectral measurements. The calibration curve was built using wet chemistry values from a subset of samples. The wet chemistry was performed by Dairyland Laboratories, Inc. (Arcadia, WI, USA) and the Quality Assurance Laboratory at Murphy-Brown, LLC (Warsaw, NC, USA). The details on the prediction curves and wet chemistry can be found in Boyles et al. [30].

4.3. Genotypic Data

Genotyping-by-sequencing (GBS) was used for genetic characterization of the GSDP and RILs populations [30,38,41]. Sequenced reads were aligned to the BTx623 v3.1 reference assembly (phytozome) using Burrows-Wheeler aligner [42]. SNP calling, imputation and filtering was done using TASSEL 5.0 pipeline [43]. The TASSEL plugin FILLIN for GSDP and FSFHap for RILs population were used to impute for missing genotypes. Following imputation SNPs with minor allele frequency (MAF) < 0.01, and sites missing in more than 10% and 30% of the genotypes in GSDP and RILs, respectively, were filtered. The number of genotypes studied for each population represent those with at least 70% of SNP sites. The genotype matrix with 224,007 SNPs from GSDP and 56,142 SNPs from RILs population was used for whole genome regression.

4.4. Statistical Analysis

The statistical software environment 'R' was used for model building and analysis [44]. The phenotypic values of the traits were adjusted for random effects of replications within environment using 'lme4' package in R [45]. Principal component analysis was done using the R package 'factoextra' [46]. Marker-based estimates of narrow sense (genomic) heritabilities were calculated using the SNP genotype matrix and phenotypic values using the R package 'heritability' [47]. A matrix with dummy variables '1' and '0' representing combinations of environmental variables (replication and year for GSDP, and replication, year and location for RILs) was used as co-variate in heritability estimation.

4.4.1. Single-Trait Single-Environment (STSE) Model:

The following genomic best linear unbiased prediction (GBLUP) model was used to assess prediction performance of an individual trait from a single environment:

$$y_j = \mu + g_j + e_j, \tag{1}$$

where y_j is a vector of adjusted phenotypic mean of the jth line ($j = 1, 2, ..., J$). μ is the overall mean which is assigned a flat prior, g_j is a vector of random genomic effect of the jth line, with $\boldsymbol{g} = (g_1, ..., g_j)^T \sim N(0, \boldsymbol{G}\sigma_1^2)$, σ_1^2 is a genomic variance, $\boldsymbol{G}$ is the genomic relationship matrix in the order $J \times J$ and is calculated [12] as $\boldsymbol{G} = \frac{ZZ^T}{2\sum p_j q_j}$, where q_j and p_j denote major and minor allele frequency of jth line respectively, and Z is the design matrix for markers of order $J \times p$ (p is total number of markers). Further, e_j is residual error assigned the normal distribution $e \sim N(0, I\sigma_e^2)$ where I is identity matrix and σ_e^2 is the residual variance with a scaled-inverse Chi-square density.

4.4.2. Bayesian Multi-Environment (BME) GBLUP Model

Considering genotype $\times$ environment interaction can contribute to substantial amount of phenotypic variance in complex traits, we fit the following univariate linear mixed model to account for environmental effects in prediction performance:

$$y_{ij} = E_i + g_j + gE_{ij} + e_{ij}, \tag{2}$$

where y_j is a vector of adjusted phenotypic mean of the jth line in the ith environment (i = 1, 2, ..., I, j = 1, 2, ..., J). E_i represents the effect of ith environment and g_j represents the genomic effect of the jth line as described in Equation (1). The term gE_{ij} represents random interaction between the genomic effect of jth line and the ith environment with $gE = (gE_{11}, ..., gE_{IJ})^T \sim N(0, \sigma_2^2\,\mathbf{I}_I \otimes \mathbf{G})$, where σ_2^2 is an interaction variance, and e_{ij} is a random residual associated with the jth line in the ith environment distributed as N(0, σ_e^2) where σ_e^2 is the residual variance.

