

## RESEARCH

# The Accuracy of Genomic Prediction between Environments and Populations for Soft Wheat Traits

Mao Huang, Brian Ward, Carl Griffey, David Van Sanford, Anne McKendry, Gina Brown-Guedira, Priyanka Tyagi, and Clay Sneller\*

## ABSTRACT

Genomic selection (GS) uses training population (TP) data to estimate the value of lines in a selection population. In breeding, the TP and selection population are often grown in different environments, which can cause low prediction accuracy when the correlation of genetic effects between the environments is low. Subsets of TP data may be more predictive than using all TP data. Our objectives were (i) to evaluate the effect of using subsets of TP data on GS accuracy between environments, and (ii) to assess the accuracy of models incorporating marker  $\times$  environment interaction (MEI). Two wheat (*Triticum aestivum* L.) populations were phenotyped for 11 traits in independent environments and genotyped with single-nucleotide polymorphism markers. Within each population-trait combination, environments were clustered. Data from one cluster were used as the TP to predict the value of the same lines in the other cluster(s) of environments. Models were built using all TP data or subsets of markers selected for their effect and stability. The GS accuracy using all TP data was  $>0.25$  for 9 of 11 traits. The between-environment accuracy was generally greatest using a subset of stable and significant markers; accuracy increased up to 48% relative to using all TP data. We also assessed accuracy using each population as the TP and the other as the selection population. Using subsets of TP data or the MEI models did not improve accuracy between populations. Using optimized subsets of markers within a population can improve GS accuracy by reducing noise in the prediction data set.

M. Huang and C. Sneller, The Ohio State Univ., Ohio Agriculture Research and Development Center, 1680 Madison Ave., Wooster, OH 44691; B. Ward and P. Tyagi, Dep. of Crop and Soil Sciences, North Carolina State Univ., Raleigh, NC 27695; C. Griffey, Dep. of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA 24061; D. Van Sanford, Univ. of Kentucky, 1405 Veterans Dr., Lexington, KY 40546; A. McKendry, Univ. of Missouri, Columbia, MO 65211; G. Brown-Guedira, USDA-ARS Eastern Regional Small Grains Genotyping Lab., Raleigh, NC. Received 26 Oct. 2017. Accepted 23 Apr. 2018. \*Corresponding author (sneller.5@osu.edu). Assigned to Associate Editor Jesse Poland.

**Abbreviations:** AMMI, additive main effects and multiplicative interaction; BLUP, best linear unbiased prediction; EM, expectation maximization algorithm; EP, elite panel population; ERR, Eberhart and Russell regression; FP, flour protein; FY, flour yield; GEBV, genomic estimated breeding value; GEI, genotype by environment interaction; GS, genomic selection; HD, heading date; HGT, plant height; IPC, interactive principal component; LD, linkage disequilibrium; MEI, marker  $\times$  environment interactions; PCA, principal component analysis; QTL, quantitative trait locus; rrBLUP, ridge regression best linear unbiased prediction; SNP, single-nucleotide polymorphism; TP, training population; TW, test weight; VP, validation population; YLD, grain yield; YP, yield panel population.

**GENOMIC SELECTION** (GS) is a marker-assisted selection tool first proposed for animal breeding by Meuwissen et al. (2001). In GS, individuals with existing phenotypic and genotypic data are used as a training population (TP) to build a prediction model. This model is used to calculate the genomic estimated breeding values (GEBVs) of individuals based solely on their genotypic data. Breeders can then select individuals with desirable GEBVs and intermate them to initiate the next breeding cycle. This can greatly reduce the duration of a breeding cycle (Meuwissen et al., 2001; Heffner et al., 2009; Jannink et al., 2010).

Published in *Crop Sci.* 58:2274–2288 (2018).  
doi: 10.2135/cropsci2017.10.0638

© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA  
All rights reserved.

As genotyping costs drastically decrease with the advent of new technology, GS can be cheaper than phenotypic selection (Meuwissen et al., 2001; Goddard and Hayes, 2007; Heffner et al., 2010; Jannink et al., 2010; Elshire et al., 2011) and thus facilitate the use of larger populations of related lines in selections.

Many studies have assessed prediction accuracy of GS in crops (Denis and Bouvet, 2011; Heffner et al., 2011a; Rutkoski et al., 2012; Battenfield et al., 2016). The accuracy of GS has been assessed in wheat (*Triticum aestivum* L.) for many traits including grain yield (YLD), test weight (TW), thousand-kernel weight, quality traits (flour yield [FY], flour protein [FP], grain hardness, grain length, grain width, sodium dodecyl sulfate sedimentation), resistance to Fusarium head blight and stem rust (caused by *Fusarium acuminatum* Ellis & Everh. and *Puccinia graminis* subsp. *graminis* Pers.:Pers, respectively), and trait stability for YLD, TW, plant height (HGT), heading date (HD), and quality traits (Heffner et al., 2011a, 2011b; Rutkoski et al., 2011, 2012, 2014; Poland et al., 2012; Crossa et al., 2016; Hoffstetter et al., 2016; Huang et al., 2016; Michel et al., 2016a).

Most studies in wheat use subsets of TP lines as the validation population (VP) and evaluate GS accuracy through cross-validation (Crossa et al., 2010; Heffner et al., 2011b; Heslot et al., 2012; Poland et al., 2012; Zhao et al., 2014; Arruda et al., 2016; Hoffstetter et al., 2016; Huang et al., 2016; Jiang et al., 2016; Michel et al., 2016a; Rutkoski et al., 2016). Compared with making predictions for different populations, the accuracy of GS will be high when using cross-validation within the same population, as lines in the TP and VP are often tested in the same set of environments. When implementing GS in actual plant breeding, the TP and VP (or selection population) will always be phenotyped in different years or seasons, so the accuracy of GS between environments is a crucial issue. There are some studies that use a separate TP and VP phenotyped in different environments to assess the efficacy of GS. Sallam and Smith (2016) used a VP that consisted of progeny that were directly derived from the TP and reported that GS accuracy for barley (*Hordeum vulgare* L.) yield ranged from 0.36 to 0.66. Michel et al. (2016b) reported the GS accuracy for wheat YLD ranged from 0.14 to 0.75 when TP and VP consisted of different lines grown in different years.

The GS accuracy between environments or populations is affected by the relatedness of the TP to the VP (Asoro et al., 2011; Heslot et al., 2013; Rutkoski et al., 2015; He et al., 2016; Zhang et al., 2016), the linkage disequilibrium (LD) patterns across the TP and VP, the size of the TP, marker density (Asoro et al., 2011; Habier et al., 2007, 2009), and the genotype  $\times$  environment interaction (GEI) patterns in the TP and VP (Heslot et al., 2013; Jarquín et al., 2014; Crossa et al., 2016). In general,

many studies have reported that GS accuracy increases as the number of markers and the size of TP increase (Bernardo and Yu, 2007; Asoro et al., 2011; Heffner et al., 2011b; Lorenz et al., 2012; Spindel et al., 2015), yet some studies have shown that systematically sampling TP data can yield prediction accuracies nearly equivalent to or higher than when using all TP data (Moser et al., 2010; Schulz-Streeck et al., 2013; Akdemir et al., 2015; Ametz, 2015; Hoffstetter et al., 2016). Similarly, GS accuracy can be improved when the TP consists of an optimized subset of lines instead of all lines (Akdemir et al., 2015; Ametz, 2015; Isidro et al., 2015; Rutkoski et al., 2015; Hoffstetter et al., 2016).

Oakey et al. (2016) suggested that to reduce genotyping costs when implementing GS, one could possibly use a smaller set of markers that could give similar or better predictive ability than using a large set of markers. Studies report that systematically selecting a subset of markers to build GS models can produce accuracies that are either comparable with using all markers (Weigel et al., 2009; Moser et al., 2010; Vazquez et al., 2010; Schulz-Streeck et al., 2013) or are even superior to using all markers (Akdemir et al., 2015; Hoffstetter et al., 2016). Using only a subset of significant set of markers improved GS accuracy in predicting wheat traits including yield, Fusarium head blight resistance, and quality traits (Hoffstetter et al., 2016) and in predicting the sire breeding values in dairy cattle breeding (Weigel et al., 2009).

Genotype  $\times$  environment interaction (GEI) or marker  $\times$  environment interaction (MEI) effects can have a profound effect on predicting line performance. A fundamental issue for plant breeding is to obtain estimated genetic values for lines using past data that can accurately reflect genetic values that will occur in future environments; GEI diminishes the predictive value of past data. Assessing GEI patterns is critical for making selections in phenotypic selection, as well as in GS (Spindel and McCouch, 2016).

Including the GEI term in GS models improved accuracy when TP and VP data were collected in different environments (Burgueño et al., 2012; Heslot et al., 2013; Jarquín et al., 2014; Lopez-Cruz et al., 2015; Yao et al., 2016; Cuevas et al., 2017; Tiezzi et al., 2017). Jarquín et al. (2014) proposed GS models that incorporate MEI effects with environment effects and/or with environment covariates that were based on climatic records or soil characteristics. They reported that GS accuracy using cross-validation within the same set of environments for wheat YLD improved up to 35% compared with models that only included main effects. Crossa et al. (2016) assessed models with the effects of accessions, markers, environments, and GEI or MEI in two populations of wheat landraces. They reported that GS models with interaction effects provided better predictions for

days to heading and days to maturity in wheat landraces than models without interaction terms. Other studies also reported that the GS accuracy was generally higher when using models with GEI effects vs. without it (Yao et al., 2016; Cuevas et al., 2017; Tiezzi et al., 2017). Most of these studies only assessed GS accuracies through cross-validation and did not predict breeding values between populations (Yao et al., 2016; Cuevas et al., 2017; Tiezzi et al., 2017). Heslot et al. (2013) used marker effects estimated within environments to cluster the environments so that similar environments could be grouped together to minimize MEI effects within each cluster of environments. They proposed that the TP could be optimized by deleting data from the least predictive environments, although this study did not perform predictions between clusters of environments.

