

Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding

Jared Crain, Suchismita Mondal, Jessica Rutkoski, Ravi P. Singh, and Jesse Poland*

Abstract

Genomics and phenomics have promised to revolutionize the field of plant breeding. The integration of these two fields has just begun and is being driven through big data by advances in next-generation sequencing and developments of field-based high-throughput phenotyping (HTP) platforms. Each year the International Maize and Wheat Improvement Center (CIMMYT) evaluates tens-of-thousands of advanced lines for grain yield across multiple environments. To evaluate how CIMMYT may utilize dynamic HTP data for genomic selection (GS), we evaluated 1170 of these advanced lines in two environments, drought (2014, 2015) and heat (2015). A portable phenotyping system called 'Phenocart' was used to measure normalized difference vegetation index and canopy temperature simultaneously while tagging each data point with precise GPS coordinates. For genomic profiling, genotyping-by-sequencing (GBS) was used for marker discovery and genotyping. Several GS models were evaluated utilizing the 2254 GBS markers along with over 1.1 million phenotypic observations. The physiological measurements collected by HTP, whether used as a response in multivariate models or as a covariate in univariate models, resulted in a range of 33% below to 7% above the standard univariate model. Continued advances in yield prediction models as well as increasing data generating capabilities for both genomic and phenomic data will make these selection strategies tractable for plant breeders to implement increasing the rate of genetic gain.

Core Ideas

- Wheat breeding
- High throughput phenotyping
- Genomic selection
- Yield prediction modeling

TO MEET THE FUTURE GLOBAL DEMAND for food, the current levels of crop production need to increase between 59 and 98% by 2050 compared to 2005 levels (Valin et al., 2013). This is up to a 2.4% yield increase per year for grain crops (Ray et al., 2013). Genetic gains in wheat yield, however, are currently estimated to be less than 1% per year (Reynolds et al., 2012; Ray et al., 2013); much lower than the 2.4% target. To increase genetic gains and maintain a stable food supply plant breeders and geneticist need to utilize contemporary methods to enhance current breeding strategies. Genomics has long promised to revolutionize plant breeding by characterizing germplasm and allowing

J. Crain, Dep. Plant Pathology, Kansas State Univ., Manhattan KS, 66506; S. Mondal, R.P. Singh, International Maize and Wheat Improvement Center (CIMMYT), Int. Apdo. Postal 6-641, 06600 Mexico, DF, Mexico; J. Rutkoski, International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila, 1301, Philippines; J. Poland, Wheat Genetics Resource Center, Dep. Plant Pathology, Kansas State Univ., Manhattan KS, 66506. *Corresponding author (jpoland@ksu.edu). Received 24 May 2017. Accepted 20 Nov. 2017.

Abbreviations: BLUP, best linear unbiased predictor; CENEB, Campo Experimental Norman E. Borlaug; CIMMYT, International Maize and Wheat Improvement Center; CT, canopy temperature; EN, elastic net; GBLUP, genomic best linear unbiased predictor; GS, genomic selection; GBS, genotyping-by-sequencing; HTP, high-throughput phenotyping; H^2 , broad sense heritability; MAF, minor allele frequency; MAS, marker assisted selection; NDVI, normalized difference vegetation index; NGS, next generation sequencing; PLSR, partial least squares regression; QTL, quantitative trait loci; RR-BLUP, random regression best linear unbiased prediction; SNP, single nucleotide polymorphism

Plant Genome 11:170043
doi: 10.3835/plantgenome2017.05.0043

© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

individual loci to be dynamically manipulated and selected for crop improvement (Beckmann and Solter, 1986). This promise is begging to be realized through implementation of molecular breeding and genomic selection (Poland, 2015).

Studies have shown that genomics and marker assisted selection (MAS) can be incorporated into breeding programs often resulting in a near two-fold rate of genetic gain compared to standard phenotypic selection (Eathington et al., 2007; Battenfield et al., 2016). Next generation sequencing (NGS) techniques have been developed that radically reduce the amount of time for marker discovery, one of the historical challenges of MAS (Xu and Crouch, 2008). Methods such as genotyping-by-sequencing (GBS) (Elshire et al., 2011), reduced representation libraries (Altshuler et al., 2000), and restriction site-associated DNA sequencing (Miller et al., 2007) can be used to quickly discover thousands of markers and genotype individuals. Additionally, these approaches avoid ascertainment bias that can be present in more costly, high-throughput SNP arrays (Davey et al., 2011). Along with providing a high number of markers, NGS methods are also very cost effective for genomic profiling of thousands of breeding lines (Poland and Rife, 2012). In comparison to ever increasing cost associated with field trials (Bernardo and Yu, 2007; Heffner et al., 2010), NGS provides opportunities for crop breeders and geneticist to genotype entire breeding programs (Poland, 2015) and generate dense genomic information that can be used for plant improvement (Morrell et al., 2011).

One method utilizing this wealth of genomic data is genomic selection (GS) as first proposed by Meuwissen et al. (2001). Genomic selection simultaneously estimates all marker effects to predict total genetic value. Genomic selection works by leveraging dense marker data that covers the entire target genome so that each quantitative trait loci (QTL) is in linkage disequilibrium with a marker (Goddard and Hayes, 2007). By simultaneously estimating all marker effects, variation can be captured that may not have identified above a significance threshold using traditional statistical approaches (Meuwissen and Goddard, 2001). Genomic selection has been shown to be more accurate in predicting estimated breeding values in relationship to true breeding values in simulation (Meuwissen et al., 2001; Zhong et al., 2009), and GS is estimated to be twice as efficient as MAS in winter wheat (Heffner et al., 2010).

Genomic selection requires accurate phenotyping, which has long been the key to enhancing genetic gains through classical plant breeding. Phenotypic observations are collected on the training population that is then used by the GS model for predictions. The phenotypic information is used not to select individual plants *per se*, but to train prediction models and predict the performance of non-phenotyped individuals from their marker scores (Meuwissen et al., 2001). Thus, phenotyping plays an essential role in the success of standard phenotypic selection as well as genomic selection models. However, the ability to assess phenotypes has lagged behind

advancements in genomics (Campos et al., 2004; White et al., 2012; Cobb et al., 2013).

Recently several field-based high-throughput phenotyping (HTP) platforms have been developed to alleviate the phenotyping bottleneck. Some of the platforms have been push carts (White and Conley, 2013; Crain et al., 2016), tractor mounted systems (Busemeyer et al., 2013a; Andrade-Sanchez et al., 2014) and aerial vehicles (Liebisch et al., 2015; Haghighattalab et al., 2016). White et al. (2012) and Deery et al. (2014) provide detailed reviews of potential phenotyping platforms and the benefits and challenges associated with each system. While each phenotyping system is varied in its capabilities and cost, they all have the option to provide dense phenotypic data that can be used to understand crop growth. Measurements provided by HTP systems have been shown to be highly correlated to manual measurements suggesting high accuracy relative to current approaches, and many HTP platforms can measure multiple traits simultaneously (Andrade-Sanchez et al., 2014; Crain et al., 2016). For example, Busemeyer et al. (2013b) mapped temporal genetic dynamics for biomass accumulation using hyperspectral imaging and time of flight cameras and showed that these measurement were highly correlated to empirical measurements of biomass.

The addition of HTP platforms provides opportunities to increase genomic selection through enhancing prediction models, but the question persists of how best to incorporate this new information. Currently, most GS models are single trait models, incorporating phenotypic information on only the target trait (Jia and Jannink, 2012). While single trait models have proven to be useful (Heffner et al., 2009, 2011), they do not take advantage of correlations between traits (Jia and Jannink, 2012). Often grain yield is predicted by GS models, but there are a myriad of physiological processes that culminate in grain yield (Pask et al., 2012) and well documented cases of physiological phenotypes correlated to yield (Amani et al., 1996; Gutiérrez-Rodríguez et al., 2004; Babar et al., 2006b). In particular, HTP methodologies have been very amenable to measuring traits such as spectral reflectance and canopy temperature (CT) that have potential for use as selection tools because of their high correlations to grain yield (Amani et al., 1996; Gutiérrez-Rodríguez et al., 2004; Babar et al., 2006a). Multiple-trait genomic selection has been proposed to leverage the shared information between correlated traits (Jia and Jannink, 2012). However, there are few examples in literature where multiple-trait or multivariate GS models have been deployed.

Jia and Jannink (2012) showed that prediction accuracies could be increased significantly for traits with low heritability through multivariate GS models that include correlated traits to the trait of interest. Calus and Veerkamp (2011) also found that higher genomic prediction accuracy could be achieved by using multi-trait GS models. Their study found that multi-trait models performed better when there were high genetic correlations between traits resulting in up to a 0.14 increase in accuracy, with accuracy being the correlation between

estimated and observed phenotypic values. Even with low genetic correlations they found a small increase in GS model accuracy that incorporated multiple traits. While both examples show enhanced predictive ability for the multi-trait model, it is difficult to ascertain how the models would perform in a large-scale wheat breeding program, particularly because of the large use of simulated data.

Recent work by Rutkoski et al. (2016) evaluated how to incorporate vegetation indexes and CT measured on a plot level by aerial hyperspectral and thermal imagery into GS models. Their findings showed that incorporation of phenotypic traits could improve prediction accuracy by as much as 70% compared to using univariate models alone when using a multivariate genomic best linear unbiased predictor (GBLUP). Their work shows the promise of combining genomic selection and HTP measurements. Thus, we examined how to 1) best incorporate dynamic HTP data into different GS models through comparison of single and multi-trait GS models in abiotic stress environments based on model prediction accuracy, 2) utilize a single or multiple different sensors for HTP measurements in GS models, and 3) assess the potential for using ground-based high-throughput phenotyping platforms for selection decisions large breeding nurseries.