4.4.3. Bayesian Multi-Output Regressor Stacking (BMORS)

BMORS is the Bayesian version of multi-trait (or multi-target) regressor stacking method [48]. The multi-target regressor stacking (MTRS) was proposed by Spyromitros-Xioufis et al. [20,21] based on multi-labeled classification approach of Godbole and Sarawagi [49]. In BMORS or MTRS, the training is done in two stages. First, L univariate models are implemented using the multi-environment GBLUP model given in Equation (2), then instead of using these models for prediction, MTRS performs the second stage of training using a second set of L meta-models for each of the L traits. The following model is used to implement each meta-model:

$$y_{ij} = \beta_1 \hat{Z}_{1ij} + \beta_2 \hat{Z}_{2ij} + ... + \beta_L \hat{Z}_{Lij} + e_{ij}, \tag{3}$$

where the covariates $\hat{Z}_{1ij}$, $\hat{Z}_{2ij}$, ..., $\hat{Z}_{Lij}$ represent the scaled prediction from the first stage training with the GBLUP model for L traits, and β_1, ..., β_L are the regression coefficients for each covariate in the model. The scaling of each prediction was performed by subtracting its mean (μ_{lij}) and dividing by its corresponding standard deviation (σ_{lij}), that is, $\hat{Z}_{lij} = (\hat{y}_{lij} - \mu_{lij})\sigma_{lij}^{-1}$, for each l = 1, ..., L. The scaled predictions of its response variables yielded by the first-stage models as predictor information by the BMORS model. Simply put, the multi-trait regression stacking model is based on the idea that a second stage model is able to correct the predictions of a first-stage model using information about the predictions of other first-stage models [20,21].

4.4.4. Performance of Prediction Model:

All prediction models were fit using Bayesian approach in statistical program 'R'. The STSE model (1) was fit using the R package 'BGLR' [50], BME model (2) and BMORS model (3) were fit using the R package 'BMTME' [19]. A minimum of 20,000 iterations with 10,000 burn-in steps was used for each Bayesian run.

The evaluation of prediction performance of models was done using a five-fold cross validation (CV), which means 80% of the samples were used as training set and testing was done on the remaining 20% for each cross-validation fold. The individuals were randomly assigned into five mutually exclusive folds. Four folds were used to train prediction models and to predict the genomic estimated breeding values (GEBVs) of the individuals in fifth fold (validation/test set). The accuracy of prediction for each fold was calculated as Pearson's correlation coefficient (r) between predicted values and adjusted phenotypic means for the individuals in validation set. Each cross validation run, therefore, resulted in five estimates of prediction accuracy. The same set of individuals were assigned to training and validation across different traits and models tested by using *set.seed()* function in R. In order to avoid bias due to sampling, we performed 10 different cross-validation runs to calculate the mean and dispersion of the prediction accuracies.

5. Conclusions

Phenotyping of grain compositional traits using near-infrared spectroscopy is labor-intensive, generally destructive, and time limiting. Therefore, the use of genomic selection for these traits will be extremely valuable. This study establishes the potential to improve genomics-assisted selection of grain composition traits by using multi-trait multi-environment model. The phenotypic measurements obtained from NIRS prediction were amenable to genomic selection as shown by moderate to high