Our objectives were (i) to assess the effect of optimizing marker sets on the accuracy of genomic prediction between environments within a population, (ii) to evaluate the accuracy of genomic predictions between populations, and (iii) to assess the accuracy of genomic prediction models that incorporate MEI effects.

## MATERIALS AND METHODS

### Populations

Two populations were used in this study. The first population was termed the elite population (EP) and was described in Huang et al. (2016). Briefly, it consisted of 273 soft winter wheat lines grown in the 2011–2012 and 2012–2013 seasons in 12 to 14 environments per season (Huang et al., 2016). An augmented design was used within each environment, with ‘Branson’ being the check cultivar. The EP was grown with high and low N treatments in Wooster, OH, and Warsaw, VA. Both N treatments received 20 kg N ha<sup>-1</sup> in the fall. The low N treatment then received 45 kg N ha<sup>-1</sup> and the high N treatment received 101 kg N ha<sup>-1</sup> in the spring (at approximately Feekes Stage 6). The second population was termed the yield population (YP), and it consisted of 294 lines (Supplemental Table S1). No lines were shared between the EP and YP except for the check cultivar Branson. In the YP, 95% of the lines came from breeding programs in five states (Kentucky, Maryland, Missouri, Ohio, and Virginia). In the EP, 66% of the lines came from the same five states, with an additional 27% coming from Illinois and Indiana. The YP was grown during the 2013–2014 and the 2014–2015 seasons each at six locations: Wooster, Custar, and Fremont, OH; Warsaw, VA; Columbia, MO; and Lexington, KY. The YP was grown with high and low N in Wooster and Columbia. The N treatments were the same as described for the EP. At each location, the low N treatment was tested with two replications, and high N treatments had one replication except for Columbia, which had two replications for high N. Fungicide was applied in both populations during flag leaf stage (at approximately Feekes Stage 9) in all environments to minimize disease. Each replication in each location was conducted as an augmented design. We defined an environment as a year–location–N treatment combination.

Data were collected for both populations for YLD (kg ha<sup>-1</sup>), TW (kg m<sup>-3</sup>), HGT (cm), HD (Julian d), and seven quality traits: FP content (%) and solvent retention capacities for sucrose content (%), lactic acid (%), water content (%), and NaCO<sub>3</sub> (%); softness equivalent (%), and FY (%), as described by Huang et al. (2016). Data for the EP were collected in 12, 12, 12, 14, and 5 environments for YLD, TW, HGT, HD, and quality traits, respectively (Table 1). For the YP, data were collected in 12, 9, 11, 11, and 4 environments for YLD, TW, HGT, HD, and quality traits, respectively (Table 1).

### Genotypic Data

The EP and YP were genotyped with the Illumina iSelect array for wheat having ~90,000 single-nucleotide polymorphism (SNP) markers (Wang et al., 2014) at the USDA-ARS Biosciences Research Laboratory, Fargo, ND. All marker data were first filtered based on minor allele frequency (<10%) and missing values (>5%; Huang et al., 2016). In the EP, the missing marker scores were first imputed using fastPHASE (Scheet and Stephens, 2006), which provided similar results as using the expectation maximization (EM) algorithm (Huang et al., 2016). The EM approach is less computationally intensive (Huang et al., 2016). It was designed for high-dimensional marker data sets with large amounts of missing values and was implemented in the R package “rrBLUP” (Endelman, 2011). Hence, missing values were imputed in the YP via EM algorithm. To avoid a specific chromosome region being overly represented, a SNP tagging approach was used to select markers that were relatively evenly distributed across the genome (Huang et al., 2016). A final set of 3919 markers was retained and used in the EP analysis. Of the 3919 EP markers, 3537 were also scored in the YP, and these were used in all YP analyses. These 3537 were also used within the EP when comparing the EP with the YP.

### Structure and Linkage Disequilibrium of the Populations

A principal component analysis (PCA) of the genotypic data was performed within the EP (Huang et al., 2016), within the YP and with both the EP and YP. A matrix of the LD *r*<sup>2</sup> values between markers were generated within each population using the “genetics” (Warnes and Leisch, 2005) and “LDheatmap” (Shin et al., 2006) packages in R. The correlation between the two LD matrices was assessed using the R package “ade4” (Dray and Dufour, 2007). The genetic similarity and diversity between and within EP and YP were estimated by calculating a simple matching coefficient, defined as the proportion of loci that had identical genotype scores between two lines, using a marker matrix implemented in SAS using Proc IML (SAS Institute, 2008).

### Phenotypic Data and Trait Stability

A two-stage approach was used to generate the best linear unbiased predictions (BLUPs) for all traits in the EP (Huang et al., 2016) and YP. We used BLUPs of genotype effects because treating genotype effects as random (e.g., shrinkage applied) or fixed had no impact on GS accuracy despite it representing a double shrinkage in our study (Huang et al., 2016). Such results are not surprising, especially for high-heritability traits (Piepho et al., 2008). Data from an environment were first adjusted

**Table 1. Cluster assignment of environments using Wards minimum variance method and the matrix of genotype × environment interaction values.**

Population†	Trait‡	Cluster	Environments§
EP	YLD	1	13VAL, 13VAM
		2	12KYM, 12MDM, 12MOM, 13MOM, 13OWM
		3	12OWL, 12OWM, 12VAM
		Outlier	12VAL, 13OWL
	TW	1	13OWL, 13OWM
		2	12KYM, 13MOM, 12MDM, 12MOM, 12VAL, 12VAM, 13VAL, 13VAM
		3	12OWL, 12OWM
	HGT	Outlier	None
		1	13VAL, 13VAM
		2	12KYM, 12MOM, 12OWL, 12OWM, 12VAM, 13MOM
		3	13OWL, 13OWM
	HD	Outlier	12MDM, 12VAL
		1	12OWL, 12OWM, 13MOM, 13ONM, 13OVM, 13OWL, 13OWM
		2	12KYM, 12MDM, 12MOM, 12VAL, 12VAM, 13VAL, 13VAM
	YP	YLD	1
2			14OWL, 14OWM, 15OVM, 14VAM, 15VAM
3			14MOL, 14MOM, 15MOL, 15MOM
Outlier			15ONM
TW		1	14OWL, 14VAM, 15ONM, 15OVM, 15OWL, 15OWM, 15VAM
		Outlier	14OWM, 14KYM
HGT		1	14OWM, 15OWL, 15OWM
		2	14MOL, 14MOM, 14VAM, 15VAM
HD		Outlier	14OWL, 14KYM, 15MOL, 15MOM
		1	15MOL, 15MOM
		2	14KYM, 14OWL, 14OWM, 14VAM, 15OWL, 15OWM, 15VAM
		3	14MOL, 14MOM
Outlier		None	

† EP, elite panel population; YP, yield panel population.

‡ YLD, yield (kg ha<sup>-1</sup>); TW, test weight (kg m<sup>-3</sup>); HGT, plant height (cm); HD, heading date (Julian d).

§ The environment codes consist of the last two numbers of the year, followed by a two-letter abbreviation for the location (VA = Virginia, KY = Kentucky, MD = Maryland, MO = Missouri, OW = Ohio [Wooster], ON = Ohio [Custar], OV = Ohio [Fremont Station]), followed by “L” for low N and “M” for moderate N.

for block effects within each replication based on means of a common check cultivar. This step was done in SAS 9.2 (SAS Institute, 2008). Then, we used the following model to obtain across-environment BLUPs:

$$Y_{ijk} = \mu + G_i + E_j + R_k(E_j) + GE_{ij} + \varepsilon_{ijk}$$

where the effects of genotype ( $G_i$ ), environment ( $E_j$ ), replicate nested within environment [ $R_k(E_j)$ ], GEI ( $GE_{ij}$ ), and residual ( $\varepsilon_{ijk}$ ) were treated as random effects. The analysis was conducted in R software (R Development Core Team, 2008) using the “lme4” package (Bates et al., 2015). Heritabilities of all traits were estimated using entry means as described in Huang et al. (2016).

The BLUPs within each cluster of environments were obtained using the same model as the one being used above for across all environments. Some clusters consisted of environments that only had one replication, so the  $R_k(E_j)$  term was not used for those environments (Table 1).

We placed the environments into different clusters from each population using Ward’s minimum variance criteria (Betrán et al., 2003; Huang et al., 2016) such that GEI variance was minimized within a cluster and maximized between clusters. In the analyses over all environments, the GEI variance was partitioned into between- and within-cluster components.

Trait stability indices of the genotypes were estimated with Eberhart and Russell regression (ERR; Eberhart and Russell,

1966) and additive main effects and multiplicative interaction (AMMI; Zobel et al., 1988) methods, as described by Huang et al. (2016). This was done in SAS and R. For predictions between populations, stability indices were only calculated over all environments. For AMMI stability, we used the first 10, 3, 6, and 1 interaction principal component (IPC) for YLD, TW, HGT, and HD in the YP, respectively, based on their scree plots. Detailed methods and results can be found in Huang et al. (2016). We obtained data on quality traits from just three environments in the YP, so we did not estimate stability of these traits.