Materials and Methods

Field Trial Design and Management

The wheat breeding program at the International Maize and Wheat Improvement center (CIMMYT) evaluates elite lines for grain yield performance across multiple environments in Campo Experimental Norman E. Borlaug (CENEB), Ciudad Obregon, Sonora, Mexico. We utilized three of these elite yield experiments conducted under drought and high temperature stress conditions in 2014 and 2015 for our study. Each year, the experiments consisted on 1092 advanced lines, which were further subdivided into 39 trials to manage the spatial variability. Each trial had three replications of 28 entries and two checks that were sown in an a lattice design with five subblocks and six entries per block. The drought environments were sown with a plot size of 4 m × 1.3 m and received 180 mm of irrigation through drip irrigation throughout the growing season. In the high temperature stress environment, the trials were sown on raised beds of 80 cm with a plot length of 2.8 m. For clarity within this study, environment refers to the drought or heat stress imposed experimental conditions, experiment refers to each year-environment combination (three experiments; 2014 Drought, 2015 Drought, and 2015 Heat) and trials refer to subsets of plots within each experiment that contained three replicates of breeding lines arranged in an a lattice designs.

Phenotypic Data Collection

Plant height and days to heading (Zadok's Growth Stage 55, Zadoks et al., 1974) were collected in all plots, and days to maturity (Zadok's Growth Stage 87, Zadoks et al., 1974) was

only collected in the first replicate. Grain yield was determined at maturity by combine harvesting the whole plot.

We used the Phenocart (Crain et al., 2016) to collect normalized difference vegetation index (NDVI) and CT from heading until physiological maturity (Table 1). The Phenocart collects georeferenced NDVI and CT data, and after assigning data to plots resulted in an average of 34 subsample measurements per plot for each sample date. To collect data from all of the plots required five to seven hours depending on walking speed. Because of the time limitation, most often the entire experiment area was taken over the course of two consecutive days for a given time point during the growing season. To efficiently record data, the Phenocart was pushed along columns of plots rather than following individual trial layout (Fig. 1). With the GreenSeeker being an active sensor, the fluctuations in NDVI values due to changing ambient conditions were insignificant during the course of the day (Kipp et al., 2014). Canopy temperature, however, was corrected as described below to compensate for ambient temperature changes resulting from the data collection pattern as noted in the data processing section.

HTP Data Processing and Analysis

For each experiment, we collected three to four sets of observations from heading until physiological maturity. For each time point of data collection, we processed the data as follows: 1) assigned data to a single plot using methods similar to Crain et al. (2016), 2) calculated heritability for each individual trial for each measurement day, and 3) calculated genotype best linear unbiased predictors (BLUPs).

Broad sense heritability (H^2) was calculated as the ratio of the genetic variance to the phenotypic variance, also known as repeatability (Piepho and Möhring, 2007). A high H^2 is indicative of higher precision, and has been related to higher predictive ability for secondary traits correlated to grain yield (Crain et al., 2017). We computed heritability for each individual trait (e.g. single trait on given date) on a trial basis as (Fehr, 1987):

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\epsilon^2}{r} + \frac{\sigma_\delta^2}{rs}} \quad [1]$$

where σ_g^2 is the genotypic variance, σ_ϵ^2 is the variance between plots, and σ_δ^2 is the residual variance (within plot variance), r is the number of replications and s is the number of subsamples per plot. For traits with only a single measurement (e.g., grain yield) the model simplifies to:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\epsilon^2}{r}} \quad [2]$$

because there is no variance attributed to subsampling.

Because CT can be affected by ambient temperature (Pask et al., 2012), we assessed several methods with which to compensate for this observed problem. We

Table 1. Days of phenotypic observation, Ciudad Obregon, Mexico, for normalized difference vegetation index (NDVI) and canopy temperature (CT) for each experiment along with the total number of collected data points and average number of data per plot. The average number is sum of NDVI and CT data combined.

Experiment	Type of observation	First observation	Second observation	Third observation	Fourth observation	Total observations
2014-Drought	Date	15–16 Feb.	22–23 Feb.	28 Feb.– 1 Mar.	NA	
	Number of data points	109,545	107,520	135,154		352,219
	Average data per plot (subsamples)	31	37	39		
2015-Drought	Date	20–21 Feb.	26 Feb.	5–6 Mar.	NA	
	Number of data points	110,699	142,795	140,490		393,984
	Average data per plot (subsamples)	37	41	40		
2015-Heat	Date	23–24 Apr.	30 Apr.	7 May	18–19 May	
	Number of data points	95,223	98,186	98,796	101,844	394,049
	Average data per plot (subsamples)	27	28	28	29	

tested spatial correction of the raw data by 1) normalized by pass of data collection, 2) modeling the data with a time of observation covariate, and 3) modeling data with a covariate for days to heading. For each of these correction approaches we assessed heritability of the corrected data compared to the heritability of the raw data. We chose the method with CT data normalized by pass of data collection as this method gave the highest heritability (repeatability) of CT data which suggested that ambient environmental fluctuation had been minimized. After performing the data normalization, the transformed data was used to calculate BLUPs.

Best linear unbiased predictors for each genotype were calculated by fitting the following mixed model to the a lattice design for each trial as:

$$y_{ijkl} = \mu + a + g_i + r_j + b_{k(j)} + \epsilon_{ijk} + \delta_{l(ijk)} \quad [3]$$

where y_{ijkl} is the phenotype observation for the trait of interest, μ is the overall mean of the population, a is a fixed effect covariate for days to heading, g_i is the random genotype effect of entry i distributed as iid $g_i \sim N(0, \sigma_i^2)$, r_j is the random effect for the j replication distributed as iid $r_j \sim N(0, \sigma_j^2)$, $b_{k(j)}$ is the random effect for the k block nested within replication distributed as iid $b_{k(j)} \sim N(0, \sigma_k^2)$, ϵ_{ijk} is the line residual distributed as iid $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$, and $\delta_{l(ijk)}$ is the l subsample nested within i genotype, j replicate, k block and is the model error variance distributed as iid $\delta_{l(ijk)} \sim N(0, \sigma_l^2)$. For traits measured without subsamples (e.g., yield), the term $\delta_{l(ijk)}$ is removed from the model. Missing phenotypic data was imputed with an expectation maximization algorithm using the Amelia R package (Honaker et al., 2011) and Pearson correlations between BLUPs for HTP traits and grain yield were calculated.

Genotypic Information

Wheat lines were genotyped using GBS following protocols by Poland et al. (2012). Single nucleotide polymorphisms (SNPs) were called using the TASSEL GBSv2 pipeline (Glaubitz et al., 2014) using the Chinese Spring reference genome (International Wheat Genome Sequencing Consortium, 2014). Initial SNP calling resulted in 2079 individuals from the three experiments

and 19,583 markers. Individuals with more than 85% missing data were excluded from the final data set. Filtering the genetic loci consisted of: 1) excluding markers with MAF less than 0.05, 2) excluding markers with greater than 5% heterozygosity, 3) excluding markers with more than 30% missing data. We imputed the data set with Beagle version 4.1 (Browning and Browning, 2016) and then removed markers that were in complete linkage disequilibrium with another marker. After filtering, the final genotypic marker set included 2031 individuals and 2254 markers.

Genomic Prediction

We evaluated genomic prediction for grain yield using several statistical models. A univariate-single trait model (uniGS), a model using only HTP traits as predictors (HTPr), a genomic model with phenotypic covariates (GS+HTP), and a multi-trait model that included grain yield, NDVI and CT for responses (multiGS) were fit for each experiment individually. The GS models were evaluated using partial least squares regression (PLSR), elastic net (EN), and GBLUP. The R packages pls (Mevik et al., 2013), glmnet (Friedman et al., 2010), and asreml (Butler, 2009) were used to fit the PLSR, EN, and GBLUP models, respectively. The EN and PLSR models used functions that were adapted for the GSwGBS package (Gaynor, 2015). A detailed review of the mathematical models for GS is given by Lorenz et al. (2011) and Heslot et al. (2012). All models were fit in a two-step process, first fitting BLUPs for each genotype and trait, and second fitting the GS model with the calculated BLUPs and genomic markers.

The general form that markers and predictors entered the models are: uniGS—A univariate formula for grain yield

$$GY = \mu + \mathbf{Z}u + \epsilon \quad [4]$$

where GY is the BLUP for grain yield from the mixed model Eq. [3], μ is the overall mean, \mathbf{Z} is an $(n \times m)$ matrix assigning markers to genotypes and u is a $(1 \times n)$ array of random effects of markers, and ϵ is the residual error.

The HTPr (HTP with multiple regression)—A univariate model predicting grain yield with only the measured HTP traits.

