

Prospects and Challenges of Applied Genomic Selection—A New Paradigm in Breeding for Grain Yield in Bread Wheat

Philomin Juliana*, Ravi P. Singh*, Jesse Poland, Suchismita Mondal, José Crossa, Osval A. Montesinos-López, Susanne Dreisigacker, Paulino Pérez-Rodríguez, Julio Huerta-Espino, Leonardo Crespo-Herrera, and Velu Govindan

P. Juliana, R.P. Singh, S. Mondal, J. Crossa, S. Dreisigacker, L. Crespo-Herrera, V. Govindan, CIMMYT, Apdo, Postal 6-641, 06600 Mexico, D.F., Mexico; J. Poland, Wheat Genetics Resource Center, Dep. of Plant Pathology, Kansas State Univ., Manhattan, KS 66506; J. Poland, Dep. of Agronomy, Kansas State Univ., Manhattan, KS 66506; O.A. Montesinos-López, Facultad de Telemática, Univ. de Colima, Colima 28040, México; P. Pérez-Rodríguez, Colegio de Postgraduados, CP 56230, Montecillos, Edo. de México, México; J. Huerta-Espino, Campo experimental Valle de México Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 56230, Chapingo, Edo. de México, México.

ABSTRACT Genomic selection (GS) has been promising for increasing genetic gains in several species. Therefore, we evaluated the potential integration of GS for grain yield (GY) in bread wheat (*Triticum aestivum* L.) in CIMMYT's elite yield trial nurseries. We observed that the genomic prediction accuracies within nurseries (0.44 and 0.35) were substantially higher than across-nursery accuracies (0.15 and 0.05) for GY evaluated in the bed and flat planting systems, respectively. The accuracies from using only a subset of 251 genotyping-by-sequencing markers were comparable to the accuracies using all 2038 markers. We also used the item-based collaborative filtering approach for incorporating other related traits in predicting GY and observed that it outperformed genomic predictions across nurseries, but was less predictive when trait correlations with GY were low. Furthermore, we compared GS and phenotypic selections (PS) and observed that at a selection intensity of 0.5, GS could select a maximum of 70.9 and 61.5% of the top lines and discard 71.5 and 60.5% of the poor lines selected or discarded by PS within and across nurseries, respectively. Comparisons of GS and pedigree-based predictions revealed that the advantage of GS over the pedigree was moderate in populations without full-sibs. However, GS was less advantageous for within-family selections in elite families with few full-sibs and minimal Mendelian sampling variance. Overall, our results demonstrate the importance of applying GS for GY at the appropriate stage of the breeding cycle, and we speculate that gains can be maximized if it is implemented in early-generation within-family selections.

Abbreviations: BLUE, best linear unbiased estimate; DTHD, days to heading; DTMT, days to maturity; EYT, elite yield trials; EYT 13–14, elite yield trial nursery in 2013–2014; EYT 14–15, elite yield trial nursery in 2014–2015; EYT 15–16, elite yield trial nursery in 2015–2016; EYT 16–17, elite yield trial nursery in 2016–2017; GBLUP, genomic best linear unbiased prediction; GBS, genotyping-by-sequencing; $G \times E$, genotype \times environment interaction; GEBVs, genomic-estimated breeding values; GS, genomic selection; GY, grain yield; IBCF, item-based collaborative filtering; LD, linkage disequilibrium; PS, phenotypic selection.

CORE IDEAS

- Genomic prediction of grain yield across nurseries or years is challenging, because of genotype \times environment interactions.
- Prediction accuracies can be improved by having at least one full-sib in the training population.
- Genomic selection (GS) was less advantageous for within-family selections in elite families with few full-sibs and minimal Mendelian sampling variance.
- It is important to apply GS at the appropriate stage of the breeding cycle.

THE INTEGRATION of innovative and cutting-edge technologies in breeding programs to accelerate genetic gains and boost efficiency is critical for sustaining global bread wheat production. Over the last few years, tremendous increase in genetic gains for economically important traits in dairy cattle has been achieved by implementing GS that uses genome-wide marker

Juliana, P., R.P. Singh, J. Poland, S. Mondal, J. Crossa, O.A. Montesinos-López, S. Dreisigacker, P. Pérez-Rodríguez, J. Huerta-Espino, L. Crespo-Herrera, and V. Govindan. 2018. Prospects and challenges of applied genomic selection—A new paradigm in breeding for grain yield in bread wheat. *Plant Genome* 11:180017. doi: 10.3835/plantgenome2018.03.0017

Received 23 Mar. 2018. Accepted 11 July 2018.

*Corresponding authors (r.singh@cgiar.org, p.juliana@cgiar.org).

This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

information to predict the genetic merit or genomic-estimated breeding values (GEBVs) of individuals prior to phenotyping (Meuwissen et al., 2001; Hayes et al., 2013). In GS, individuals that have been genotyped and phenotyped (referred to as the ‘training population’) are used to train prediction models and predict the performance of genotyped individuals that have not been phenotyped (referred to as the ‘validation population’ or ‘selection candidates’). Genomic selection is an efficient selection tool for shortening the cycle time, predicting individuals in unobserved environments (sparse testing), increasing the selection accuracy for traits of low heritability (Muir, 2007; Calus et al., 2008; Crossa et al., 2017), and eliminating poor-performing progenies before the next generation of costly, time-consuming, and laborious field testing (Heffner et al., 2009). Although GS has been investigated for several traits in wheat via cross-validations (Crossa et al., 2014; Battenfield et al., 2016; Hayes et al., 2017; Juliana et al., 2017a; 2017b; Pérez-Rodríguez et al., 2017), assessments of its routine implementation in a wheat breeding program are lacking.

Breeding for high GY is the primary target for most wheat breeding programs. However, it is challenging because of the complex genetic nature of GY (controlled by many loci with small effects), low heritability, the poor stability of genotypes across different environments resulting from genotype \times environment interactions ($G \times E$), a lack of clear understanding of the genetic basis of GY, inconsistent GY quantitative trait loci detected across different environments, and the role of epistatic and non-genetic effects (Quarrie et al., 2005; Kuchel et al., 2007a; Snape et al., 2007; Griffiths et al., 2015; Jiang et al., 2017). Hence, the benefits of marker-assisted selection for GY are limited and extensive yield testing remains crucial for the development of well-adapted, stable, and high yielding wheat varieties. However, yield testing is expensive and several lines are discarded in each yield trial. Therefore, one approach to minimize the phenotyping costs and leverage genomic information would be to integrate GS with phenotypic selection (PS) and evaluate lines that are only predicted to have some favorable alleles. This integrated approach is expected to result in better responses over PS, especially for a trait with low heritability (Dekkers, 2007; Calus et al., 2008).

A critical aspect of GS is determining the appropriate prediction model. Several models that differ in their assumption about the underlying genetic architecture of traits have been proposed. Among these, the best linear unbiased prediction and its extensions [genomic best linear unbiased prediction (GBLUP) that uses the genomic relationships based on markers] have been the most widely used parametric models for genomic-enabled prediction (VanRaden, 2008; Habier et al., 2013). Several Bayesian models like the Bayesian ridge regression, BayesA, BayesB, BayesCpi, and the Bayesian least absolute shrinkage and selection operator (Meuwissen et al., 2001; Park and Casella, 2008; Kizilkaya et al., 2010; Gianola, 2013), which use different prior assumptions for marker effects, are also used for GS. In addition, the semiparametric model called

‘reproducing kernel Hilbert spaces’, which is expected to capture nonadditive effects (Gianola et al., 2006; Gianola and van Kaam, 2008), has been evaluated in wheat (Heslot et al., 2012; Pérez-Rodríguez et al., 2012; Rutkoski et al., 2012; Crossa et al., 2014; Juliana et al., 2017b).

Grain yield is known to be affected by variability in days to heading (DTHD), days to maturity (DTMT), plant height, and lodging (Fischer and Stapper, 1987; Flinham et al., 1997; Quarrie et al., 2005; Kuchel et al., 2007b; Addison et al., 2016) that should be considered in prediction models. Although statistical models have been developed for incorporating multiple traits in genomic-enabled prediction (Montesinos-López et al., 2016), the time needed to fit these models continues to be a significant challenge. Recently, a new algorithm implemented in GS proved to be competitive with respect to accuracy and computing time (Montesinos-López et al., 2018). The algorithm called ‘item-based collaborative filtering’ (IBCF) is very popular in electronic commerce websites for recommending items and products. Here, the customer’s interests are used as input to generate a list of recommended items. The basic idea in the IBCF algorithm lies in building a database of preferences for items by users. For a specific user, u_s , we look for the set of items that the user has rated in the database, compute how similar they are to the target item, and then select the k most similar items. Simultaneously, the corresponding similarities between the k most similar items are also computed. Once the most similar items are found, prediction for the specific user and target items is computed via a simple weighted average of the similar items. Thus, one of our objectives was to evaluate the IBCF approach for predicting GY by incorporating information from other traits that can affect GY. In this case, users are comparable to GY and items are comparable to traits like DTHD, DTMT, plant height, and lodging, which are expected to have some correlation with GY.

For GS to have a real impact in breeding programs, it is essential to look beyond prediction accuracies. Though various micro and macro environmental factors like soils, irrigation, management, weather patterns, etc. play a role in determining GY, modeling these effects might not be very beneficial to a breeder who is interested in the line’s performance across a range of different environments and would want to exploit this variation to develop varieties with wide and specific adaptation (Snape et al., 2007). In addition, the variable extent of plasticity (Bradshaw, 1965) (the range of phenotypes produced by a single genotype in different environments) among genotypes for GY in wheat (Sadras et al., 2009) will hamper accurate predictions of GY. Hence, without a good understanding of stability, phenotypic plasticity, and $G \times E$ interactions, GY predictions across environments will be challenging. However, a breeder is also not interested exclusively in prediction accuracies, but how genomics-assisted breeding can be leveraged in selecting lines with high breeding values and discarding lines that might not contribute to any future genetic gains. Hence, a key objective of this study was to look beyond prediction accuracies and compare selections made from PS and GS

for GY. In addition, our objectives were to (i) assess how marker subsets and the training population schemes can affect the accuracy of genomic predictions for GY, and (ii) compare genomic and pedigree-based predictions for different nurseries and subsets of lines with and without full-sibs in the training population.

MATERIALS AND METHODS

Populations and Phenotyping Data

For this study, we used four elite yield trial (EYT) nurseries from CIMMYT's bread wheat breeding program. The EYTs are second-year yield trial nurseries, each comprising 1092 lines. The selected bulk breeding method was used to develop these lines, where all the selections were bulked in early generations until the head rows or individual plants derived from F_5 , F_6 , or F_7 lines. About 70,000 individual plant-derived lines were then grown in small 1-by 1-m plots (preyield testing stage) and selected visually for agronomic characteristics, spike health, and a few diseases. These selections resulted in the first-year yield trial nurseries comprising about 9000 lines, which were further selected for GY to form the EYT nurseries. Although the 1092 lines in the four EYT nurseries were different, some level of genetic relatedness is expected between the lines, because of several common parents used across the years.

The EYT nurseries were planted during mid-November (which is the optimum planting time for CIMMYT's yield trials) in optimally irrigated environments (receiving 500 mm of water) under bed and flat planting systems at the Norman E. Borlaug Research station, Ciudad Obregon, Sonora, Mexico. They were sown in 39 trials, each comprising 28 lines and two high-yielding checks ('Kachu' and 'Borlaug') that were arranged in an α -lattice design, with three replications and six blocks. The nurseries were evaluated for GY on a plot basis during the 2013–2014 (EYT 13–14), 2014–2015 (EYT 14–15), 2015–2016 (EYT 15–16), and 2016–2017 (EYT 16–17) seasons. In addition, traits like DTHD (number of days from germination to 50% of spike emergence), DTMT (number of days from germination to 50% physiological maturity), plant height (in cm), and lodging (measured in an ordinal scale from 0 to 5) were also recorded.