## Genomic Selection

For all GS analysis, the ridge regression BLUP (rrBLUP) model (Endelman, 2011; Lopez-Cruz et al., 2015; Crossa et al., 2016) was used and computed in R (R Development Core Team, 2008):

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{v} + \mathbf{e}$$

where  $\mu$  is the overall mean,  $\mathbf{Z}$  matrix represents the marker scores for each genotype,  $\mathbf{v}$  is the vector of random marker effects,  $\mathbf{e}$  is the residual matrix, and  $\mathbf{y}$  is a vector of the observations. In the rrBLUP model, the fixed effect is the intercept and a penalization parameter  $\lambda = \sigma_e^2 / \sigma_v^2$  was used, in which  $\sigma_e^2$  is the residual variance and  $\sigma_v^2$  is the variance for  $\mathbf{v}$ . This penalization parameter allows for shrinkage while equally estimating the marker effects (Endelman 2011).

The accuracy of GS was estimated (i) overall environments within a population, (ii) between clusters of environments within a population, and (iii) between populations across environments. This was done using (i) data from all markers (the full set of 3919 markers after filtering and tagging), (ii) subsets of markers selected for significance, or (iii) markers selected for stability.

Hence, for between-cluster predictions within each population, traits were analyzed within each separate cluster. The GS models were built using data from one cluster of environments to predict performance of lines in a different cluster.

For prediction within each population, we analyzed data and performed GS both over all environments and within each cluster of environments. For predicting between the EP and YP, we used trait values over all environments.

## Developing Marker Subsets within Each Cluster within a Population

To select markers based on significance, we conducted association analyses for each trait using the “GAPIT” package in R (Lipka et al., 2012). Both the PCA and kinship matrices were used. Subsets of markers were formed by selecting markers that were significant at  $p < 0.05$  (P0.05 subsets). This was done using data only from environments within a cluster. The 0.05 probability level was chosen subjectively, although Hoffstetter et al. (2016) reported that a subset of markers significant at  $p < 0.05$  increased GS accuracy by up to 76% compared with using all markers.

To select subsets of markers based on stability, we first did an association analysis within each environment. For example, there were 12 environments for YLD in the EP, and so we performed 12 independent association analyses for YLD. We used the marker effects to develop a MEI matrix within a cluster. The MEI was derived using the formula

$$y_{ij} = M_i + E_j + MEI_{ij}$$

where  $y_{ij}$  is the value of the  $i$ th marker effect in the  $j$ th environment,  $M_i$  is the main effect of the  $i$ th marker,  $E_j$  is the main effect of the  $j$ th environment, and  $MEI_{ij}$  is the marker  $\times$  environment interaction effect. We assessed the percentage of markers to use in a subset by randomly selecting between 5 and 100% of the markers in increments of 5%, running rrBLUP, and assessing accuracy using cross-validation. Accuracy did not increase with  $>40\%$  of the markers. Thus, we chose to investigate sets with 5 or 40% of the markers. Two methods were used to select subsets of stable markers. We calculated the variance of the MEI effects for each marker and selected the 10 (VAR10 subsets) and 40% (VAR40 subsets) of markers with the lowest MEI variance (Table 2). In the second method, we conducted an AMMI analysis of the MEI matrix and selected the 10% of the markers that had low principal components scores (AMMI10 subsets). The AMMI analyses involved conducting a singular value decomposition of the MEI matrix to extract the IPC scores for each marker. The stability of each marker was obtained by summing the absolute value of the first  $n$  IPC scores for that marker. A low sum indicated high stability. The number of IPC scores used for calculating the marker stability for a trait was chosen based on the scree plot and ranged from two to eight. These VAR10, VAR40, and AMMI10 subsets (Table 2) were formed using the MEI matrices based on marker effects from environments only within one cluster; thus, the

**Table 2. Description of the subsets of training population data.**

Data subsetting method	Description
Ran5	Randomly select 5% of the markers
Ran40	Randomly select 40% of the markers
Var40	Select 40% of the markers with the lowest marker $\times$ environment interaction (MEI) variance
Var10	Select 10% of the markers with the lowest MEI variance
AMMI10	Select 10% of the markers with the lowest MEI variance as estimated using additive main effects and multiplicative interaction (AMMI)
P0.05	Select all markers that were significant in the association analysis at $p < 0.05$
PVar10	Select 15% markers that were most significant, and then choose the ones with the lowest MEI for a total of just 10% of all markers. based on AMMI

phenotypic data used to select stable markers within one cluster of environments was independent of the phenotypic data from the other cluster.

For each trait and set of environments, we constructed one subset of markers based on both their significance and their stability (PVAR10 subset). For the PVAR10 subset, we first selected 15% of the markers with the largest effects (absolute value of allelic effect) from the association analysis. Then, we selected the markers with the lowest MEI variance, producing a subset with 10% of the total markers. The subsets P0.05, VAR10, PVAR10, and VAR40 contained either 5, 10, or 40% of the markers based on different selection criteria. We also generated marker subsets containing a random selection of 5 (RAN5 subset) or 40% (RAN40) of the markers to serve as controls.

## Evaluating the Optimization Methods

Within each population and for each trait, we first used all phenotypic and genotypic data as a TP and then set up TPs that used the different markers subsets (using RAN5, RAN40, VAR10, VAR40, AMMI10, P0.05, or PVAR10). Within a population, we conducted 10-fold cross-validation to assess the accuracy of GS, as described by Huang et al. (2016). The cross-validations were run 500 times with the RAN5 and RAN40 subsets. The difference in accuracy was minimal between these 500 iterations, so we selected 10 runs for the other data subsets to be less computationally intensive.

## Use of Marker $\times$ Environment Interactions in Predictions

For the analyses between populations, we used BLUPs over all environments from each population. The ability of data from one population to predict the phenotypes in the other population was assessed. Either all data or a subset of TP data from one population was used to build a GS model that was then used to calculate the GEBVs of lines from a different population. Cross-validation was not used in these analyses, as each population was independent of the other. To predict between two populations, by using all data, we also assessed two additional models that incorporate either environment and/or MEI effects (Lopez-Cruz et al., 2015; Crossa et al., 2016). The first model incorporated genomic and environment main effects:

$$y_{ij} = \mu + g_i + e_j + \varepsilon_{ij}$$

where  $y_{ij}$  is the observation for the  $i$ th genotype at the  $j$ th environment,  $e_j$  is the effect of the  $j$ th environment, and  $\varepsilon_{ij}$  is the residual effects associated with the  $ij$ th observation. In this model,  $g_i = \sum_{k=1}^m z_{ik}u_k$ , in which  $z_{ik}$  is the score (-1, 0, or 1) for the  $k$ th marker for the  $i$ th genotype and  $u_k$  is the effect for the  $k$ th marker (Lopez-Cruz et al., 2015). In this model, the  $u_k$  values were assumed to be the main marker effects and were the same across environments. The intercept and  $e_j$  were fixed effects in this model (Lopez-Cruz et al., 2015; Crossa et al., 2016).

The second model allows for borrowing information across environments to estimate marker effects across different environments, and it includes MEI effects (Lopez-Cruz et al., 2015; Crossa et al., 2016):

$$y_{ij} = \mu + g_i + e_j + ge_{ij} + \varepsilon_{ij}$$

The definitions of  $y_{ij}$ ,  $g_i$ ,  $e_j$ , and  $\varepsilon_{ij}$  are the same as in the abovementioned main effects model. The interaction term  $ge_{ij}$  is the interaction of genomic values of the  $i$ th individual and the  $j$ th environment. In this model, the marker effects consisted of two parts—the marker main effect, which is a constant across all environments, and an environment-specific effect for each marker (Crossa et al., 2016). The fixed effects in this model included the intercept and environment main effects. The random effects included marker effects, GEI effects, and residual effects (Lopez-Cruz et al., 2015). The  $g + e$  main effects model and the  $g + e + ge$  model were all computed in R using the package “BGLR” (de los Campos and Pérez-Rodríguez, 2013).

## RESULTS

### Phenotypic Analysis

Analysis of the EP phenotypic data was reported in Huang et al. (2016). In both the EP and YP, heritability was >0.73 for all traits, and the effects of environment, genotype, and GEI were significant ( $p < 0.05$ ) for all traits (Huang et al., 2016; Table 3).

### Clustering of Environments within Each Population

We identified three clusters of environments for YLD and HD in the YP and for YLD, TW, and HGT in the EP. Two clusters of environments were identified for HGT in the

YP and for HD in the EP (Table 1). In the YP, there was one cluster of environments and one outlier for TW. Some environments were outliers and were excluded from the analyses between clusters (Table 1). In the EP, the correlation of the phenotypic BLUPs between the most divergent clusters of environments was low for YLD ( $r = 0.34$ ) and TW ( $r = 0.33$ ) but was high for HGT ( $r = 0.78$ ) and HD ( $r = 0.82$ ); these correlations in YP were also low for YLD ( $r = 0.27$ , Table 4). The percentage of GEI variance attributed to between clusters was >50% for all traits except YLD in both populations (Supplemental Table S2), suggesting that clustering was least effective for YLD. The GEI variance as a percentage of genotype + GEI variance was also greater for YLD than for the other traits (Supplemental Table S2).

