Theoretical and Applied Genetics (2021) 134:381–398
https://doi.org/10.1007/s00122-020-03703-z

ORIGINAL ARTICLE

Genomic prediction of agronomic traits in wheat using different models and cross-validation designs

Teketel A. Haile[1] · Sean Walkowiak[2] · Amidou N'Diaye[1] · John M. Clarke[1] · Pierre J. Hucl[1] · Richard D. Cuthbert[3] · Ron E. Knox[3] · Curtis J. Pozniak[1]

Received: 7 December 2019 / Accepted: 8 October 2020 / Published online: 1 November 2020

Abstract

Key message **Genomic predictions across environments and within populations resulted in moderate to high accuracies but across-population genomic prediction should not be considered in wheat for small population size.**

Abstract Genomic selection (GS) is a marker-based selection suggested to improve the genetic gain of quantitative traits in plant breeding programs. We evaluated the effects of training population (TP) composition, cross-validation design, and genetic relationship between the training and breeding populations on the accuracy of GS in spring wheat (*Triticum aestivum* L.). Two populations of 231 and 304 spring hexaploid wheat lines that were phenotyped for six agronomic traits and genotyped with the wheat 90 K array were used to assess the accuracy of seven GS models (RR-BLUP, G-BLUP, BayesB, BL, RKHS, GS + de novo GWAS, and reaction norm) using different cross-validation designs. BayesB outperformed the other models for within-population genomic predictions in the presence of few quantitative trait loci (QTL) with large effects. However, including fixed-effect marker covariates gave better performance for an across-population prediction when the same QTL underlie traits in both populations. The accuracy of prediction was highly variable based on the cross-validation design, which suggests the importance to use a design that resembles the variation within a breeding program. Moderate to high accuracies were obtained when predictions were made within populations. In contrast, across-population genomic prediction accuracies were very low, suggesting that the evaluated models are not suitable for prediction across independent populations. On the other hand, across-environment prediction and forward prediction designs using the reaction norm model resulted in moderate to high accuracies, suggesting that GS can be applied in wheat to predict the performance of newly developed lines and lines in incomplete field trials.

Introduction

Wheat is an important cereal crop that accounts for more than 20% of the total calories consumed by humans globally and is a staple food for about 35% of the world's population (Breiman and Graur 1995). Canada is the sixth largest wheat producing country in the world with a total production of 31.7 million tonnes in 2018 (FAO 2020). In Canada, most of the wheat is produced in the prairie provinces of Alberta, Saskatchewan, and Manitoba, and a small proportion is grown in British Columbia and eastern Canada (McCallum and DePauw 2008). Canada is the second largest exporter of wheat after the Russian Federation; 22.8 million tonnes of the wheat grain produced in Canada in 2018 was exported (FAO 2020). Canadian wheat is recognized globally for its high end-use quality.

Wheat breeding involves the creation of new genetic variation through controlled hybridization of two or more

Communicated by Hiroyoshi Iwata.

John M. Clarke: Deceased 01 February 2020.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00122-020-03703-z) contains supplementary material, which is available to authorized users.

✉ Curtis J. Pozniak
curtis.pozniak@usask.ca

[1] Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada

[2] Canadian Grain Commission, Grain Research Laboratory, Winnipeg, MB, Canada

[3] Semiarid Prairie Agricultural Research Centre, Agriculture and Agri-Food Canada, Swift Current, SK, Canada

parents followed by selfing and advancing generations by selecting offspring with desirable agronomic, disease resistance, and end-use quality traits. These advanced wheat lines then undergo repeated field testing, and if they meet accepted standards are released as new cultivars. This process normally takes 10 to 15 years and is resource intensive. Traditionally, selection of desirable plants within segregating early generation populations is generally based on visual assessments of agronomic traits and laboratory tests of end-use quality traits, which are laborious and expensive. For quantitative traits, selection based on the phenotype alone is subject to confounding effects from the environment, so entries are evaluated over multiple locations and years. This makes phenotypic selection time-consuming and expensive. Moreover, the short growing season of the Canadian prairies (a frost-free period of 90–120 days) presents a challenge for large scale field evaluation and selection of breeding material. The application of molecular markers greatly improves the precision and speed of the breeding cycle through marker-assisted selection (MAS) (Collard and Mackill 2008; Randhawa et al. 2013). Marker-assisted selection has been successful to identify and select QTL with moderate to large effects; however, MAS has limited application to improve complex traits controlled by many QTL with small effects (Heffner et al. 2009).