The best linear unbiased estimates (BLUEs) of the breeding lines for each planting system within each nursery were estimated with the ASREML statistical package (Gilmour, 1997) via the following mixed linear model:

$$y_{ijkl} = \mu + g_i + t_j + r_{k(j)} + b_{l(jk)} + \varepsilon_{ijkl}, \quad [1]$$

where y_{ijkl} is the GY, μ is the mean, g_i is the fixed effect of the genotype, t_j is the random effect of the trial that is independent and identically distributed $t_j \sim N(0, \sigma_t^2)$, $r_{k(j)}$ is the random effect of the replicate within the trial that is independent and identically distributed $r_{k(j)} \sim N(0, \sigma_r^2)$, $b_{l(jk)}$ is the random effect of the incomplete block within the trial and the replicate that is independent and identically distributed $b_{l(jk)} \sim N(0, \sigma_b^2)$, and ε_{ijkl} is the residual that is independent and identically distributed $\varepsilon_{ijkl} \sim N(0, \sigma_e^2)$. Since DTHD and

lodging were also associated with GY, we included them as fixed effect covariates in the model [Eq. 1]. For across-nursery predictions, we obtained BLUEs for each line by including the random effect of the environment in Model 1 and the model with DTHD and lodging as covariates.

The square root of the heritability (H) for GY within each year for both the planting systems was calculated on a line-mean basis across the replicates using the formula:

$$H = \sqrt{\frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{nreps}}}, \quad [2]$$

where σ_g^2 is the genetic variance, σ_e^2 is the error variance, and $nreps$ is the number of replications. The estimates of the genetic and residual variances were obtained via the average information-restricted maximum likelihood algorithm (Gilmour et al., 1995) implemented in the R package 'heritability' (Kruijer et al., 2015). Tests of analysis of variance (ANOVA) were performed with JMP statistical software (www.jmp.com, accessed 5 Sept. 2018) to detect if the families and sibs nested within families contributed significantly to GY variation in the different nurseries.

Genotyping and Principal Component Analysis

All 4368 lines in the four nurseries were genotyped with genotyping-by-sequencing (GBS) (Elshire et al., 2011; Poland et al., 2012) for obtaining genome-wide markers. Genotyping was done at Kansas State University, on an Illumina HiSeq2500 (Illumina Inc., San Diego, CA) with 190 samples pooled per lane. The initial read length was 100 bp, which was trimmed to 64 bp after removing the barcode. Marker polymorphisms were called with the TASSEL version 5 GBS pipeline (Glaubitz et al., 2014) and anchored to the International Wheat Genome Sequencing Consortium's first version of the reference sequence (RefSeq version 1.0) assembly of the bread wheat variety 'Chinese Spring'. Markers with more than 60% missing data, less than 5% minor allele frequency, and more than 10% heterozygosity were removed, and 2038 markers were obtained from an initial set of 34,900 markers. Missing marker data were imputed with LinkImpute (Money et al., 2015), implemented in TASSEL version 5 (Bradbury et al., 2007). The lines were also filtered and those with more than 50% missing data were removed, resulting in 3485 lines (767 lines from EYT 13–14, 775 lines from EYT 14–15, 964 lines from EYT 15–16, and 980 lines from EYT 16–17). We then performed a principal component analysis for the 3485 lines with 2038 markers to determine population structure (the genotyping and phenotyping data for these lines are available in Supplemental File S1).

Genomic and Pedigree-Based Predictions

Since several genomic prediction models are known to result in similar prediction accuracies (Heslot et al., 2012; Rutkoski et al., 2012; Juliana et al., 2017b), we have only used the GBLUP model, which can be represented as:

$$y_i = \mu + u_i + \varepsilon_i, \quad [3]$$

where y_i is the response variable for individual i , μ is the general mean, u_i is an additive genetic effect for individual i , [we assume that the joint distribution of the vector of additive genetic effects \mathbf{u} is $N(\mathbf{0}, \mathbf{G}\sigma_g^2)$, where \mathbf{G} is the additive relationship matrix computed from markers according to Endelman and Jannink (2012), and σ_g^2 is the variance component associated with markers] and ε_i is the error term [we assume that the joint distribution of ε is $N(\mathbf{0}, \mathbf{I}\sigma_e^2)$ where σ_e^2 is the residual variance]. The GBLUP model was fitted with the BGLR package in R (Pérez and de los Campos, 2014).

Since CIMMYT maintains an excellent pedigree of the breeding lines, we also fitted the previous model by replacing the genomic relationship matrix with the additive relationship matrix derived from the pedigree (\mathbf{A} matrix), which is twice the coefficient of parentage ($\mathbf{A} = 2\mathbf{f}_{xy}$, where \mathbf{f}_{xy} is the coefficient of parentage) and represents identical-by-descent relationships. In addition, we fitted a model that included the relationship matrices derived from markers and pedigree (the \mathbf{G} and \mathbf{A} matrices) jointly. The BLUEs estimated in Eq. [1] (no covariates) and the BLUEs with DTHD and lodging as fixed-effect covariates (with covariates) were used in the prediction models. Prediction accuracies were calculated as the Pearson's correlation between the phenotypic BLUEs and the predicted values.

Marker Subsets

The effects of the number of markers and linkage disequilibrium (LD) on genomic predictions were evaluated with different marker subsets created from the estimated intermarker correlations (r^2) among the 2038 markers. After randomly retaining only one of the correlated markers, we had marker subsets of 810 markers ($r^2 = 0.8$), 504 markers ($r^2 = 0.6$), 251 markers ($r^2 = 0.4$), and 29 markers ($r^2 = 0.2$).

Training Population Schemes

We evaluated three training population schemes for implementing GS in the EYTs. In Scheme 1, the datasets for each EYT were divided into five sets; and four of them were used as the training population (611 lines in EYT 13–14, 620 lines in EYT 14–15, 771 lines in EYT 15–16, and 784 lines in EYT 16–17) to predict the remaining lines in the fifth set or the validation population. Here, the prediction accuracies are merely the cross-validation accuracies within nurseries. In Scheme 2, we used one nursery as the training population to predict another for every possible nursery combination (forward and backward predictions). In Scheme 3, we used all the previous nurseries as the training population to predict a given nursery. Here, EYT 13–14 and EYT 14–15 were used to predict EYT 15–16, whereas EYT 13–14, EYT 14–15, and EYT 15–16 were used to predict EYT 16–17.

Genomic and Pedigree-Based Predictions in Populations with and without Full-Sibs in the Training Population

The predictive abilities of the \mathbf{G} and \mathbf{A} matrices in populations without full-sibs were assessed via cross-validations on a subset of lines with only one line per cross in the nurseries (343 lines in EYT 13–14, 226 lines in EYT 14–15, 243 lines in EYT 15–16, and 201 lines in EYT 16–17). We also evaluated the advantage of having full-sibs in the training population by using a subset of lines in each nursery that had at least one full-sib in that nursery (424 lines in EYT 13–14, 549 lines in EYT 14–15, 721 lines in EYT 15–16, and 779 lines in EYT 16–17). With the full-sib subsets, we performed two different analyses: (i) all other full-sibs in the training population were used to predict one random full-sib in the validation population; (ii) one random full-sib in the training population was used to predict all other full-sibs in the validation population.

Item-Based Collaborative Filtering

In the IBCF approach, the rating $P_{i,j}$, for user i in item j can be predicted from the following expression (Sarwar et al., 2001):

$$P_{i,j} = \frac{\sum_{j \in N} y_{i,j} w_{j,j'}}{\sum_{j \in N} |w_{j,j'}|}, \quad [4]$$

where the summation is over all other rated items ($j \in N$) for user i (N is the number of rated items), $w_{j,j'}$ is the weight between items j and j' , and $y_{i,j}$ is the rating for user i of item j . The weights used in Eq. [4] were obtained from an item-to-item similarity matrix built from the cosine similarity that provides information on how similar an item is to another item (Montesinos-López et al., 2018):

$$\cos(\theta) = \frac{\sum_{j=1}^n x_j y_j}{\sqrt{\sum_{j=1}^n x_j^2} \sqrt{\sum_{j=1}^n y_j^2}}. \quad [5]$$

We implemented IBCF with the GY BLUPs obtained by incorporating marker information and then used them for prediction along with similarities to correlated traits like DTHD, DTMT, height, and lodging. The rating matrix used for implementing IBCF is shown in Supplemental Table S1 and all the traits were standardized before the IBCF method was applied.

Phenotypic Selection and GS

We evaluated different scenarios for the potential use of GEBVs in making selections: (i) selections within nurseries (and years) using the GEBVs estimated from random cross-validations and (ii) selections across nurseries (and years) by including only the previous year's data in the training population. We then compared the PS and GS of the lines at selection thresholds of 0.1, 0.2, 0.3, and 0.5 by classifying them as: selected by PS only (S_{PS}), selected by GS only (S_{GS}), selected by GS and PS (S_{GSPS}), and not selected by either GS or PS (NS). The percentage of poor

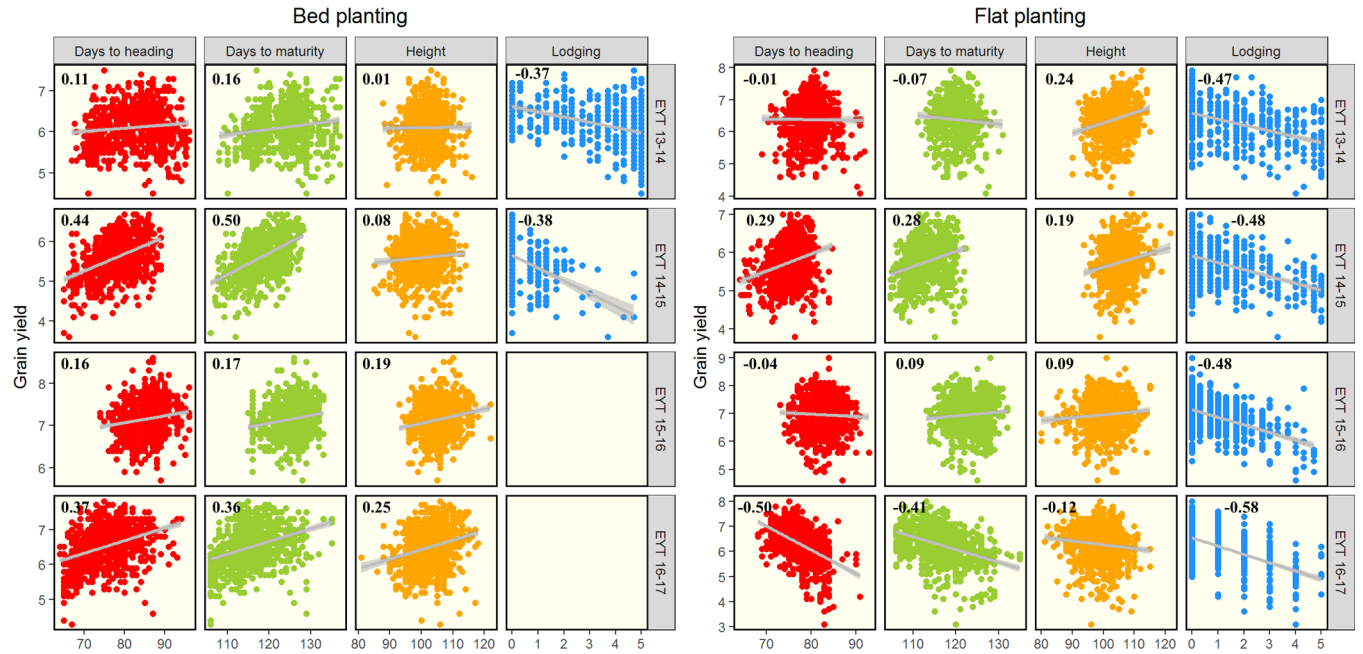


Fig. 1. Correlations among wheat grain yield, days to heading, days to maturity, height, and lodging in bed and flat planting systems in the elite yield trial (EYT) nurseries.

lines discarded by GS that was also discarded by PS was calculated as:

$$\frac{NS}{NS + S_{GS}} \quad [6]$$

The percentage of top lines selected by GS that was also selected by PS (Supplemental Fig. S1) was calculated as:

$$\frac{S_{GSPS}}{S_{GSPS} + S_{PS}} \quad [7]$$

RESULTS

Phenotypic Data and Principal Component Analysis

The mean GY in EYT 13–14, EYT 14–15, EYT 15–16 and EYT 16–17 was 6.1 ± 0.5 , 5.6 ± 0.5 , 7.2 ± 0.4 , and 6.5 ± 0.6 t ha⁻¹, respectively in the bed planting system and 6.4 ± 0.6 , 5.7 ± 0.5 , 6.9 ± 0.6 , and 6.3 ± 0.7 t ha⁻¹, respectively in the flat planting system. The line-mean heritabilities for GY in these nurseries were 0.58, 0.73, 0.66, and 0.69 in the bed planting system and 0.8, 0.6, 0.66, and 0.68 in the flat planting system. The correlations between GY in the bed and flat planting systems were low to moderate: 0.40 (EYT 13–14), 0.57 (EYT 14–15), 0.12 (EYT 15–16), and 0.12 (EYT 16–17).