### Population Structure and Relationship between the Elite Panel and Yield Panel Populations

The population structure in the YP was visually similar to that in the EP, as each was characterized by one large group of lines and one smaller group that was mainly composed of lines from the University of Missouri (Fig. 1a). Results from PCA for both populations together showed that lines from both populations were interspersed in all regions of the graph (Fig. 1b). The genetic similarities via simple matching coefficient between the two populations (0.60) and within the EP (0.60) or within the YP (0.60) were nearly identical, suggesting that these two populations were very similar. The LD matrices of the two populations were tested via mental test, and they were significantly correlated ( $r = 0.87$ ,  $p < 0.01$ ), indicating that they have a similar LD pattern among the common markers.

### Genomic Selection Accuracy between Clusters of Environments within a Population Using Subsets of Markers

There was a variable degree of commonality between the marker subsets developed for any trait. For all traits, the scheme used to create the AMMI10 and VAR10 marker

**Table 3. Variance components, their significance, and entry mean broad-sense heritability ( $H$ ) from the ANOVA of the yield panel (YP).**

Trait†	Genotype (G)	Environment (E)	Replication(E)	G × E	Residual	$H$
YLD	64,220.0**	682,555.0*	10,435.0	64,745.0**	179,122.0	0.83
TW	175.8**	812.1	150.2	526.9**	103.0	0.73
HGT	31.9**	29.9*	0.8	4.7**	19.2	0.96
HD	2.5**	40.9*	0.0	0.9**	2.0	0.94
FP	0.1**	1.0	–	0.2**	–	0.75
SU	9.7**	0.2	–	5.9**	–	0.89
LA	92.6**	68.5	–	36.2**	–	0.93
WA	2.9**	1.1	–	1.3	–	0.91
SO	7.9**	2.0	–	2.6**	–	0.94
SE	6.9**	0.3	–	3.5**	–	0.91
FY	1.4**	0.5	–	1.0**	–	0.88

\*\*, \* Significant at the 0.05 and 0.01 probability levels, respectively.

† YLD, yield (kg ha<sup>-1</sup>); TW, test weight (kg m<sup>-3</sup>); HGT, plant height (cm); HD, heading date (Julian d); Solvent retention capacity (%) for flour protein (FP), sucrose content (SU), lactic acid (LA), water content (WA), NaCO<sub>3</sub> (SO), softness equivalent (SE), and flour yield (FY).

**Table 4. Accuracy of genomic selection between cluster of environments (data from one cluster was used to predict performance in the other cluster) within the elite panel (EP) and yield panel (YP). The phenotypic correlation (*r*) between clusters (labeled C1–C3) is shown, as is accuracy using all training population data (all data) and subsets of data as described in Table 2.**

Population	Trait†	Data set‡	C1–C2	C2–C1	C1–C3	C3–C1	C2–C3	C3–C2	Avg.
EP	YLD	<i>r</i>	0.51	0.51	0.34	0.34	0.41	0.41	0.42
		All data	0.45	0.49	0.25	0.27	0.31	0.25	0.34
		Ran5	0.37	0.42	0.18	0.22	0.18	0.17	0.26
		Ran40	0.44	0.48	0.24	0.26	0.29	0.23	0.32
		Var40	0.44	0.46	0.24	0.28	0.23	0.24	0.32
		Var10	0.38	0.42	0.19	0.24	0.15	0.18	0.26
		AMMI10	0.38	0.41	0.19	0.24	0.15	0.20	0.26
		P0.05	0.49	0.48	0.32	0.33	0.36	0.37	0.39
		PVar10	0.49	0.51	0.32	0.33	0.38	0.38	0.40
		Var40-L	0.50	0.47	0.30	0.32	0.33	0.37	0.38
		Var10-L	0.48	0.49	0.31	0.32	0.38	0.39	0.40
		P0.05-L	0.50	0.50	0.35	0.34	0.46	0.43	0.43
		PVar10-L	0.51	0.52	0.34	0.33	0.41	0.41	0.42
YP	YLD	<i>r</i>	0.50	0.50	0.27	0.27	0.48	0.48	0.42
		All data	0.47	0.44	0.24	0.21	0.46	0.43	0.37
		P0.05	0.48	0.46	0.29	0.24	0.47	0.46	0.40
		P0.05-L	0.48	0.52	0.26	0.25	0.47	0.47	0.41
EP	HGT	<i>r</i>	0.84	0.84	0.78	0.78	0.86	0.86	0.83
		All data	0.77	0.72	0.67	0.66	0.75	0.80	0.73
		Ran5	0.68	0.61	0.57	0.56	0.64	0.69	0.62
		Ran40	0.76	0.71	0.66	0.65	0.74	0.79	0.72
		Var40	0.75	0.62	0.65	0.63	0.66	0.77	0.68
		Var10	0.72	0.57	0.62	0.59	0.60	0.73	0.64
		AMMI10	0.72	0.58	0.62	0.59	0.61	0.73	0.64
		P0.05	0.78	0.74	0.70	0.66	0.79	0.78	0.74
		PVar10	0.79	0.77	0.71	0.70	0.80	0.82	0.76
		Var40-L	0.81	0.77	0.73	0.72	0.81	0.85	0.78
		Var10-L	0.82	0.79	0.74	0.74	0.84	0.87	0.80
		P0.05-L	0.84	0.81	0.79	0.79	0.84	0.87	0.82
		PVar10-L	0.89	0.88	0.83	0.85	0.88	0.89	0.87
YP	HGT	<i>r</i>	0.85	0.85	–	–	–	–	0.85
		All data	0.82	0.83	–	–	–	–	0.83
		P0.05	0.82	0.83	–	–	–	–	0.83
		P0.05-L	0.85	0.84	–	–	–	–	0.85
EP	HD	<i>r</i>	0.82	0.82	–	–	–	–	0.82
		All data	0.75	0.75	–	–	–	–	0.75
		Ran5	0.60	0.57	–	–	–	–	0.59
		Ran40	0.73	0.72	–	–	–	–	0.73
		Var40	0.58	0.64	–	–	–	–	0.61
		Var10	0.54	0.55	–	–	–	–	0.55
		AMMI10	0.55	0.57	–	–	–	–	0.56
		P0.05	0.78	0.77	–	–	–	–	0.78
		PVar10	0.76	0.80	–	–	–	–	0.78
		Var40-L	0.76	0.78	–	–	–	–	0.77
		Var10-L	0.78	0.80	–	–	–	–	0.79
		P0.05-L	0.80	0.81	–	–	–	–	0.81
		PVar10-L	0.82	0.82	–	–	–	–	0.82
YP	HD	<i>r</i>	0.70	0.70	0.52	0.52	0.53	0.53	0.58
		All data	0.71	0.66	0.57	0.51	0.51	0.50	0.58
		P0.05	0.70	0.67	0.51	0.52	0.50	0.54	0.57
		P0.05-L	0.67	0.69	0.49	0.55	0.52	0.59	0.58
EP	TW	<i>r</i>	0.68	0.68	0.33	0.33	0.47	0.47	0.49
		All data	0.71	0.67	0.31	0.32	0.46	0.52	0.50
		Ran5	0.64	0.61	0.26	0.28	0.41	0.47	0.45
		Ran40	0.70	0.67	0.30	0.32	0.45	0.52	0.49
		Var40	0.69	0.66	0.30	0.33	0.44	0.51	0.49
		Var10	0.66	0.61	0.29	0.32	0.41	0.49	0.46
		AMMI10	0.66	0.61	0.29	0.32	0.41	0.49	0.46
		P0.05	0.68	0.66	0.31	0.34	0.45	0.51	0.49
		PVar10	0.68	0.67	0.34	0.36	0.47	0.52	0.51
		Var40-L	0.72	0.65	0.36	0.32	0.49	0.47	0.50
		Var10-L	0.70	0.66	0.37	0.31	0.48	0.48	0.50
		P0.05-L	0.70	0.71	0.29	0.34	0.48	0.47	0.50
		PVar10-L	0.73	0.68	0.37	0.34	0.48	0.47	0.51

† YLD, yield (kg ha<sup>-1</sup>); TW, test weight (kg m<sup>-3</sup>); HGT, plant height (cm); HD, heading date (Julian d).

‡ All data, predictions using all training population data; the other abbreviations represent the subsets of training population data define in Table 2.

