

Multienvironment and Multitrait Genomic Selection Models in Unbalanced Early-Generation Wheat Yield Trials

Brian P. Ward,^{*} Gina Brown-Guedira, Priyanka Tyagi, Frederic L. Kolb, David A. Van Sanford, Clay H. Sneller, and Carl A. Griffey

ABSTRACT

The majority of studies evaluating genomic selection (GS) for plant breeding have used single-trait, single-site models that ignore genotype \times environment interaction (GEI) effects. However, such studies do not accurately reflect the complexities of many applied breeding programs, and previous papers have found that models that incorporate GEI effects and multiple traits can increase the accuracy of genomic estimated breeding values (GEBVs). This study's goal was to test GS methods for prediction in scenarios that simulate early-generation yield testing by correcting for field spatial variation, and fitting multienvironment and multitrait models on data for 14 traits of varying heritability evaluated in unbalanced designs across four environments. Corrections for spatial variation increased across-environment trait heritability by 25%, on average, but had little effect on model predictive ability. Results between all models were generally equivalent when predicting the performance of newly introduced genotypes. However, models incorporating GEI information and multiple traits increased prediction accuracy by up to 9.6% for low-heritability traits when phenotypic data were sparsely collected across environments. The results suggest that GS models using multiple traits and incorporating GEI effects may best be suited to predicting line performance in new environments when phenotypic data have already been collected across a subset of the total testing environments.

B.P. Ward and C.A. Griffey, Dep. of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA 24061; B.P. Ward, current address, Dep. of Crop and Soil Sciences, North Carolina State Univ., Raleigh, NC 27695; G. Brown-Guedira, USDA-ARS Plant Science Research Unit, Raleigh, NC 27695; P. Tyagi, Dep. of Crop and Soil Sciences, North Carolina State Univ., Raleigh, NC 27695; F.L. Kolb, Dep. of Crop Sciences, Univ. of Illinois Urbana, IL 61801; D.A. Van Sanford, Dep. of Plant and Soil Sciences, Univ. of Kentucky, Lexington, KY 40546; C.H. Sneller, Ohio Agricultural Research and Development Center, Ohio State Univ., Wooster, Ohio 44691. Received 19 Mar. 2018. Accepted 1 Dec. 2018. ^{*}Corresponding author (boward2@ncsu.edu). Assigned to Associate Editor Jesse Poland.

Abbreviations: AR1, first-order autoregressive residual structure; BLUP, best linear unbiased prediction; CV, cross validation; EMMA, efficient mixed model association; FLSG, flag leaf stay green; GBLUP, genomic best linear unbiased prediction; GEBV, genomic estimated breeding value; GEI, genotype \times environment interaction; GGE, genotype + genotype \times environment; GRM, genomic relationship matrix; GS, genomic selection; GSQM, grains per square meter; GW, grain weight; HD, heading date; HT, plant height; LD, linkage disequilibrium; MAT, physiological maturity date; MEI, marker \times environment interaction; NDVI, normalized difference vegetation index at Zadok's growth stage 25; NIR, near-infrared; PROT, whole-grain protein content; PYT, preliminary yield test; QTL, quantitative trait locus; SNP, single nucleotide polymorphism; SPH, seeds per head; SSQM, spikes per square meter; STARCH, whole-grain starch content; TKW, thousand-kernel weight; TP, training population; TWT, test weight; VP, validation population; YLD, grain yield.

GENOMIC SELECTION (GS) is a method of marker-based selection that uses a large number of markers spread throughout the genome, such that every quantitative trait locus (QTL) affecting a trait is assumed to be in high linkage disequilibrium (LD) with at least one marker. As such, it is an extension

Published in *Crop Sci.* 59:491–507 (2019).
doi: 10.2135/cropsci2018.03.0189

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

of the earlier method of marker-assisted selection, in which a few markers associated with trait QTLs are used to perform selection in combination with phenotypic information (Meuwissen et al., 2001). Genomic selection models typically use a training population (TP), which is both genotyped and phenotyped, to calculate genomic estimated breeding values (GEBVs) for individuals in a validation population (VP), which is only genotyped. The accuracy of a GS model could hypothetically be measured by the correlation coefficient between the VP's GEBVs and true breeding values. However, the true breeding value is not known, and hence the correlation between the GEBVs and estimated breeding values (most commonly phenotypic means) is calculated instead. This correlation is referred to as a model's *predictive ability* (Ould Estaghirou et al., 2013). There are several properties of the TP, and its relationship to the VP, which are critical to achieving high prediction accuracy. These include the TP's census (i.e., actual) population size, its effective population size, and the degree of relatedness between the TP and VP (Pszczola et al., 2013). The effective population size is almost always lower than the census population size and may be defined as the number of individuals that would give rise to the observed level of inbreeding in an idealized population with random mating (Falconer and Mackay, 1996).

Various methods for increasing the predictive ability of GS models have been explored. One is the correction of field spatial variation. Methods for spatial correction include the use of blocking, with resolvable incomplete block designs such as the α -lattice (Patterson and Williams, 1976) being popular in early-generation testing. Additional methods use models to perform post hoc spatial corrections at a finer scale for field experiments employing a variety of designs. Cullis et al. (1998) reported that these methods can produce better estimates of genotypic effects than the use of blocking designs alone. A popular class of models for performing spatial corrections are the first-order autoregressive (AR1) models (Gilmour et al., 1997). First-order autoregressive functions model an exponential decay in the correlation between adjacent plots' phenotypic values as a function of the number of intervening plots plus one (Piepho et al., 2015). Models incorporating an AR1 residual structure may correct for spatial variation in a single dimension, or in a separable two-dimensional fashion (AR1 \times AR1), and may optionally include a "nugget" or unit residual term to partition out variation due to random measurement error (Cullis et al., 1998). Studies have presented varying results on the effects of spatial corrections on GS model predictive ability. Lado et al. (2013) found that correcting for spatial variation was crucial for increasing predictive ability. Bernal-Vasquez et al. (2014) reported that AR1 models generally did not significantly increase predictive ability and recommended

simpler models fitting row and column effects. Elias et al. (2018) explored the use of multiple spatial parametric kernels in a cassava (*Manihot esculenta* Crantz) breeding trial and observed a median increase in predictive ability of 3.4% when accounting for spatial trends.

Several studies have also examined the possibility of increasing GS model predictive ability by incorporating genotype \times environment (GEI) information to model relationships between environments. Early GS studies performed in plants typically used multi-environment ANOVA models to calculate adjusted means for genotypes. Critically, an ANOVA model fit in this context ignores GEI by assuming a uniform covariance between pairs of environments (Isik et al., 2017). Burgueño et al. (2012) first incorporated GEI effects into a multi-environment GS model using pedigree and genome-wide marker data. Subsequent studies have further investigated potential gains in GS accuracy enabled by the incorporation of GEI information. Guo et al. (2013) studied the traits leaf width and leaf length in a maize (*Zea mays* L.) nested association mapping panel and found that median multi-environment model predictive ability was higher than the predictive ability of single-environment models both across and within populations. Jarquín et al. (2014) introduced a reaction-norm model modeling GEI as functions of markers and environmental covariates. Their results demonstrated that the incorporation of GEI effects into the model could increase both across- and within-environment predictive ability, depending on the specific model specification and cross-validation procedure. Zhang et al. (2015) studied GS across multiple environments using a biparental maize population and found that complex traits such as grain yield benefitted the most from incorporation of GEI effects, whereas simpler traits such as anthesis date demonstrated more modest gains in predictive ability. Lado et al. (2016) evaluated GS for grain yield in wheat (*Triticum aestivum* L.) in a total of 35 environments. They used a strategy of performing GS prediction within mega-environments to increase the ability to predict genotype performance in new environments.

Several studies have also performed GS using multiple traits. Studies using simulated datasets have suggested that multitrait models can be used to increase predictive ability for low-heritability traits that are correlated with higher-heritability traits, or when a trait is simply too difficult or expensive to measure in all individuals within a population (Calus and Veerkamp, 2011; Jia and Jannink, 2012; Hayashi and Iwata, 2013; Guo et al., 2014). Several studies have subsequently assessed multitrait GS in datasets consisting of nonsimulated phenotypic data (Jia and Jannink, 2012; Pszczola et al., 2013; dos Santos et al., 2016; Rutkoski et al., 2016; Schulthess et al., 2016; Wang et al., 2016).

Thus far, the application of GS methods to preliminary yield tests (PYTs) has been understudied. Preliminary

yield tests present several common challenges to the accurate assessment of genotype performance, including (i) a limited number of total testing environments, (ii) few or possibly no replications within each environment, (iii) unbalanced designs across environments (especially across years) due to annual selections and advancement, and (iv) limited seed availability. In this study, the predictive ability of different GS models was evaluated under the conditions listed above for a variety of quantitative traits with varying heritability and genetic architectures. Endelman et al. (2014) examined the question of optimal PYT designs when genome-wide marker data are available by evaluating both genomic predictive ability and response to selection for a range of experimental designs. However, their study only focused on the trait grain yield and used biparental populations. Lado et al. (2016) studied the use of GS models incorporating GEI data to predict grain yield in highly unbalanced datasets, although they used a relatively large number of environments. The objective of the present study was to evaluate the effects of several methods on GS model predictive ability, including correction for spatial variation, adoption of models incorporating GEI effects, and multitrait analyses. We additionally sought to evaluate the utility of GS models in panels representing more ad hoc assemblies of genotypes, rather than assemblies of large families of full or half-sibs.

MATERIALS AND METHODS

Germplasm Selection

The study was conducted over 2 yr and included a total of 329 genotypes. Of these, 41 genotypes were tested in both years. Of the remaining genotypes, half (144) were tested only in the first year, and the other half (144) were tested only in the second year. All genotype-year combinations are listed in Supplemental Table S1. Within each year, genotypes were sourced from breeding programs in Illinois (31), Kentucky (30), Missouri (2), and Virginia (122). Seven genotypes were removed during quality filtering of the genotypic data, leaving a total of 322 genotypes among both years used for further analysis. Five checks were included in the study, including ‘Bess’, ‘Branson’, IL00-8250, ‘Roane’, and ‘Shirley’. These lines were selected as they are soft wheat lines that have frequently been grown in production and used as parents in the mid-Atlantic and Midwest regions. With the exception of checks and several older cultivars, the majority of genotypes were either F₄ or F₅ filial generation.

Experimental Design and Field Management

Experimental plots were planted in a total of four environments (two locations in 2 yr) in the 2013–2014 and 2014–2015 winter wheat growing seasons. Within each year, trials were planted at Kentland Farm near Blacksburg, VA (Guernsey [fine, mixed, superactive, mesic Aquic Hapludalfs]–Hayter [fine-loamy, mixed, active, mesic Ultic Hapludalfs] silt loams; 37.1965° N, 80.5718° W; 531 m asl) and the Eastern Virginia Agricultural

Research and Extension Center in Warsaw, VA (Kempsville sandy loam [fine-loamy, siliceous, subactive, thermic Typic Hapludults]; 37.9879° N, 76.7770° W; 40 m asl). A generalized randomized complete block design with two replications was used in each environment.

Each experimental unit consisted of a seven-row plot with a length of 2.74 m, width of 0.91 m, row spacing of 15.2 cm, and a harvested area of 2.49 m². All plots were sown with 70 g of seed. Seed was treated with Raxil MD fungicide (0.48% tebuconazole, 0.64% metalaxyl; Bayer CropScience) at a rate of 2.95 mL a.i. kg⁻¹ seed, and Gaucho 600 flowable insecticide (48.7% imidacloprid, Bayer CropScience) at a rate of 0.7 mL a.i. kg⁻¹ seed. At each location, seeds were planted to roughly coincide with the average date of first frost (Supplemental Table S2).