In the bed planting system, the correlation of GY with DTHD and DTMT was positive, although it was slightly higher in EYT 14–15 and EYT 16–17 (Fig. 1). We also observed that the average DTHD was higher in EYT 15–16 (85 d) and EYT 13–14 (82 d). A similar trend was also observed with DTMT, with higher averages in EYT 15–16 (125 d) and EYT 13–14 (123 d). Plant height had a positive correlation with GY in all the nurseries (ranging between 0.01 and 0.25), and lodging had a negative correlation with GY in EYT 13–14 and EYT 14–15 (-0.37 and -0.38).

In the flat planting system, GY had a negative correlation with DTHD in all the nurseries except EYT 14–15. Considering DTMT, the correlation with GY was positive in EYT 14–15 and EYT 15–16 (0.28 and 0.09), but it was negative in EYT 13–14 and EYT 16–17 (-0.07 and -0.41). The average DTHD was highest in EYT 15–16 (83 d), followed by EYT 13–14 (80 d). Similarly, with DTMT, the lines in EYT 15–16 (123 d) and EYT 13–14 (120 d) had higher average DTMT. We also observed that GY had a positive correlation with plant height in three nurseries (ranging between 0.09 and 0.24), but a negative correlation in EYT 16–17 (-0.12). Lodging was negatively correlated with GY in all the nurseries and the correlations ranged between -0.47 and -0.58.

ANOVA indicated that the families contributed significantly to the GY variation in the bed and flat planting systems, but the sibs nested within families had no significant contribution to GY variability in the different nurseries (Supplemental Table S2). Principal component analysis indicated that there was no grouping of nurseries or population structure when Principal Component 1 (explained 7.8% of the variation) was plotted against Principal Component 2 (explained 6.3% of the variation) (Fig. 2).

Genomic and Pedigree Relationships

The **G** matrices and **A** matrices from all the nurseries were rescaled between zero and one, to facilitate comparisons and were visualized by heat maps (Supplemental Fig. S2). We observed that several lines with zero to low relationships (0–0.2) with the **A** matrix had relationships of 0.2 or slightly higher with the **G** matrix. This difference in the degree of relationships captured by the two matrices could be caused by (i) the realized relationships between the lines captured by the **G** matrix and (ii) the

high identity-by-state similarities in genomic regions that have been highly selected for various traits (e.g., heading and height) captured by the **G** matrix.

Considering family sizes, the lines in EYT 13–14 were from 488 crosses that included 335 crosses with only one line per cross and several other crosses with 2 to 11 full-sibs per cross. In EYT 14–15, the lines were from 386 crosses that included 226 crosses with one line per cross and several crosses with family sizes ranging from 2 to 12. In EYT 15–16, the lines were from 444 crosses that included 239 crosses with one line per cross and several families with 2 to 15 full-sibs. The lines in EYT 16–17 were from 408 crosses that included 199 crosses with one line per cross and several large families with 27, 20, 17, 15, 11, and 10 full-sibs each. The hotspots of high relationships seen in Supplemental Fig. S2 result from the families with a large number of full-sibs.

Genomic Prediction Accuracies within and across Nurseries

Bed Planting

In the bed planting system, the average cross-validation accuracies within nurseries (scheme 1) ranged between 0.37 and 0.51 in the four nurseries (Table 1). However, when EYT 13–14 was predicted using other nurseries as training populations (Scheme 2), the accuracies dropped down to 0.20, 0.02, and 0.13. When EYT 14–15 was predicted from other nurseries, EYT 16–17 (0.27) and EYT 13–14 (0.24) resulted in high accuracies. We also observed that EYT 15–16 was best predicted from EYT 16–17 (0.18), and EYT 16–17 was best predicted from EYT 15–16 (0.32). The mean genomic prediction accuracy within-nurseries was 0.44 ± 0.08 and the mean prediction accuracy across nurseries was 0.15 ± 0.12 .

In Scheme 3, where all the lines from previous nurseries were used in the training population, the genomic prediction accuracies for EYT 15–16 and EYT 16–17 were 0.03 and 0.08 lower than the accuracies obtained from just the lines from the previous nursery. Genomic predictions with BLUEs where DTHD and lodging were included as covariates yielded within-nursery accuracies of 0.19, 0.50, 0.25, and 0.39, for the four nurseries, respectively. These accuracies were 0.18, 0.12, and 0.12 lower than those from the model without covariates in EYT 13–14, EYT 15–16, and EYT 16–17, respectively, and similar in EYT 14–15. The across-nursery predictions, including the covariates, resulted in the same accuracies to a 0.06 decrease in accuracy.

Flat Planting

In the flat planting system, the average within-nursery prediction accuracies ranged between 0.28 and 0.56 in the four nurseries (Table 1). However, when EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17 were predicted from other nurseries, the highest accuracies achieved were 0.07, 0.08, 0.02, and 0.13, respectively. The mean within-nursery genomic prediction accuracy was 0.35 ± 0.06 and the mean across-nursery prediction accuracy was 0.05 ± 0.05 for GY in the flat planting system.

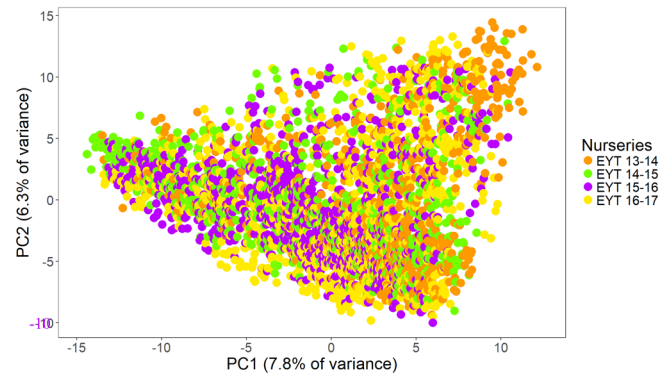


Fig. 2. Principal component analysis of the four elite yield trial (EYT) nurseries showing the plot of Principal Component 1 (PC1) vs. Principal Component 2 (PC2).

In Scheme 3, the accuracies from using all the lines in the previous nurseries resulted in similar to a 0.06 decrease in accuracies in EYT 15–16 and EYT 16–17, respectively. The GY BLUEs adjusted for covariates resulted in within-nursery prediction accuracies ranging between 0.21 and 0.40, that were 0.07, 0.03, 0.07, and 0.16 lower than the accuracies obtained from the GY BLUEs without covariates in EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17, respectively. In across-nursery predictions with covariates, accuracies that were similar to 0.11 lower were obtained.

Genomic Prediction Accuracies with Marker Subsets

The effect of marker number on prediction accuracies was assessed and we observed that subsets of 810, 504, and 251 markers (corresponding to r^2 values of 0.8, 0.6, and 0.4, respectively) did not lead to a significant loss in prediction accuracies in either planting system (Table 1). Although the accuracy of predictions reduced negligibly with a reduced number of markers, the average decrease with these subsets was only 0.01. While the highest reduction (–0.18) was observed when EYT 14–15 was predicted from EYT 16–17, a slight increase in accuracy (up to 0.07) was obtained in several datasets. However, when 29 markers corresponding to an r^2 of 0.2 were used, it led to a 0.11 loss in accuracy on average, as expected with very few markers for a complex trait. While the maximum decrease in accuracy was observed in the within-nursery prediction of EYT 15–16 (–0.33), a slight increase in accuracies (ranging between 0.01 and 0.08) was obtained in six datasets.

Genomic and Pedigree-Based Prediction Accuracies within and across Nurseries

Bed Planting

In the bed planting system, the mean within-nursery prediction accuracies with the pedigree model ranged between 0.39 and 0.60 in the four nurseries (Fig. 3). Although these accuracies were similar to the mean genomic prediction accuracies in EYT 13–14 and EYT 14–15, they were 0.07 and 0.09 higher in EYT 15–16 and EYT 16–17, respectively. In across-nursery pedigree-based predictions, the pedigree-based prediction accuracies were, in general, lower than the

Table 1. Genomic prediction accuracies for grain yield in bed and flat planting systems from the genomic best linear unbiased prediction (GBLUP) model with all the markers and marker subsets

Planting system	Bed planting system						Flat planting system					
	All markers		0.8	0.6	0.4	0.2	All markers		0.8	0.6	0.4	0.2
	Marker set†											
Covariates‡	NC	WC		NC			NC	WC		NC		
EYT 13–14	0.37	0.19	0.40	0.39	0.34	0.24	0.28	0.21	0.28	0.31	0.31	0.16
EYT 13–14 from EYT 14–15	0.20	0.14	0.20	0.18	0.18	0.05	0.07	0.02	0.11	0.14	0.12	0.01
EYT 13–14 from EYT 15–16	0.02	-0.03	0.02	0.01	0.01	0.09	0.02	0.05	0.02	0.02	0.02	0.06
EYT 13–14 from EYT 16–17	0.13	0.08	0.15	0.09	0.06	0.04	0.02	0.01	0.02	0.003	0.01	0.02
EYT 14–15	0.51	0.50	0.50	0.52	0.47	0.29	0.43	0.40	0.41	0.44	0.41	0.22
EYT 14–15 from EYT 13–14	0.24	0.22	0.23	0.23	0.21	0.07	0.08	0.04	0.09	0.11	0.10	0.03
EYT 14–15 from EYT 15–16	0.05	0.05	0.07	0.07	0.10	0.10	0.03	-0.01	0.01	0.07	0.05	-0.01
EYT 14–15 from EYT 16–17	0.27	0.27	0.21	0.18	0.09	0.02	0.03	0.002	-0.01	-0.02	-0.06	-0.08
EYT 15–16	0.37	0.25	0.37	0.32	0.23	0.04	0.46	0.39	0.45	0.42	0.37	0.25
EYT 15–16 from EYT 13–14	-0.03	-0.07	-0.02	-0.02	-0.02	0.05	0.004	0.08	-0.01	0.01	0.01	0.05
EYT 15–16 from EYT 14–15	0.08	0.08	0.10	0.09	0.10	0.04	-0.01	-0.02	-0.03	-0.01	-0.01	-0.05
EYT 15–16 from EYT 16–17	0.18	0.13	0.17	0.13	0.12	0.03	0.02	-0.04	0.01	-0.01	0.05	0.03
EYT 15–16 from EYT 13–14 and EYT 14–15	0.05	0.03	0.03	0.02	0.02	-0.02	-0.03	-0.01	-0.01	-0.01	0	-0.03
EYT 16–17	0.51	0.39	0.50	0.49	0.46	0.26	0.56	0.40	0.57	0.55	0.53	0.25
EYT 16–17 from EYT 13–14	0.04	0.02	0.07	0.07	0.06	0.02	0.11	0.11	0.13	0.12	0.12	0.002
EYT 16–17 from EYT 14–15	0.28	0.28	0.26	0.25	0.22	-0.01	0.12	0.01	0.17	0.15	0.12	-0.03
EYT 16–17 from EYT 15–16	0.32	0.29	0.31	0.31	0.29	0.08	0.13	0.09	0.14	0.10	0.17	0.07
EYT 16–17 from EYT 13–14, EYT 14–15 and EYT 15–16	0.24	0.20	0.18	0.17	0.15	0.08	0.07	0.05	0.04	0.04	0.02	-0.01

† “All markers” refers to the complete set of 2038 markers, 0.8 is a marker subset comprising 810 markers ($r^2 = 0.8$), 0.6 is a marker subset comprising 504 markers ($r^2 = 0.6$), 0.4 is a marker subset comprising 251 markers ($r^2 = 0.4$), and 0.2 is a marker subset comprising 29 markers ($r^2 = 0.2$).