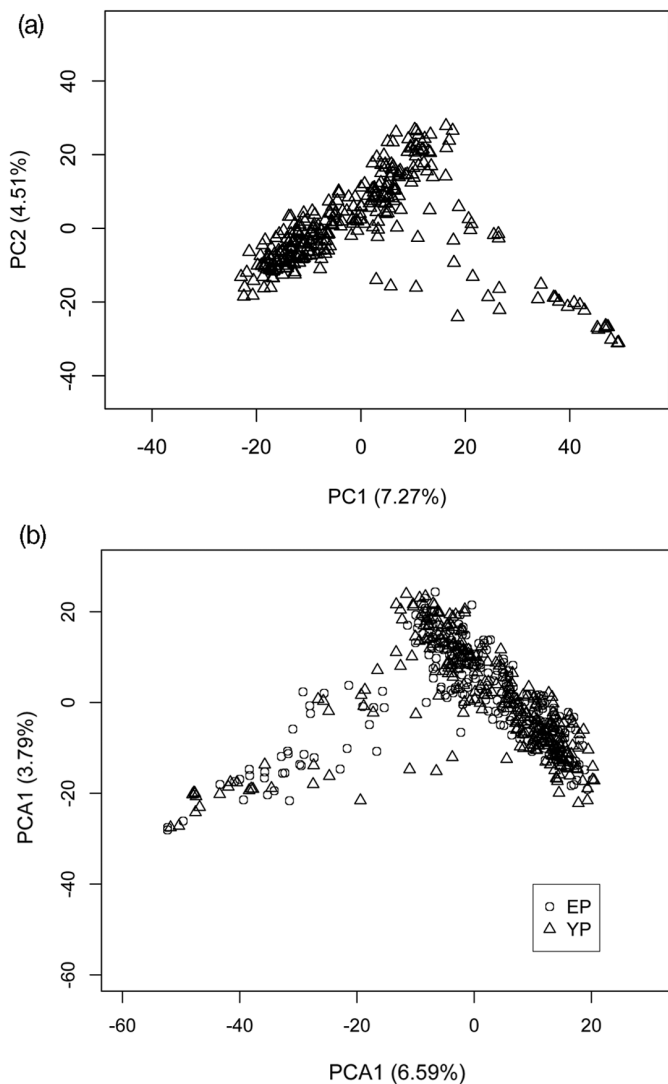


Fig. 1. Plot of the first two principal component scores for (a) the lines within the yield panel population (YP) and (b) within the elite panel (EP) and YP populations using data from 3537 single-nucleotide polymorphism markers. The plot of principal component analysis (PCA) for within the EP was reported by Huang et al. (2016).

subsets provided almost identical subsets of markers (results not shown). This indicates that using total MEI variance or the AMMI approach produced similar sets of markers.

The GS accuracy between clusters of environments varied by population, trait, the subset of data being used, and the particular comparison (Table 4). The accuracy of GS predictions between clusters using all TP data (e.g., all markers and all lines) was, on average, 7.2% lower than the phenotypic correlation, ranging from 39.0% lower to 10.9% greater (Fig. 2, Table 4, Supplemental Fig. S1). For YLD, HGT, and HD, the average GS accuracy between clusters of environments was 16.2, 9.0, and 3.1% lower than the average phenotypic correlation between clusters (Table 4). The average phenotypic correlation and GS accuracy between clusters were nearly equal for TW (Table 4). We assessed the relationship of the correlation of phenotypic means between

environment clusters with the ability of a GS model built using data from one cluster of environments to predict the phenotypes of a different cluster. The accuracy of GS predictions between clusters was highly associated ( $r = 0.97$ ) with the correlation of phenotypes between clusters.

For most traits and between-cluster comparisons, the Ran5 subsets gave the lowest GS accuracy (Table 4). The accuracy with the Ran40 subsets was 16.9% better than with the Ran5 subsets when averaged over all comparisons, and its accuracy was similar to the GS accuracy using all markers (Table 4).

The markers in the VAR10, VAR40, and AMMI10 subsets were selected based on the stability of their effects across environments within a particular cluster. For all comparisons, GS accuracy for VAR10 and AMMI10 were nearly identical—when averaged over all comparisons, both were 9% less accurate than the VAR40 subsets (Table 4). The accuracy of the VAR40 subsets was similar to the accuracy obtained with the RAN40 subsets, both of which produced accuracies that were only 7% less accurate than using all data when averaged over all comparisons.

The P0.05 and PVAR10 subsets of markers were selected considering the magnitude of marker effects (P0.05), as well as their stability (PVAR10). Averaged over all comparisons, the accuracy of the P0.05 subsets was 5.1% greater than using data from all markers, whereas the accuracy of the PVAR10 subsets was 7.7% more accurate than using all data (Table 4). The increase in accuracy for these two subsets was most notable for YLD, where P0.05 and PVAR10 were, on average, 14.8 and 17.6% more accurate than using all marker data. These marker subsets had minimal impact on increasing between-cluster accuracy compared with using all data for TW, HGT, and HD (Table 4). The P0.05 and PVAR10 marker subsets had a similar number of markers as the RAN5 subset but produced greater GS accuracy than the RAN5 subsets or when using all markers (Table 4).

### Between Populations: Genomic Selection Accuracy Using All Training Population Data or Subsets of Data

From the results within the EP, we determined that the P0.05 (significant markers) subset of markers was the best, and thus we only used all data and this subset of data in the between-population analyses. For 13 out of the 19 traits and trait stability indices, using all data from one population to predict the phenotypes of lines in the other population gave higher accuracy than using the best P0.05 subsets of marker data (Supplemental Table S3). Thus, only results from using all TP data will be discussed.

When using either all data or subsets of marker data, GS accuracy between populations was considerably lower than the accuracy between clusters of environments within a population (Fig. 2, Tables 4 and 5), being 95, 11,



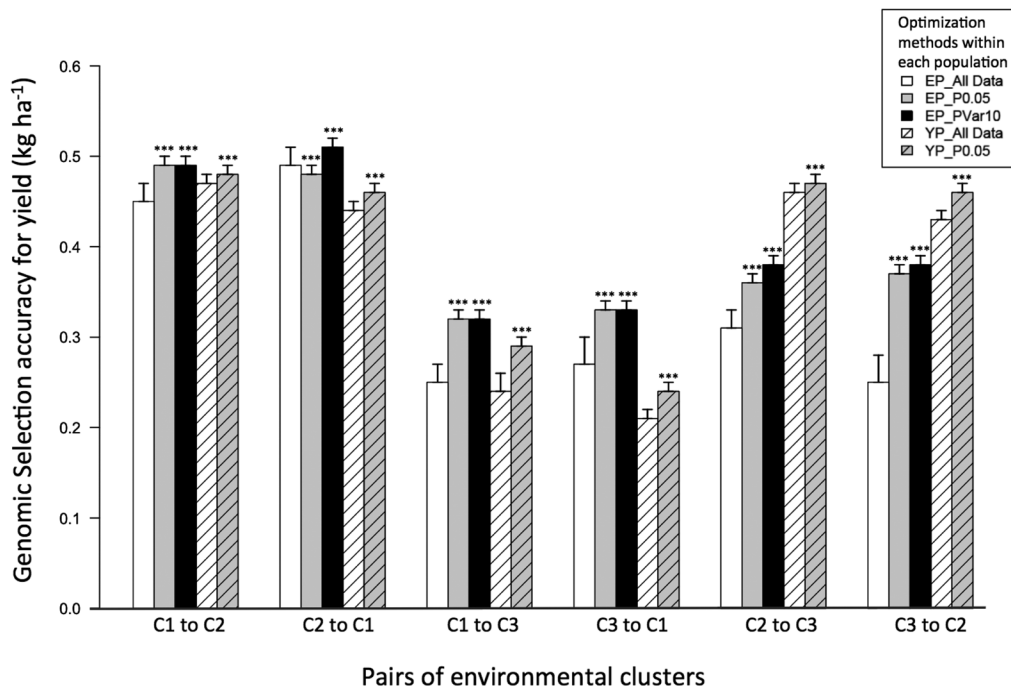


Fig. 2. Plot of the genomic selection accuracy between clusters of environment within a population for grain yield, using subset of data (P0.05 = set of only significant markers; PVAR10 = set of significant and stable markers) compared with using all marker data within the elite panel (EP) and yield panel (YP) populations. \*\*\* An estimate of accuracy that is significantly ( $p < 0.001$ ) different from the accuracy obtained using all marker data. The  $p$  values were adjusted to account for multiple comparisons within each population.

Table 5. Accuracy of genomic selection within the elite population (EP), within the yield population (YP), and between populations where all data from one population was used as the training population (TP) and the other served as the validation population (VP). All accuracies were estimated using data from all TP markers and lines. The last column shows the correlation of marker effects estimated within the EP or the YP for all markers that were significant ( $p < 0.05$ ) in at least one of the populations.

Trait†	Accuracy via cross-validation within the		TP = EP	TP = YP	Avg. between-population accuracy	$r$ of significant marker effects
	EP	YP	VP = YP	VP = EP		
YLD	0.33	0.49	-0.03	0.06	0.02	0.12
TW	0.66	0.41	0.39	0.50	0.44	0.34
HGT	0.54	0.74	0.62	0.48	0.55	0.41
HD	0.56	0.68	0.41	0.32	0.37	0.21
AYLD	0.44	0.18	0.13	0.12	0.12	-
ATW	0.17	0.22	0.00	-0.03	-0.02	-
AHGT	0.28	0.13	-0.01	-0.02	-0.02	-
AHD	0.35	0.48	0.29	0.15	0.22	-
BYLD	0.36	0.53	0.32	0.13	0.23	-
BTW	0.42	0.34	0.23	0.23	0.23	-
BHGT	0.27	0.43	0.37	0.25	0.31	-
BHD	0.65	0.37	0.23	0.33	0.28	-
FP	0.41	0.46	0.19	0.22	0.20	0.09
SU	0.6	0.58	0.39	0.42	0.41	0.35
LA	0.64	0.67	0.58	0.53	0.56	0.47
WA	0.65	0.60	0.46	0.47	0.47	0.37
SO	0.57	0.63	0.45	0.42	0.43	0.31
SE	0.37	0.59	0.31	0.26	0.28	0.34
FY	0.49	0.61	0.31	0.37	0.34	0.32

† YLD, yield ( $\text{kg ha}^{-1}$ ); TW, test weight ( $\text{kg m}^{-3}$ ); HGT, plant height (cm); HD, heading date (Julian d); Solvent retention capacity (%) for flour protein (FP), sucrose content (SU), lactic acid (LA), water content (WA),  $\text{NaCO}_3$  (SO), softness equivalent (SE), and flour yield (FY). A and B before trait abbreviations indicate the additive main effects and multiplicative interaction index stability and the Eberhart and Russell regression index stability, respectively, for each trait.