At Blacksburg and Warsaw, several tiller counts representative of the test area as a whole were used to calculate ideal N application rates at Zadok’s growth stage 25 (Zadoks et al., 1974) in the spring. Once plants reached Zadok’s growth stage 30, tissue samples were collected by cutting handfuls of leaf material from roughly 25 randomly selected sites per environment. For each environment, collected leaf tissue was then mixed thoroughly and analyzed for N content by the Dumas method (reviewed in Muñoz-Huerta et al., 2013) to determine ideal additional N application rates, per standard regional recommendations from the Virginia Cooperative Extension Service (Alley et al., 1993). Chemicals were applied as needed to control lodging and pest pressure in each environment (Supplemental Table S2).

Phenotyping

Table 1 summarizes the phenotypic traits that were assessed across all environments, with their abbreviations, units of measure, and trait ontologies. Normalized difference vegetative index (NDVI) was measured for each plot at Zadok’s growth stage 25 as described by Phillips et al. (2004) using a Greens-eeker handheld crop sensor (Trimble Agriculture). Heading date (HD) was recorded as the Julian date at which 50% of plant tillers within a plot had extruded heads from the boot. Physiological maturity date (MAT) was defined as the date at which 50% of peduncles within a plot had turned yellow. After plants reached maturity, a 0.914-m cutting of all aboveground plant material was taken from one of the three inner rows of each plot and placed inside a paper bag. The number of spikes per cutting were counted manually to derive an estimate of spikes per square meter (SSQM, count). Cuttings were then threshed on a plot combine (Wintersteiger) with settings optimized to recover as much threshed seed as possible. Threshed seeds were weighed to derive an estimate of grain weight per meter of row (GW, g). The total number of seeds threshed from each cutting were then counted on a Count-A-Pak optical seed counter (Seedburo Equipment) to derive an estimate of grains per square meter (GSQM, count). The number of heads was divided by the number of seeds to derive an estimate of seeds per head (SPH, count). Thousand-kernel weight (TKW, g) was then calculated as

$$\text{TKW} = \frac{\text{net weight of threshed seed}}{\text{number of seeds}} \times 1000 \quad [1]$$

Table 1. Phenotypic traits assessed in Blacksburg and Warsaw, VA, during the 2013–2014 and 2014–2015 growing seasons.

Trait	Abbreviation	Units	Trait ontology†	Ontology description
Heading date	HD	Julian d	TO:0000137	Days required for 50% of heads to emerge from boot
Physiological maturity	MAT	Julian d	TO:0000469	Days required for 50% of peduncles to lose green color
Flag leaf stay green	FLSG	d	TO:0000249	Days between heading and flag leaf senescence
NDVI at Zadok's growth stage 25	NDVI	–	CO_321:0000301	Wheat canopy normalized difference vegetation index (NDVI) trait
Mature plant height	HT	cm	TO:0000207	Height of plant from soil surface to tip of spike excluding awns
Grain weight	GW	g dry wt. m ⁻¹ row	TO:0000589	Grain yield (g dry weight) from 1-m row cut at physiological maturity
Grains per square meter	GSQM	grains m ⁻²	CO_321:0000017	Number of grains threshed from 1-m row cutting, converted to square meters
Spikes per square meter	SSQM	spikes m ⁻²	CO_321:0000166	Number of spikes (fertile culms) per unit area, sample or plant
Seeds per head	SPH	count	TO:0002759	Number of grains within an inflorescence
Thousand-kernel weight	TKW	g	TO:0000382	Seed weight estimated by weighing 1000 seeds
Test weight	TWT	g L ⁻¹	TO:0000612	Weight per unit volume of grain at standard moisture level
Yield, 13.5% moisture	YLD	kg ha ⁻¹	TO:0000396	Grain yield standardized to 13.5% moisture equivalence
Near-infrared whole-grain starch	STARCH	%	TO:0000696	Amount of starch present in the plant or plant part
Near-infrared whole-grain protein	PROT	%	CO_321:0000073	Content of protein (corrected to 13.5% moisture basis) in whole wheat grain samples

† Matching trait ontologies in the Planteome database: <http://browser.planteome.org/amigo/search/ontology>.

Plant height (HT, cm) was averaged from two measurements within each plot and was recorded as the distance from the soil surface to the tip of the heads (excluding any awns if present). Plots were harvested at maturity using a Wintersteiger plot combine. Moisture content and test weight of harvested grain was measured using a GAC 2500-AGRI grain analysis computer (Dickey-John Corporation). Grain yield (YLD, kg ha⁻¹) was derived from the yield of each harvested plot at 13.5% moisture equivalence.

Whole grain starch (STARCH, %), and protein (PROT, %) content were estimated via near-infrared (NIR) spectroscopy for subsamples from each plot using an XDS Rapid Content Analyzer (FOSS NIR Systems). Thirty samples, each consisting of 25 g of seed, were sent to Cumberland Valley Analytical Services (Hagerstown, MD) for wet-chemistry analysis of protein content (AOAC International, 2000) to generate calibration curves for the NIR data. Samples were chosen to represent the full range of observed NIR protein content values. The coefficient of determination of a regression between NIR and wet-chemistry-derived protein content values was 0.78 (data not shown). After phenotyping, the R package 'gge' (Wright and Laffont, 2016) was used to generate genotype + genotype-by-environment (GGE) interaction plots (Yan et al., 2000) for each trait using the 41 genotypes tested across all environments (Supplemental Figs. S6a–S6n).

Genotyping, Marker Filtering, and Imputation

Seeds of each genotype were germinated, and genomic DNA was isolated from the leaf tissue of 10-d-old seedlings on an LGC Genomics Oktopure robotic extraction platform, using sbeadex magnetic microparticle reagent kits. Genotyping-by-sequencing was performed on an Illumina HiSeq after a *PstI-MseI* double restriction digest of genomic DNA (Poland et al., 2012). Single nucleotide polymorphism (SNP) calling was performed using TASSEL-GBS in TASSEL 5.2.24 (Bradbury et al., 2007; Glaubitz et al., 2014). Illumina reads were aligned to the International Wheat Genome Sequencing Consortium 'Chinese Spring' v1.0 reference sequence (IWGSC et al., 2018)

using the Burrows–Wheeler Aligner (Li and Durbin, 2009). The genotypic datasets for the 2013–2014 and 2014–2015 material were then jointly filtered to remove SNPs with missing data frequencies >20%, mean sequencing depth less than four, heterozygous call frequencies >15%, and minor allele frequency <5%. After the initial filtering, missing data in the genotypic dataset were imputed and phased using Beagle 4.1 (Browning and Browning, 2007, 2016). After imputation, the dataset was once again filtered to remove SNPs with minor allele frequencies <5%. The imputed genotypic dataset was then filtered in PLINK 1.9 (Chang et al., 2015) to remove all but one SNP in groups of SNPs within 64 bp of each other, to attempt to lessen the inclusion of SNPs misaligned between subgenomes, which typically exhibit low LD with surrounding SNPs. Finally, all but one SNP was removed from groups of SNPs in high LD ($r^2 > 0.8$) using a 250-SNP sliding window, advancing by 10 SNPs with each step. After filtering, a total of 7748 SNPs remained for further analysis. A random subset of 1000 SNPs were selected from the filtered genotypic data to calculate the effective population size of the testing panel using the formula from Hedrick (2011):

$$N_e = 1/(2r^2) \quad [2]$$

where the effective population size (N_e) is a function of the average linkage disequilibrium between SNPs located on different chromosomes (r^2).

Spatial Corrections and Heritability Estimation

For each individual trait–environment combination, the following baseline model was fit:

$$Y_i = \mu + G_i + \epsilon_i \quad [3]$$

where phenotypic response (Y_i) is a function of the within-environment mean (μ), the fixed effect of the i th genotype (G_i), and residual error (ϵ_i). This model was then compared against several AR1 models accounting for spatial variation (Table 2). Models were compared based on several metrics, including the Akaike information criterion (Akaike, 1974), log-likelihood ratio tests,

Table 2. Baseline and spatial-correction models tested for each trait–environment combination.

Model	Fixed term	Random terms	Residual structure†
Baseline	Genotype	–	IID
Randomized complete block	Genotype	Replication	IID
Two-dimensional first-order autoregressive	Genotype	–	AR1 (rows) × AR1 (columns)
Two-dimensional first-order autoregressive plus units	Genotype	–	AR1 (rows) × AR1 (columns) + UNITS
Row-wise first-order autoregressive plus units	Genotype	–	AR1 (rows) × ID (columns) + UNITS
Column-wise first-order autoregressive plus units	Genotype	–	ID (rows) × AR1 (columns) + UNITS

† IID, independently and identically distributed in row and column directions; AR1, first-order autoregressive; UNITS, “nugget” unit residual variance; ID, identically distributed.

the standard error of the variance for differences among genotypes within an environment, and visual inspection of plot-level heatmaps of residuals. The selected best-fitting models were often consistent across environments for a given trait, though this was not always the case (Supplemental Table S3).

After model selection for spatial trend correction, heritability was estimated for each trait–environment combination using the method of Cullis et al. (2006), treating genotype as a random effect:

$$h_g^2 = 1 - \frac{A_{tr}}{2\sigma_G^2} \quad [4]$$

where the generalized heritability (h_g^2) is a function of the average pairwise prediction error between pairs of genotypes within an environment (A_{tr}), and the genotypic variance (σ_G^2). This procedure was performed for both the baseline model described in Eq. [3], and for the selected best-fitting spatially adjusting model selected for each trait–environment combination.

Phenotypic Modelling and Genomic Selection

Four models were fit to test various methods of predicting phenotypic performance using data across multiple environments or traits. For all models, genomic prediction was performed using the genomic best linear unbiased predictor (GBLUP) method (VanRaden, 2007, 2008). Each location–year combination was considered a separate environment, and models were fit using both raw plot-level data, and plot-level fitted values generated from the spatial-adjusting models described above. We followed the advice of Piepho et al. (2015) and retained a model term for replication effect when using the spatially adjusted data. The four models are described below.

Two-Step Adjusted Means Model

The first model is hereafter abbreviated as GBLUP_M and included a first step implementing a mixed linear model in which adjusted means (i.e., genotypic best linear unbiased predictions [BLUPs]) across all environments were calculated, followed by a second mixed linear model in which GEBVs were calculated. For this model, genotypic BLUPs were first estimated with the R package ‘lme4’ (Bates et al., 2015) using an ANOVA model:

$$Y_{ijk} = \mu + G_i + E_j + R_{k(E_j)} + GE_{ij} + \varepsilon_{ijk} \quad [5]$$

where the phenotypic response (Y_{ijk}) is a function of the overall mean (μ), and the random effects of the i th genotype (G_i), the fixed effect of the j th environment (E_j), the random effect of the k th replication nested within the j th environment [$R_{k(E_j)}$], the random effect of the genotype × environment

interaction (GE_{ij}) and the residual error (ε_{ijk}). Note that this model assumes independence between each factor (Isik et al., 2017). The across-environment genotypic BLUPs were then used as inputs to perform GBLUP using the R package ‘rrBLUP’ (Endelman, 2011). Briefly, GBLUP solves the following mixed model for \mathbf{u} :

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad [6]$$

where \mathbf{y} is a vector ($n \times 1$) of phenotypic observations, \mathbf{b} is a vector ($p \times 1$) of fixed effects, \mathbf{u} is a vector ($q \times 1$) of random effects, and $\boldsymbol{\varepsilon}$ is a vector ($n \times 1$) of residuals. \mathbf{X} and \mathbf{Z} are incidence matrices that relate the elements of \mathbf{b} and \mathbf{u} to \mathbf{y} , with dimensions ($n \times p$) and ($n \times q$), respectively. Note that in the case of no additional fixed effects being supplied, $\mathbf{X}\mathbf{b}$ is equivalent to an $n \times 1$ vector of genotypic means ($\boldsymbol{\mu}$), such that Eq. [6] is equivalent to

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad [7]$$

The variance structure of \mathbf{u} is $\sim N(0, \mathbf{G}\sigma_u^2)$, where \mathbf{G} is an $n \times n$ matrix of additive genetic relationships estimated from marker data, and σ_u^2 is the variance of the random effects. The \mathbf{G} matrix was estimated using the R package ‘pedigree’ (Coster, 2012). The variance structure of $\boldsymbol{\varepsilon}$ is $\sim N(0, \mathbf{I}\sigma_\varepsilon^2)$, where \mathbf{I} is an $n \times n$ identity matrix, and σ_ε^2 is the error variance. Package ‘rrBLUP’ uses the efficient mixed model association (EMMA) method for solving mixed linear models (Kang et al., 2008), which is computationally efficient, but limited to the use of a single random term (u , the random genotype effect) in addition to the residual error.