‡ NC, no covariates; WC, days to heading and lodging used as covariates; EYT, elite yield trial.

corresponding genomic prediction accuracies. Overall, the mean prediction accuracy with the pedigree model was 0.48 ± 0.09 within nurseries and 0.09 ± 0.07 across nurseries.

In the markers plus pedigree model, the within-nursery accuracies ranged between 0.41 and 0.61. These accuracies were slightly higher (ranging between 0.04 and 0.1) than the accuracies from the GBLUP model. In across-nursery predictions, accuracies that were similar or slightly lower (a maximum decrease of 0.08) than the genomic prediction accuracies were obtained. Overall, the mean prediction accuracy within nurseries was 0.51 ± 0.09 and the mean prediction accuracy across nurseries was 0.14 ± 0.10 with the markers plus pedigree model.

Flat Planting

In the flat planting system, the mean within-nursery prediction accuracies with the pedigree model ranged between 0.30 and 0.62 (Fig. 3). Although these accuracies were similar to the genomic prediction accuracies for EYT 13–14 and EYT 14–15, they were 0.12 and 0.06 higher for EYT 15–16 and EYT 16–17, respectively. When EYT 13–14, EYT 14–15 and EYT 15–16 were predicted from other nurseries via the pedigree model, negative to zero accuracies were obtained. The across-nursery genomic prediction accuracies for these three nurseries were also very poor and meaningful comparisons with the pedigree-based model accuracies could not be made. However, when EYT 16–17 was predicted from other nurseries, the accuracies were comparatively better, with EYT 13–14 resulting in the highest prediction accuracy (0.17). Overall, the mean within-nursery prediction

accuracy was 0.48 ± 0.15 and the mean across-nursery prediction accuracy was 0.02 ± 0.07 with the pedigree-based model for GY in the flat planting system.

In the markers plus pedigree model, the within-nursery GY prediction accuracies were 0.04, 0.07, 0.12, and 0.06 higher than the genomic prediction accuracies in EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17, respectively. However, the across-nursery predictions resulted in poor accuracies that were similar or slightly lower or higher than the genomic prediction accuracies. The mean within-nursery prediction accuracy for GY in the flat planting system was 0.51 ± 0.13 , and the mean across-nursery prediction accuracy was 0.04 ± 0.05 with the markers plus pedigree model.

Genomic and Pedigree-Based Prediction Accuracies in Populations with and without Full-Sibs in the Training Population

The cross-validation genomic prediction accuracies in populations without full-sibs were 0.18, 0.17, 0.22, and 0.15 lower than the corresponding genomic prediction accuracies for the complete set of lines in EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17 respectively, in the bed planting system (Table 2). Similarly, in the flat planting system, the accuracies in populations without full-sibs were similar or were 0.11, 0.24, and 0.24 lower than the corresponding genomic prediction accuracies for all the lines in the four nurseries, respectively. The prediction accuracies obtained from the set of lines that had one or more full-sibs in the training population were either similar or slightly better (a maximum increase of

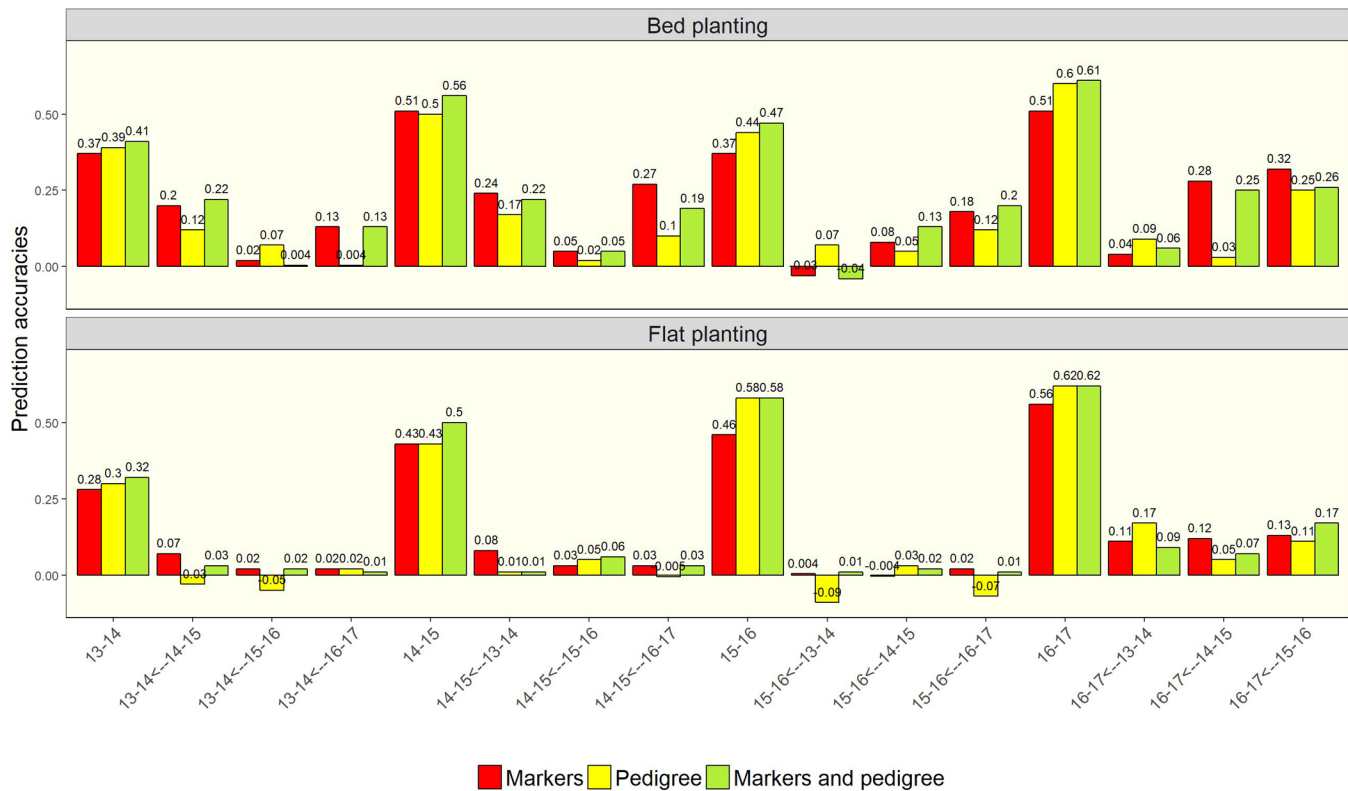


Fig. 3. Prediction accuracies for wheat grain yield with the genomic best linear unbiased prediction (GBLUP) model with markers, pedigree model, and the markers plus pedigree model in the bed and flat planting systems.

0.08) than the accuracies obtained from the complete set of lines in both the planting systems. In general, using all other full-sibs in a cross to predict one random full-sib resulted in slightly better prediction accuracies (a maximum increase of 0.14) than having one random full-sib to predict all other full-sibs.

In pedigree-based predictions, the accuracies in populations without full-sibs were 0.16 to 0.44 lower than the corresponding accuracies for the complete set of lines in both the planting systems. The prediction accuracies from using one or more full-sibs in the training population were generally superior to the prediction accuracies in the complete set, with a maximum increase in accuracy of 0.09. Finally, including all other full-sibs in the training population to predict one random full-sib resulted in a maximum increase of 0.13 in accuracy over using one random full-sib to predict others.

A clear advantage of the **G** matrix over the **A** matrix was seen in the case of predicting lines with no full-sibs in the training population. Here, the realized relationships captured by the markers led to similar, a 0.23 increase, a 0.04 decrease, and a 0.17 increase in accuracy over those obtained from the pedigree for GY in the bed planting system in EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17 respectively. Similarly, we obtained a 0.15 increase, 0.16 increase, 0.12 decrease and 0.14 increase in the accuracies for GY in the flat planting system in EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17, respectively, with the **G** matrix. However, when the training population had at least one full-sib, the **A** matrix

performed similar to or gave a maximum increase of 0.13 in accuracy over the genomic prediction accuracy.

Prediction Accuracies from the IBCF Approach

We compared genomic prediction accuracies from the GBLUP model with prediction accuracies from the IBCF approach for the training population Scheme 1 and with only lines from the previous nursery as the training population (Table 3). In Scheme 1, GY in the bed planting system was predicted at a higher accuracy via the IBCF approach, with 0.04, 0.06, and 0.16 increases in accuracy in EYT 13–14, EYT 14–15, and EYT 15–16, respectively. However, in EYT 16–17, a decrease (0.07) in accuracy compared with the GBLUP model was observed. When the lines in the previous nursery were used as the training population, EYT 14–15, EYT 15–16, and EYT 16–17 were predicted at better accuracies with the IBCF approach than with the GBLUP model, resulting in 0.12, 0.15, and 0.07 increases in accuracy, respectively. Overall, a maximum increase of 0.16 in within-nursery prediction accuracies and a 0.15 increase in across-nursery prediction accuracies were obtained with the IBCF approach over the GBLUP model in the bed planting system.

In the flat planting system, the IBCF accuracies for Scheme 1 were 0.09, 0.08, 0.28, and 0.14 lower than the corresponding GBLUP accuracies in EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17, respectively. However, when the previous year's data was used as the training population, the IBCF approach led to similar (EYT 15–16) and slightly better accuracies, resulting

Table 2. Genomic prediction accuracies for grain yield in the bed and flat planting systems with the genomic best linear unbiased prediction (GBLUP) model in populations with and without full-sibs

Dataset		Lines with no full-sibs		Lines with one or more full-sibs			
Training population				All other full-sibs		One random full-sib	
Validation population				One random full-sib		All other full-sibs	
Nursery	Planting system	Bed	Flat	Bed	Flat	Bed	Flat
EYT 13–14†	No. of individuals	343		424		424	
	No. of individuals in the training population	274		273		151	
	No. of individuals in the validation population	69		151		273	
	Genomic prediction accuracy	0.19	0.29	0.44	0.3	0.42	0.22
	Pedigree prediction accuracy	0.19	0.14	0.48	0.36	0.45	0.31
EYT 14–15	No. of individuals	226		549		549	
	No. of individuals in the training population	181		390		159	
	No. of individuals in the validation population	45		159		390	
	Genomic prediction accuracy	0.34	0.32	0.54	0.51	0.49	0.43
	Pedigree prediction accuracy	0.11	0.16	0.59	0.53	0.49	0.47
EYT 15–16	No. of individuals	243		721		721	
	No. of individuals in the training population	194		517		204	
	No. of individuals in the validation population	49		204		517	
	Genomic prediction accuracy	0.15	0.22	0.4	0.45	0.26	0.4
	Pedigree prediction accuracy	0.19	0.34	0.47	0.58	0.34	0.5
EYT 16–17	No. of individuals	201		779		779	
	No. of individuals in the training population	161		486		293	
	No. of individuals in the validation population	40		293		486	
	Genomic prediction accuracy	0.36	0.32	0.38	0.51	0.5	0.55
	Pedigree prediction accuracy	0.19	0.18	0.45	0.47	0.6	0.57

† EYT, elite yield trial.

Table 3. Genomic prediction accuracies for grain yield in bed and flat planting systems from the genomic best linear unbiased prediction (GBLUP) model and prediction accuracies from the item-based collaborative filtering (IBCF) approach

Planting system	Predictor(s)	Genomic relationships		Genomic relationships, days to heading, days to maturity, height, and lodging	
	Approach	GBLUP		IBCF	
	Training population†	Scheme 1	Lines from the previous nursery	Scheme 1	Lines from the previous nursery
Bed	EYT 13–14‡	0.37		0.41	
	EYT 14–15	0.51	0.24	0.57	0.36
	EYT 15–16	0.37	0.08	0.53	0.23
	EYT 16–17	0.51	0.32	0.44	0.39
Flat	EYT 13–14	0.28		0.19	
	EYT 14–15	0.43	0.08	0.35	0.21
	EYT 15–16	0.46	-0.01	0.18	0.02
	EYT 16–17	0.56	0.13	-0.42	-0.48

† Scheme 1 refers to using some lines from the same nursery as the training population.