28, and 41% lower for YLD, TW, HGT, and HD, respectively. The between-population GS accuracy for a trait was similar (correlation  $r = 0.89$ , calculated using values in Columns 4 and 5 in Table 5) whether the EP or the YP was used as the TP. The average accuracy of GS between populations was negligible for YLD and AMMI stability indices for TW and HGT and was  $<0.25$  for AMMI stability indices for YLD, TW, and HD; ERR indices for YLD, TW, and HD; and FP (Table 5). On average, the stability estimated using the regression method gave higher between-population GS accuracy than using the AMMI method (Table 5).

The GS accuracy between populations was associated with GS accuracy within each population obtained using cross-validation (Table 5, using values in Column 2 or 3 correlating to values in Column 6). The correlation of the average accuracy for within-population prediction and the average accuracy for between-population prediction was 0.91 (Fig. 3). If accuracy for a trait was low within a population, then the data from that population had a low ability to predict phenotypes in the other population (Fig. 3). The accuracy between populations was also associated with the correlation of genetic effects of significant markers between the two populations (Table 5). Prediction accuracy was low when the correlation of marker effects was low.

### Accuracy of Genomic Selection Using the $g + e$ and $g + e + ge$ Models for Between-population Prediction

The model including both marker and environment main effects and a model containing those effects plus the MEI effects were assessed for between-population accuracy for each trait. These two models generally produced lower between-environment accuracy for TW, HGT, and HD (Table 6) than the standard GS model rrBLUP (Table 5). Compared with using the rrBLUP model, the  $g + e$  and  $g + e + ge$  models gave the largest improvement of GS accuracy for YLD when using YP to predict EP, but these accuracies remained low ( $r < 0.15$ , Table 6).

## DISCUSSION

One of the most fundamental issues in plant breeding is obtaining estimates of genetic values from one data set that can predict phenotypic performance in future environments. This issue remains whether using phenotypic data or GEBVs (Spindel and McCouch, 2016). We investigated whether using subsets of marker data could improve the ability of GS to predict performance in different environments. Markers were selected based on their effects, the stability of their effects, or both criteria.

Our results show that when the correlation of phenotypes between two environments was low, the ability of data from one environment to predict the other using

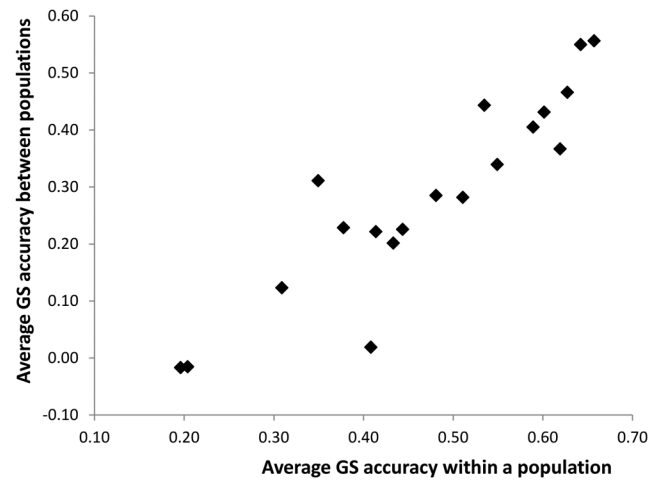


Fig. 3. Plot of genomic selection (GS) accuracy between populations using data from all traits regressed onto the average accuracy of the trait within the elite and yield populations.

GS was also low, as has been reported before (Dekkers, 2007). On average, for all traits across all between-cluster comparisons within a population, randomly selecting markers (Ran5, Ran40) or subsetting markers based on their stability alone (AMMI10, Var10, and Var40) did not improve the GS accuracy compared with using all the TP data (Table 4). These results suggest that selecting markers based solely on stability of marker effects will not improve GS accuracy of between-environment predictions. This may be because uninformative markers with no effect on the trait can be stable if they may have little to no effect in any environment, and thus the markers were included in these subsets along with predictive markers that do have an effect. We later confirmed that using significant markers could increase GS accuracy (Fig. 2, Supplemental Fig. S1).

Table 6. Accuracy of genomic selection between the elite population (EP) and the yield population (YP) using two models,  $g + e$  and  $g + e + ge$ . For each trait the EP and the YP were each used as either the training population (TP) and as the validation population (VP).

Trait†	Model			
	$g + e$		$g + e + ge$	
	TP = EP	TP = YP	TP = EP	TP = YP
	VP = YP	VP = EP	VP = YP	VP = EP
YLD	-0.03	0.10	-0.02	0.14
TW	0.37	0.45	0.38	0.40
HGT	0.59	0.45	0.55	0.42
HD	0.39	0.32	0.40	0.31
FP	0.17	0.19	0.16	0.18
SU	0.29	0.42	0.27	0.34
LA	0.58	0.50	0.56	0.45
WA	0.44	0.45	0.42	0.37
SO	0.37	0.40	0.35	0.35
SE	0.31	0.25	0.32	0.22
FY	0.27	0.37	0.24	0.34

† YLD, yield ( $\text{kg ha}^{-1}$ ); TW, test weight ( $\text{kg m}^{-3}$ ); HGT, plant height (cm); HD, heading date (Julian d); Solvent retention capacity (%) for flour protein (FP), sucrose content (SU), lactic acid (LA), water content (WA),  $\text{NaCO}_3$  (SO), softness equivalent (SE), and flour yield (FY).

The results indicate that, within a population, a GS model built using data from one set of environments can predict the performance of lines in an independent set of environments. The accuracy of GS for YLD between environments within a population for all traits was significantly improved, relative to using all TP data, when using subsets of markers. The GS accuracy increased 8.1% in the YP and 14.7% in EP on average by using the P0.05 subset, and by 17.6% in the EP by using the PVAR10 subset (Fig. 2, Supplemental Fig. S1). Minimal increase in GS accuracy was observed for the other traits (Table 4). Grain yield was the trait that had the greatest GEI, suggesting that greater gains in GS accuracy from using subsets of significant and stable markers may be realized for traits with extensive GEI than for traits with low GEI. Our results were similar to those of a previous wheat study (Hoffstetter et al., 2016) where using subsets of significant markers for YLD, Fusarium head blight resistance, and quality traits increased GS accuracy by 41 to 76% compared with using all markers. Hoffstetter et al. (2016) also reported that GS accuracy within and between environments was greatest with a TP composed of a subset of lines with low GEI variance and significant markers. In animal breeding, it was also reported that when using the subset of markers with the largest effects, GS accuracy was either similar to or was somewhat improved compared using all markers in the TP (Abdollahi-Arpanahi et al., 2014; Weigel et al., 2009; Moser et al., 2010; Vazquez et al., 2010; Yao et al., 2016). It is likely that by using only markers with stable significant effects across environments, we are excluding markers from the least predictive portion of the genome.

Using data from one population to predict the performance of lines in a different, yet highly related, population that was phenotyped in different environments has proven to be more difficult than predicting the performance of the same lines in different environments (Schulz-Streeck et al., 2013; Beaulieu et al., 2014; Jarquín et al., 2014; Lorenz and Smith, 2015; Lado et al., 2016). In our study, GS accuracy between the EP and YP populations was lower than GS accuracy between a cluster of environments within the same population (Fig. 2, Tables 5). In addition, using data subsets did not improve GS accuracy between populations compared with using all TP data. Data from one population was poor at predicting YLD, FP, and AMMI-based trait stability in the other population (Table 5).

The EP and YP were closely related to each other, as shown by the PCA graph (Fig. 1b) and the simple matching coefficients. The LD matrices of the two populations were highly correlated, suggesting that there could also be a similar LD pattern between markers and quantitative trait loci (QTLs) in each population. This is one requirement in order for GS to work, and this consistency between populations of LD between markers and QTLs appeared to

occur for TW, HGT, HD, and solvent retention capacities for sucrose, lactic acid, water, and  $\text{NaCO}_3$ . This suggests that the LD patterns between markers and QTLs for YLD and FP would also likely be consistent between the two populations. A second requirement for GS to work is for QTLs to have similar effects in each population or set of environments. In our study, the ability to use data from one population to predict performance in the other was related to the correlation of the marker effects estimated in the two populations (Table 5). The correlation of markers effects between the EP and YP for YLD and FP was low. A low correlation of marker effects would lead to high GEI, which would lead to a low phenotypic correlation between environments. Unsurprisingly, as shown here and by others (Dekkers, 2007), when the correlation of phenotypes between environments is low, GS will not be very predictive. We also report that GS accuracy within a population was strongly associated with GS accuracy between populations (Fig. 2, Table 4, Supplemental Fig. S1). If GS accuracy was low within a population, then GS accuracy between populations was also low.

In our study, low GS accuracy between populations for YLD stability and several other stability parameters indicate that stability estimated in one set of environments is not very predictive of stability in another set of lines and environments. The ERR stability indices were predicted with a higher accuracy than the AMMI-based stability indices. This could be because the ERR measures are based on environmental indices (i.e., average trait value within an environment) whose values and impact may be more repeatable across sets of environments, whereas the AMMI IPC scores may be very specific to the set of environments from which they were calculated (Eberhart and Russell, 1966; Zobel et al., 1988). Hickey et al. (2014) suggested that for closely related populations, effective GS accuracy could be obtained with a small number of markers (200–500) and perhaps 1000 lines. Our populations are considerably smaller, so maybe greater accuracy could have been attained if the size of each population was increased by adding lines that are closely related.