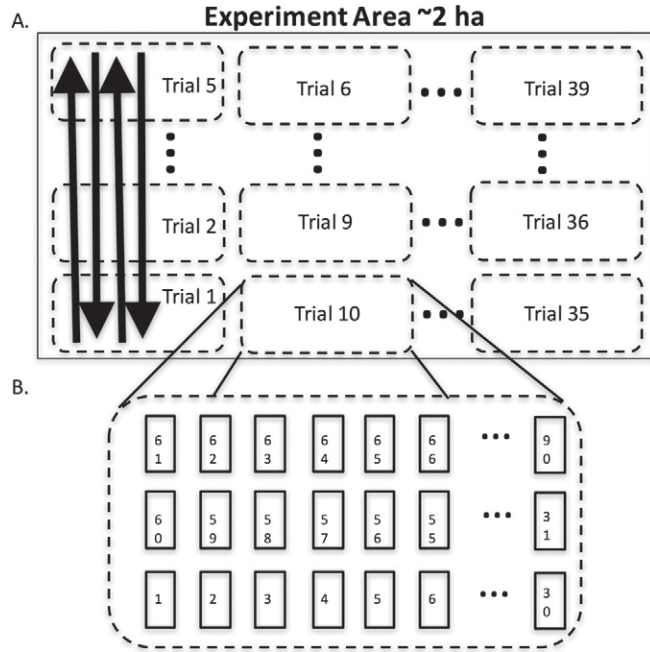


Fig. 1. Example field layout and design of each experiment area (2014 and 2015 Drought, and 2015 Heat). Panel A represent the entire experiment area which was approximately two ha in size which was subdivided into 39 individual trials that each had 30 unique entries in an a lattice design. Bold, solid lines with arrows in the first 5 trials represent the pattern of data collection, with CT data normalized based on each pass (each bold arrow is a unique pass) for all trials. Panel B is an enlarged picture of a single trial that includes 90 plots (30 genotypes replicated 3 times) in an a lattice design.

$$GY = \mu + X\beta + \epsilon \quad [5]$$

where HTP traits are fixed effects represented by X , which is a matrix of individual observations for each time of CT and NDVI measurements and β are the fixed effect of HTP measurements.

GS+HTP—A univariate model for grain yield that included covariate predictor traits.

$$GY = \mu + X\beta + Zu + \epsilon \quad [6]$$

which combines Eq. [4] (genetic markers as random effects) with Eq. [5] (phenotypic observations as fixed effects).

MultiGS—A multivariate model predicting grain yield along with HTP traits NDVI and CT.

$$\begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} X & 0 \\ \vdots & \vdots \\ 0 & X_n \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} Z & 0 \\ \vdots & \vdots \\ 0 & Z_n \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad [7]$$

where n is the number of traits (grain yield, CT, and NDVI), t_i is a vector of BLUPs that have been predicted using Eq. [3]. Matrix X is a design matrix of fixed effects, which simplifies to a vector of 1 for each trait representing the mean as only the markers were entered in this equation. The Z matrix represents the random

marker matrix for each trait, with each trait also having a random error term represented by $\epsilon_{1..n}$.

In addition to fitting predictor traits for each individual day, BLUPs for NDVI and CT were averaged across the different time points of measurement as this reduced the computational time in fitting the multivariate models. For the multivariate GBLUP, this resulted in fitting the model as

$$\begin{bmatrix} t_{gy} \\ t_{CT_AVG} \\ t_{NDVI_AVG} \end{bmatrix} = \begin{bmatrix} X & 0 \\ \vdots & \vdots \\ 0 & X_n \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} Z & 0 \\ \vdots & \vdots \\ 0 & Z_n \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad [8]$$

where t_{gy} , t_{CT_AVG} , and t_{NDVI_AVG} are predicted grain yield, averaged CT, and averaged NDVI respectively, with average CT and NDVI consisting of three to four measurement dates. This compares to using each individual HTP observation as:

$$\begin{bmatrix} t_{gy} \\ t_{CT_1} \\ \vdots \\ t_{CT_4} \\ t_{NDVI_1} \\ \vdots \\ t_{NDVI_4} \end{bmatrix} = \begin{bmatrix} X & 0 \\ \vdots & \vdots \\ 0 & X_n \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} Z & 0 \\ \vdots & \vdots \\ 0 & Z_n \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad [9]$$

where each $t_{Trait_observation}$ corresponds to each measurement date. Using Eq. [8] made the model both feasible and significantly faster to compute, in comparison to fitting each observation date. Along with fitting models using both NDVI and CT, we also fit the same models with only NDVI or CT to compare the effect of individual predictor traits.

To evaluate the accuracy of the GS models, we used cross-validation where the data from each experiment was divided by individual trials within the experiment. This division followed from the experimental design and represented 39 trials and subsequently a 39 fold cross-validation. The most closely related lines (e.g., full-sibs) are arranged within the same trial, thus cross-validation on trials presents a realistic application of GS when predicting into new sets of breeding lines. If a random sample of genotypes was used for the cross-validation, it is likely genetically similar lines would be used to predict line performance possibly biasing the results. Thirty-eight of the divided data sets were used to train the prediction model, and from the trained model predictions were made on the final fold. The cross-validation was repeated for each trial predicting all genotypes one time.

Along with developing GS models within each experiment we evaluated the models for prediction across experiments, using each experiment to predict the other experiments. For the prediction across experiments, the cross-validation proceeded similarly to within experiment evaluation, however, if the same

genotype was in the training and testing set, the line was removed from the training set before prediction. This ensured that no individual line would contribute to predicting its own performance.

For the MultiGS model, grain yield was masked from the prediction set similar to the methods in Rutkoski et al. (2016). For this model, CT and NDVI entered the model for all genotypes, with only the prediction set having a masked grain yield value. In application, this would represent a breeding program that had CT and NDVI collected throughout the growing season, but had not yet harvested the plants, and could select and harvest plants based on prediction values. The prediction accuracy was determined by Pearson correlation coefficient between the observed values and the cross-validated genomic estimated breeding value across all folds of cross-validation. We present the correlations directly and did not divide the correlation coefficient by heritability, as this introduces additional error from the heritability calculation (Heslot et al., 2012). All data sets and scripts are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.7f138>.

Results

Phenotypic Data Collection and Processing

Using the Phenocart (Crain et al., 2016), we generated dense phenotypic data on three experiments. Over the growing season we collected nearly 400,000 data points for each experiment with three or four observations per growing season (Table 1). Many of the observation days for CT had clear patterns that were highly associated with the time of measurement (Fig. 2a). The CT data was normalized to account for the ambient environment effects resulting in the average heritability for CT across all trials increasing from 0.34 to 0.55 (Fig. 3) with a range of H^2 from 0.00 to 0.91 for the normalized data. Additionally, we observed a negative correlation with grain yield which is consistent with previous studies (Balota et al., 2008; Crain et al., 2016), further confirming the data transformation was suitable (Fig. 4). The range of H^2 for NDVI was 0.00 to 0.99 with the average H^2 over all trials at 0.81 (Supplemental Table S1). The overall high heritability of NDVI indicated that these measurements were less influenced by the ambient conditions than CT.

Genomic Prediction Accuracy within Experiments

To effectively utilize the HTP data both in terms of data collection and computational efficiency, we fit the GS models in two ways: 1) with the HTP BLUPs averaged by trait for all measurements (Eq. [8]) and 2) each individual HTP measurement (Eq. [9]). Using the averaged HTP traits (Table 2) resulted in predictions that were similar to, or higher than, using all data in the prediction for most cases except for the HTP_r models (Supplemental Table S2). Even though there were large differences for HTP_r models that used averaged data or all observations, on building more complex models through

incorporation of genotypic markers model performance between average data and all data converged to similar prediction accuracy. Based on the fact that complex models including genotypic effects were similar in accuracy with averaged and all data, we used the averaged data for all subsequent analysis.

Across the three experiments there were large differences in prediction accuracy among and within experiments (Fig. 5). We used three distinct models (PLS, EN, GBLUP) for genomic prediction and assessed their accuracy using cross-validation with accuracy reported as the Pearson correlation coefficient between BLUP for grain yield and the predicted value for grain yield. The prediction models tended to perform similarly across experiments with EN and GBLUP consistently outperforming PLSR models. With the exception of the PLSR models, the addition of HTP traits to the genomic models tended toward higher accuracy although not statistically different than the univariate GS model. The GBLUP models had the highest increase in accuracy when HTP data was incorporated in the model. The multiGS GBLUP resulted in an average of 7% increase in accuracy compared to marker (uniGS) selection across the three experiments with the increase ranging from -5% to 20%. The EN methods showed positive increases in accuracy prediction compared to univariate GS, and GBLUP results varied with the particular model used, although none of the predictions were statistically different from the univariate GS model. The HTP_r models were the most variable from providing the highest prediction for the Heat 2015 experiment to having negative correlations for the Drought 2015 experiment (Fig. 5).

Genomic Prediction Accuracy of Single Predictors

Along with fitting the GS models with both NDVI and CT, we fit the models using each predictor trait (NDVI or CT) by itself. Across all experiments, the data followed similar trends to fitting the full models with both NDVI and CT. The PLSR models had lower predictive ability, and HTP_r models show more variability than the more complex GS models. For the GS+HTP models there was a slight trend that CT provided higher predictive ability than NDVI, although this was within the confidence intervals for the models. Excluding PLSR, the multiGS models with both NDVI and CT provided equal or better predictive power over the univariate models across all environments with an average 3% gain for each trait (Table 2). Within each environment the accuracies were not statistically different and ranged from -5 to 20% of the univariate models.