prediction accuracy for single trait prediction. While multi-environment model alone did not lead to much improvement over single environment model, stacking of regression from multiple traits showed substantial improvement in prediction accuracy. The prediction accuracy increased by 32% and 41% in the RILs and GSDP, respectively, when using the Bayesian multi-output regressor stacking (BMORS) model compared to a single trait single environment model. The ability to predict line performance in an unobserved environment is of great importance to breeding programs, and results show high accuracy for predicting whole environments using BMORS.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4395/10/9/1221/s1. The supplementary file contains six figures: Figure S1. Phenotypic distribution of grain composition traits in the RILs. In the x-axes, SC: South Carolina, TX: Texas, numbers represent years. Values are percentage dry basis for protein, fat and starch; gross energy is in KCal/lb; and amylose is in percent of starch. Figure S2. Phenotypic distribution of grain composition traits in the GSDP. Numbers in x-axes represent years. Values are percentage dry basis for protein, fat and starch; gross energy is in Cal/g; and amylose is in percent of starch. Figure S3. PCA analysis of correlation matrix between traits. a. GSDP, and b. RILs. Ams: amylose, GE: gross energy, Prt: protein, Sta: starch, SC: South Carolina, TX: Texas. The numbers in the text represent years of the environment. Figure S4. Overall prediction accuracy of traits across all the environment for the three prediction methods in the two populations. The y-axis shows prediction accuracy calculated as Pearson's correlation between observed values and predicted values of phenotypes. Legend represents the environment/years. SC: South Carolina, TX: Texas, GSDP: Grain sorghum diversity panel, RILs: recombinant inbred lines. Figure S5. Prediction accuracy using five-fold CV in Bayesian multi-environment (BME) model. a. GSDP, and b. RILs. Legend represents the environment/years. SC: South Carolina, TX: Texas. Pale blue dots represent the mean of prediction accuracy. Figure S6. Heatmap for genomic relationship matrix calculated using vanRaden (2008). a. GSDP, b. RILs. Trees show hierarchical clustering using Euclidean distance.

Author Contributions: S.S. conceptualized the study, performed data analysis, and wrote the manuscript; R.B. helped with experimental design and field phenotyping; J.L.B. helped in computation and data analysis; K.J. helped in near infra-red phenotyping; S.K. helped with conception of the study, acquisition of fund, and management of the study. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Data Availability: The codes and phenotypic data used in the study can be accessed through github at sirjansapkota/GrainComp_GS.

Abbreviations

The following abbreviations are used in this manuscript:

GSDP	Grain sorghum diversity panel
RIL	Recombinant inbred line
NIRS	Near infra-red spectroscopy
GP	Genomic prediction
GBLUP	Genomic best linear unbiased prediction
STSE	Single trait single environment
BME	Bayesian multi-environment
BMORS	Bayesian multi-output regressor stacking
MTME	Multi-trait multi-environment
CV	Cross validation
SC	South Carolina
TX	Texas
QTL	Quantitative trail loci
SNP	Single nucleotide polymorphism

References

1. Awika, J.M. Major cereal grains production and use around the world. In *Advances in Cereal Science: Implications to Food Processing and Health Promotion*; ACS Publications: Washington, DC, USA, 2011; pp. 1–13.
2. Mace, E.S.; Tai, S.; Gilding, E.K.; Li, Y.; Prentis, P.J.; Bian, L.; Cruickshank, A. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **2013**, *4*, 2320. [CrossRef]
3. Taylor, J.R.; Schober, T.J.; Bean, S.R. Novel food and non-food uses for sorghum and millets. *J. Cereal Sci.* **2006**, *44*, 252–271. [CrossRef]
4. Taylor, J. Food product development using sorghum and millets: opportunities and challenges. *Qual. Assur. Saf. Crop. Foods* **2012**, *4*, 151. [CrossRef]
5. Zhu, F. Structure, physicochemical properties, modifications, and uses of sorghum starch. *Compr. Rev. Food Sci. Food Saf.* **2014**, *13*, 597–610. [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*, 1819–1829.
7. Bernardo, R.; Yu, J. Prospects for genomewide selection for quantitative traits in maize. *Crop. Sci.* **2007**, *47*, 1082–1090. [CrossRef]
8. Habier, D.; Fernando, R.L.; Dekkers, J.C. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **2007**, *177*, 2389–2397. [CrossRef] [PubMed]
9. Zhong, S.; Dekkers, J.C.; Fernando, R.L.; Jannink, J.L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. *Genetics* **2009**, *182*, 355–364. [CrossRef] [PubMed]
10. Combs, E.; Bernardo, R. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome* **2013**, *6*, 1–7. [CrossRef]
11. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Dreisigacker, S. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [CrossRef]
12. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef] [PubMed]
13. Burgueño, J.; de los Campos, G.; Weigel, K.; Crossa, J. Genomic prediction of breeding values when modeling genotype×environment interaction using pedigree and dense molecular markers. *Crop. Sci.* **2012**, *52*, 707–719. [CrossRef]
14. Jarquín, D.; Crossa, J.; Lacaze, X.; Du Cheyron, P.; Daucourt, J.; Lorgeou, J.; Burgueño, J. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* **2014**, *127*, 595–607. [CrossRef] [PubMed]
15. Jia, Y.; Jannink, J.L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* **2012**, *192*, 1513–1522. [CrossRef]
16. Guo, G.; Zhao, F.; Wang, Y.; Zhang, Y.; Du, L.; Su, G. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* **2014**, *15*, 30. [CrossRef]
17. Lado, B.; Vázquez, D.; Quincke, M.; Silva, P.; Aguilar, I.; Gutiérrez, L. Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor. Appl. Genet.* **2018**, *131*, 2719–2731. [CrossRef]
18. 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]
19. 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]
20. 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.
21. Spyromitros-Xioufis, E.; Tsoumakas, G.; Groves, W.; Vlahavas, I. Multi-label classification methods for multi-target regression. *arXiv* **2012**, arXiv:12116581.