For between-population predictions, our results are well aligned with those of others who have used a TP and VP that were composed of different lines and/or were tested in different environments. Jarquín et al. (2014) reported that GS accuracy for wheat YLD was greater within a population than between populations that consist of different lines tested in different environments. It was reported that the GS accuracy through cross-validation was higher than those obtained from predicting between populations, as is also shown in this study. Jarquín et al. (2016) also compared GS accuracies for YLD by predicting through cross-validation vs. by predicting between locations, years, and location–year combination schemes, and they confirmed that GS accuracy via cross-validation is

higher than that using the other schemes. Michel et al. (2016b) assessed GS accuracy with data from lines from five breeding cycles of winter wheat. They reported the average GS accuracy via cross-validation within a population was 0.43 for YLD and 0.57 for FP but was just 0.29 and 0.41 for YLD and FP, when the TP and VP consisted of different lines tested in different environments. Battenfield et al. (2016) reported that GS accuracy for wheat TW between populations tested in different years ranged from 0.12 to 0.35, whereas GS accuracy was 31.8% greater within a population. Dawson et al. (2013) assessed the accuracy of GS for wheat YLD using a TP of lines from a previous year to predict performance of VP lines in future years using data on 622 lines from CIMMYT that were phenotyped over 17 yr. When the TP and VP were tested under different environments, the adjusted GS accuracy for YLD ranged from -0.10 to 0.80 across 17 yr of testing environments (Dawson et al., 2013). Battenfield et al. (2016) reported that when the TP and VP consisted of lines grown in different environments and had a few lines in common, GS accuracy ranged from 0.33 to 0.44 for FY and FP but ranged from 0.82 to 0.90 within populations.

When the performance of new lines needs to be predicted, Jarquín et al. (2014) recommend using GS models with GEI terms. They assessed two scenarios: the first one was to predict the value of lines that were tested in some but not all environments, and the second one was to predict the value of new lines that were not tested in any environments (Jarquín et al., 2014). Their model worked better under the first scenario, which allowed borrowing information from one set of environments for the same line to predict its performance in a different set of environments, whereas the latter uses data from one set of lines to predict performance of a different line in different environments (Crossa et al., 2014; Jarquín et al., 2014, 2016; Saint-Pierre et al., 2016). Cuevas et al. (2017) reported that using models incorporating the *ge* term always produced higher GS accuracy than models without *ge* for one maize (*Zea mays* L.) and four wheat CIMMYT data sets. In our study, however, using the *g + e* and *g + e + ge* models did not significantly improve GS accuracy for most traits (Table 6). Similar to our results, Dawson et al. (2013) reported that modeling MEI effects did not improve GS accuracy for YLD across a wide range of environments compared with the model without GEI. They suggested that there were inconsistent marker effects among the mega environments.

The same trait being evaluated in two environments could be viewed as two traits, and thus selection based on data from one environment to improve that trait in another environment could be viewed as indirect selection (Falconer and Mackay, 1996). Our results show that indirect selection using GS should be effective where direct selection using phenotypes is effective. The advantage of

GS then is that a cycle of GS can be completed in much less time than a cycle of phenotypic selection (Meuwissen et al., 2001; Goddard and Hayes, 2007; Heffner et al., 2010; Jannink et al., 2010). Gain per unit of time is a key measure of plant breeding efficiency. One of the major limitations to plant breeding efficiency is the amount of time required to complete a cycle of phenotypic selection. This problem could be addressed and improved by using GS, as GS could rapidly improve selection efficiency per unit time and cost (Meuwissen et al., 2001; Heffner et al., 2009; Jannink et al., 2010; Sallam and Smith 2016). The results of this study suggest that GEI may now be one of the major factors limiting genetic gain per unit of time.

### Author Contribution Statement

M. Huang wrote the manuscript and analyzed the phenotypic and genotypic data. M. Huang also collected the data at Ohio location. B. Ward, C. Griffey, D. Van Sanford, and A. McKendry contributed to data collection in Virginia, Kentucky, and Missouri, in addition to editing the manuscript. G. Brown-Guedira and P. Tyagi assisted the genotypic data for the YP. C. Sneller led the project, edited the manuscript, and helped with data analysis.

### Conflict of Interest

The authors declare that there is no conflict of interest.

### Compliance with Ethical Standards

This research complies with the current laws of the United States of America.

### Supplemental Material Available

Supplemental material for this article is available online.

### Acknowledgments

We sincerely thank the members at Dr. C. Sneller's laboratory for their help with field data collection in Ohio. Funding support was from Triticeae Coordinated Agricultural Project (2011-68002-30029) of the USDA National Institute of Food and Agriculture.

### References

- Abdollahi-Arpanahi, R., A. Pakdel, A. Nejat-Javaremi, M. Moradi Shahrabak, G. Morota, B. Valente, et al. 2014. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. *J. Anim. Breed. Genet.* 131:183-193. doi:10.1111/jbg.12079
- Akdemir, D., J.I. Sanchez, and J.-L. Jannink. 2015. Optimization of genomic selection training populations with a genetic algorithm. *Genet. Sel. Evol.* 47:38. doi:10.1186/s12711-015-0116-6
- Ametz, C. 2015. Genomic selection in bread wheat. Ph.D. diss., Univ. für Bodenkultur, Vienna. <http://permalink.obvsg.at/bok/AC10777353> (accessed 5 June 2018).
- Arruda, M., A. Lipka, P. Brown, A. Krill, C. Thurber, G. Brown-Guedira, et al. 2016. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol. Breed.* 36:84. doi:10.1007/s11032-016-0508-5

- Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome* 4:132–144. doi:10.3835/plantgenome2011.02.0007
- Bates, D., M. Maechler, B. Bolker, S. Walker, R.H.B. Christensen, H. Singmann et al. 2015. Package ‘lme4’. R Found. Stat. Comput., Vienna.
- Battenfield, S.D., C. Guzmán, R.C. Gaynor, R.P. Singh, R.J. Peña, S. Dreisigacker, et al. 2016. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* 9(2). doi:10.3835/plantgenome2016.01.0005
- Beaulieu, J., T.K. Doerksen, J. MacKay, A. Rainville, and J. Bousquet. 2014. Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics* 15:1048. doi:10.1186/1471-2164-15-1048
- Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47:1082–1090. doi:10.2135/cropsci2006.11.0690
- Betrán, F., D. Beck, M. Bänziger, and G. Edmeades. 2003. Genetic analysis of inbred and hybrid grain yield under stress and non-stress environments in tropical maize. *Crop Sci.* 43:807–817. doi:10.2135/cropsci2003.8070
- Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype  $\times$  environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52:707–719. doi:10.2135/cropsci2011.06.0299
- Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724. doi:10.1534/genetics.110.118521
- Crossa, J., D. Jarquín, J. Franco, P. Pérez-Rodríguez, J. Burgueño, C. Saint-Pierre, et al. 2016. Genomic prediction of gene bank wheat landraces. *G3: Genes, Genomes, Genet.* 6:1819–1834. doi:10.1534/g3.116.029637
- Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Omella, J. Cerón-Rojas, et al. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112:48–60. doi:10.1038/hdy.2013.16
- Cuevas, J., J. Crossa, O.A. Montesinos-López, J. Burgueño, P. Pérez-Rodríguez and G. de los Campos. 2017. Bayesian genomic prediction with genotype  $\times$  environment interaction kernel models. *G3: Genes, Genomes, Genet.* 7: 41–53. doi:10.1534/g3.116.035584
- Dawson, J.C., J.B. Endelman, N. Heslot, J. Crossa, J. Poland, S. Dreisigacker, et al. 2013. The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Res.* 154:12–22. doi:10.1016/j.fcr.2013.07.020
- de los Campos, G., and P. Pérez-Rodríguez. 2013. BGLR: Bayesian generalized linear regression. R package version 1. R Found. Stat. Comput., Vienna.
- Dekkers, J. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* 124:331–341. doi:10.1111/j.1439-0388.2007.00701.x
- Denis, M., and J.M. Bouvet. 2011. Genomic selection in tree breeding: Testing accuracy of prediction models including dominance effect. *BMC Proc.* 5:O13. doi:10.1186/1753-6561-5-S7-O13
- Dray, S., and A.-B. Dufour. 2007. The ade4 package implementing the duality diagram for ecologists. *J. Stat. Softw.* 22(4). doi:10.18637/jss.v022.i04
- Eberhart, S.A., and W.A. Russell. 1966. Stability parameters for comparing varieties. *Crop Sci.* 6:36–40. doi:10.2135/cropsci1966.0011183X000600010011x
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi:10.1371/journal.pone.0019379
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4:250–255. doi:10.3835/plantgenome2011.08.0024
- Falconer, D.S., and T.F. Mackay. 1996. Introduction to quantitative genetics. Longman, Harlow, UK.
- Goddard, M.E., and B. Hayes. 2007. Genomic selection. *J. Anim. Breed. Genet.* 124:323–330. doi:10.1111/j.1439-0388.2007.00702.x
- Habier, D., R. Fernando, and J. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397. doi:10.1534/genetics.107.081190
- Habier, D., R.L. Fernando, and J.C. Dekkers. 2009. Genomic selection using low-density marker panels. *Genetics* 182:343–353. doi:10.1534/genetics.108.100289
- He, S., A.W. Schulthess, V. Mirdita, Y. Zhao, V. Korzun, R. Bothe, et al. 2016. Genomic selection in a commercial winter wheat population. *Theor. Appl. Genet.* 129:641–651. doi:10.1007/s00122-015-2655-1
- Heffner, E.L., J.-L. Jannink, H. Iwata, E. Souza, and M.E. Sorrells. 2011a. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci.* 51:2597–2606. doi:10.2135/cropsci2011.05.0253
- Heffner, E.L., J.-L. Jannink, and M.E. Sorrells. 2011b. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4:65–75. 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.* 50:1681–1690. doi:10.2135/cropsci2009.11.0662
- Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. *Crop Sci.* 49:1–12. doi:10.2135/cropsci2008.08.0512
- Heslot, N., J.-L. Jannink, and M.E. Sorrells. 2013. Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Sci.* 53:921–933. doi:10.2135/cropsci2012.07.0420
- Heslot, N., H.-P. Yang, M.E. Sorrells, and J.-L. Jannink. 2012. Genomic selection in plant breeding: A comparison of models. *Crop Sci.* 52:146–160. doi:10.2135/cropsci2011.06.0297
- Hickey, J.M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, et al. 2014. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* 54:1476–1488. doi:10.2135/cropsci2013.03.0195
- Hoffstetter, A., A. Cabrera, M. Huang and C. Sneller. 2016. Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat. *G3: Genes, Genomes, Genet.* 6:2919–2928. doi:10.1534/g3.116.032532
- Huang, M., A. Cabrera, A. Hoffstetter, C. Griffey, D. Van Sanford, J. Costa, et al. 2016. Genomic selection for wheat traits and trait stability. *Theor. Appl. Genet.* 129:1697–1710. doi:10.1007/s00122-016-2733-z