Genomic Prediction Accuracy across Experiments

In addition to examining within experiment predictions, we also assessed how across environment predictions would be influenced by the addition of HTP data. We used each experiment to predict the other experiments for all models (Fig. 6). For all across experiment predictions, the accuracies were low with $r = 0.1$ as the highest average for any set of models across the six prediction sets. The HTP_r models were much more inconsistent providing the highest predictions

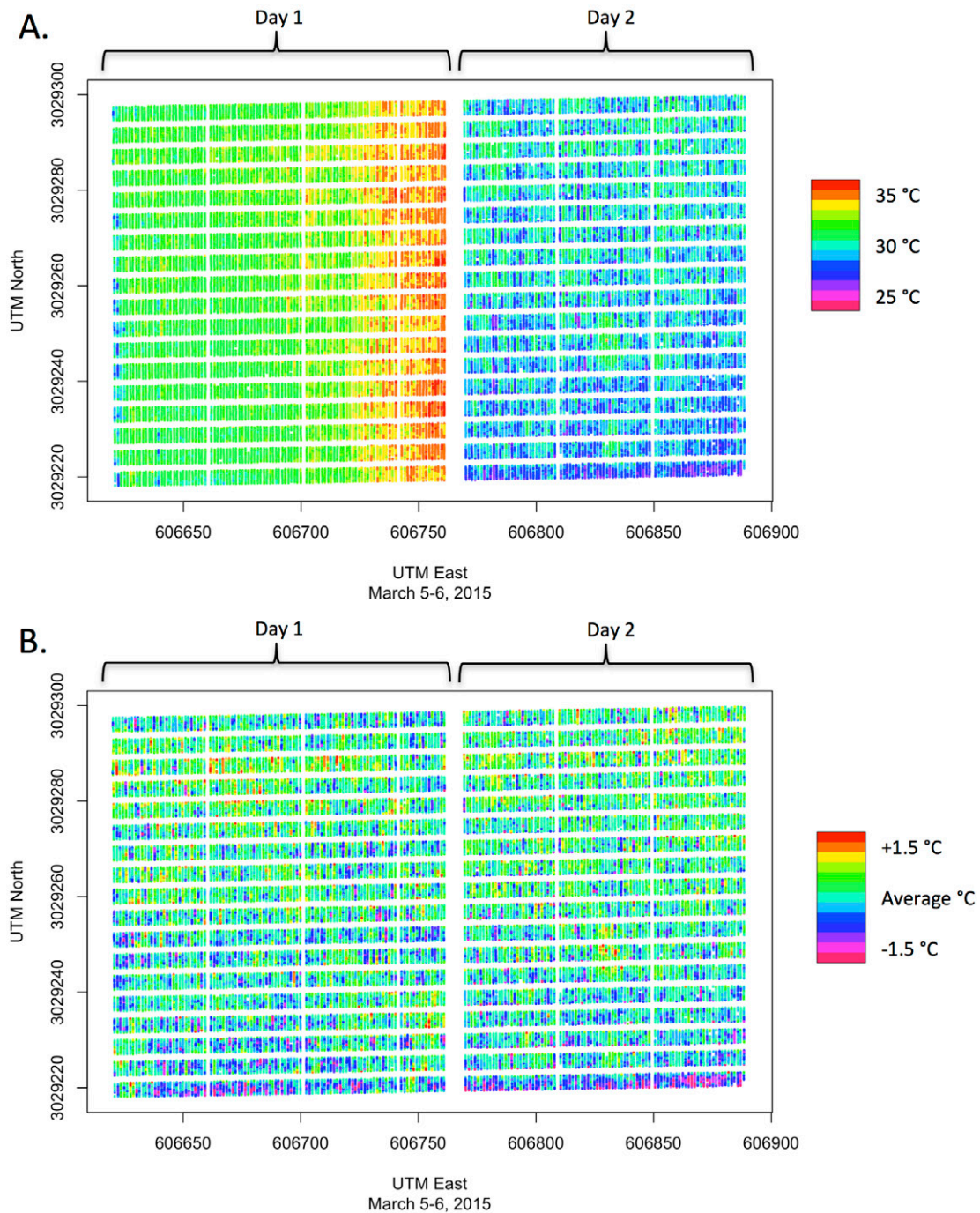


Fig. 2. Canopy temperature (CT) data for 3510 plots in the Drought 2015 experiment color coded by temperature reading. Panel A shows the raw CT data that was obtained collecting data column by column (Fig. 1). The data collection was over 2 d, and there are clear differences in the right and left portions showing environmental differences due to day as well as daily changes. Day 1 shows an increasing temperature progression from left to right corresponding to increasing temperatures over the period of time during measurement. Panel B is the temperature differences normalized by column. The inter- and intra-daily gradients have been effectively removed, along with the minor spatial patterns, there are also differences among genotypes.

for three of the data sets and then having low or even negative correlation for the other three data sets. While the accuracy varied among prediction sets, models using genetic data (uniGS, GS+HTP, multiGS) tended to consistently provide positive predictions. The PLSR models tended to have lower accuracies than the EN and GBLUP models continuing the general trend observed within experiment predictions.

Discussion

Phenotypic Data Collection and Processing

With the development of effective field-based HTP platforms (Busemeyer et al., 2013a; Andrade-Sanchez et al., 2014; Barker et al., 2016; Crain et al., 2016) scientists have access to more data than ever before to predict line

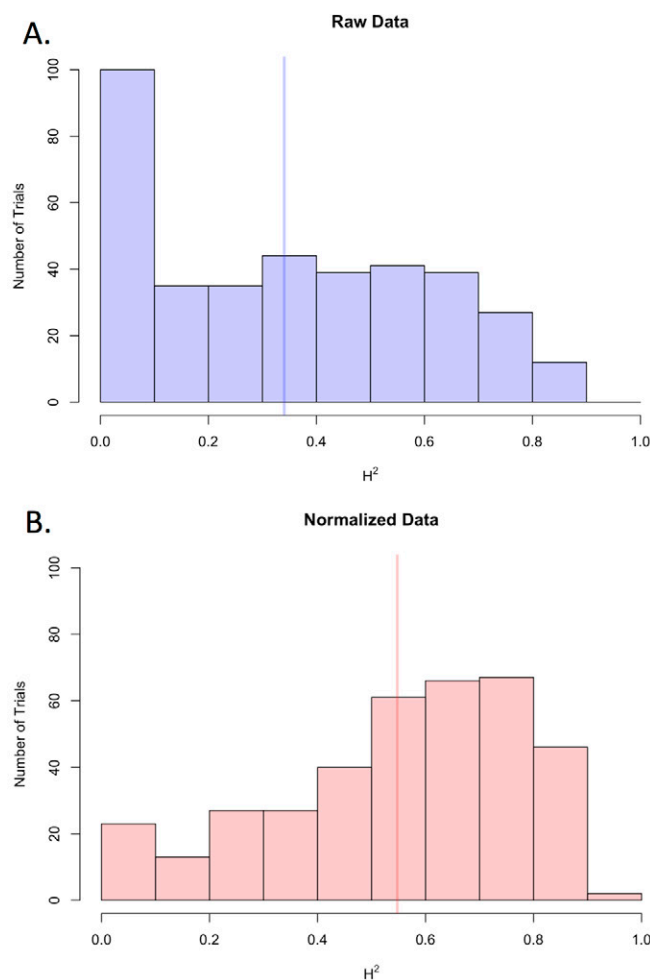


Fig. 3. Broad sense heritability (H^2) for canopy temperature (CT) for each individual trial. Raw data is shown in blue with a vertical line at 0.34 representing the mean, Panel A. Normalizing the data resulted in much higher H^2 values with a mean of 0.55 shown by the red vertical line, Panel B.

performance. The most effective utilization of this data remains challenging, in particular for large field trials where daily environmental changes can occur throughout data collection. Due to the length of time to collect the data and the large data sets obtained we investigated how best to use CT as it is influenced by diurnal environmental conditions. We evaluated several models to remedy this problem, and we choose to normalize data by pass of data collection (Fig. 1) as this model gave the highest broad-sense heritability across all trials. This normalization aggregated small time periods (1–3 minutes) of data collection during which ambient changes should be minimal and then normalized the data so that observations were deviations from the mean. This effectively compensated for nonrandom diurnal temperature variation in CT (Fig. 2a and 2b). The NDVI measurements from each trial were used without correction, as the GreenSeeker was an active sensor which limits effects due to environmental variation (Solari et al., 2008), and the observed values changed little within the day or from 1 d to the next. Based on our data, it appears that sensor

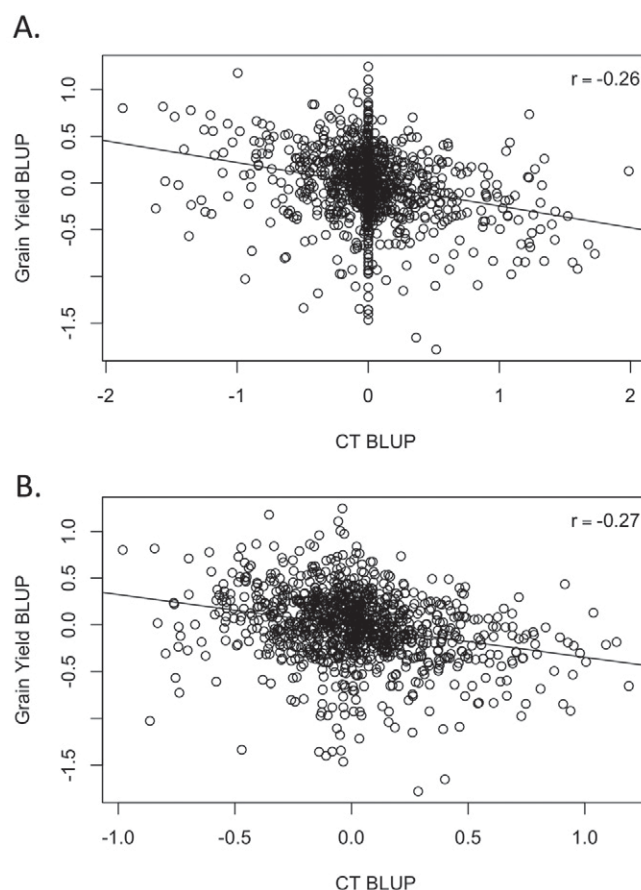


Fig. 4. Pearson relationship between best linear unbiased predictors (BLUPs) for canopy temperature (CT) and grain yield using BLUPs from the Heat 2015 experiment, measured April 23 to 24 ($n = 1058$). BLUPs calculated from raw CT data is shown in Panel A, and from normalized data in Panel B. Trend line is line of best fit with correlation included in the panels.

readings can be standardized to better reflect the true biological information that is present in the field. While we have used a normalization technique, other sensors may require a different type of calibration. Nevertheless, this finding is useful for scientists trying to collect dynamic measurements over large field trials.