‡ EYT, elite yield trial.

in increases of 0.13 (EYT 14–15) and 0.35 (EYT 16–17). With EYT 16–17, the IBCF approach resulted in negative within and across-nursery prediction accuracies with both training population schemes because of the negative correlation of GY with DTHD in this nursery.

Phenotypic Selection and GS

Phenotypic Selection and GS within Nurseries by Including some Lines from the Nursery in the Training Population (Scheme 1)

We evaluated the scenario of making selections within nurseries when some lines from each nursery were

included in the training population. In the bed planting system, on average, 93, 86, 80, and 65% of the poor lines were discarded by GS and 34, 45, 53, and 65% of the top lines were selected by GS at thresholds of 0.1, 0.2, 0.3, and 0.5, respectively, in EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17 (Table 4). However, Table 4 clearly indicates inflated estimates of the lines discarded, when fewer lines were selected than discarded (i.e., 92.5% of the poor lines were discarded on average at a selection intensity of 0.1, 85% of the poor lines were discarded on average at a selection intensity of 0.2, and 80% of the poor lines were discarded on average at a selection intensity of 0.3). Hence, we considered it appropriate

Table 4. Phenotypic selection (PS) and genomic selection (GS) using the genomic best linear unbiased prediction (GBLUP) model for grain yield in the bed and flat planting systems

Planting system		Bed planting			Flat planting	
<u>Training population: Scheme 1 (lines from the same nursery)</u>						
Nursery	Phenotypic selection threshold	0.1	0.2	0.3	0.5	0.5
	Genomic selection threshold	0.1	0.2	0.3	0.5	0.5
EYT	Poor lines discarded (%)	92.0	85.9	80.5	63.4	61.1
13–14†	Top lines selected (%)	31.3	44.3	50.8	64.6	63.7
EYT	Poor lines discarded (%)	93.1	86.5	81.4	67.8	64.1
14–15	Top lines selected (%)	39.0	46.5	57.3	68.6	63.9
EYT	Poor lines discarded (%)	92.7	84.6	77.1	60.1	65.7
15–16	Top lines selected (%)	34.3	40.6	46.9	59.1	65.5
EYT	Poor lines discarded (%)	92.2	86.3	80.7	68.1	71.5
16–17	Top lines selected (%)	29.3	46.9	55.9	68.5	70.9
<u>Training population: Lines from the previous nursery</u>						
Nursery	Phenotypic selection threshold	0.1	0.2	0.3	0.5	0.5
	Genomic selection threshold	0.1	0.2	0.3	0.5	0.5
EYT	Poor lines discarded (%)	89.4	79.9	72.9	60.5	49.8
14–15	Top lines selected (%)	5.0	23.6	39.7	61.5	49.4
EYT	Poor lines discarded (%)	91.1	80.6	72.3	54.4	50.4
15–16	Top lines selected (%)	18.0	27.9	33.8	53.8	50.4
EYT	Poor lines discarded (%)	92.1	84.3	77.6	57.6	55.7
16–17	Top lines selected (%)	22.2	41.6	47.5	57.8	56.5

† EYT, elite yield trial.

to use only a selection intensity of 0.5 (where an equal number of lines are selected and discarded) for PS and GS comparisons. Thus, in the flat planting system, on average, 66% of the poor lines were discarded and 66% of the top lines were selected by both PS and GS at a selection intensity of 0.5 (Table 4). Overall, an average of 33% of the lines were selected by both PS and GS, 17.3% of the lines were selected by PS only, 33% of the lines were discarded by both GS and PS, and 17.1% of the lines were selected by GS only in the bed and flat planting systems at a selection intensity of 0.5 (Fig. 4).

Phenotypic Selection and GS across Nurseries by Including Lines from the Previous Nursery in the Training Population

We evaluated the scenario of making selections when only the lines from the previous nursery were used as a training population. In the bed planting system, an average of 29% of the lines were discarded by GS and PS, 29% of the lines were selected by GS and PS, 21% of the lines were selected by PS only, and 21% of the lines were selected by GS only (Fig. 5). Overall, genomic predictions were successful in discarding, on average, 58% of the poor lines and selecting 56% of the top lines. In the flat planting system, on average, 26% of the lines were discarded by GS and PS, 26% of the lines were selected by GS and PS, 24% of the lines were selected by PS only, and 24% of the lines were selected by GS only (Fig. 5). Hence, GS was successful in selecting 52% of the top lines and discarding 52% of the poor lines.

DISCUSSION

Genomic Predictions within and across Nurseries

Our results indicated moderate to high genomic prediction accuracies for GY within nurseries, but a severe decrease in accuracies was observed when predictions were made across nurseries (and also across years). When GY was adjusted for covariates like DTHD and lodging, lower prediction accuracies were obtained because some markers that had significant effects on GY were also associated with these traits in some nurseries (unpublished results). We also observed differences in the GY predictive abilities of the same nurseries in the bed and flat planting systems, which can be attributed to the moderate to low correlations for GY in these planting systems and also the high incidence of lodging in the flat planting system.

The poor across-nursery GY predictions pose a serious challenge to implementing GS that can be attributed to decaying family relationships, decaying LD, and the effect of $G \times E$ interactions. In our results, it is unclear whether decaying LD or family relationships contributed to poor across-nursery predictions because these two factors are entangled and indistinguishable (Wientjes et al., 2013). However, the key player is $G \times E$ interactions because the validation populations were not evaluated in the same year as the training populations and the response of genotypes to different weather patterns that prevailed in the growing seasons (the average temperatures during the 2013–2014, 2014–2015, 2015–2016, and 2016–2017 seasons were 17.7 ± 1.7 , 18.4 ± 2.7 , 16.8 ± 2.3 , and $17.7 \pm 2^\circ\text{C}$, respectively) might have resulted in low prediction accuracies. Rutkoski et al. (2015) also concluded that forward predictions might be largely driven by the level of $G \times E$ interactions between the training and validation populations, rather than the relationships between the two populations, which is in accordance with the results of this study.

The ‘year’ effect was also reflected in the poor phenotypic GY correlations of the lines in the EYT nurseries and the same lines previously evaluated as first-year yield trials (two replicates) in the bed planting system (0.23, 0.30, 0.23, and 0.18 for EYT 13–14, EYT 14–15, EYT 15–16, and EYT 16–17, respectively). Therefore, when GY correlations across years are low, our ability to use data from 1 yr to predict a related line’s performance in another year will be limited. In addition, extraneous field and management variations also hamper accurate predictions of GY. We also observed that the correlations of GY with traits like DTHD, DTMT, and height were variable across years and even between the two planting systems within a year, which had a substantial effect on GY and the predictive abilities. For example, in the 2014–2015 season where the temperatures were higher than usual, lower GY, early heading, and early maturity were observed.

Effect of Marker Number on Genomic Predictions

The effect of marker number on prediction accuracies was assessed, and the use of subsets of markers did not lead to a significant loss in prediction accuracies. This result is in

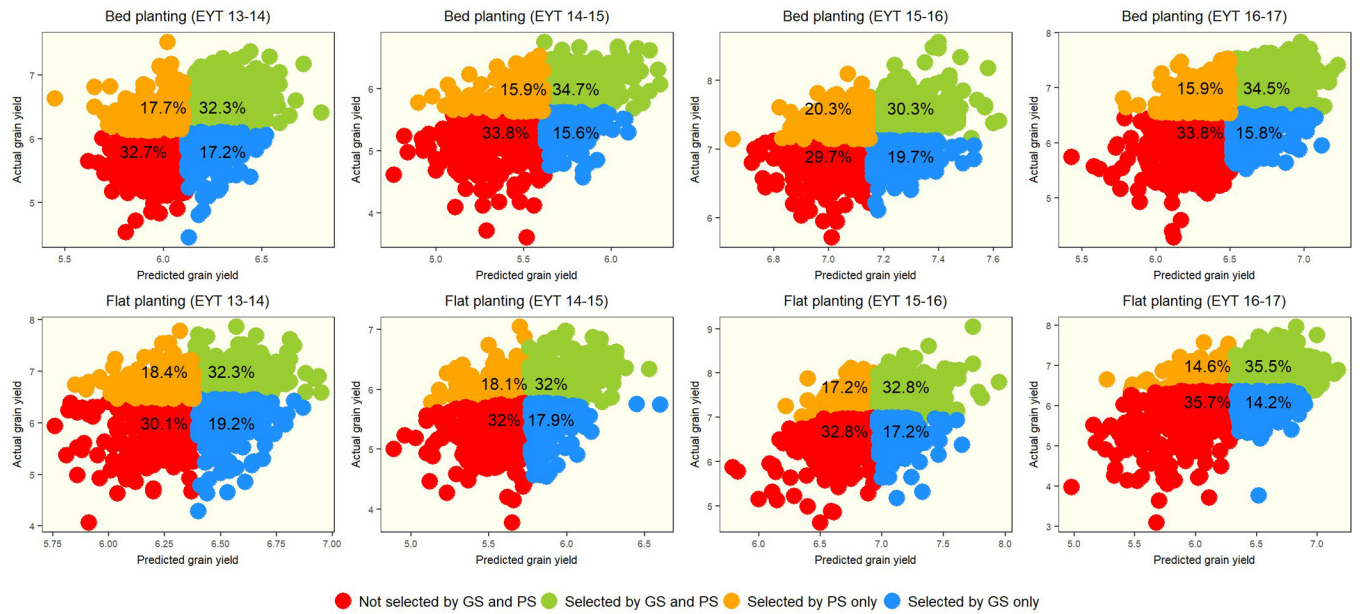


Fig. 4. Phenotypic selection (PS) and genomic selection (GS) for wheat grain yield within nurseries in the bed and flat planting systems.

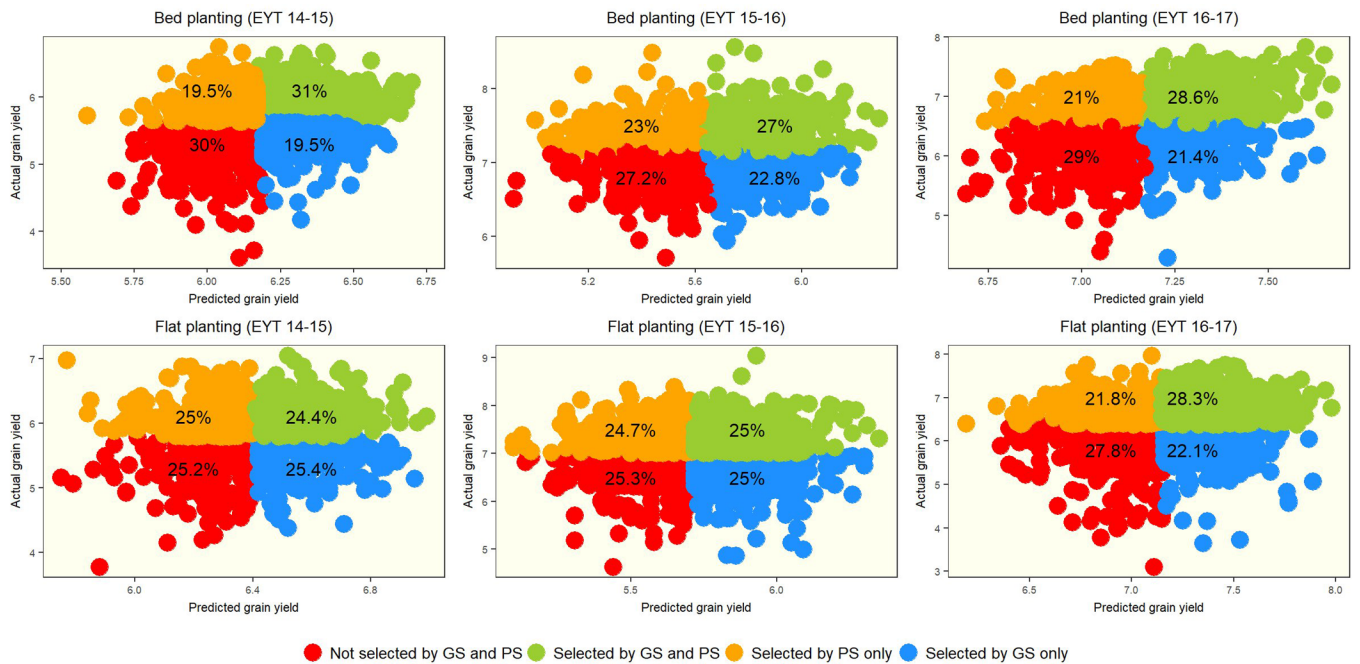


Fig. 5. Phenotypic selection (PS) and genomic selection (GS) for wheat grain yield across nurseries in the bed and flat planting systems, with lines from the previous nursery as the training population.

agreement with previous studies that have reported only a small loss in accuracy with sparse markers or low-density marker panels (Luan et al., 2009; Weigel et al., 2009; Moser et al., 2010; Lenz et al., 2017). While the average decrease in accuracy with these subsets was only 0.01, about 251 markers were sufficient to obtain accuracies equivalent to those observed with all the 2038 markers, indicating the existence of extensive LD in wheat and high collinearity among the GBS markers. Although this implies that genomic selection for GY in these populations can be implemented with a limited number of markers, it is also a concern, because 251 informative markers might only capture the broad

relationships among the lines and not the minor differences among full-sibs. This could also be the reason why the pedigree-based predictions outperformed the genomic predictions in populations with full-sibs.