Two-Step Stratified Model

The stratified model, hereafter abbreviated GBLUP_S, is a simplistic approach to producing within-environment performance predictions when limited to a numerical method for solving mixed linear models that can only support a single random term, such as EMMA. The GBLUP_S model handles each environment separately and includes a first step of generating within-environment means, followed by a second step of calculating environment-specific GEBVs. As entries within each environment were balanced with low levels of missing data, the within-environment means were simply calculated as the simple arithmetic mean of the replications. The within-environment means were then used as input to perform GBLUP within each environment separately using package ‘rrBLUP’ as described by Eq. [7]. Note that a separate \mathbf{G} matrix was fit for each environment for the GBLUP_S model, as not all genotypes were present in all environments.

Single-Step Multienvironment Model

A multienvironment model incorporating GEI effects, abbreviated as GBLUP_{GEI}, was fit to plot-level data in a single step using ASreml-R (Butler et al., 2009). This model treats measurements of the same trait across different environments as separate traits (Falconer and Mackay, 1996), and has the general form

$$\mathbf{Y}_{ijk} = \mu + E_j + GE_{ij} + R_{k(E_j)} + \varepsilon_{ijk} \quad [8]$$

The phenotypic response (\mathbf{Y}_{ijk}) is a vector consisting of concatenated subvectors ($\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$) for phenotypes measured in each of m environments and is a function of the overall mean (μ), the fixed effect of the j th environment (E_j), and the random effects of the genotype \times environment interaction term (GE_{ij}), the k th replication nested within the j th environment [$R_{k(E_j)}$], and the residual error (ε_{ijk}). Note that genotypic main effects are still estimable, despite not being explicitly included in the model formula. For the multienvironment model, $\text{var}(GE)$ is defined as the Kronecker product (\otimes) between the \mathbf{G} matrix and the genetic variance–covariance matrix between environments ($\boldsymbol{\rho}$), such that $\text{var}(GE) \sim N[0, (\mathbf{G} \otimes \boldsymbol{\rho}) \sigma_{GE}^2]$ where σ_{GE}^2 is the variance due to genotype \times environment interaction. The resulting square matrix has dimensions $(n \times m) \times (n \times m)$. Various forms of $\boldsymbol{\rho}$, entailing different model assumptions and numbers of parameters to estimate, may be fit. For instance, defining a $\boldsymbol{\rho}$ matrix with a single uniform variance for all environments (σ_E^2) on diagonal elements, and a single uniform covariance between all pairs of environments [$\text{cov}(E)$] on off-diagonal elements yields a “compound symmetry” model, which is equivalent to a traditional multisite ANOVA model as described in Eq. [5] (Isik et al., 2017):

$$\boldsymbol{\rho} = \begin{bmatrix} \sigma_E^2 & \text{cov}(E) & \cdots & \text{cov}(E) \\ \text{cov}(E) & \sigma_E^2 & \cdots & \text{cov}(E) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(E) & \text{cov}(E) & \cdots & \sigma_E^2 \end{bmatrix} \quad [9]$$

In addition to the compound symmetry model, we evaluated several other variance–covariance structures. Fitting a heterogeneous variance model relaxed the assumption of homoscedasticity of the compound symmetry model, allowing for unequal variances for each environment from 1 to m within the $\boldsymbol{\rho}$ matrix:

$$\boldsymbol{\rho} = \begin{bmatrix} \sigma_{E1}^2 & \text{cov}(E) & \cdots & \text{cov}(E) \\ \text{cov}(E) & \sigma_{E2}^2 & \cdots & \text{cov}(E) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(E) & \text{cov}(E) & \cdots & \sigma_{Em}^2 \end{bmatrix} \quad [10]$$

as well as allowing unequal variances for the $R_{k(E_j)}$ replication nested within environment effect. An unstructured model is similar to the heterogeneous variance model except that it allows for fitting unequal covariances between environments in the $\boldsymbol{\rho}$ matrix in addition to unequal variances within environments. However, multienvironment models containing unstructured variance–covariance matrices can easily become overparameterized, causing restricted maximum likelihood

variance component estimates to fail to converge, particularly as the number of environments increases (Oakey et al., 2016). Although unstructured models were selected as the best-fitting models for several traits, they often led to sporadic convergence failures during cross-validation. Multiplicative models using a factor–analytic variance–covariance structure (Smith et al., 2001) are often used as a substitute for unstructured models, and a factor–analytic model with a single factor (FA1) was used for the traits HD and STARCH. Factor analysis seeks to identify common factors which give rise to correlations between variables; in a GEI context, there may be up to $(n \text{ genotypes}) - 1$ or $(m \text{ environments}) - 1$ factors, whichever is less (reviewed in Meyer, 2009). To calculate the covariance matrix $\boldsymbol{\rho}$, let $\boldsymbol{\Gamma}$ be an $m \times k$ matrix of factor loadings l , where k is the number of factors, and let $\boldsymbol{\psi}$ be a heterogeneous diagonal matrix of specific variances. For the case of a single factor, as used in this study, $\boldsymbol{\Gamma}$ becomes a vector:

$$\boldsymbol{\Gamma} = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{bmatrix}, \boldsymbol{\psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_m \end{bmatrix} \quad [11]$$

The $\boldsymbol{\rho}$ matrix may then be calculated as

$$\boldsymbol{\rho} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\psi} = \begin{bmatrix} l_1^2 + \psi_1 & l_1 l_2 & \cdots & l_1 l_m \\ l_1 l_2 & l_2^2 + \psi_2 & \cdots & l_2 l_m \\ \vdots & \vdots & \ddots & \vdots \\ l_1 l_m & l_2 l_m & \cdots & l_m^2 + \psi_m \end{bmatrix} \quad [12]$$

Residual variance for the GEI model is $\sim N[0, (\mathbf{R}_0 \otimes \mathbf{I}) \sigma_\varepsilon^2]$, where \mathbf{R}_0 is an $m \times m$ matrix of residuals across environments, with diagonal elements as the residual variance (σ_ε^2) within each environment, and off-diagonal elements equal to zero:

$$\mathbf{R}_0 = \begin{bmatrix} \sigma_{\varepsilon 1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{\varepsilon 2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{\varepsilon m}^2 \end{bmatrix} \quad [13]$$

In the case of the compound symmetry model, all variances in \mathbf{R}_0 are assumed equal ($\sigma_{\varepsilon 1}^2 = \sigma_{\varepsilon 2}^2 = \cdots = \sigma_{\varepsilon m}^2$). The heterogeneous variance and factor analytic models relaxed the assumption to allow for different residual variances within each environment. \mathbf{I} is an identity matrix of dimensions $n \times n$. Thus, the full matrix modeling the residual variance has dimensions $(m \times n) \times (m \times n)$. Models were selected for each trait based on multiple criteria, including the Akaike information criterion, log likelihood estimates, and model parsimony (Supplemental Table S4).

Single-Step Multienvironment, Multitrait Model

Finally, a multienvironment, multitrait model, abbreviated as GBLUP_{MV}, was fit to plot-level data using ASreml-R. This was an ANOVA model as described in Eq. [5], except that the response, Y_{ijk} , consisted of a matrix of dimensions $n \times t$,

with n being the number of plot-level observations, and t being the number of traits included in the analysis. Unstructured variance–covariance matrices were used to model the relationships between traits for residuals, genotypic main effects (G_j), and genotype \times environment interaction (GE_{ij}). The replication nested within environment effect [$R_{k(E_j)}$] was assumed to be independent between traits. Note that mixed linear modeling software such as ASReml cannot directly estimate separable three-way variance–covariance structures (i.e., genotype \otimes environment \otimes trait) when each of the contributing factors itself has a complex variance–covariance structure. Modeling systems with this degree of complexity is a more involved process. Bayesian methods for handling complex three-way variance–covariance structures have been proposed for continuous data (Montesinos-López et al., 2016) and count data (Montesinos-López et al., 2017). Subsequently, a more computationally efficient method for producing ratings of genotypes for multitrait, multienvironment datasets using recommender systems has been proposed (Montesinos-López et al., 2018).

Cross-Validation

Random subset cross-validation was used to assess model predictive ability by correlating the GEBVs generated by all GBLUP models for lines in the VP against their phenotypic adjusted means (i.e., BLUPs), which were calculated by first running each model with all available data, assuming independence between genotypes. For each model, the cross-validation process was then repeated 50 times, randomly dividing the total phenotypic observations into a TP and VP, omitting phenotypic data for observations in the VP, predicting GEBVs for the VP from the training set, correlating GEBVs with genotypic BLUPs, and finally averaging across the cross-validation replicates. Note that random subset cross-validation is nonexhaustive (i.e., a single cross-validation replication consists of a single random training–validation split). For the GBLUP_S model, the cross validation entailed 50 replications within each environment. For the two-step models (GBLUP_M and GBLUP_S), observations consisted of genotypes, and only the predictive ability for main genotypic effects were recorded.

For the one-step models (GBLUP_{GEI} and GBLUP_{MV}), observations consisted of plot-level data, and predictive ability for both genotypic main effects and genotype \times environment interaction effects were recorded. For these models, two separate cross-validation schemes introduced by Burgueño et al. (2012), CV1 and CV2, were used (Supplemental Table S5a). For CV1, genotypes were assigned to either the TP or the VP across all environments. Therefore, CV1 simulates the introduction of new genotypes into a breeding program. For the CV2 cross-validation scheme, specific genotype–environment combinations were assigned to the TP, ensuring that for each genotype only the phenotypic data from at most a single trait or environment would be assigned to the validation population. Therefore, the CV2 cross-validation scheme simulates a scenario in which data on a particular trait is collected in some environments but not others. Given the unbalanced experimental design used, the CV2 cross-validation scheme primarily simulated a scenario in which phenotypic data is available for a particular genotype across some locations within a year, but not across years (Supplemental Tables S5a and S5b). For both the GBLUP_{GEI} and GBLUP_{MV} models, nomenclature

will be used to designate the model and cross-validation scheme combination, so that, for instance, GBLUP_{GEI}/CV1 will refer to the use of the GEI model with the CV1 cross-validation scheme, whereas GBLUP_{MV}/CV2 will refer to the use of the multitrait, multienvironment model with the CV2 cross-validation scheme.