Advances in low cost, high-throughput genotyping technologies have resulted in the availability of abundant molecular markers in wheat spanning the whole genome (Cavanagh et al. 2013; Wang et al. 2014a; Winfield et al. 2016). In the context of genomic selection, these dense genome-wide markers could be used to predict genomic estimated breeding values (GEBVs) of individuals without phenotypic records (Meuwissen et al. 2001). Genomic selection involves estimating the effects of markers based on the genotypic and phenotypic data of a TP and predicting GEBVs of individuals in a breeding population (BP) by combining their marker genotypes with the marker effects estimated from the TP (Meuwissen 2009). Selection decisions will then be based on GEBVs estimated by whole-genome prediction models.

Several approaches and statistical models have been developed to implement GS (Bernardo 2014; Burgueño et al. 2012; de los Campos et al. 2009a, 2010; Gianola et al. 2006; Gianola and van Kaam 2008; Habier et al. 2011; Jia and Jannink 2012; Meuwissen et al. 2001; Park and Casella 2008; VanRaden 2008; Yang and Tempelman 2012). Most studies on GS evaluated the predictive performance of models through a cross-validation approach by systematically partitioning the same population into training and validation sets. Genomic estimated breeding values are predicted for individuals in the validation set, and the prediction accuracy of the model is usually determined by assessing the correlation between GEBVs and actual phenotypes of the individuals in the validation set. This technique is useful to compare the predictive performance of different statistical models and model parameters with respect to particular traits, but genomic predictions based on independent populations need to be evaluated for practical application of GS in crop breeding programs. An attractive application of GS for wheat breeding is the use of data that are routinely generated in a breeding program to train a model that can be used to estimate GEBVs of a BP that comprise progeny of multiple crosses. However, such an application can be constrained by low-moderate levels of genetic relatedness between the TP and BP. Genomic prediction models utilize genetic relationships among individuals as well as information from linkage disequilibrium (LD) between markers and QTL (Habier et al. 2007). Genetic relationships are influenced by generations of descent or population stratification (Asoro et al. 2011). The degree of genetic relationship between the training and breeding populations has been reported as an important factor that affects the accuracy of GS prediction (Clark et al. 2012; Habier et al. 2007, 2010; Hayes et al. 2009; Riedelsheimer et al. 2013). However, there are still unanswered questions related to the diversity of the TP to produce reliable predictions across populations, whether including few highly related lines in a TP of diverse germplasm could lead to acceptable prediction performance in breeding lines, and the performance of GS when some unrelated material is incorporated in a breeding program. Therefore, the objective of this study was to evaluate the effects of TP composition, cross-validation design, and genetic relationship between the training and validation populations on GS accuracy using Canadian spring hexaploid wheat germplasm.

Materials and methods

Plant material and phenotypic data

Training population (TP)

A TP of 231 spring hexaploid wheat genotypes was used to estimate the effects of markers using seven GS models (Online Resource 1). Phenotypic data for these lines were obtained from two different experiments. The first experiment was composed of 100 commercial wheat varieties, hereafter called diversity panel one (DP1). This population was composed of both contemporary and historic Canadian wheat varieties. These varieties were evaluated at Kernen Crop Research Farm, Saskatoon, SK, (lat 52° 08′, long 106° 32′) from 2011 to 2014 and at Swift Current, SK (lat 50° 16′, long 107° 44′) from 2012 to 2014. The field experiments were laid out in 200 plots, each plot having an area of 4.25 m^2 with five seeded rows at Kernen, and an area of 3.65 m^2 with four seeded rows at Swift Current. A seeding

rate of 300 and 275 seeds per m^2 was used at Kernen and Swift Current, respectively. The second experiment was composed of 200 spring hexaploid wheat varieties and advanced breeding lines, hereafter called diversity panel two (DP2), selected from breeding programs across western Canada. Each of these lines was evaluated in 2014 at the Seed Farm of the Crop Development Centre (CDC) in Saskatoon, SK (lat 52° 08′ long 106° 36′) on a 0.74 m^2 plot area, with two seeded rows and again in 2015 at Kernen and Rosthern, SK (lat 52° 41′ long 106° 19′) on a 4.25 m^2 plot area, with five seeded rows. The seeding rate of plots was 300 seeds per m^2. The two experiments were connected through 27 common lines. Both experiments were arranged in alpha-lattice experimental designs with two replications in each site-year. The field experiments were seeded in early to mid-May and harvested in mid to late September in each year. The 231 lines are subsets of the 273 lines from DP1 and DP2 that were genotyped and used as training set.