Prediction of GY via the IBCF Approach

The IBCF approach was used to predict GY incorporating information from multiple traits and we observed that it gave a maximum of 0.16 and 0.15 increase in within and across-nursery accuracies over the GBLUP model in the bed planting system. However, in the flat planting system, the GBLUP model outperformed the IBCF approach

within nurseries (maximum increase in accuracy of 0.28), whereas the IBCF resulted in similar to slightly higher accuracies across nurseries. The negative accuracies obtained with the IBCF approach in EYT 16–17 occurred because of the negative correlation of DTHD with GY in this nursery, which was not observed for any other nursery. This indicates that changes in correlations of GY with related traits will affect the predictive ability, especially when one particular year is different from the others. While the ability of the IBCF approach to yield accuracies higher than the genomic prediction accuracies across nurseries is encouraging, it is also not desirable for GY to have very high correlations with these traits, because we could just be selecting for these traits rather than real GY.

Effect of the Training Population Schemes on Genomic Prediction

The composition, size, and optimization of the training population are critical in GS. We investigated three training population schemes, and our results indicated that including lines from within nurseries (and years) in the training population always resulted in the highest accuracy. When only the previous year's data were included in the training population, we obtained slightly higher accuracies than when we used all the data from previous years. Several studies have reported an increase in accuracy with large training population sizes and it has also been suggested that combining phenotypes from several populations in a large training population or adding a few individuals from the unrelated population to the training population might increase accuracy (Muir, 2007; de Roos et al., 2009; Daetwyler et al., 2010; Asoro et al., 2011; Jarquín et al., 2014b). However, our results indicate that increasing the training population size might not necessarily result in high accuracies, as also observed by Dawson et al. (2013) and Lorenz and Smith (2015). It was also suggested by de Roos et al. (2009) that combining populations in a training population may be less advantageous when there is a substantial quantitative trait locus \times environment interaction, as marker effects will differ across populations, which might explain why using all the lines from the previous nurseries was not very successful in our study. However, our results are based on inferences from limited datasets and in diverged populations, higher marker densities may be required for exploiting the advantage of larger training populations and achieving high accuracies (de Roos et al., 2009). Nevertheless, the forward prediction scenario from a correlated previous nursery or year is realistic for applied breeding programs and will be ideal in cases where there are more related sibs in the previous nursery.

Comparison of Genomic and Pedigree-Based Predictions

Pedigree-based prediction accuracies were similar or slightly better (a maximum increase of 0.12) than the genomic prediction accuracies within nurseries and the model with markers and pedigree resulted in the highest accuracy, consistent with previous reports (de los Campos et al., 2009; Pérez et al., 2010; Burgueño et al., 2012; Bartholomé et al., 2016; Juliana et al.,

2017a). However, the advantage of considering both the markers and pedigree was not very high, because of some redundancy in the information captured by both, as also reported by Crossa et al. (2010).

Although genomic predictions are expected to result in higher accuracies than pedigree-based predictions, because the Mendelian sampling term and allele sharing between sibs are exploited (Daetwyler et al., 2007; Hayes et al., 2009), the high within-nursery accuracies from the pedigree-based predictions were a consequence of the family structure in these nurseries. In other words, about 50% of the lines in each nursery had no full-sibs and were represented by only one individual per cross, but the remaining lines that had at least one full-sib in the nursery had minimal variance for GY, because of being selected previously for high GY. In this case, genomic predictions will have limited scope for being advantageous over pedigree-based predictions, as the Mendelian sampling variance is not a significant source of the GY variability (clearly observed from the insignificant effect of sibs nested within families in the ANOVA). However, the clear advantage of the genomic relationship matrix over the pedigree relationship matrix was seen in the case of predicting lines with no full-sibs, where the realized relationships captured by markers led to a moderate increase in accuracy.

The genetic relatedness between individuals in the training and validation populations has to be maximized for obtaining precise GEBV estimates (Habier et al., 2010; Clark et al., 2012; Pszczola et al., 2012; Thorwarth et al., 2017). Although smaller training populations can be used when the validation population is closely related to the training population (Mackay et al., 2015), it was suggested that at least one closely related line should be present in the training and validation populations to achieve higher accuracies with distantly related individuals (Daetwyler et al., 2014). In our study, we observed that when the training population had at least one full-sib from a cross, the pedigree-based predictions performed similar to or gave a maximum increase in accuracy of 0.13 over the genomic prediction accuracies. This implies that although markers account for genomic relationships that occurred among the founders of the population, the narrow variance among full-sibs and the insignificant contribution of more distant relationships further back in time (before the five generations of pedigree considered), as also observed by Luan et al. (2012), might have contributed to high accuracies with the pedigree. In addition, the imperfect estimation of the identity-by-state relationships by the limited number of markers available for this study might have resulted in high pedigree-based prediction accuracies. Vela-Avitúa et al. (2015) reported that the accuracy of genomic predictions based on the identity-by-state relationships declined rapidly as marker densities declined and, at the lowest densities, genomic predictions based on the identity-by-state relationships was even outperformed by the pedigree-based model. Nevertheless, our results emphasize that in highly pedigreed breeding populations where across-family selection is combined with selection among the top within-family performers in advanced lines, the predictive ability of the genomic relationship

matrix compared with the pedigree relationship matrix will be underestimated.

Comparison of PS and GS

While increasing the selection intensity increases the rate of genetic gain (Pryce and Daetwyler, 2012), we observed that at a stringent selection intensity of 0.1, we would be at the risk of losing of $66.5 \pm 4.2\%$ of the top performers on average in within-nursery predictions and $84.4 \pm 6.4\%$ of the top performers in across-nursery predictions with GS. This risk should be taken into concern to avoid losing top performers that may be at a low frequency, but have rare alleles with substantial effects that the models were not trained to predict. Overall, a maximum of 72% of the poor lines was discarded and 71% of the top lines was selected within nurseries via GS. In across-nursery predictions, a maximum of 61% of the poor lines was discarded and 62% of the top lines was selected by GS.

Prospects of Implementing GS for GY in an Applied Wheat Breeding Program

The successful implementation of GS in a breeding program will depend largely on the stage at which it is used and the gains per unit of time and cost compared with conventional breeding. Our results clearly indicate that GS will be less advantageous at a stage where there is minimal Mendelian sampling variance, as there is no scope for outperforming pedigree-based predictions. This demonstrates the importance of applying GS at the appropriate stage of the breeding cycle to obtain maximum gains and some prospects of implementing GS for GY in bread wheat are discussed below.

Integrating GS with PS and Increasing Gains per Unit Cost by Substituting Replications with GS

One or more within-year GY replications can be substituted by GEBV-based selections, since the within-nursery genomic prediction accuracies were high and only slightly lower than the line-mean heritabilities. While this would enhance the accuracy of within-year selections and capitalize on the genetic variance for GY, gains per unit cost can also be maximized if the genotyping data are used for multiple trait selections. Assuming that the cost for yield-testing one unit of replication on a plot-basis is US\$10 and the cost of genotyping a line is US\$20, the gains per unit cost can be maximized if the genotyping cost is shared for 20 key traits (GY in 6 environments, 4 diseases, and 10 quality traits) and the genotyping cost per trait is US\$1. This integrated strategy would result in a reduction of yield-testing costs for a breeding program while also minimizing the risks associated with complete abstinence from phenotyping.

Looking beyond Prediction Accuracies

While prediction accuracies are key to successful evaluation of GS, it is important to consider a trait like GY as a unique case and rethink how precise the predicted values would be, given the low heritability and high $G \times E$ interactions. Although pursuing accuracies close to one for such

a trait will be an impossibility, it is difficult to set a target accuracy beyond which a breeder might apply GS for GY across nurseries. Hence, the way to move forward with GS for a complex trait like GY is to not rely solely on prediction accuracies but look beyond it as to how effectively the GEBVs can be used. For this, we should take advantage of a key finding in this study that even at low prediction accuracies across nurseries, we could still select or discard a reasonable number of lines with GS. Since an average of 65.6% (716 lines) and a maximum of 71.2% (778 lines) were selected and discarded by PS and GS within each nursery, cost savings of about US\$7160 to US\$7780 from using GEBVs for selections within nurseries can be obtained.

Using Genomic Selection for Early-Generation Within-Family Selections

While this study has explored the possibility of using GS for GY in the EYT stage, we speculate that greater gains can be achieved if it is implemented in early-generation within-family selections, where yield testing is not feasible because of limited seed. However, a pilot study with some families and a reasonable number of individuals per family is required to understand the comparative advantage of GS over conventional breeding.

Challenges for Implementing Genomic Selection for Grain Yield in an Applied Wheat Breeding Program
Some of the considerable challenges in implementing GS in an applied bread wheat breeding program are discussed below.

Poor Prediction Accuracies across Nurseries and Families

The poor accuracies from across-nursery and across-family predictions remain a major challenge for implementing GS, as observed in this study and also in previous studies (Reif et al., 2013; Crossa et al., 2014). However, this problem is also prevalent in animal breeding, where prediction equations trained with one breed do not accurately predict another breed or a reproductively isolated population (Harris et al., 2008; Toosi et al., 2010; Taylor, 2014; Wientjes et al., 2015). This challenge can be addressed by including some parental information for unrelated lines in the training population and is similar to animal breeding, where inclusion of the sire's genetic information in the training population results in higher accuracies (Lund et al., 2009). In addition, including more phenotypes and markers might help in achieving higher prediction accuracies (Hickey et al., 2014), as effects estimated from a limited number of families might not detect rare alleles with a large effect in some families (Wientjes et al., 2015).

Genotype \times Environment Interactions

Though this study clearly indicates the role of $G \times E$ interactions in decreasing accuracies, genomic predictions for unknown future environments are also inherently complex. A lack of correlation among the years leading to genotype rank changes and variation in performance

across years will undoubtedly challenge GS (Cossa et al., 2017; Pérez-Rodríguez et al., 2017). Although, several studies have shown that modeling $G \times E$ can result in substantial gains in prediction accuracy (Burgueño et al., 2012; Heslot et al., 2013; Jarquín et al., 2014a; Lopez-Cruz et al., 2015; Cuevas et al., 2016), accurate estimations of $G \times E$ interactions require models trained with data from large populations that are replicated in different environments. Furthermore, some information on a genotype's performance in a particular environment is mandatory for implementing GS. For a wheat breeding program like that of CIMMYT, where selections are conducted on a limited number of locations for distribution of lines to several national programs in diverse environments or countries, selecting lines with minimal $G \times E$ interactions is key for the successful delivery of widely adapted varieties.