Satisfactory predictions of genotype performance cannot be obtained when using CV1 cross-validation without the use of a relationship matrix (estimated either from pedigree or from genome-wide markers); otherwise, predictions simply consist of the mean value of genotypes within each environment. However, preliminary analyses suggested that in the case of using CV2 cross-validation, meaningful predictions of genotype performance across and within environments could be formed without any knowledge of relationships among genotypes. Therefore, for the GBLUP_{GEI} model using CV2 cross-validation, we compared the predictions generated from a compound symmetry model fitting the genomic relationship matrix (GRM) against the predictions generated from the same model fitting an identity matrix (**I**) in place of the GRM to assume independence among genotypes.

For the multitrait, multienvironment model, the CV2 cross-validation scheme was performed in both a “linked” manner (i.e., the same genotype–environment combinations were assigned to the VP for each trait), and in an “unlinked” manner (i.e., genotype–environment combinations were randomly assigned to the VP for each trait; Supplemental Table S5b). The CV1 cross-validation scheme was always performed in a “linked” fashion for the multitrait, multienvironment model, as it simulates a scenario in which new genotypes lacking any phenotypic data across environments or traits are introduced into a breeding program.

A TP/VP ratio of 80:20 was used for all cross-validation replications. For the CV2 cross-validation scheme, the TP/VP ratio was weighted equally, such that 80% of observations within each environment were used for training. For the GBLUP_M model and the GBLUP_{GEI} or GBLUP_{MV} models using CV1 cross-validation, the 80:20 TP/VP split entailed assigning 258 genotypes to the TP for each trait, and the remaining 64 to the VP. For the GBLUP_{GEI} and GBLUP_{MV} models using CV2 cross-validation, this entailed assigning 581 of the 726 total genotype–environment combinations to the TP, whereas the remaining 145 observations were assigned to the VP. Finally, \sim 185 genotypes were present in each environment (with some slight variation due to genotyping failures), so that using an 80:20 TP/VP ratio with the GBLUP_S model entailed assigning \sim 148 genotypes within each environment to the TP, with the remaining 37 genotypes being assigned to the VP.

Comparisons among Models

The across-environment predictive abilities were compared between the GBLUP_M model and the GBLUP_{GEI} model using the CV1 cross-validation schemes. Within-environment prediction accuracies were generated by the GBLUP_{GEI}/CV1 model, although these predictions could not be produced by the GBLUP_M model. Within-environment predictive abilities were compared for the GBLUP_{GEI} model using CV2 cross-validation without inclusion of a GRM (i.e., assuming independence between genotypes), the same model and cross-validation scheme with the inclusion of a GRM, and

the GBLUP_S model. Across-environment predictions were compared for the GBLUP_{GEI}/CV2 with and without the inclusion of a GRM; note that across-environment predictions could not be directly estimated from the GBLUP_S model. Finally, the predictive ability of the GBLUP_{MV} model was compared against the predictive ability of the corresponding individual-trait GBLUP_{GEI} models for both CV1 and CV2 cross-validation schemes, using both a set of highly-correlated traits (GW, MAT, and YLD), and a set of statistically uncorrelated traits (flag leaf stay green [FLSG], NDVI, and test weight [TWT]). Both of these sets were selected to contain traits with a range of heritability (Table 3).

RESULTS

General Genotype Performance and Correlation of Traits

For reference, trait names and their abbreviations are shown in Table 1. The GRM calculated from the genotypic data showed several groups of highly interrelated genotypes present within the Virginia and Illinois germplasm (Supplemental Fig. S1), and a principal component analysis of the SNP matrix showed that Virginia genotypes formed a large cluster, with a second, overlapping cluster forming from a combination of Kentucky and Illinois genotypes (Supplemental Fig. S2). Mean across-environment heritability ranged from 0.2 or 0.34 for GW with uncorrected data or spatially corrected data, respectively, to 0.95 or 0.96 for TKW with raw data and spatially corrected data (Table 3). Many traits demonstrated strong positive phenotypic and genetic correlation (e.g., YLD, MAT, and GW), as well as strong negative correlation (e.g., TKW and GSQM; Fig. 1).

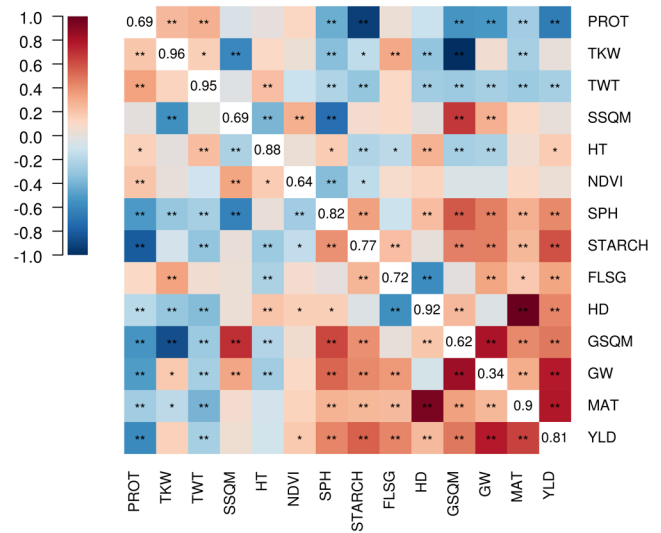


Fig. 1. Correlations among traits calculated from genotypic best linear unbiased predictions (BLUPs). Phenotypic correlations are below the diagonal; genetic correlations are above. Numbers in the diagonal are each trait's generalized heritability averaged across all environments. PROT, whole-grain protein content; TKW, thousand-kernel weight; TWT, test weight; SSQM, spikes per square meter; HT, plant height; NDVI, normalized difference vegetation index at Zadok's growth stage 25; SPH, seeds per head; STARCH, whole-grain starch content; FLSG, flag leaf stay green; HD, heading date; GSQM, grains per square meter; GW, grain weight; MAT, physiological maturity date; YLD, grain yield. * Significant at the 0.05 level. ** Significant at the 0.01 level.

Effect of Spatial Adjustments

The use of various AR1 residual structures to correct for spatial phenotypic trends (Table 2) generally resulted in an increase in trait heritabilities within environments, as well as an increase in the mean heritability across environments

Table 3. Trait descriptive statistics, within-environment and mean across-environment heritability estimates from raw phenotypic data, and within-environment and mean across-environment heritability estimates from spatially corrected data for lines grown in Blacksburg, VA, and Warsaw, VA, for the 2013–2014 and 2014–2015 growing seasons.

Trait†	Units	Descriptive statistics				Raw data heritability†					Spatially corrected data heritability†				
		Min.	Mean	Max.	SD	14Bb	14War	15Bb	15War	Mean	14Bb	14War	15Bb	15War	Mean
FLSG	d	21	28.69	38	2.76	0.26	0.80	0.75	0.55	0.59	0.55	0.83	0.77	0.74	0.72
GSQM	grains m ⁻²	8460	1.85 × 10 ⁴	3.13 × 10 ⁴	3277	0.62	0.50	0.51	0.58	0.55	0.70	0.51	0.59	0.66	0.62
GW	g dry wt. m ⁻¹ row	47.68	96.64	157.9	16.53	0.19	0.14	0.44	0.05	0.20	0.42	0.14	0.57	0.25	0.34
HD	Julian d (1 Jan.)	121	128	136	3.21	0.75	0.94	0.95	0.96	0.90	0.83	0.95	0.96	0.96	0.92
HT	cm	59.69	85.43	119.4	9.29	0.85	0.85	0.84	0.86	0.85	0.91	0.88	0.87	0.87	0.88
MAT	Julian d (1 Jan.)	151	159.3	171	4.72	0.83	0.87	0.86	0.76	0.83	0.89	0.91	0.89	0.90	0.90
NDVI	–	0.26	0.54	0.75	0.08	0.06	0.39	0.41	0.41	0.32	0.57	0.66	0.69	0.64	0.64
PROT	%	9.67	12.34	16.04	1.01	0.66	0.38	0.63	0.74	0.60	0.83	0.47	0.71	0.76	0.69
SPH	count	8.54	21.94	33.29	3.04	0.86	0.82	0.77	0.81	0.82	0.87	0.83	0.77	0.84	0.82
SSQM	spikes m ⁻²	459.3	853	1485	161.2	0.60	0.62	0.45	0.52	0.55	0.65	0.65	0.52	0.92	0.69
STARCH	%	46.88	52.51	56.49	1.41	0.46	0.71	0.19	0.69	0.51	0.81	0.72	0.72	0.84	0.77
TKW	g	24.1	34.57	91.6	3.98	0.96	0.97	0.94	0.93	0.95	0.97	0.98	0.94	0.96	0.96
TWT	g L ⁻¹	652.6	759	810.9	19.7	0.96	0.92	0.91	0.93	0.93	0.97	0.96	0.93	0.96	0.95
YLD	kg ha ⁻¹	3579	6627	9053	1027	0.63	0.45	0.73	0.46	0.57	0.83	0.80	0.78	0.82	0.81

† 14Bb, 2014 Blacksburg, VA; 14War, 2014 Warsaw, VA; 15Bb, 2015 Blacksburg, VA; 15War, 2015 Warsaw, VA.

‡ FLSG, flag leaf stay green; GSQM, grains per square meter; GW, grain weight; HD, heading date; HT, plant height; MAT, physiological maturity date; NDVI, normalized difference vegetation index at Zadok's growth stage 25; PROT, whole-grain protein content; SPH, seeds per head; SSQM, spikes per square meter; STARCH, whole-grain starch content; TKW, thousand-kernel weight; TWT, test weight; YLD, grain yield.

(Table 3). There were a few exceptions—for example, the heritability of the trait GW in Warsaw in 2014 was 0.14 with or without correction for spatial variation. Despite the increase in heritability, the use of fitted values from the AR1 models as input often did not significantly increase the predictive ability of the GBLUP_M or the GBLUP_{GEI}/CV1 models, with the exception of the traits NDVI, for which predictive ability significantly increased from using spatially adjusted values, and MAT, for which the use of spatially adjusted values led to lower predictive ability in the GBLUP_M model (Fig. 2). Confidence intervals for the GBLUP_{GEI} model using CV2 cross-validation tended to be much smaller than those produced by either the GBLUP_M or GBLUP_{GEI}/CV1 models, and the use of spatially adjusted phenotypic values significantly increased predictive ability for the traits GW and NDVI. Significant, though perhaps not meaningful, increases in predictive ability were also observed for the traits PROT, STARCH, and YLD using CV2 cross-validation.

Prediction of Genotype Effects across Environments

A main effect for genotype performance across environments could be predicted from the GBLUP_M model and the GBLUP_{GEI} model using both the CV1 and CV2 cross-validation schemes. Across-environment predictive ability

was generally equivalent between the GBLUP_M model and the GBLUP_{GEI} model using CV1 cross-validation (Table 4, Fig. 3A). In addition, the across-environment predictive ability was generally equivalent between the GBLUP_{GEI}/CV2 model run with and without a GRM (Supplemental Table S6, Fig. 3B). The exceptions were the traits GW and PROT, for which mean predictive abilities were significantly lower when including the GRM. Mean predictive ability for GW was 0.79 and 0.84 when running the model with and without a GRM, respectively; predictive ability for PROT was 0.84 and 0.88 when running the model with and without a GRM. In general, heritability explained a moderate to high portion of the accuracy of GEBVs generated by the three model–cross-validation combinations (Fig. 4), with coefficients of determination between mean across-environment heritability and mean predictive ability of 0.33, 0.42, and 0.87 for the GBLUP_M and the GBLUP_{GEI} models using the CV1 and CV2 cross-validation schemes, respectively. The trait with the highest heritability, TKW, consistently had the highest predictive ability in both the GBLUP_M and GBLUP_{GEI} models, whereas the trait with the lowest heritability, GW, had the lowest predictive ability in the GBLUP_{GEI}/CV2 model. The trait PROT had a mean, across-environment heritability nearly twice that of GW (0.69 vs. 0.34) but had the lowest predictive ability in the GBLUP_M and GBLUP_{GEI}/CV1 models.