22. Heffner, E.L.; Jannink, J.L.; Iwata, H.; Souza, E.; Sorrells, M.E. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop. Sci.* **2011**, *51*, 2597–2606. [CrossRef]

23. Battenfield, S.D.; Guzmán, C.; Gaynor, R.C.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; 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] [PubMed]

24. Haile, J.K.; N'Diaye, A.; Clarke, F.; Clarke, J.; Knox, R.; Rutkoski, J.; Pozniak, C.J. Genomic selection for grain yield and quality traits in durum wheat. *Mol. Breed.* **2018**, *38*, 75. [CrossRef]

25. 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]

26. Guo, Z.; Tucker, D.M.; Basten, C.J.; Gandhi, H.; Ersoz, E.; Guo, B.; Gay, G. The impact of population structure on genomic prediction in stratified populations. *Theor. Appl. Genet.* **2014**, *127*, 749–762. [CrossRef] [PubMed]

27. Duhnen, A.; Gras, A.; Teyssèdre, S.; Romestant, M.; Claustres, B.; Daydé, J.; Mangin, B. Genomic selection for yield and seed protein content in Soybean: A study of breeding program data and assessment of prediction accuracy. *Crop. Sci.* **2017**, *57*, 1325–1337. [CrossRef]

28. Hayes, B.; Panozzo, J.; Walker, C.; Choy, A.; Kant, S.; Wong, D.; Spangenberg, G.C. 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]

29. Rapp, M.; Lein, V.; Lacoudre, F.; Lafferty, J.; Müller, E.; Vida, G.; Leiser, W.L. Simultaneous improvement of grain yield and protein content in durum wheat by different phenotypic indices and genomic selection. *Theor. Appl. Genet.* **2018**, *131*, 1315–1329. [CrossRef]

30. Boyles, R.E.; Pfeiffer, B.K.; Cooper, E.A.; Rauh, B.L.; Zielinski, K.J.; Myers, M.T.; Kresovich, S. Genetic dissection of sorghum grain quality traits using diverse and segregating populations. *Theor. Appl. Genet.* **2017**, *130*, 697–716. [CrossRef]

31. Velazco, J.G.; Jordan, D.R.; Mace, E.S.; Hunt, C.H.; Malosetti, M.; Van Eeuwijk, F.A. Genomic prediction of grain yield and drought-adaptation capacity in sorghum is enhanced by multi-trait analysis. *Front. Plant Sci.* **2019**, *10*, 997. [CrossRef]