Low Heritability

While GS is expected to be promising for improving traits of low heritability, it should also be noted that markers will be able to explain only the heritable portion of the genetic variance. For a trait like GY, where heritability can be as low as 0.2 across years, it is not surprising that genomic predictions will result in poor accuracies. In addition, the use of BLUEs adjusted for different field designs as true breeding values for a trait like GY with high $G \times E$ interactions would be misleading, as it also results from nonheritable environmental effects, in addition to genetic effects (phenotype = genotype + environmental effects).

Marker Technologies and the Costs of Genotyping

It is unknown if the prediction accuracies obtained from GBS markers reflect the maximum accuracies that could be obtained from any whole-genome genotyping technology. Although alternative genotyping technologies are available and continue to evolve, their high cost is still prohibitive for large-scale implementation of GS. If we consider an ideal scenario of implementing GS in the F_2 stage of CIMMYT's wheat breeding scheme where maximum gains will be expected from within-family selections, the approximate cost of genotyping 1130,000 lines from F_2 simple crosses (1000 crosses \times 1130 plants per cross) and 1320,000 plants from F_2 top crosses (1200 crosses \times 1100 plants per cross) will be US\$49,000,000 at the rate of US\$20 per sample. In the F_3 stage, the cost of genotyping 415,000 lines (1000 crosses \times 415 plants per cross) will be US\$8300,000 USD. However, these costs do not reflect the cost of sampling, DNA extraction, database management, and analysis. In addition to the high cost of genotyping, other factors like the limited seed availability in early generations, the logistics and timeliness of genotyping massive populations, segregation of the populations, and the relative gain from GS compared with the low cost of line-development (US\$5.7 is the approximate cost of obtaining seed for planting F_5 to F_6 plots from individual spikes or plants), hinder implementation of GS in the early generations of a large breeding program like that of CIMMYT.

Optimizing Training Populations and the Frequency of Phenotyping Required for Model Training

Though large training populations for GY predictions exist for CIMMYT's bread wheat breeding program, leveraging historical information for prediction of yet to be observed phenotypes remains a challenge. Optimization of the training populations to identify a core subset of lines for reliable predictions is essential and has proved to be efficient in many studies (Rincent et al., 2012; Isidro et al., 2014).

Genomic selection was expected to change the role of phenotyping dramatically as in a GS-driven breeding cycle, where phenotypes will be used only to update prediction models and not for selecting lines (Heffner et al., 2009). However, our results indicate that the use of ancestral training populations subject to high environmental variations is not ideal for a GS-driven breeding scheme at CIMMYT, and the best strategy would be to use only data from immediate relatives or a correlated season, as also suggested in other studies (Lorenz and Smith, 2015; Hoffstetter et al., 2016). In addition, field and management variations across years might also result in stochasticity, thereby emphasizing the need for managed environments to collect phenotypes for training models in GS. Furthermore, if we consider the change in relationships and allele frequencies across populations, it is evident that large populations should be used for accurate estimation of marker effects (Lande and Thompson, 1990; Goddard and Hayes, 2007; Solberg et al., 2008; van der Werf, 2009; VanRaden et al., 2009). While molecular marker information can enhance the efficiency of PS (Whittaker et al., 2000), successful implementation of GS will require some amount of phenotyping (Bernardo and Yu, 2007). The frequency of phenotyping will depend largely on the genetic relatedness among the individuals in the training and validation populations and might be high for wheat breeding programs like that of CIMMYT, where new germplasm is introduced every year for novel diversity.

CONCLUSION

Designing a GS-driven wheat breeding program for GY needs careful consideration of several factors that are highlighted below:

- (i) A complex trait like GY will be challenging for GS, as much as it is challenging for PS, because GS models are trained with phenotypic data. Hence, the reliability of prediction models should be improved by using well-replicated phenotypic GY data for training, and accuracy estimates derived from single-year phenotypic performance should be considered cautiously.
- (ii) Genomic selection should be implemented within its scope; in other words, GS should not be expected to perform well when the models are trained with completely unrelated lines or with lines evaluated in noncorrelated environments.
- (iii) A significant advantage of genomic relationships over the pedigree for implementing GS in the yield testing stage was not observed in this study because of good pedigree records and the family structure in these

nurseries. However, wheat breeding programs that do not maintain pedigree records or have a reasonable number of full-sibs in the yield testing stage can effectively implement GS for sparse GY testing.

- (iv) This study has only evaluated GS for minimizing phenotyping in the advanced yield testing stage. However, it is also important to test GS for GY in early generations, where the real advantage of rapid cycling generations can be achieved. In addition, combining GS with faster breeding technologies like speed breeding (Watson et al., 2018) might lead to an acceleration of genetic gains by reducing time.

In conclusion, we have investigated and discussed several prospects and challenges of implementing GS for GY in wheat. However, the findings of this study are exclusive to the nurseries used and cannot be generalized to all breeding programs without further population-specific analysis. While GS for GY is promising in some scenarios, this is an exciting area that needs more research and understanding of the fit of GS in wheat breeding programs.

Supplemental Information

Supplemental File S1: Genotyping and phenotyping data for 3485 lines from the elite yield trial nurseries.

Supplemental Table S1: The rating matrix used for implementing the item-based collaborative filtering approach.

Supplemental Table S2: ANOVA for grain yield in the bed and flat planting systems.

Supplemental Fig. 1: Graphical representation of calculating the percentage of top lines selected by genomic selection (GS) that were also selected by phenotypic selection (PS) and the percentage of poor lines discarded by genomic selection (GS) that were also discarded by phenotypic selection (PS)

Supplemental Fig. 2: The genomic (upper panel) and pedigree (lower panel) relationship matrices for the four elite yield trial (EYT) nurseries (rescaled between 0 and 1), where the color key from blue to red represents relationships between 0 and 1. Several lines that had zero to low relationships (0 to 0.2) with the pedigree relationship matrix had relationships of 0.2 or slightly higher with the genomic relationship matrix.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

ACKNOWLEDGMENTS

This research was made possible through the support provided by the Delivering Genetic Gain in Wheat project (managed by Cornell University and funded by the Bill and Melinda Gates Foundation and the United Kingdom Department for International Development) and Feed the Future project through the US Agency for International Development (USAID), under the terms of Contract No. AID-OAA-A-13-0005. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the USAID. We also extend our sincere gratitude to the innovation lab at Kansas State University for their support in generating the genotyping data and also to Paul Tanger (USAID) for his helpful comments to improve the manuscript. PJ drafted the manuscript and performed the research. RPS, JP, and JC planned the study and supervised the analysis. OM performed the ICBF analysis. PP provided the pedigree relationships. SM, JH, LC, and VG generated the phenotyping data. SD performed the DNA extraction.

REFERENCES

- Addison, C.K., R.E. Mason, G. Brown-Guedira, M. Guedira, Y. Hao, R.G. Miller, et al. 2016. QTL and major genes influencing grain yield potential in soft red winter wheat adapted to the southern United States. *Euphytica* 209:665–677. doi:10.1007/s10681-016-1650-1
- Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome* 4:132–144. doi:10.3835/plantgenome2011.02.0007
- Bartholomé, J., J. Van Heerwaarden, F. Isik, C. Boury, M. Vidal, C. Plomion, et al. 2016. Performance of genomic prediction within and across generations in maritime pine. *BMC Genomics* 17:1–14. doi:10.1186/s12864-016-2879-8
- Battenfield, S.D., C. Guzmán, R.C. Gaynor, R.P. Singh, R.J. Peña, S. Dreisigacker, et al. 2016. Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome* 9(2). doi:10.3835/plantgenome2016.01.0005
- Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47:1082–1090. doi:10.2135/cropsci2006.11.0690
- 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
- Bradshaw, A.D. 1965. Evolutionary significance of phenotypic plasticity in plants. *Adv. Genet.* 13:115–155.
- Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52:707–719. doi:10.2135/cropsci2011.06.0299
- Calus, M.P.L., T.H.E. Meuwissen, A.P.W. de Roos, and R.F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics* 178:553–561. doi:10.1534/genetics.107.080838
- Clark, S.A., J.M. Hickey, H.D. Daetwyler, and J.H. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol.* 44:4. doi:10.1186/1297-9686-44-4
- Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, et al. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724. doi:10.1534/genetics.110.118521
- Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, G. de los Campos, et al. 2017. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* 22:961–975. doi:10.1016/j.tplants.2017.08.011
- Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, et al. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity (Edinb)* 112:48–60. doi:10.1038/hdy.2013.16
- Cuevas, J., J. Crossa, S. Víctor, S. Pérez-Elizalde, P. Pérez-Rodríguez, G. de los Campos, et al. 2016. Bayesian genomic prediction with genotype × environment interaction kernel models. *G3 (Bethesda)* 7(1):41–53. doi:10.1534/g3.116.035584
- Daetwyler, H.D., U.K. Bansal, H.S. Bariana, M.J. Hayden, and B.J. Hayes. 2014. Genomic prediction for rust resistance in diverse wheat landraces. *Theor. Appl. Genet.* 127:1795–1803. doi:10.1007/s00122-014-2341-8
- Daetwyler, H.D., J.M. Hickey, J.M. Henshall, S. Dominik, B. Gredler, J.H.J. Van Der Werf, et al. 2010. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Anim. Prod. Sci.* 50:1004–1010. doi:10.1071/AN10096
- Daetwyler, H.D., B. Villanueva, P. Bijma, and J.A. Woolliams. 2007. Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* 124:369–376.
- Dawson, J.C., J.B. Endelman, N. Heslot, J. Crossa, J. Poland, S. Dreisigacker, et al. 2013. The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crop. Res.* 154:12–22. doi:10.1016/j.fcr.2013.07.020
- de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, et al. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375–385. doi:10.1534/genetics.109.101501