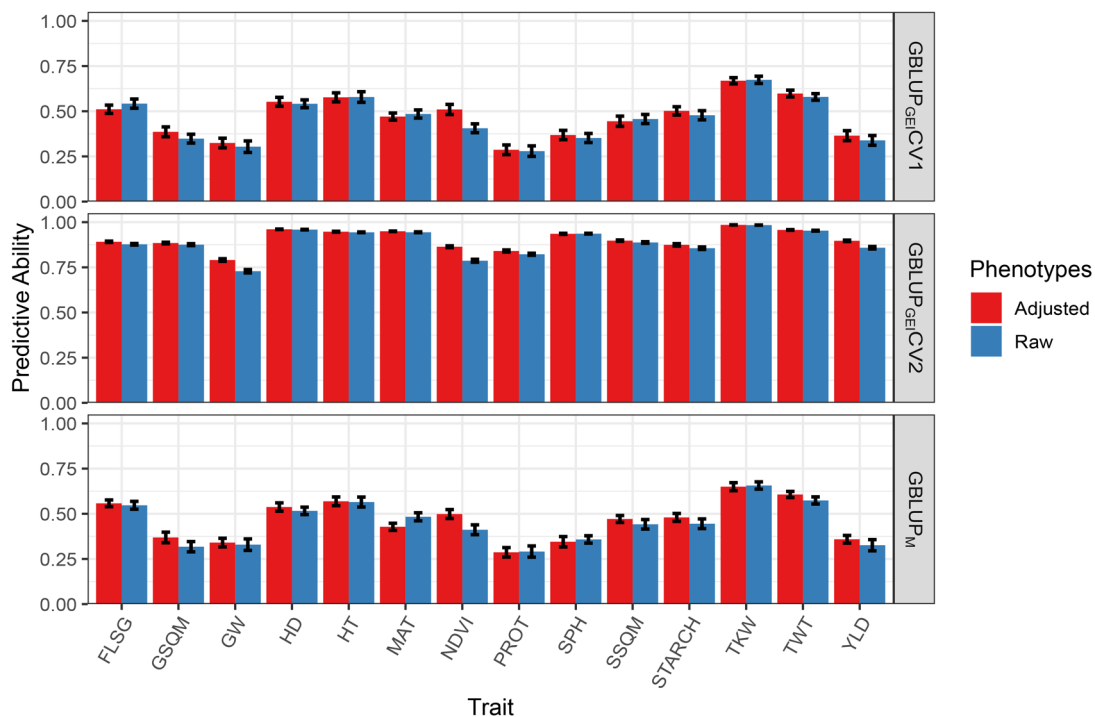


Fig. 2. Comparison of mean predictive ability across 50 replications of random-subsetting cross-validation calculated using either spatially adjusted or raw phenotypic values. Genomic estimated breeding values (GEBVs) were calculated with a model utilizing across-environment adjusted means (GBLUP_M), and a multi-environment model (GBLUP_{GEI}) using either the CV1 or CV2 cross-validation schemes. Bars represent 95% confidence intervals. FLSG, flag leaf stay green; GSQM, grains per square meter; GW, grain weight; HD, heading date; HT, plant height; MAT, physiological maturity date; NDVI, normalized difference vegetation index at Zadok's growth stage 25; PROT, whole-grain protein content; SPH, seeds per head; SSQM, spikes per square meter; STARCH, whole-grain starch content; TKW, thousand-kernel weight; TWT, test weight; YLD, grain yield.

Table 4. Mean predictive ability and 95% confidence intervals (CIs) across all environments included in the study for the two-step adjusted means model (GBLUP_M) and the multienvironment model using the CV1 cross-validation scheme (GBLUP_{GEI}/CV1). For the multienvironment model, genotypic main effect (i.e. across-environment) predictions were generated in addition to within-environment predictions. For the two-step model, only main effect predictions were generated. Summary statistics were calculated across 50 replications of random subsetting cross validation, with 80% of observations used for training, and the remaining 20% used for validation.

Trait‡	GBLUP _M †		GBLUP _{GEI} /CV1†									
	Across envs		Across envs		14Bb		14War		15Bb		15War	
	Mean	CI§	Mean	CI	Mean	CI	Mean	CI	Mean	CI	Mean	CI
FLSG	0.56	0.019	0.51	0.023	0.30	0.035	0.50	0.032	0.53	0.032	0.55	0.026
GSQM	0.37	0.029	0.39	0.027	0.49	0.033	0.33	0.037	0.25	0.045	0.47	0.034
GW	0.34	0.024	0.32	0.027	0.30	0.041	0.21	0.034	0.40	0.033	0.38	0.036
HD	0.54	0.023	0.54	0.025	0.37	0.042	0.47	0.038	0.64	0.022	0.62	0.022
HT	0.57	0.024	0.57	0.025	0.49	0.038	0.45	0.038	0.67	0.025	0.57	0.030
MAT	0.43	0.020	0.45	0.019	0.32	0.029	0.34	0.030	0.58	0.030	0.55	0.030
NDVI	0.50	0.025	0.50	0.029	0.43	0.036	0.43	0.038	0.39	0.046	0.52	0.033
PROT	0.29	0.026	0.29	0.027	0.30	0.044	0.19	0.045	0.39	0.038	0.36	0.036
SPH	0.35	0.029	0.37	0.026	0.45	0.040	0.37	0.041	0.35	0.037	0.40	0.037
SSQM	0.47	0.019	0.45	0.028	0.41	0.039	0.42	0.037	0.45	0.036	0.48	0.036
STARCH	0.48	0.022	0.52	0.021	0.53	0.028	0.52	0.030	0.55	0.028	0.53	0.028
TKW	0.65	0.022	0.67	0.017	0.66	0.027	0.71	0.026	0.66	0.028	0.67	0.026
TWT	0.61	0.017	0.60	0.013	0.63	0.023	0.52	0.024	0.65	0.017	0.65	0.015
YLD	0.36	0.021	0.37	0.027	0.43	0.032	0.34	0.036	0.40	0.036	0.48	0.033

† Across envs, across environments; 14Bb, Blacksburg, VA, 2014; 14War, Warsaw, VA, 2014; 15Bb, Blacksburg, VA, 2015; 15War, Warsaw, VA, 2015.

‡ FLSG, flag leaf stay green; GSQM, grains per square meter; GW, grain weight; HD, heading date; HT, plant height; MAT, physiological maturity date; NDVI, normalized difference vegetation index at Zadok's growth stage 25; PROT, whole-grain protein content; SPH, seeds per head; SSQM, spikes per square meter; STARCH, whole-grain starch content; TKW, thousand-kernel weight; TWT, test weight; YLD, grain yield.

§ CI radius (i.e. margin of error) of 95% confidence interval.

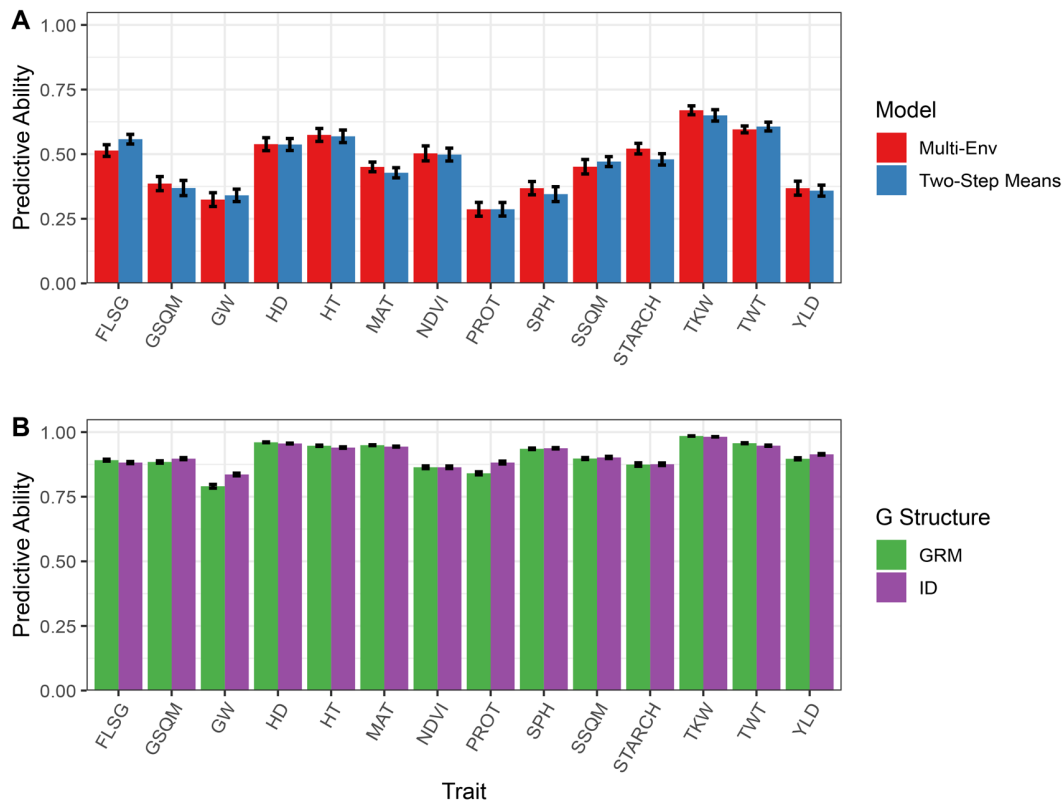


Fig. 3. Predictive abilities for main genotypic effects across environments for (A) CV1 cross-validation using the two-step adjusted means model (GBLUP_M) vs. the one-step multienvironment model (GBLUP_{GEI}), and (B) CV2 cross-validation using the one-step multienvironment model, either assuming independence between genotypes (ID) or using a realized genomic relationship matrix (GRM) to model relationships between genotypes. FLSG, flag leaf stay green; GSQM, grains per square meter; GW, grain weight; HD, heading date; HT, plant height; MAT, physiological maturity date; NDVI, normalized difference vegetation index at Zadok's growth stage 25; PROT, whole-grain protein content; SPH, seeds per head; SSQM, spikes per square meter; STARCH, whole-grain starch content; TKW, thousand-kernel weight; TWT, test weight; YLD, grain yield.