- 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.* 128:145–158. doi:10.1007/s00122-014-2418-4
- Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genomics* 9:166–177. doi:10.1093/bfpg/elq001
- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, et al. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127:595–607. doi:10.1007/s00122-013-2243-1
- Jarquín, D., J. Specht, and A. Lorenz. 2016. Prospects of genomic prediction in the USDA Soybean Germplasm Collection: Historical data creates robust models for enhancing selection of accessions. *G3: Genes, Genomes, Genet.* 6:2329–2341. doi:10.1534/g3.116.031443
- Jiang, Y., A.W. Schulthess, B. Rodemann, J. Ling, J. Plieske, S. Kollers, et al. 2016. Validating the prediction accuracies of marker-assisted and genomic selection of *Fusarium* head blight resistance in wheat using an independent sample. *Theor. Appl. Genet.* 130:471–482. doi:10.1007/s00122-016-2827-7
- Lado, B., P.G. Barrios, M. Quincke, P. Silva, and L. Gutiérrez. 2016. Modeling genotype  $\times$  environment interaction for genomic selection with unbalanced data from a wheat breeding program. *Crop Sci.* 56:2165–2179. doi:10.2135/cropsci2015.04.0207
- Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, et al. 2012. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. doi:10.1093/bioinformatics/bts444
- Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.-L. Jannink et al. 2015. Increased prediction accuracy in wheat breeding trials using a marker $\times$  environment interaction genomic selection model. *G3: Genes, Genomes, Genet.* 5:569–582. doi:10.1534/g3.114.016097
- Lorenz, A.J., and K.P. Smith. 2015. Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. *Crop Sci.* 55:2657–2667. doi:10.2135/cropsci2014.12.0827
- Lorenz, A., K. Smith, and J.-L. Jannink. 2012. Potential and optimization of genomic selection for *Fusarium* head blight resistance in six-row barley. *Crop Sci.* 52:1609–1621. doi:10.2135/cropsci2011.09.0503
- Meuwissen, T., B. Hayes, and M. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Michel, S., C. Ametz, H. Gungor, B. Akgöl, D. Epure, H. Grausgruber, et al. 2016a. Genomic assisted selection for enhancing line breeding: Merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials. *Theor. Appl. Genet.* 130:363–376. doi:10.1007/s00122-016-2818-8
- Michel, S., C. Ametz, H. Gungor, D. Epure, H. Grausgruber, F. Löschenberger, et al. 2016b. Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theor. Appl. Genet.* 129:1179–1189. doi:10.1007/s00122-016-2694-2
- Moser, G., M.S. Khatkar, B.J. Hayes, and H.W. Raadsma. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genet. Sel. Evol.* 42:37. doi:10.1186/1297-9686-42-37
- Oakey, H., B. Cullis, R. Thompson, J. Comadran, C. Halpin and R. Waugh. 2016. Genomic selection in multi-environment crop trials. *G3: Genes, Genomes, Genet.* 6:1313–1326. doi:10.1534/g3.116.027524
- Piepho, H., J. Möhring, A. Melchinger, and A. Büchse. 2008. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161:209–228. doi:10.1007/s10681-007-9449-8
- Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, et al. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5:103–113. doi:10.3835/plantgenome2012.06.0006
- R Development Core Team. 2008. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.
- Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J.-L. Jannink, and M. Sorrells. 2012. Evaluation of genomic prediction methods for *Fusarium* head blight resistance in wheat. *Plant Genome* 5:51–61. doi:10.3835/plantgenome2012.02.0001
- Rutkoski, J., J. Poland, S. Mondal, E. Autrique, L.G. Pérez, J. Crossa, et al. 2016. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genet.* 6:2799–2808. doi:10.1534/g3.116.032888
- Rutkoski, J., R. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.-L. Jannink, et al. 2015. Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. *Plant Genome* 8(2). doi:10.3835/plantgenome2014.10.0074
- Rutkoski, J.E., E.L. Heffner, and M.E. Sorrells. 2011. Genomic selection for durable stem rust resistance in wheat. *Euphytica* 179:161–173. doi:10.1007/s10681-010-0301-1
- Rutkoski, J.E., J.A. Poland, R.P. Singh, J. Huerta-Espino, S. Bhavani, H. Barbier, et al. 2014. Genomic selection for quantitative adult plant stem rust resistance in wheat. *Plant Genome* 7(3). doi:10.3835/plantgenome2014.02.0006
- Saint Pierre, C., J. Burgueño, J. Crossa, G.F. Dávila, P.F. López, E.S. Moya, et al. 2016. Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Sci. Rep.* 6:27312. doi:10.1038/srep27312
- Sallam, A.H., and K.P. Smith. 2016. Genomic selection performs similarly to phenotypic selection in barley. *Crop Sci.* 56:2871–2881. doi:10.2135/cropsci2015.09.0557
- SAS Institute. 2008 SAS/STAT user's guide, version 9.2. SAS Inst., Cary, NC
- Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78:629–644. doi:10.1086/502802
- Schulz-Streeck, T., J.O. Ogotu, A. Gordillo, Z. Karaman, C. Knaak, and H.P. Piepho. 2013. Genomic selection allowing for marker-by-environment interaction. *Plant Breed.* 132:532–538. doi:10.1111/pbr.12105
- Shin, J.-H., S. Blay, B. McNeney, and J. Graham. 2006. LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* 16:1–10. doi:10.18637/jss.v016.c03
- Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard, E. Redoña, et al. 2015. Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11:e1004982. doi:10.1371/journal.pgen.1004982 [erratum: 11:e1005350].

- Spindel, J.E., and S.R. McCouch. 2016. When more is better: How data sharing would accelerate genomic selection of crop plants. *New Phytol.* 212:814–826. doi:10.1111/nph.14174
- Tiezzi, F., G. de Los Campos, K.P. Gaddis, and C. Maltecca. 2017. Genotype by environment (climate) interaction improves genomic prediction for production traits in US Holstein cattle. *J. Dairy Sci.* 100:2042–2056. doi:10.3168/jds.2016-11543
- Vazquez, A., G. Rosa, K. Weigel, G. De los Campos, D. Gianola, and D. Allison. 2010. Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. *J. Dairy Sci.* 93:5942–5949. doi:10.3168/jds.2010-3335
- Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, et al. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12:787–796. doi:10.1111/pbi.12183
- Warnes, G. and F. Leisch. 2005. *Genetics: Population genetics*. R package version 1:1–5. R Found. Stat. Comput., Vienna.
- Weigel, K., G. De Los Campos, O. González-Recio, H. Naya, X. Wu, N. Long, et al. 2009. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. *J. Dairy Sci.* 92:5248–5257. doi:10.3168/jds.2009-2092
- Yao, C., G. de los Campos, M. VandeHaar, D. Spurlock, L. Armentano, M. Coffey, et al. 2016. 0307 Use of marker  $\times$  environment interaction whole genome regression model to incorporate genetic heterogeneity for residual feed intake, dry matter intake, net energy in milk, and metabolic body weight in dairy cattle. *J. Anim. Sci.* 94:146–147. doi:10.2527/jam2016-0307
- Zhang, X., A. Sallam, L. Gao, T. Kantarski, J. Poland, L.R. DeHaan, et al. 2016. Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. *Plant Genome* 9(1). doi:10.3835/plantgenome2015.07.0059
- Zhao, Y., M. Mette, M. Gowda, C. Longin, and J. Reif. 2014. Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity* 112:638–645. doi:10.1038/hdy.2014.1
- Zobel, R.W., M.J. Wright, and H.G. Gauch. 1988. Statistical analysis of a yield trial. *Agron. J.* 80:388–393. doi:10.2134/agronj1988.00021962008000030002x