Genomic Prediction

We evaluated several different GS models to identify the best approaches to leverage genomic information and HTP measurements. As GS can be applied to enhance plant breeding by reducing the time per breeding cycle (Meuwissen et al., 2001; Heffner et al., 2009) and HTP data can increase the accuracy of phenotypic selection and training models for GS prediction (Cobb et al., 2013), the combination of GS and HTP should translate to higher genetic gains in the field. Overall, we found several trends including 1.) the addition of HTP data increased, or was equal to (e.g., not significantly lower than), GS model performance. 2.) environments that had low heritability for grain yield (Drought 2015 average $H^2 = 0.68$) had worse prediction performance than environments with higher heritability for grain yield (Drought 2014 and

Table 2. Genomic prediction and 95% confidence intervals (CI) for prediction accuracies for three different genomic selection (GS) models, in three experiments and two HTP sensors, fit with average normalized difference vegetation index and canopy temperature best linear unbiased predictors (BLUPs) from all observations. Accuracy is given as Pearson correlation between BLUPs for grain yield and the observed value for grain yield. uniGS is a univariate model with only genetic markers, HTP_r is a univariate model for grain yield predicted with only high-throughput phenotyping traits, GS+HTP is a univariate model for grain yield with high-throughput phenotyping traits as covariates, and multiGS is a multi-response GS model with both grain yield and high throughput-phenotyping traits as responses.

Method	Model	Sensor	2015 Heat		2015 Drought		2014 Drought		Avg. accuracy	Increase from HTP trait†
			r	95% CI	r	95% CI	r	95% CI		
Partial Least Squares Regression	GS		0.22	0.16–0.27	0.17	0.12–0.23	0.43	0.38–0.48	0.27	1
		Both	0.30	0.24–0.35	-0.04	-0.10–0.02	0.31	0.26–0.37	0.19	0.70
		NDVI	0.31	0.25–0.36	-0.15	-0.21–0.09	0.03	-0.03–0.09	0.06	0.23
	GS+HTP	CT	0.19	0.13–0.25	-0.01	-0.07–0.05	0.29	0.23–0.35	0.16	0.57
		Both	0.22	0.16–0.28	0.16	0.10–0.22	0.42	0.36–0.47	0.27	0.98
		NDVI	0.23	0.17–0.28	0.17	0.11–0.23	0.43	0.37–0.48	0.28	1.01
	multiGS	CT	0.24	0.18–0.29	0.19	0.13–0.24	0.43	0.37–0.48	0.29	1.05
		Both	0.20	0.14–0.26	0.13	0.07–0.19	0.23	0.17–0.29	0.19	0.68
		NDVI	0.14	0.08–0.20	0.18	0.12–0.24	0.23	0.17–0.29	0.18	0.67
	GS	CT	0.20	0.14–0.26	0.21	0.15–0.26	0.39	0.34–0.44	0.27	0.98
		Both	0.21	0.15–0.27	0.21	0.15–0.27	0.45	0.40–0.50	0.29	1
		NDVI	0.31	0.25–0.36	-0.10	-0.16–0.04	0.03	-0.04–0.09	0.08	0.28
Elastic Net	GS+HTP	CT	0.19	0.13–0.25	-0.07	-0.13–0.01	0.29	0.23–0.35	0.14	0.47
		Both	0.22	0.16–0.28	0.22	0.16–0.28	0.45	0.39–0.50	0.30	1.02
		NDVI	0.21	0.15–0.26	0.22	0.17–0.28	0.44	0.39–0.49	0.29	1
	multiGS	CT	0.23	0.17–0.29	0.23	0.17–0.28	0.45	0.39–0.50	0.30	1.05
		Both	0.22	0.16–0.27	0.23	0.17–0.29	0.45	0.40–0.50	0.30	1.03
		NDVI	0.21	0.16–0.27	0.22	0.16–0.27	0.44	0.38–0.49	0.29	1
	GS	CT	0.23	0.17–0.29	0.23	0.17–0.28	0.44	0.39–0.49	0.30	1.03
		Both	0.24	0.18–0.29	0.20	0.14–0.26	0.45	0.40–0.50	0.30	1
		NDVI	0.31	0.25–0.36	-0.03	-0.09–0.03	0.31	0.26–0.37	0.19	0.65
	HTPr	CT	0.19	0.13–0.25	-0.01	-0.07–0.05	0.29	0.23–0.35	0.16	0.53
		Both	0.22	0.16–0.27	0.20	0.14–0.26	0.44	0.39–0.49	0.29	0.97
		NDVI	0.22	0.16–0.27	0.20	0.14–0.26	0.44	0.39–0.49	0.29	0.97
GBLUP	multiGS	CT	0.23	0.17–0.28	0.20	0.14–0.26	0.44	0.39–0.49	0.29	0.98
		Both	0.29	0.23–0.34	0.19	0.13–0.25	0.47	0.42–0.52	0.32	1.07
		NDVI	0.29	0.23–0.34	0.20	0.14–0.25	0.45	0.40–0.50	0.31	1.06
	GS+HTP	CT	0.25	0.20–0.31	0.19	0.14–0.25	0.47	0.42–0.52	0.30	1.02
		Both	0.24	0.18–0.29	0.20	0.14–0.26	0.45	0.40–0.50	0.30	1
		NDVI	0.31	0.25–0.36	-0.15	-0.21–0.09	0.03	-0.03–0.09	0.06	0.21
	GS	CT	0.19	0.13–0.25	-0.01	-0.07–0.05	0.29	0.23–0.35	0.16	0.53
		Both	0.22	0.16–0.27	0.20	0.14–0.26	0.44	0.39–0.49	0.29	0.97
		NDVI	0.22	0.16–0.27	0.20	0.14–0.26	0.44	0.39–0.49	0.29	0.97
	HTPr	CT	0.23	0.17–0.28	0.20	0.14–0.26	0.44	0.39–0.49	0.29	0.98
		Both	0.29	0.23–0.34	0.19	0.13–0.25	0.47	0.42–0.52	0.32	1.07
		NDVI	0.29	0.23–0.34	0.20	0.14–0.25	0.45	0.40–0.50	0.31	1.06

†Increase from HTP accuracy is as the average model across the three environments, with increase or decrease of HTP methods considered as a deviation from the GS model.

Heat 2015, $H^2 = 0.94$ and $H^2 = 0.82$ respectively), and 3.) the HTP_r models exhibited the most fluctuation in model performance from highest accuracy in one experiment to no accuracy or even negative correlations in another. To try and understand this range of differences, we observed the correlation between BLUPs for grain yield and HTP traits, Table 3 and Supplemental Table S3. For NDVI relationship to grain yield in Drought 2015, the average of the three dates had a correlation of $r = 0.01$; however, by individual date the correlation ranged from -0.20 to 0.37. This trend was also seen with Drought 2014, but not in the Heat 2015 environment. Understanding the physiological factors leading to this contrast could help with improved genotype-by-environment predictions and predictions into new environments.