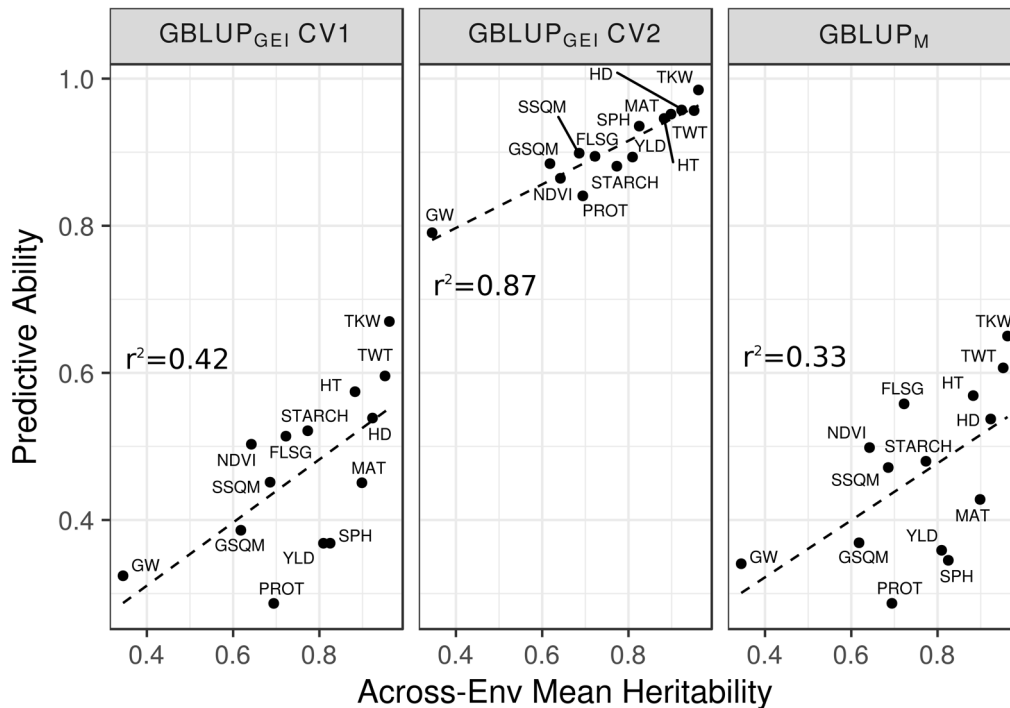


Fig. 4. Mean across-environment heritability estimates vs. mean predictive ability for the multienvironment model (GBLUP_{GEI}) using CV1 and CV2 cross-validation, and the two-step adjusted means model (GBLUP_M). FLSG, flag leaf stay green; GSQM, grains per square meter; GW, grain weight; HD, heading date; HT, plant height; MAT, physiological maturity date; NDVI, normalized difference vegetation index at Zadok's growth stage 25; PROT, whole-grain protein content; SPH, seeds per head; SSQM, spikes per square meter; STARCH, whole-grain starch content; TKW, thousand-kernel weight; TWT, test weight; YLD, grain yield.

Prediction of Genotype Effects within Environments

Predictions of phenotypic performance within environments are not estimable using the GBLUP_M model but can be obtained from the GBLUP_{GEI} model, whether using CV1 or CV2 cross-validation. The within-environment predictions generated by the GBLUP_{GEI}/CV1 model for a given trait were generally closely related to the corresponding across-environment predictions (Table 4), with some occasional larger deviations. For instance, the across-environment prediction accuracy for MAT was 0.43, whereas the prediction accuracies within environments for this trait ranged from 0.34 in Warsaw in 2014, to 0.58 in Blacksburg in 2015.

Within-environment model prediction accuracies were also tested for the CV2 scenario in which data is collected for a particular trait in some environments but not others. Specifically, the within-environment prediction accuracies generated by the GBLUP_S model run with a GRM were compared against those produced by the GBLUP_{GEI}/CV2 model, run either with or without the inclusion of a GRM (Supplemental Table S6). The GBLUP_S model underperformed the GBLUP_{GEI} for nearly all trait-environment combinations, whether the GBLUP_{GEI} model was run with or without the inclusion of the GRM (Fig. 5). Exceptions to this were the trait STARCH tested in Blacksburg in 2015, where the GBLUP_S model outperformed the GBLUP_{GEI}/CV2 model run without a GRM, but

underperformed the GBLUP_{GEI}/CV2 model when a GRM was included. In addition, the GBLUP_S model's predictive ability was statistically equivalent to that of the GBLUP_{GEI}/CV2 model run without a GRM for STARCH and NDVI in Warsaw in 2015, and for GW in Blacksburg in 2015, although the GBLUP_{GEI}/CV2 model run with a GRM once again outperformed the GBLUP_S model for this trait. In general, the GBLUP_{GEI}/CV2 model run with a GRM produced statistically higher prediction accuracies than the same model assuming independence between genotypes for roughly two-thirds of trait-environment combinations, and the two produced statistically equivalent prediction accuracies for the remaining trait-environment combinations. The exceptions were GSQM and PROT tested in Warsaw in 2014, and GSQM, SSQM, and YLD tested in Blacksburg in 2015, for which the use of the GRM led to significantly lower prediction accuracies.

As with the across-environment predictions, heritability explained a moderate to high portion of within-environment predictive ability. For the GBLUP_{GEI}/CV2 model fitting a GRM, the coefficients of determination between within-environment prediction accuracies and corresponding within-environment heritability estimates (generated from spatially adjusted phenotypes) were 0.56, 0.77, 0.70, and 0.45 for the environments Blacksburg in 2014, Warsaw in 2014, Blacksburg in 2015, and Warsaw in 2015, respectively (Supplemental Fig. S3).

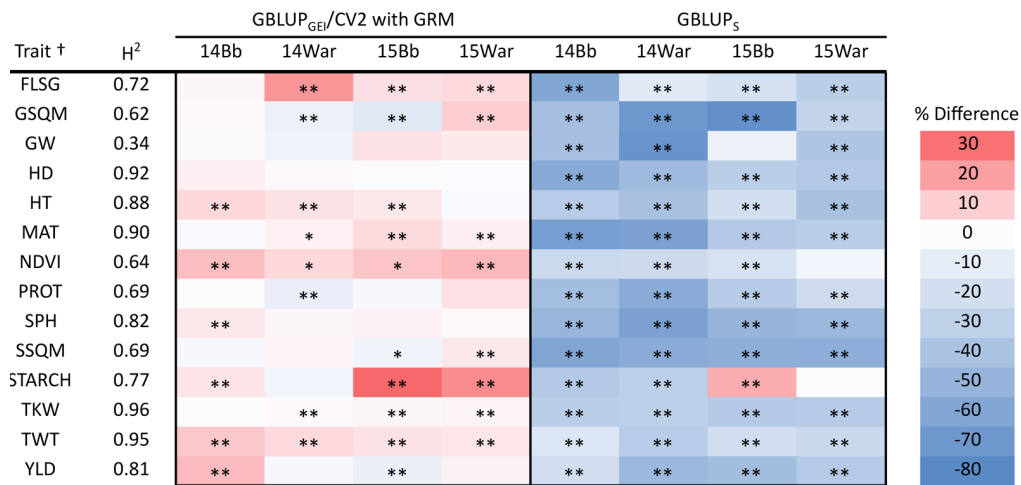


Fig. 5. Difference in mean within-environment prediction accuracies between a multienvironment model using a CV2 cross-validation scheme and assuming independence between genotypes, the same model including a genomic relationship matrix to estimate relationships between genotypes, and a stratified model. Numeric values represent the percentage difference between the listed model, environment, and trait combination and the corresponding environment and trait combination as evaluated by the multienvironment model assuming independence. H^2 , generalized heritability averaged across environments; GBLUP_{GEI}/CV2, multienvironment model using CV2 cross-validation scheme, GBLUP_S stratified model; 14Bb, Blacksburg, VA, 2014; 14War, Warsaw, VA, 2014; 15Bb, Blacksburg, VA, 2015; 15War, Warsaw, VA, 2015. † FLSG, flag leaf stay green; GSQM, grains per square meter; GW, grain weight; HD, heading date; HT, plant height; MAT, physiological maturity date; NDVI, normalized difference vegetation index at Zadok's growth stage 25; PROT, whole-grain protein content; SPH, seeds per head; SSQM, spikes per square meter; STARCH, whole-grain starch content; TKW, thousand-kernel weight; TWT, test weight; YLD, grain yield. * Significant at the 0.05 level. **Significant at the 0.01 level.

Multitrait Genomic Prediction

In this study, the multitrait, multienvironment model offered no advantage over corresponding single-trait, multienvironment models when using the CV1 cross-validation scheme to predict genotype performance across environments, whether using highly correlated traits or uncorrelated traits (data not shown). In contrast, the use of a multiple-trait model did significantly increase predictive ability compared with corresponding single-trait models when using CV2 cross-validation if traits were correlated (Fig. 6). The degree of this increase in predictive ability was proportional to trait heritability, with lower-heritability traits (e.g., GW) exhibiting a larger increase. The high-heritability trait MAT did not exhibit any statistically significant increase in accuracy from the use of the multitrait model. This was not the case for uncorrelated traits, for which the multitrait model offered no benefits over single-trait models. The use of a multitrait model also significantly increased within-environment predictive accuracies for lower-heritability traits in the set of highly correlated traits (Supplemental Fig. S4). In addition, the use of an unlinked CV2 cross-validation scheme led to slightly lower prediction accuracies compared with the use of a linked CV2 scheme for the traits GW (0.85 for unlinked CV2 cross-validation vs. 0.87 for linked CV2) and YLD (0.91 for unlinked CV2 cross-validation vs. 0.92 for linked CV2).

DISCUSSION

The results of this study suggest various scenarios in preliminary wheat breeding trials for which multienvironment and multitrait GS models are and are not well suited. The

introduction of new genotypes into a breeding program was simulated through cross-validation using either the two-step adjusted means model (GBLUP_M), or the multi-environment or multitrait + multienvironment models using the CV1 cross-validation scheme (GBLUP_{GEI}/CV1 and GBLUP_{MV}/CV1). The present study suggests little reason to favor more complicated and computationally intensive multienvironment or multitrait models when attempting to predict the main genotypic effects of untested genotypes, as the GBLUP_M, GBLUP_{GEI}, and GBLUP_{MV} models all performed nearly equivalently under this scenario. As newly introduced genotypes are unphenotyped across all environments, multienvironment models have no additional information to leverage for increasing prediction accuracies in CV1 scenarios (Burgueño et al., 2012). Jarquín et al. (2014) similarly observed no significant increase in CV1 predictive ability when adding a GEI effect to a model incorporating genotype, environment, and environmental covariate main effects, though they did observe a significant increase when genotype × environmental covariate effects were incorporated in addition to the GEI effects. Multienvironment models do have the obvious practical advantage of allowing for direct prediction of genotype performance across and within environments simultaneously, and they generally produced within-environment predictions that were on par with corresponding across-environment predictions in the CV1 scenario. There were certain trait-environment combinations for which prediction accuracy was far higher or lower than the across-environment accuracy

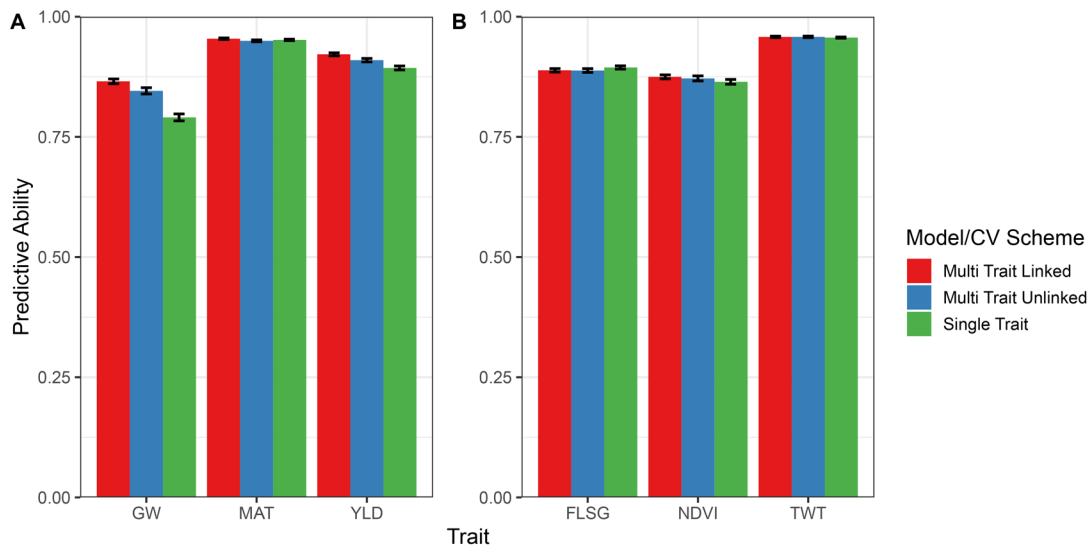


Fig. 6. Mean predictive ability of single trait and multitrait models using CV2 cross-validation schemes for (A) highly correlated traits, and (B) statistically uncorrelated traits. All models shown fit data from multiple environments in a single step. Cross-validation for the multitrait model was performed either in a linked fashion, in which the genotype–environment combinations assigned to the validation population (VP) were constant across traits, or in an unlinked fashion, in which the genotype–environment combinations assigned to the VP varied randomly between traits. All models were tested using 50 cross-validation replications with a 80:20 training population (TP)/VP ratio. Bars represent 95% confidence intervals. GW, grain weight; MAT, physiological maturity date; YLD, grain yield; FLSG, flag leaf stay green duration; NDVI, normalized difference vegetation index at Zadok's growth stage 25; TWT, test weight.