32. Murray, S.C.; Sharma, A.; Rooney, W.L.; Klein, P.E.; Mullet, J.E.; Mitchell, S.E.; Kresovich, S. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop. Sci.* **2008**, *48*, 2165–2179. [CrossRef]

33. Sukumaran, S.; Xiang, W.; Bean, S.R.; Pedersen, J.F.; Kresovich, S.; Tuinstra, M.R.; Yu, J. Association mapping for grain quality in a diverse sorghum collection. *Plant Genome* **2012**, *5*, 126–135. [CrossRef]

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

35. Sapkota, S.; Boyles, R.; Cooper, E.; Brenton, Z.; Myers, M.; Kresovich, S. Impact of sorghum racial structure and diversity on genomic prediction of grain yield components. *Crop. Sci.* **2020**, *60*, 132–148. [CrossRef]

36. Bhatta, M.; Gutierrez, L.; Cammarota, L.; Cardozo, F.; Germán, S.; Gómez-Guerrero, B.; Castro, A.J. Multi-trait Genomic Prediction Model Increased the Predictive Ability for Agronomic and Malting Quality Traits in Barley (*Hordeum vulgare* L.). *G3 Genes Genomes Genet.* **2020**, *10*, 1113–1124. [CrossRef] [PubMed]

37. Casa, A.M.; Pressoir, G.; Brown, P.J.; Mitchell, S.E.; Rooney, W.L.; Tuinstra, M.R.; Kresovich, S. Community resources and strategies for association mapping in sorghum. *Crop. Sci.* **2008**, *48*, 30–40. [CrossRef]

38. Boyles, R.E.; Cooper, E.A.; Myers, M.T.; Brenton, Z.; Rauh, B.L.; Morris, G.P.; Kresovich, S. Genome-wide association studies of grain yield components in diverse sorghum germplasm. *Plant Genome* **2016**, *9*, 1–17. [CrossRef]

39. Kays, S.E.; Barton, F.E. Rapid prediction of gross energy and utilizable energy in cereal food products using near-infrared reflectance spectroscopy. *J. Agric. Food Chem.* **2002**, *50*, 1284–1289. [CrossRef]

40. De Alencar Figueiredo, L.F.; Sine, B.; Chantereau, J.; Mestres, C.; Fliedel, G.; Rami, J.F.; Courtois, B. Variability of grain quality in sorghum: Association with polymorphism in Sh2, Bt2, SssI, Ae1, Wx and O2. *Theor. Appl. Genet.* **2010**, *121*, 1171–1185. [CrossRef]

41. Morris, G.P.; Ramu, P.; Deshpande, S.P.; Hash, C.T.; Shah, T.; Upadhyaya, H.D.; Riera-Lizarazu, O.; Brown, P.J.; Acharya, C.B.; Mitchell, S.E.; et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 453–458. [CrossRef]

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

43. Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J.; Elshire, R.J.; Sun, Q.; Buckler, E.S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* **2014**, *9*, e90346. [CrossRef] [PubMed]

44. R Core Team. R: A Language and Environment for Statistical Computing. 2019. Available online: https://www.R-project.org/ (accessed on 26 April 2019).

45. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]

46. Kassambara, A.; Mundt, F. Factoextra: extract and visualize the results of multivariate data analyses. *R Package Version* **2017**, *1*, 337–354.

47. Kruijer, W.; Boer, M.P.; Malosetti, M.; Flood, P.J.; Engel, B.; Kooke, R.; van Eeuwijk, F.A. Marker-based estimation of heritability in immortal populations. *Genetics* **2015**, *199*, 379–398. [CrossRef] [PubMed]

48. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Cuevas, J.; Montesinos-López, J.C.; Gutiérrez, Z.S.; 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] [PubMed]

49. Godbole, S.; Sarawagi, S. Discriminative methods for multi-labeled classification. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 26–28 May 2004; Springer: Berlin/Heidelberg, Germany, 2004; p. 22–30.

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