While the PLSR and EN models were computationally efficient, the GBLUP methods required much more computation time, thus we fit each model with all of the HTP traits and the same models with the average of the HTP traits, Eq. [9] and [8], respectively. While the models utilizing genetic data did not change significantly from using averaged or all data, the simple HTP_r models fluctuated greatly in performance. Using all data observations in the HTP_r models may have allowed one or two highly predictive measurement sets to influence the final prediction. This may be best exemplified by the fact that using all HTP data (8 HTP variables + grain yield) resulted in higher HTP_r accuracy for PLSR prediction; however, the multiGS PLSR prediction using averaged data (average NDVI, average CT, and grain yield) had a 0.12 increase in correlation

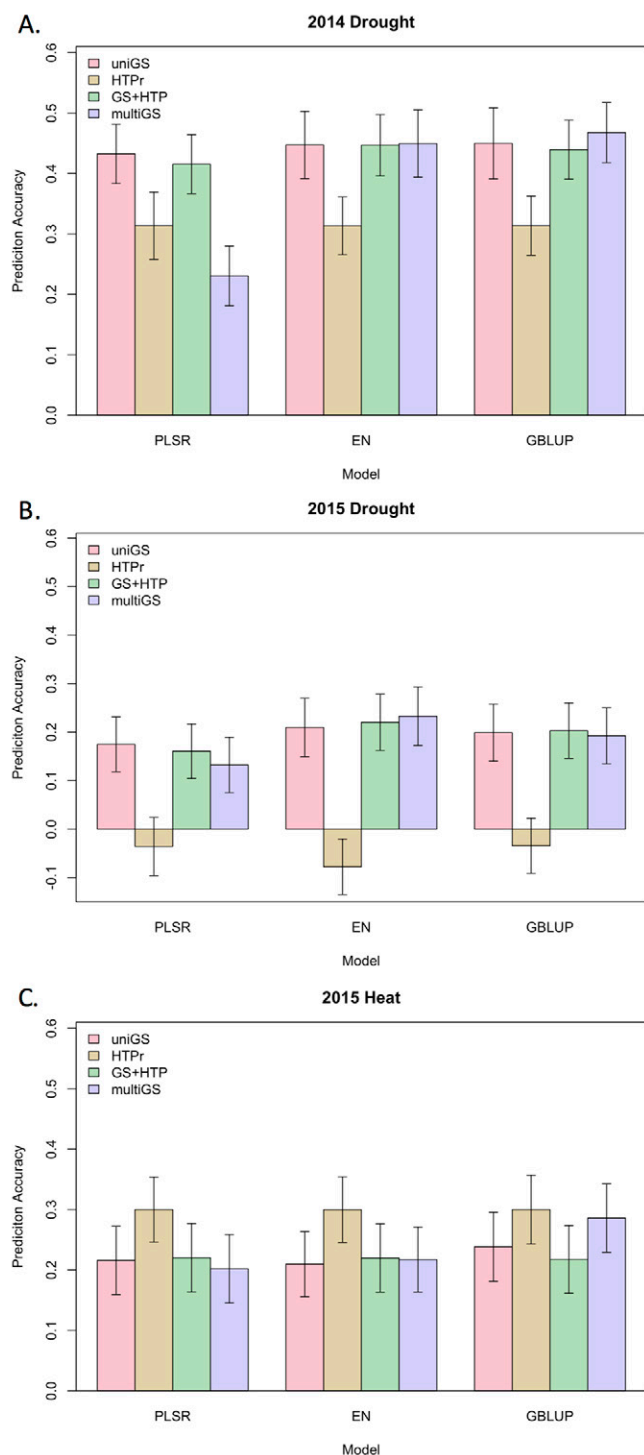


Fig. 5. Performance of genomic selection (GS) across three experiments. Each panel represents one experiment area with the three different GS models. The x axis is grouped by statistical model, PLSR partial least squares regression, EN elastic net, and GBLUP genomic best linear unbiased predictor. The colored bar plots represent four model formulations that are only genetic data (uniGS), only phenotypic data (HTPr), genetic data and phenotypic covariates (GS+HTP), and multi-response for grain yield and HTP (multiGS). Model accuracy are given on the y axis in terms of the correlation coefficient between the predicted best linear unbiased predictor (BLUP) for grain yield and the observed BLUP for grain yield with error bars representing 95% confidence interval.

coefficient accuracy compared to the same model using all data (Table 2 and Supplementary Table 2).

In addition to evaluating model performance, we also examined how individual predictor traits (NDVI or CT) affected model accuracy. This allowed for an assessment of which traits may be better suited for data collection, thus guiding future HTP research and data collection. While including a single trait resulted in gains over univariate models, using both traits together resulted in an average 7% gain for the multiGS GBLUP model (Table 2) suggesting that the HTP traits could act in an additive fashion. These results provide evidence for collecting multiple phenotypic traits if possible within breeding programs.

Finally, we applied across experiment prediction to all environments, representing a realistic application as breeding programs would likely leverage breeding nursery data across years and locations and into new, untested environments. Overall, the prediction models had low accuracy, which would be expected as the introduction of genotype-by-environment interactions increase and were not modeled here. Both the GBLUP and EN multiGS tended to provide a higher accuracy over uniGS models, but the low nature of all predictions would be challenging to apply across environment prediction within a breeding program.

Model Assessment

Genomic selection has the potential to revolutionize plant breeding programs (Desta and Ortiz, 2014). Determining how best to utilize this information in concert is needed to allow breeders to maximize resources and achieve maximum genetic gains. In this study, we have examined several possibilities to incorporate HTP information into genomic selection models, based on our future vision that breeding programs will have genomic profiles available for all breeding lines and that high-throughput phenotyping will become routine. While our findings show that HTP traits can be used to improve model performance over univariate GS models alone, there may be additional ways to increase model accuracy. We have used all data, representative of a breeding program, but application of GS model optimization could ensure that maximum variation is captured in the training set (Isidro et al., 2015) or that low predictive phenotypes are removed from the training set (Rutkoski et al., 2016).

Across environments, the models displayed several trends. Partial least squares regression models usually had the lowest prediction accuracy. The GBLUP models performed the best when using multiGS compared to other models. Model performance may be due to the assumptions that each model inherently makes and how the models utilize the data. For example, partial least squares regression is a model that extracts latent variables and is useful when there are a number of variables and their relationships are not clearly understood (Tobias, 1995). Latent variables are extracted as linear combinations from the original data sets (both predictors and response) and the latent variables are chosen to maximize the predicted response (Lorenz

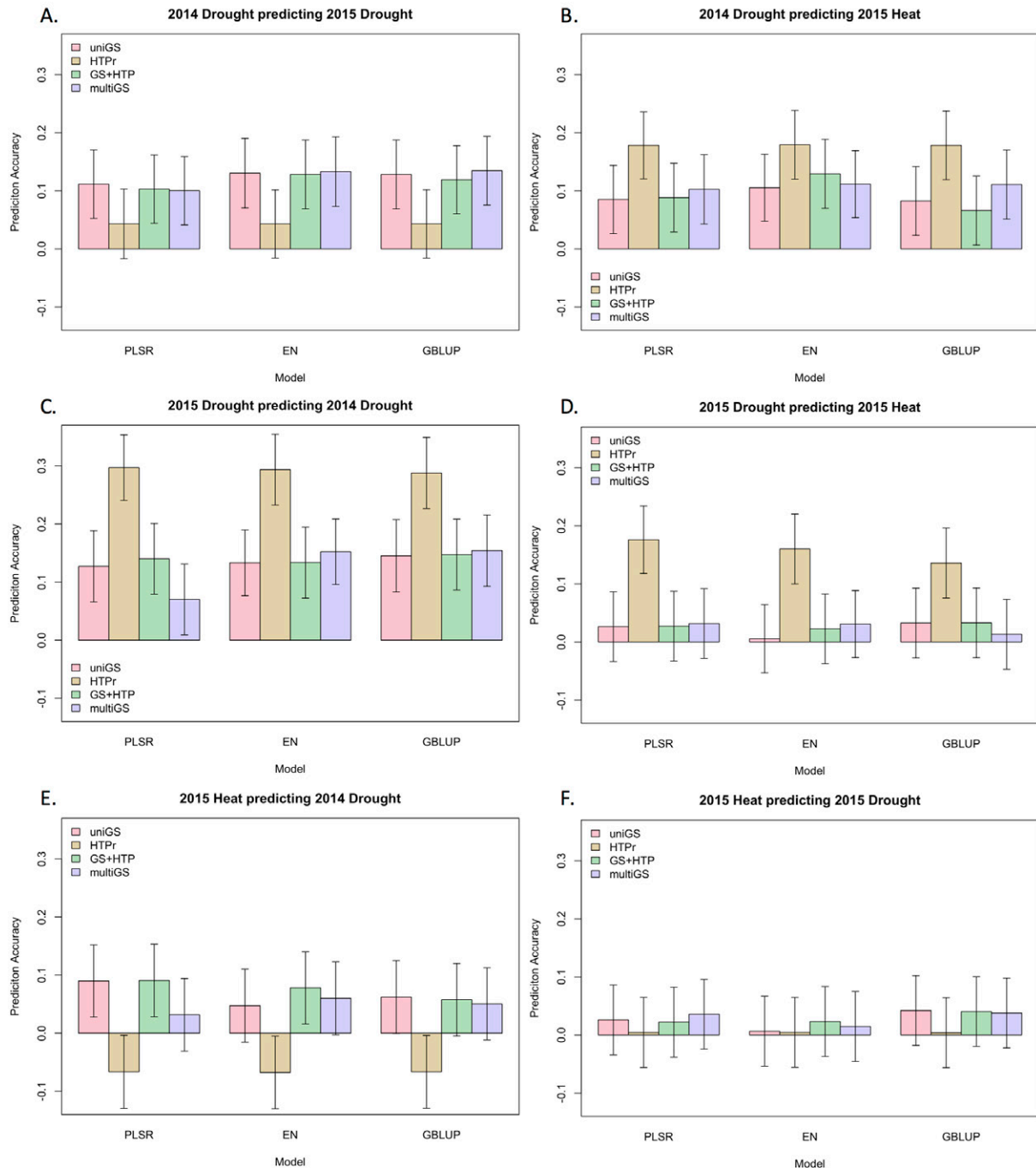


Fig. 6. Performance of genomic selection (GS) models using cross environment prediction, using one environment to predict other environments. Each panel represents one experiment predicted by another experiment, for example Panel A uses the 2014 Drought information to predict the 2015 Drought environment. The x axis is grouped by statistical model, PLSR partial least squares regression, EN elastic net, and GBLUP genomic best linear unbiased predictor. The colored bar plots represent four model formulations that are only genetic data (uniGS), only phenotypic data (HTPr), genetic data and phenotypic covariates (GS+HTP), and multi-response for grain yield and HTP (multiGS). Model accuracy are given on the y axis in terms of the correlation coefficient between the predicted best linear unbiased predictor (BLUP) for grain yield and the observed BLUP for grain yield with error bars representing 95% confidence interval.

et al., 2011). Within our data set, we have variables which are expected to have relationships to yield from previous research (Amani et al., 1996; Gutiérrez-Rodríguez et al., 2004; Babar et al., 2006b), and the GS method like PLSR may be identifying these known associations. Elastic net penalizes the models when the number of predictors (marker set) is much larger than the number of observations

(genotypes). It functions by both selecting variables (markers are dropped from the model) as well as imposing shrinkage on the variables that remain in the model (Zou and Hastie, 2005). Using HTP variables, would allow the EN method to single out a few predictors and drop other terms (markers) while still maintaining or improving model performance over the uniGS models. The EN model could

Table 3. Pearson correlation between best linear unbiased predictors (BLUPs) for grain yield (GY), canopy temperature (CT), and normalized difference vegetation index (NDVI) across three experiments, 2015 Heat, 2015 Drought, and 2014 Drought. The CT and NDVI data have been averaged across collection dates.