- de Roos, A.P.W., B.J. Hayes, and M.E. Goddard. 2009. Reliability of genomic predictions across multiple populations. *Genetics* 183:1545–1553. doi:10.1534/genetics.109.104935
- Dekkers, J.C.M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* 124:331–341. doi:10.1111/j.1439-0388.2007.00701.x
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi:10.1371/journal.pone.0019379
- Endelman, J.B., and J.-L. Jannink. 2012. Shrinkage estimation of the realized relationship matrix. *G3 (Bethesda)* 2:1405–1413. doi:10.1534/g3.112.004259
- Fischer, R.A., and M. Stapper. 1987. Lodging effects on high-yielding crops of irrigated semidwarf wheat. *Field Crop. Res.* 17:245–258. doi:10.1016/0378-4290(87)90038-4
- Flintham, J.E., A. Borner, A.J. Worland, and M.D. Gale. 1997. Optimizing wheat grain yield : Effects of *Rht* (gibberellin-insensitive) dwarfing genes. *J. Agric. Sci. Cambridge* 128:11–25. doi:10.1017/S0021859696003942
- Gianola, D. 2013. Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* 194:573–596. doi:10.1534/genetics.113.151753
- Gianola, D., R.L. Fernando, and A. Stella. 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173:1761–1776. doi:10.1534/genetics.105.049510
- Gianola, D., and J.B.C.H.M. van Kaam. 2008. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178:2289–2303. doi:10.1534/genetics.107.084285
- Gilmour, A. 1997. ASREML for testing fixed effects and estimating multiple trait variance components. *Proc. Assoc. Adv.* 1:386–390.
- Gilmour, A.R., R. Thompson, and B.R. Cullis. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51:1440–1450. doi:10.2307/2533274
- Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS : A high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:e90346. doi:10.1371/journal.pone.0090346
- Goddard, M.E., and B.J. Hayes. 2007. Genomic selection. *J. Anim. Breed. Genet.* 124:323–330. doi:10.1111/j.1439-0388.2007.00702.x
- Griffiths, S., L. Wingen, J. Pietragalla, G. Garcia, A. Hasan, D. Miralles, et al. 2015. Genetic dissection of grain size and grain number trade-offs in CIMMYT wheat germplasm. *PLoS One* 10:1–18. doi:10.1371/journal.pone.0118847
- Habier, D., R.L. Fernando, and D.J. Garrick. 2013. Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics* 194:597–607. doi:10.1534/genetics.113.152207
- Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 42:5. doi:10.1186/1297-9686-42-5
- Harris, B.L., D.L. Johnson, and R.J. Spelman. 2008. Genomic Selection in New Zealand and the implications for national genetic evaluation. In: J.D. Sattler, editor, Identification, breeding, production, health and recording of farm animals. Proceedings of the 36th ICAR Biennial Session, Niagara Falls, NY. 16–20 June 2008. ICAR, Rome. p. 325–330
- Hayes, B.J., P.J. Bowman, A.J. Chamberlain, and M.E. Goddard. 2009. Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443. doi:10.3168/jds.2008-1646
- Hayes, B.J., H.A. Lewin, and M.E. Goddard. 2013. The future of livestock breeding: Genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 29:206–214. doi:10.1016/j.tig.2012.11.009
- Hayes, B.J., J. Panozzo, C.K. Walker, A.L. Choy, S. Kant, D. Wong, et al. 2017. Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. *Theor. Appl. Genet.* 130:2505–2519. doi:10.1007/s00122-017-2972-7
- Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. *Crop Sci.* 49:1–12. doi:10.2135/cropsci2008.08.0512
- Heslot, N., D. Akdemir, M.E. Sorrells, and J.L. Jannink. 2013. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127:463–480. doi:10.1007/s00122-013-2231-5
- Heslot, N., H. Yang, M.E. Sorrells, and J.-L. Jannink. 2012. Genomic selection in plant breeding: A comparison of models. *Crop Sci.* 52:146–160. doi:10.2135/cropsci2011.06.0297
- Hickey, J.M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, et al. 2014. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* 54:1476–1488. doi:10.2135/cropsci2013.03.0195
- Hoffstetter, A., A. Cabrera, M. Huang, and C. Sneller. 2016. Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat. *G3 (Bethesda)* 6:2919–2928. doi:10.1534/g3.116.032532
- Isidro, J., J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M.E. Sorrells. 2014. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* 128:145–158. doi:10.1007/s00122-014-2418-4
- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, et al. 2014a. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127:595–607. doi:10.1007/s00122-013-2243-1
- Jarquín, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz. 2014b. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15:740. doi:10.1186/1471-2164-15-740
- Jiang, Y., R.H. Schmidt, Y. Zhao, and J.C. Reif. 2017. A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* 49(12):1741–1746. doi:10.1038/ng.3974
- Juliana, P., R.P. Singh, P.K. Singh, J. Crossa, J. Huerta-Espino, C. Lan, et al. 2017a. Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theor. Appl. Genet.* 130. doi:10.1007/s00122-017-2897-1
- Juliana, P., R.P. Singh, P.K. Singh, J. Crossa, J.E. Rutkoski, J.A. Poland, et al. 2017b. Comparison of models and whole-genome profiling approaches for genomic-enabled prediction of *Septoria tritici* blotch, *Stagonospora nodorum* blotch, and tan spot resistance in wheat. *Plant Genome* 10(2). doi:10.3835/plantgenome2016.08.0082
- Kizilkaya, K., R.L. Fernando, and D.J. Garrick. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J. Anim. Sci.* 88:544–551. doi:10.2527/jas.2009-2064
- Kruijer, W., M.P. Boer, M. Malosetti, P.J. Flood, B. Engel, R. Kooke, et al. 2015. Marker-based estimation of heritability in immortal populations. *Genetics* 199:379–398. doi:10.1534/genetics.114.167916
- Kuchel, H., K. Williams, P. Langridge, H.A. Eagles, and S.P. Jefferies. 2007a. Genetic dissection of grain yield in bread wheat. II. QTL-by-environment interaction. *Theor. Appl. Genet.* 115:1015–1027. doi:10.1007/s00122-007-0628-8
- Kuchel, H., K.J.J. Williams, P. Langridge, H.A.A. Eagles, and S.P.P. Jefferies. 2007b. Genetic dissection of grain yield in bread wheat. I. QTL analysis. *Theor. Appl. Genet.* 115:1029–1041. doi:10.1007/s00122-007-0629-7
- Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756.
- Lenz, P.R.N., J. Beaulieu, S.D. Mansfield, S. Clément, M. Despons, and J. Bousquet. 2017. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genomics* 18:335. doi:10.1186/s12864-017-3715-5
- 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 (Bethesda)* 5:569–582. doi:10.1534/g3.114.016097
- Lorenz, A.J., and K.P. Smith. 2015. Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. *Crop Sci.* 55:2657–2667. doi:10.2135/cropsci2014.12.0827
- Luan, T., J.A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T.H.E. Meuwissen. 2009. The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genetics* 183:1119–1126. doi:10.1534/genetics.109.107391
- Luan, T., J.A. Woolliams, J. Ødegård, M. Dolezal, S.I. Roman-Ponce, A. Baginato, et al. 2012. The importance of identity-by-state information for the

- accuracy of genomic selection. *Genet. Sel. Evol.* 44:28. doi:10.1186/1297-9686-44-28
- Lund, M.S., G. Su, U.S. Nielsen, and G.P. Aamand. 2009. Relation between accuracies of genomic predictions and ancestral links to the training data. *Interbull Bull.* 40:1–5.
- Mackay, I., E. Ober, and J. Hickey. 2015. GpluSE: Beyond genomic selection. *Food Energy Secur.* 4:25–35. doi:10.1002/fes3.52
- 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.
- Money, D., K. Gardner, Z. Migicovsky, H. Schwaninger, G. Zhong, and S. Myles. 2015. LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3 (Bethesda)* 5:2383–2390. doi:10.1534/g3.115.021667
- 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 (Bethesda)* 8:131–147. doi:10.1534/g3.117.300309
- Montesinos-López, O.A., A. Montesinos-López, J. Crossa, F.H. Toledo, O. Pérez-Hernández, K.M. Eskridge, et al. 2016. A genomic Bayesian multi-trait and multi-environment model. *G3 (Bethesda)* 6:2725–2744. doi:10.1534/g3.116.032359
- Moser, G., M.S. Khatkar, B.J. Hayes, and H.W. Raadsma. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genet. Sel. Evol.* 42:37. doi:10.1186/1297-9686-42-37
- Muir, W.M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.* 124:342–355. doi:10.1111/j.1439-0388.2007.00700.x
- Park, T., and G. Casella. 2008. The Bayesian Lasso. *J. Am. Stat. Assoc.* 103:681–686. doi:10.1198/0162145080000000337
- Pérez-Rodríguez, P., J. Crossa, J. Rutkoski, J. Poland, R. Singh, A. Legarra, et al. 2017. Single-step genomic and pedigree genotype \times environment interaction models for predicting wheat lines in international environments. *Plant Genome* 10(2). doi:10.3835/plantgenome2016.09.0089
- Pérez-Rodríguez, P., D. Gianola, J.M. Gonzalez-Camacho, J. Crossa, Y. Manes, and S. Dreisigacker. 2012. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 (Bethesda)* 2:1595–1605. doi:10.1534/g3.112.003665
- Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495. doi:10.1534/genetics.114.164442
- Pérez, P., G. de los Campos, J. Crossa, and D. Gianola. 2010. Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome* 3:106–116. doi:10.3835/plantgenome2010.04.0005
- 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
- Pryce, J.E., and H.D. Daetwyler. 2012. Designing dairy cattle breeding schemes under genomic selection: A review of international research. *Anim. Prod. Sci.* 52:107–114. doi:10.1071/AN11098
- Pszczola, M., T. Strabel, H.A. Mulder, and M.P.L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J. Dairy Sci.* 95:389–400. doi:10.3168/jds.2011-4338
- Quarrie, S.A.A., A. Steed, C. Calestani, A. Semikhodskii, C. Lebreton, C. Chinoy, et al. 2005. A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring \times SQ1 and its use to compare QTLs for grain yield across a range of environments. *Theor. Appl. Genet.* 110:865–880. doi:10.1007/s00122-004-1902-7
- Reif, J.C., Y. Zhao, T. Würschum, M. Gowda, and V. Hahn. 2013. Genomic prediction of sunflower hybrid performance. *Plant Breed.* 132:107–114. doi:10.1111/pbr.12007
- Rincint, R., D. Laloe, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, et al. 2012. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192:715–728. doi:10.1534/genetics.112.141473
- Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J.-L. Jannink, and M. Sorrells. 2012. Evaluation of genomic prediction methods for *Fusarium* head blight resistance in wheat. *Plant Genome* 5:51. doi:10.3835/plantgenome2012.02.0001
- Rutkoski, J., R.P. Singh, S. Bhavani, J. Poland, J.L. Jannink, and M.E. Sorrells. 2015. Efficient use of historical data for genomic selection: A case study of stem rust resistance in wheat. *Plant Genome* 8:1–10. doi:10.3835/plantgenome2014.09.0046
- Sadras, V.O., M.P. Reynolds, A.J. de la Vega, P.R. Petrie, R. Robinson, A.J. De Vega, et al. 2009. Phenotypic plasticity of yield and phenology in wheat, sunflower and grapevine. *Field Crop. Res.* 110:242–250. doi:10.1016/j.fcr.2008.09.004
- Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In: V.Y. Shen, N. Saito, M.R. Lyu, M.E. Zurko, editors, *Proceedings of the 10th International Conference on the World Wide Web*, Hong Kong. 1–5 May 2001. ACM, New York. p. 285–295.
- Snape, J.W., J. Foulkes, J. Simmonds, M.L. Waite, L.J. Fish, Y. Wang, et al. 2007. Dissecting gene \times environmental effects on wheat yields via QTL and physiological analysis. *Euphytica* 154:401–408. doi:10.1007/s10681-006-9208-2
- Solberg, T.R., A.K. Sonesson, J.A. Woolliams, and T.H.E. Meuwissen. 2008. Genomic selection using different marker types and densities. *J. Anim. Sci.* 86:2447–2454. doi:10.2527/jas.2007-0010
- Taylor, J.F. 2014. Implementation and accuracy of genomic selection. *Aquaculture* 420–421:S8–S14. doi:10.1016/j.aquaculture.2013.02.017
- Thorwarth, P., J. Ahlemeyer, A.M. Bochard, K. Krumnacker, H. Blümel, E. Laubach, et al. 2017. Genomic prediction ability for yield-related traits in German winter barley elite material. *Theor. Appl. Genet.* 130:1669–1683. doi:10.1007/s00122-017-2917-1
- Toosi, A., R.L. Fernando, and J.C.M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. *J. Anim. Sci.* 88:32–46. doi:10.2527/jas.2009-1975
- van der Werf, J.H.J. 2009. Potential benefit of genomic selection in sheep. *Proc. Assoc. Adv. Anim. Breed. Genet.* 18:38–41.
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. doi:10.3168/jds.2007-0980
- VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, et al. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24. doi:10.3168/jds.2008-1514
- Vela-Avitúa, S., T.H. Meuwissen, T. Luan, and J. Ødegård. 2015. Accuracy of genomic selection for a sib-evaluated trait using identity-by-state and identity-by-descent relationships. *Genet. Sel. Evol.* 47:9. doi:10.1186/s12711-014-0084-2
- Watson, A., S. Ghosh, M.J. Williams, W.S. Cuddy, J. Simmonds, M.D. Rey, et al. 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. Plants* 4:23–29. doi:10.1038/s41477-017-0083-8
- Weigel, K.A., G. de los Campos, O. González-Recio, H. Naya, X.L. Wu, N. Long, G.J.M. Rosa, and D. Gianola. 2009. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. *J. Dairy Sci.* 92:5248–5257. doi:10.3168/jds.2009-2092
- Whittaker, J.C., R. Thompson, and M.C. Denham. 2000. Marker-assisted selection using ridge regression. *Genet. Res.* 75:249–252. doi:10.1017/S0016672399004462
- Wientjes, Y.C.J., M.P.L. Calus, M.E. Goddard, and B.J. Hayes. 2015. Impact of QTL properties on the accuracy of multi-breed genomic prediction. *Genet. Sel. Evol.* 47:1–16. doi:10.1186/s12711-015-0124-6
- Wientjes, Y.C.J., R.F. Veerkamp, and M.P.L. Calus. 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics* 193:621–631. doi:10.1534/genetics.112.146290