(e.g., FLSG and SPH in Blacksburg in 2014). However, the patterns of GEI calculated from the genotypes tested across all environments (Supplemental Figs. S6a–S6n) did not appear to closely reflect the observed differences in prediction accuracy between environments.

Multienvironment models offered more tangible advantages when genotypes were tested in some environments and not others, a scenario simulated in the CV2 cross-validation scheme. This is a clear benefit in the setting of early-generation yield trials, as it allows for the optimization of sparse resource allocation. It should be noted that in this setting a multienvironment model can provide quite high predictive ability without the inclusion of marker data. The present study found that a multienvironment model including a GRM provided genotype main effect estimates that were approximately equivalent to those produced by the same model assuming independence between genotypes. Rutkoski et al. (2016) reported similar findings in the context of multitrait models (note that the multitrait model used in the present study was only tested with the inclusion of a GRM). This suggests that within a setting in which sparse testing of genotypes is occurring across environments, reasonably accurate across-environment predictions can be generated without any genotyping. However, the use of a GRM to model relationships among genotypes increased within-environment prediction accuracies for many traits. The question of a potential tradeoff point for early-generation yield trials, where the added cost of genotyping is offset by additional genetic gain, is explored by Endelman et al. (2014). Their work suggests that differences in predictive

ability between models incorporating a GRM and those assuming independence between genotypes will be more pronounced as TP size increases. Therefore, it is possible that the similar predictive ability observed in this study when using a GRM vs. assuming independence was simply due to sample size.

The multienvironment models for several traits (GSQM, GW, PROT, and SPH) could only reliably converge when using a compound symmetry variance–covariance structure (Supplemental Table S4). As this model is equivalent to a classical across-environment ANOVA model, we would likewise expect equivalent results to using the two-step approach of running the ANOVA model followed by the model described in Eq. [7]. Despite more complex models being selected as better fitting for several traits, the use of heterogeneous variance and factor analytic models rarely produced an improvement in predictive ability over the simpler compound symmetry model (Supplemental Figs. S5a and S5b). Burgueño et al. (2011) found that the use of factor-analytic variance–covariance structures could increase prediction accuracy up to 6% compared with simpler linear models when GEI is significant. Otherwise, factor-analytic models offered no advantage over the simpler models. This suggests that the environments used in this study may have been similar enough for most traits to safely ignore GEI. More importantly, it suggests that in an applied setting, cross-validation of the TP should be used for the purposes of model selection, as suggested by Bernal-Vasquez et al. (2014), in addition to classical model selection criteria such as the Akaike information criterion.

Several previous studies reported that multitrait GS models could be used to increase predictive ability for low-heritability traits that are highly correlated with auxiliary, higher-heritability traits (Jia and Jannink, 2012; Rutkoski et al., 2016; Schulthess et al., 2016; Wang et al., 2016). In the present study, the predictive ability of the GBLUP_{MV} model was greater than that of corresponding single-trait models for intercorrelated traits when using the CV2 cross-validation scheme (Fig. 6). It is possible that model predictive ability could be further increased by including additional traits. Wang et al. (2016) observed much higher predictive accuracies for some traits when the number of traits included in the model was increased from two to eight. However, the addition of each trait greatly increases computing time and model complexity. One finding that was against expectation was that the use of an unlinked CV2 cross-validation scheme with the GBLUP_{MV} model did not affect predictive ability compared with using a linked CV2 cross-validation scheme. In addition, the use of unlinked CV2 led to slightly lower prediction accuracies for the lower-heritability traits (GW and YLD) included in the GBLUP_{MV} model when using intercorrelated traits. The reason for this is not immediately clear, and we are not aware of other studies that have tested this exact cross-validation scenario. Montesinos-López et al. (2016) tested one cross-validation scheme that was equivalent to our CV2 linked scheme. The second cross-validation scheme they used did not resemble either of our CV2 methods, as it simulated a scenario in which a trait is not assessed in all lines within one environment but is assessed in all other environments. The GBLUP_{MV} model offered few advantages over the GBLUP_{GEI} model when using uncorrelated traits and/or the CV1 cross-validation scheme. Rutkoski et al. (2016) and Pszczola et al. (2013) reported similar findings, in which the use of a multitrait model including secondary traits had no influence on prediction accuracy if secondary trait phenotypes were not collected for the VP. Jia and Jannink (2012) found some scenarios in simulated datasets in which a multitrait model could outperform corresponding single-trait models in a CV1 scenario, in particular when a very low-heritability trait ($h^2 = 0.1$) exhibited moderate genetic correlation ($r = 0.5$) with a moderate-heritability trait ($h^2 = 0.5$), when these traits were each influenced by a low number of QTLs. In practice, it is not clear how often similar conditions would be encountered in applied settings, and indeed the same paper reported little difference in prediction accuracies between multitrait and single-trait models under a CV1 scenario when using real pine tree (*Pinus* spp.) disease resistance data.

The present study also examined the effect of post hoc spatial corrections on predictive ability. Although this study selected the best-fitting model for each individual trait-environment combination, the separable two-dimensional AR1 model with “nugget” unit residual effect (AR1 × AR1 + units) was often selected as the best-fitting model.

For practical reasons, it may be beneficial to select a single, robust model accounting for spatial variability for use across all environments, thereby allowing for fitting the residual structure as part of the one-step model fitting process. Gilmour et al. (1997) suggested that no single model is appropriate for correcting spatial trends in all environments but recommended using AR1 × AR1 models as a default starting point to assess spatial variation. The present study found that heritability was usually increased by accounting for spatial variation, but that there was often little corresponding effect on predictive ability. Bernal-Vasquez et al. (2014) found that various autoregressive models correcting for spatial variation generally did not increase predictive ability in a rye (*Secale cereale* L.) dataset. They instead found that a model fitting simple row and column effects produced better improvements in accuracy and noted that AR1 models may often present convergence difficulties. Our findings suggest that although AR1 models have the appealing effect of increasing heritability, the modest gains in predictive ability may not justify the added complexity of fitting them in a GS setting.

For achieving prediction accuracies >0.5 in wheat, Bassi et al. (2016) recommended a TP size of at least 50 individuals if entries of the VP are full-sibs of entries in the TP, 100 individuals for half-sibs, and 1000 individuals for less related TPs and VPs. The models used in the present study produced predictive accuracies exceeding 0.5 for many traits, despite a relatively modest overall panel size of 329. The panel used in this study consisted of an ad hoc assembly of elite soft winter wheat breeding germplasm; although it was not created specifically for performing GS, it contained some sets of full and half-sibs, and the genotypes included in the panel generally were highly interrelated due to historical germplasm exchanges between public soft wheat breeding programs. Using Eq. [2], the effective population size for the testing panel was estimated at 70 (roughly one-fifth the census panel size). Several studies suggest that GS prediction accuracies in small grains plateau for many traits at relatively small TP sizes. For example, a genomic prediction study for multiple Fusarium head blight-related traits in advanced midwestern and eastern US wheat breeding lines found that predictive ability generally plateaued at a nominal TP size of 192 genotypes (Arruda et al., 2015). Lorenz et al. (2012) similarly found that prediction accuracies for Fusarium head blight-related traits in barley (*Hordeum vulgare* L.) plateaued at ~200 genotypes. Breeding programs will likely steadily expand TPs over time as additional genotypes are tested, but our findings suggest that satisfactory predictions can be generated for many moderate and high-heritability traits with relatively modest numbers of genotyped lines.

Problems experienced with convergence for several traits highlight one practical limitation to the technique

of using mixed linear models with data collected across environments and traits. As the number of environments increases, so too does the number of parameters that must be estimated during model fitting. This problem is further compounded in multitrait, multienvironment models. There are several strategies to address this issue. First, this study indicated that many traits did not require complex covariance structures to model GEI; our recommendation is to favor more parsimonious models so long as model fit is not compromised. In addition, as previously mentioned, factor-analytic covariance structures can be used to capture much of the covariance modeled by an unstructured matrix while requiring the estimation of far fewer parameters. The complexity reduction due to adopting a factor-analytic structure was limited in this study due to the low number of environments, but this benefit increases with the number of environments included. Another strategy is to group environments into mega-environments and subsequently perform GS within mega-environments where GEI is more limited. Lado et al. (2016) identified mega-environments in a large, unbalanced, multienvironment dataset prior to performing GS by using multiplicative models including the additive main and multiplicative interactive (AMMI) model (Gauch and Zobel, 1988; Zobel et al., 1988) and the GGE model (Yan et al., 2000). Similar analyses to identify mega-environments may be performed using factor-analytic mixed models (Burgueño et al., 2008). Finally, models that fit a marker \times environment interaction (MEI) term may succeed where traditional GEI models fail to converge. Such MEI models were first used by Heslot et al. (2013) to characterize environments and were extended by Lopez-Cruz et al. (2015) and Crossa et al. (2016). An advantage of MEI models is that marker effects are always present in every environment, in contrast with the extreme imbalance that is often witnessed in multienvironment breeding trials when considering GEI (Rutkoski et al., 2017).

CONCLUSION

This study examined methods for increasing the predictive ability of GS methods in early-generation yield trials representing ad hoc assemblies of genotypes from various crosses, producing several key findings:

1. The use of AR1 models for correcting spatial variability led to large increases in trait heritability, but generally modest or nonsignificant increases in predictive ability. However, the correction of spatial effects only produced lower accuracies for one trait in the two-step model, so there is generally little risk in attempting to use AR1 models for spatial correction.
2. Prediction of across-environment genotype main effects were generally equivalent between a two-step model using adjusted means across environments and one-step

multienvironment models using CV1 cross-validation, although the latter has the advantage of also producing within-environment performance predictions.

3. The use of a realized genomic relationship matrix with one-step multienvironment models in a CV2 cross-validation scenario can increase within-environment prediction accuracies over similar models that assume independence among genotypes. Across-environment prediction accuracies will be similar in this setting, with or without the inclusion of a realized relationship matrix.
4. The use of a multitrait, multienvironment model can increase predictive ability for low-heritability traits in a CV2 cross-validation setting.

This experiment additionally suggests that there is room for further work studying the optimization of early-generation field breeding trials using a variety of germplasm panel compositions beyond biparental crosses.

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Material Available

Supplemental material is available online for this article.

Acknowledgments

Genotyping costs for this project were supported by the Agriculture and Food Research Initiative Competitive Grant 2017-67007-25939 from the USDA National Institute of Food and Agriculture. The lead author was supported by the Virginia Agricultural Council Grant 617, and by funding from the Virginia Small Grains Board. The authors thank Professor Jim Holland of the USDA-ARS for his assistance in revising this manuscript.