	2015 Heat NDVI	2015 Heat CT	2015 Drought GY	2015 Drought NDVI	2015 Drought CT	2014 Drought GY	2014 Drought NDVI	2014 Drought CT
2015 Heat GY	0.31***	-0.20***	0.11***	0.00	0.02	†	-	-
2015 Heat NDVI	1	-0.69***	-0.08**	0.38***	-0.32***	-	-	-
2015 Heat CT		1	0.22***	-0.30***	0.42***	-	-	-
2015 Drought GY			1	0.01	-0.04	-	-	-
2015 Drought NDVI				1	-0.58***	-	-	-
2015 Drought CT					1	-	-	-
2014 Drought GY						1	-0.06*	-0.30***
2014 Drought NDVI							1	-0.22***

* Significant at the 0.05 probability level.

** Significant at the 0.01 probability level.

*** Significant at the < 0.001 probability level.

† Genotypes differed between years preventing comparison of same genotypes.

effectively choose variables that result in the highest prediction using markers or HTP data as relevant. For example, in Fig. 5 Panel A, 2014 Drought, the HTPr model is similar to the uniGS model suggesting that much of the predictive power is included in the HTP traits. In Panel B, 2015 Drought, the HTPr model has no predictive power and thus in the GS+HTP model the information used for prediction is contained in the genetic markers. The GBLUP methods are a popular statistical technique to perform GS using random regression best unbiased prediction (RR-BLUP; Lorenz et al., 2011). In RR-BLUP, the typical least squares estimates are shrunk toward zero using a penalty term that reduces collinearity between predictors but keeps all predictors (Lorenz et al., 2011). Keeping all predictors could account for the difference in EN and GBLUP model performance for the HTP+GS methods (Fig. 5). The implicit assumptions that each method imposes on the data also assert practical limits on the utility of the methods.

Due to computational constraints, we only fit the multiGS model with HTP data that had been averaged across all days; however, its prediction was often the highest or not significantly different from the highest performing models. Application of dimensional reduction techniques such as matrix decomposition (Mrode, 2014) may allow for faster model fitting and for the addition of more traits given that the number of effects (variance and covariance parameters) rise linearly with the addition of more traits and observation time points. As phenotyping programs expand the number of traits measured and the frequency of measurements making efficient computational use of the data will be an active area of research. Currently, the models selected by a breeding program may depend on the use of the predictions. For a breeding program HTPr models may work well for decisions in advancing lines to further generations while providing fast computation. This selection method would be best suited within individual environments as two out of three environments had high HTPr prediction accuracy (Fig. 5). The application of HTPr models is limited to field trials as the model provides no information about the underlying genetic architecture and can only be assessed on the same experiment that

the line per se is evaluated. However, if a breeder wanted to select lines to add in the crossing block GBLUP methods may be more appropriate in representing the potential breeding value based on genomic composition.

Conclusions

Sustaining food production into the future will be challenging with expected population increases and limited availability of land and resources (Hawkesford et al., 2013). To meet the expected increase in demand, a marked increase in genetic gain is needed. We investigated several GS methods and the how best to model the phenotypic information. Utilizing HTP traits can improve model performance, thus enabling more accurate selection of superior breeding lines from larger populations. Regardless of the specific models used for HTP, the incorporation of HTP into prediction models seems effective in increasing selection accuracy. The advances in genotyping combined with efficient phenotyping platforms will continue to push the integration of genomics and phenomics for breeding and genetics. By utilizing both genetic information and phenotypic data, it is expected that breeders will be able to more efficiently identify and select superior higher yielding crop varieties.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

Acknowledgments

This work was supported through the National Science Foundation- Plant Genome Research Program (IOS-1238187) and the US Agency for International Development (USAID) Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051). Field trials at CIMMYT, Mexico were supported by the Durable Rust Resistance in Wheat project funded by Bill and Melinda Gates Foundation (BMGF) and the UK Department for International Development (DFID), Cereal Systems Initiative for South Asia Project funded by BMGF and USAID, and Grain Research and Development Corporation (GRDC), Australia. A large portion of the computing for this project was performed on the Beocat Research Cluster at Kansas State University, which is funded in part by NSF grants CNS-1006860, EPS-1006860, and EPS-0919443. J. Crain was supported through a fellowship from the Monsanto Beachell-Borlaug International Scholars program. We appreciate the comments of two anonymous reviewers that helped to improve the manuscript.

References

- Altshuler, D., V.J. Pollara, C.R. Cowles, W.J. Van Etten, J. Baldwin, L. Linton, and E.S. Lander. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407(6803):513–516. doi:10.1038/35035083
- Amani, I., R.A. Fischer, and M.P. Reynolds. 1996. Canopy temperature depression association with yield of irrigated wheat cultivars in a hot climate. *J. Agron. Crop Sci.* 176:119–129. doi:10.1111/j.1439-037X.1996.tb00454.x
- Andrade-Sanchez, P., M.A. Gore, J.T. Heun, K.R. Thorp, A.E. Carmo-Silva, A.N. French, M.E. Salvucci, and J.W. White. 2014. Development and evaluation of a field-based high-throughput phenotyping platform. *Funct. Plant Biol.* 41(1):68–79. doi:10.1071/FP13126
- Babar, M.A., M. Van Ginkel, A.R. Klatt, B. Prasad, and M.P. Reynolds. 2006a. The potential of using spectral reflectance indices to estimate yield in wheat grown under reduced irrigation. *Euphytica* 150(1–2):155–172. doi:10.1007/s10681-006-9104-9
- Babar, M.A., M.P. Reynolds, M. Van Ginkel, A.R. Klatt, W.R. Raun, and M.L. Stone. 2006b. Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. *Crop Sci.* 46(3):1046–1057. doi:10.2135/cropsci2005.0211
- Balota, M., W.A. Payne, S.R. Evett, and T.R. Peters. 2008. Morphological and physiological traits associated with canopy temperature depression in three closely related wheat lines. *Crop Sci.* 48(5):1897–1910. doi:10.2135/cropsci2007.06.0317
- Barker, J., N. Zhang, J. Sharon, R. Steeves, X. Wang, Y. Wei, and J. Poland. 2016. Development of a field-based high-throughput mobile phenotyping platform. *Comput. Electron. Agric.* 122:74–85. doi:10.1016/j.compag.2016.01.017
- Battenfield, S.D., C. Guzmán, R.C. Gaynor, R.P. Singh, R.J. Peña, S. Dreisigacker, A.K. Fritz, and J.A. Poland. 2016. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* 9(2):1–12. doi:10.3835/plantgenome2016.01.0005
- Beckmann, J.S., and M. Soller. 1986. Restriction fragment length polymorphism and genetic improvement of agricultural species. *Euphytica* 35(1):111–124. doi:10.1007/BF00028548
- Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47(3):1082–1090. doi:10.2135/cropsci2006.11.0690
- Browning, B.L., and S.R. Browning. 2016. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98(1):116–126. doi:10.1016/j.ajhg.2015.11.020
- Busemeyer, L., D. Mentrup, K. Möller, E. Wunder, K. Alheit, V. Hahn, H.P. Maurer, J.C. Reif, T. Würschum, J. Müller, F. Rahe, and A. Ruckelshausen. 2013a. BreedVision—a multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors (Basel)* 13(3):2830–2847. doi:10.3390/s130302830
- Busemeyer, L., A. Ruckelshausen, K. Möller, A.E. Melchinger, K. V Alheit, H.P. Maurer, V. Hahn, E.A. Weissmann, J.C. Reif, and T. Würschum. 2013b. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Sci. Rep.* 3:2442. doi:10.1038/srep02442
- Butler, D. 2009. *Asreml: asreml() fits the linear mixed model*. R package version 3.0. www.vsnl.co.uk. (accessed 29 Jan. 2018).
- Calus, M.P.L., and R.F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. *Genet. Sel. Evol.* 43(1):26. doi:10.1186/1297-9686-43-26
- Campos, H., M. Cooper, J.E. Habben, G.O. Edmeades, and J.R. Schussler. 2004. Improving drought tolerance in maize: A view from industry. *F. Crop. Res.* 90(1):19–34. doi:10.1016/j.fcr.2004.07.003
- Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 2013. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* 126(4):867–887. doi:10.1007/s00122-013-2066-0
- Crain, J.L., M.P. Reynolds, and J.A. Poland. 2017. Utilizing high-throughput phenotypic data for improved phenotypic selection of stress adaptive traits in wheat. *Crop Sci.* 57:648–659. doi:10.2135/cropsci2016.02.0135
- Crain, J.L., Y. Wei, J. Barker, S.M. Thompson, P.D. Alderman, M. Reynolds, N. Zhang, and J. Poland. 2016. Development and deployment of a portable field phenotyping platform. *Crop Sci.* 56(3):965–975. doi:10.2135/cropsci2015.05.0290
- Davey, J., P. Hohenlohe, P. Etter, J. Boone, J. Catchen, and M. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12(7):499–510. doi:10.1038/nrg3012
- Deery, D., J. Jimenez-Berni, H. Jones, X. Sirault, and R. Furbank. 2014. Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy* 4(3):349–379. doi:10.3390/agronomy4030349
- Desta, Z.A., and R. Ortiz. 2014. Genomic selection: Genome-wide prediction in plant improvement. *Trends Plant Sci.* 19(9):592–601. doi:10.1016/j.tplants.2014.05.006
- Eathington, S.R., T.M. Crosbie, M.D. Edwards, R.S. Reiter, and J.K. Bull. 2007. Molecular markers in a commercial breeding program. *Crop Sci.* 47(S3):S154–S163. doi:10.2135/cropsci2007.04.0015IPBS
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5): e19379. doi:10.1371/journal.pone.0019379
- Fehr, W.R. 1987. *Principles of cultivar development: Theory and technique*. Macmillan Publishing Co., New York.
- Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33(1):1–22. doi:10.18637/jss.v033.i01
- Gaynor, R.C. 2015. GSWGBS: An R package genomic selection with genotyping-by-sequencing. *Genomic selection for Kansas wheat*. K-State Res. Exch. Manhattan, KS.
- Glaubitz, J.C., T.M.T. Casstevens, F. Lu, J. Harriman, R.R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9(2):e90346. doi:10.1371/journal.pone.0090346
- Goddard, M.E., and B.J. Hayes. 2007. Genomic selection. *J. Anim. Breed. Genet.* 124:323–330.
- Gutiérrez-Rodríguez, M., M.P. Reynolds, J.A. Escalante-Estrada, and M.T. Rodríguez-González. 2004. Association between canopy reflectance indices and yield and physiological traits in bread wheat under drought and well-irrigated conditions. *Aust. J. Agric. Res.* 55(11):1139–1147. doi:10.1071/AR04214
- Haghighattalab, A., L. González Pérez, S. Mondal, D. Singh, D. Schinstock, J. Rutkoski, I. Ortiz-Monasterio, R.P. Singh, D. Goodin, and J. Poland. 2016. Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods* 12(1):35. doi:10.1186/s13007-016-0134-6
- Hawkesford, M.J., J.L. Araus, R. Park, D. Calderini, D. Miralles, T. Shen, J. Zhang, and M.A.J. Parry. 2013. Prospects of doubling global wheat yields. *Food Energy Secur.* 2:34–48. doi:10.1002/fes3.15
- Heffner, E.L., J. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4(1):65–75. doi:10.3835/plantgenome2010.12.0029
- Heffner, E.L., A.J. Lorenz, J.L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: Gain per unit time and cost. *Crop Sci.* 50(5):1681–1690. doi:10.2135/cropsci2009.11.0662
- Heffner, E.L., M.E. Sorrells, and J.-L.L. Jannink. 2009. Genomic selection for crop improvement. *Crop Sci.* 49(1):1. doi:10.2135/cropsci2008.08.0512
- Heslot, N., H.P. Yang, M.E. Sorrells, and J.L. Jannink. 2012. Genomic selection in plant breeding: A comparison of models. *Crop Sci.* 52(1):146–160. doi:10.2135/cropsci2011.06.0297
- Honaker, J., G. King, and M. Blackwell. 2011. *Amelia II: A program for missing data*. *J. Stat. Softw.* 45(7):1–47. doi:10.18637/jss.v045.i07
- International Wheat Genome Sequencing Consortium. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345(6194):1251788.
- Isidro, J., J.L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M.E. Sorrells. 2015. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* 128(1):145–158. doi:10.1007/s00122-014-2418-4
- Jia, Y., and J.L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192(4):1513–1522. doi:10.1534/genetics.112.144246

- Kipp, S., B. Mistele, and U. Schmidhalter. 2014. The performance of active spectral reflectance sensors as influenced by measuring distance, device temperature and light intensity. *Comput. Electron. Agric.* 100:24–33. doi:10.1016/j.compag.2013.10.007
- Liebisch, F., N. Kirchessner, D. Schneider, A. Walter, and A. Hund. 2015. Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant Methods* 11(1):9. doi:10.1186/s13007-015-0048-8
- Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J.L. Jannink. 2011. Genomic selection in plant breeding: Knowledge and prospects. *Adv. Agron.* 110:77–123. doi:10.1016/B978-0-12-385531-2.00002-5
- Meuwissen, T.H., and M.E. Goddard. 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol.* 33(6):605–634. doi:10.1186/1297-9686-33-6-605
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829.
- Mevik, B.H., R. Wehrens, and K.H. Liland. 2013. pls: Partial least squares and principal component regression. <http://cra>.
- Miller, M.R., J.P. Dunham, A. Amores, W.A. Cresko, and E.A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17(2):240–248. doi:10.1101/gr.5681207
- Morrell, P.L., E.S. Buckler, and J. Ross-Ibarra. 2011. Crop genomics: Advances and applications. *Nat. Rev. Genet.* 13(2):85–96.
- Mrode, R.A. 2014. Linear models for the prediction of animal breeding values. 3rd ed. CABI, Boston, MA. doi:10.1079/9781780643915.0000
- Pask, A., J. Pietragalla, D. Mullan, and M.P. Reynolds. 2012. Physiological breeding II: A field guide to wheat phenotyping. Cimmyt, Mexico City, Mexico.
- Piepho, H.P., and J. Möhring. 2007. Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177(3):1881–1888. doi:10.1534/genetics.107.074229
- Poland, J. 2015. Breeding-assisted genomics. *Curr. Opin. Plant Biol.* 24:119–124. doi:10.1016/j.pbi.2015.02.009
- Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7(2):e32253. doi:10.1371/journal.pone.0032253
- Poland, J., and T. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5(3):92–102. doi:10.3835/plantgenome2012.05.0005
- Ray, D.K., N.D. Mueller, P.C. West, and J.A. Foley. 2013. Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8(6):e66428. doi:10.1371/journal.pone.0066428
- Reynolds, M., J. Foulkes, R. Furbank, S. Griffiths, J. King, E. Murchie, M. Parry, and G. Slafer. 2012. Achieving yield gains in wheat. *Plant Cell Environ.* 35(10):1799–1823.
- Rutkoski, J., J. Poland, S. Mondal, E. Autrique, L. González Pérez, J. Crossa, M. Reynolds, and R. Singh. 2016. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genetics*. 6(9): 2799–2808.
- Solari, F., J. Shanahan, R. Ferguson, J. Schepers, and A. Gitelson. 2008. Active sensor reflectance measurements of corn nitrogen status and yield potential. *Agron. J.* 100(3):571–579. doi:10.2134/agronj2007.0244
- Tobias, R.D. 1995. An introduction to partial least squares regression. *Proc. Ann. SAS Users Gr. Int. Conf.* 20th: 1250–1257.
- Valin, H., R.D. Sands, D. van der Mensbrugghe, G.C. Nelson, H. Ahammad, E. Blanc, B. Boudirsky, S. Fujimori, T. Hasegawa, P. Havlik, E. Heyhoe, P. Kyle, D. Mason-D'Croz, S. Paltsev, S. Rolinski, A. Tabeau, H. van Meijl, M. von Lampe, and D. Willenbockel. 2013. The future of food demand: Understanding differences in global economic models. *Agric. Econ.* 45(1):1–17.
- White, J., P. Andrade-Sanchez, M.A. Gore, K.F. Bronson, T.A. Coffelt, M.M. Conley, K.A. Feldmann, A.N. French, J.T. Heun, D.J. Hunsaker, M.A. Jenks, B.A. Kimball, R.L. Roth, R.J. Strand, K.R. Thorp, G.W. Wall, and G. Wang. 2012. Field-based phenomics for plant genetics research. *F. Crop. Res.* 133:101–112. doi:10.1016/j.fcr.2012.04.003
- White, J.W., and M.M. Conley. 2013. A flexible, low-cost cart for proximal sensing. *Crop Sci.* 53(4):1646–1649. doi:10.2135/cropsci2013.01.0054
- Xu, Y., and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci.* 48(2):391–407. doi:10.2135/cropsci2007.04.0191
- Zadoks, J.C., T.T. Chang, and C.F. Konzak. 1974. A decimal code for the growth stages of cereals. *Weed Res.* 14(6):415–421. doi:10.1111/j.1365-3180.1974.tb01084.x
- Zhong, S., J.C.M. Dekkers, R.L. Fernando, and J.L. Jannink. 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. *Genetics* 182(1):355–364. doi:10.1534/genetics.108.098277
- Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic-net. *J.R. Stat. Soc.* 67(2):301–320. doi:10.1111/j.1467-9868.2005.00503.x