References

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19:716–723. doi:10.1109/TAC.1974.1100705
- Alley, M.M., D.E. Brann, E.L. Stromberg, E.S. Hagood, A. Herbert, E.C. Jones, et al. 1993. Intensive soft red winter wheat production: A management guide. Virginia Tech Coop. Ext. Distribution Ctr., Blacksburg, VA.
- AOAC International. 2000. Protein (crude) in animal feed. Official Method 990.03. In: Official methods of analysis of AOAC International. 17th ed. Assoc. Official Anal. Chem., Rockville, MD.
- Arruda, M.P., P.J. Brown, A.E. Lipka, A.M. Krill, C. Thurber, and F.L. Kolb. 2015. Genomic selection for predicting *Fusarium* head blight resistance in a wheat breeding program. *Plant Genome* 8(3). doi:10.3835/plantgenome2015.01.0003
- Bassi, F.M., A.R. Bentley, G. Charmet, R. Ortiz, and J. Crossa. 2016. Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242:23–36. doi:10.1016/j.plantsci.2015.08.021
- Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. lme4: Linear mixed-effects models using Eigen and S4. R Found. Stat. Comput., Vienna. <http://CRAN.R-project.org/package=lme4> (accessed 6 Jan. 2018).

- Bernal-Vasquez, A.-M., J. Möhring, M. Schmidt, M. Schönleben, C.-C. Schön, and H.-P. Piepho. 2014. The importance of phenotypic data analysis for genomic prediction— a case study comparing different spatial models in rye. *BMC Genomics* 15:646. doi:10.1186/1471-2164-15-646
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. doi:10.1093/bioinformatics/btm308
- Browning, S.R., and B.L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084–1097. doi:10.1086/521987
- Browning, B.L., and S.R. Browning. 2016. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98:116–126. doi:10.1016/j.ajhg.2015.11.020
- Burgueño, J., J. Crossa, P.L. Cornelius, and R.-C. Yang. 2008. Using factor analytic models for joining environments and genotypes without crossover genotype \times environment interaction. *Crop Sci.* 48:1291–1305. doi:10.2135/cropsci2007.11.0632
- Burgueño, J., J. Crossa, J.M. Cotes, F.S. Vicente, and B. Das. 2011. Prediction assessment of linear mixed models for multi-environment trials. *Crop Sci.* 51:944–954. doi:10.2135/cropsci2010.07.0403
- 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
- Butler, D.G., B.R. Cullis, A.R. Gilmour, and B.J. Gogel. 2009. *ASReml-R reference manual*. VSN Int., Hemel Hempstead, UK.
- Calus, M.P., and R.F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. *Genet. Sel. Evol.* 43:26. doi:10.1186/1297-9686-43-26
- Chang, C.C., C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: R ising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi:10.1186/s13742-015-0047-8
- Coster, A. 2012. pedigree: Pedigree functions. *R Found. Stat. Comput., Vienna*. <https://CRAN.R-project.org/package=pedigree> (accessed 29 July 2018).
- Crossa, J., G. de los Campos, M. Maccaferri, R. Tuberosa, J. Burgueño, and P. Pérez-Rodríguez. 2016. Extending the marker \times environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. *Crop Sci.* 56(5):2193–2209. doi:10.2135/cropsci2015.04.0260
- Cullis, B., B. Gogel, A. Verbyla, and R. Thompson. 1998. Spatial analysis of multi-environment early generation variety trials. *Biometrics* 54:1–18. doi:10.2307/2533991
- Cullis, B.R., A.B. Smith, and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* 11:381. doi:10.1198/108571106X154443
- dos Santos, J.P.R., R.C. de Castro Vasconcellos, L.P.M. Pires, M. Balestre, and R.G.V. Pinho. 2016. Inclusion of dominance effects in the multivariate GBLUP model. *PLoS One* 11:e0152045. doi:10.1371/journal.pone.0152045
- Elias, A.A., I. Rabbi, P. Kulakow, and J.-L. Jannink. 2018. Improving genomic prediction in cassava field experiments using spatial analysis. *G3: Genes, Genomes, Genet.* 8:53–62. doi:10.1534/g3.117.300323
- 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
- Endelman, J.B., G.N. Atlin, Y. Beyene, K. Semagn, X. Zhang, M.E. Sorrells, and J.-L. Jannink. 2014. Optimal design of preliminary yield trials with genome-wide markers. *Crop Sci.* 54:48–59. doi:10.2135/cropsci2013.03.0154
- Falconer, D.S., and T.F.C. Mackay. 1996. *Introduction to quantitative genetics*. 4th ed. Addison Wesley Longman, Essex, UK.
- Gauch, H.G., and R.W. Zobel. 1988. Predictive and postdictive success of statistical analyses of yield trials. *Theor. Appl. Genet.* 76:1–10. doi:10.1007/BF00288824
- Gilmour, A.R., B.R. Cullis, and A.P. Verbyla. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric. Biol. Environ. Stat.* 2:269–293. doi:10.2307/1400446
- Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:e90346. doi:10.1371/journal.pone.0090346
- Guo, G., F. Zhao, Y. Wang, Y. Zhang, L. Du, and G. Su. 2014. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 15:30. doi:10.1186/1471-2156-15-30
- Guo, Z., D.M. Tucker, D. Wang, C.J. Basten, E. Ersoz, W.H. Briggs, et al. 2013. Accuracy of across-environment genome-wide prediction in maize nested association mapping populations. *G3: Genes, Genomes, Genet.* 3:263–272. doi:10.1534/g3.112.005066
- Hayashi, T., and H. Iwata. 2013. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics* 14:34. doi:10.1186/1471-2105-14-34
- Hedrick, P. 2011. *Genetics of populations*. Jones & Bartlett Learning, Sudbury, MA.
- 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
- Isik, F., J. Holland, and C. Maltecca. 2017. *Genetic data analysis for plant and animal breeding*. Springer Int. Publ., Cham, Switzerland. doi:10.1007/978-3-319-55177-7
- IWGSC (The International Wheat Genome Sequencing Consortium), R. Appels, K. Eversole, C. Feuillet, B. Keller, J. Rogers, et al. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi:10.1126/science.aar7191
- Jarquín, D., J. Crossa, X. Lacaze, P.D. 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
- Jia, Y., and J.-L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192:1513–1522. doi:10.1534/genetics.112.144246
- Kang, H.M., N.A. Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723. doi:10.1534/genetics.107.080101
- 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
- Lado, B., I. Matus, A. Rodríguez, L. Inostroza, J. Poland, F. Belzile, et al. 2013. Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3: Genes, Genomes, Genet.* 3:2105–2114. doi:10.1534/g3.113.007807
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. doi:10.1093/bioinformatics/btp324
- 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., K.P. 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.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Meyer, K. 2009. Factor-analytic models for genotype \times environment type problems and structured covariance matrices. *Genet. Sel. Evol.* 41:21. doi:10.1186/1297-9686-41-21
- Montesinos-López, O.A., A. Montesinos-López, J. Crossa, J.C. Montesinos-López, D. Mota-Sanchez, F. Estrada-González, et al. 2018. Prediction of multiple-trait and multiple-environment genomic data using recommender systems. *G3: Genes, Genomes, Genet.* 8:131–147. doi:10.1534/g3.117.300309
- Montesinos-López, O.A., A. Montesinos-López, J. Crossa, F.H. Toledo, J.C. Montesinos-López, P. Singh, et al. 2017. A Bayesian Poisson-lognormal model for count data for multiple-trait multiple-environment genomic-enabled prediction. *G3: Genes, Genomes, Genet.* 7:1595–1606. doi:10.1534/g3.117.039974
- Montesinos-López, O.A., A. Montesinos-López, J. Crossa, F. Toledo, O. Pérez-Hernández, K.M. Eskridge, and J. Rutkowski. 2016. A genomic Bayesian multi-trait and multi-environment model. *G3: Genes, Genomes, Genet.* 6:2725–2744. doi:10.1534/g3.116.032359
- Muñoz-Huerta, R.F., R.G. Guevara-Gonzalez, L.M. Contreras-Medina, I. Torres-Pacheco, J. Prado-Olivarez, and R.V. Ocampo-Velazquez. 2013. A review of methods for sensing the nitrogen status in plants: Advantages, disadvantages and recent advances. *Sensors (Basel)* 13:10823–10843. doi:10.3390/s130810823
- 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
- Ould Estaghrvirou, S.B., J.O. Ogutu, T. Schulz-Streeck, C. Knaak, M. Ouzunova, A. Gordillo, and H.-P. Piepho. 2013. Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. *BMC Genomics* 14:860. doi:10.1186/1471-2164-14-860
- Patterson, H.D., and E.R. Williams. 1976. A new class of resolvable incomplete block designs. *Biometrika* 63:83–92. doi:10.1093/biomet/63.1.83
- Phillips, S.B., D.A. Keahey, J.G. Warren, and G.L. Mullins. 2004. Estimating winter wheat tiller density using spectral reflectance sensors for early-spring, variable-rate nitrogen applications. *Agron. J.* 96:591–600. doi:10.2134/agronj2004.0591
- Piepho, H.-P., J. Möhring, M. Pflugfelder, W. Hermann, and E.R. Williams. 2015. Problems in parameter estimation for power and AR(1) models of spatial correlation in designed field experiments. *Commun. Biom. Crop Sci.* 10:3–16.
- 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:e32253. doi:10.1371/journal.pone.0032253
- Pszczola, M., R.F. Veerkamp, Y. de Haas, E. Wall, T. Strabel, et al. 2013. Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. *Animal* 7:1759–1768. doi:10.1017/S175173111300150X
- Rutkoski, J.E., J. Crain, J. Poland, and M.E. Sorrells. 2017. Genomic selection for small grain improvement. In: R.K. Varshney, et al., editors, *Genomic selection for crop improvement: New molecular breeding strategies for crop improvement*. Springer, Cham, Switzerland. p. 99–130 doi:10.1007/978-3-319-63170-7_5
- 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
- Schulthess, A.W., Y. Wang, T. Miedaner, P. Wilde, J.C. Reif, and Y. Zhao. 2016. Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor. Appl. Genet.* 129:273–287. doi:10.1007/s00122-015-2626-6
- Smith, A., B. Cullis, and R. Thompson. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57:1138–1147. doi:10.1111/j.0006-341X.2001.01138.x
- VanRaden, P.M. 2007. Efficient estimation of breeding values from dense genomic data. *J. Dairy Sci.* 90(Supplement 1):374–375.
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. doi:10.3168/jds.2007-0980
- Wang, X., L. Li, Z. Yang, X. Zheng, S. Yu, C. Xu, and Z. Hu. 2016. Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity* 118:302–310. doi:10.1038/hdy.2016.87
- Wright, K., and J.-L. Laffont. 2016. Package “gge”: Genotype plus genotype-by-environment biplots. *R Found. Stat. Comput.* <https://CRAN.R-project.org/package=gge> (accessed 3 Aug. 2018).
- Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavnic. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40:597–605. doi:10.2135/cropsci2000.403597x
- Zadoks, J.C., T.T. Chang, and C.F. Konzak. 1974. A decimal code for the growth stages of cereals. *Weed Res.* 14:415–421. doi:10.1111/j.1365-3180.1974.tb01084.x
- Zhang, X., P. Pérez-Rodríguez, K. Semagn, Y. Beyene, R. Babu, M.A. López-Cruz, et al. 2015. Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity* 114:291–299. doi:10.1038/hdy.2014.99
- 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