Breeding population (BP)

The BP was composed of 304 recombinant inbred lines that were developed from a three-way cross (CDC Plentiful// Pasteur/CDC Utmost) made at the CDC, University of Saskatchewan. Pasteur is a short-statured, later maturing but high yielding general-purpose wheat cultivar from Wiersum Plant Breeding in the Netherlands. CDC Utmost and CDC Plentiful are standard height, early to medium maturing and high yielding cultivars from the CDC, University of Saskatchewan. The first cross was made between Pasteur and CDC Utmost during early fall of 2011 and the second cross was made to CDC Plentiful during the winter of 2011/2012. The F_1 generation was grown in a controlled environment facility during summer 2012. Seed from the F_1 was bulked and the F_2 generation was grown in a greenhouse at the University of Saskatchewan. The F_3 generation was grown through single seed descent. The F_3 spikes were threshed individually and each were planted (F_4 generation) on single hill plots at the Seed Farm of the CDC during spring 2013. A single spike was harvested and threshed individually from each F_4 plant and the F_5 generation was grown under field conditions in a winter nursery during the winter of 2013/2014. The population was then advanced to the F_6, F_7, and F_8 generations for field trials and phenotyping. Five independent field trials were conducted across two research sites for the F_4:F_6, F_4:F_7, and F_4:F_8 generations. In the F_4:F_6 generation, 322 entries were randomly selected, and each was grown at Kernen on a 2.48 m^2 plot with four seeded rows during the spring/summer of 2014. The F_4:F_7 and F_4:F_8 generations were grown both at Kernen and Rosthern in 4.25 m^2 plots with five seeded rows during the spring/summer of 2015 and 2016. Seeding rates were 300 seeds per m^2. The field experiments were arranged in seven blocks, each containing 50 plots (46 entries and four check cultivars), for a total of 350 plots in each environment (site-year). The three parental lines and 'AC Barrie' were grown as replicated check cultivars. The field experiments were arranged in an augmented randomized complete block design, where the four check cultivars were randomly assigned to plots within each block and unreplicated entries were randomly arranged in the remaining plots (Federer 1961). Two of the three parental lines, CDC Plentiful and CDC Utmost, were also included in the TP to improve the genetic relationship between the TP and BP (Online Resource 1). Pasteur, which is distantly related to the Canadian wheat lines, was not included in the TP. The 304 lines are subsets of the 322 lines that were genotyped and used as BP.

Agronomic traits, including days to heading, plant height, days to maturity, grain yield, test weight, and kernel weight, were measured in the TP and BP. Days to heading was recorded for each plot as the number of days from seeding to when 50% of the spikes emerged out of the flag leaf sheath. Plant height was measured for each plot when the plants approached physiological maturity by taking the length of the main stem from the soil surface to the tip of the spike, while excluding the awns. Days to maturity was recorded as the number of days from seeding to when 50% of the spikes in a plot turned to a straw color. Plots were harvested using a small plot combine at physiological maturity. Grain yield was measured by taking the mass of grain harvested from each plot after the grains were air dried to constant moisture. Test weight was measured as the weight of dockage-free grain in grams required to fill a level 0.5 L container. Grain yield and test weight were reported in kg ha^{-1} and kg hL^{-1}, respectively. Kernel weight in grams was determined from a subsample of 200 kernels that were free from foreign material and broken kernels.

DNA extraction and genotyping

Genomic DNA was extracted from fresh leaves of one-week-old seedlings using a modified CTAB approach (CIMMYT 2005). For the BP, DNA was extracted from a single seedling per line in the F_7 generation. All lines in the TP and BP were genotyped using the wheat 90 K SNP array (Wang et al. 2014a). Genotype calling was performed using the GenomeStudio Polyploid Clustering Module v1.0 (Illumina, San Diego, CA). For the TP, 17,887 polymorphic SNPs with call frequency greater than 90% and minor allele frequency higher than 10% were obtained (Maccaferri et al. 2015). For the BP, 16,115 polymorphic SNPs with call frequency greater than 90% and minor allele frequency higher than 20% were obtained. A total of 9,187 SNPs were in common within the two populations. Missing marker genotypes were imputed with

the population mean for that marker using the function '*A.mat*' in R package *rrBLUP*, v4.4 (Endelman 2011).

Statistical analysis of phenotypic data

The phenotypic data were analyzed using analysis of variance with SAS Mixed models, v9.4 (SAS Institute 2015). For the TP, the DP1 and DP2 data sets were analyzed separately and combined across data sets and environments. For each data set, analyses were conducted in each environment (site-year) separately as well as combined across all environments. The linear mixed model used for analysis of variance in individual and combined environments, respectively, was in the form

$$y_{ijk} = \mu + G_i + R_j + B_{jk} + \varepsilon_{ijk} \quad (1)$$

and

$$y_{ijkl} = \mu + G_i + R_{jl} + B_{jkl} + E_l + (\mathrm{GE})_{il} + \varepsilon_{ijkl} \quad (2)$$

where y_{ijk} and y_{ijkl} denote the observed trait for i-th genotype in the k-th incomplete block within the j-th replicate of the l-th environment, μ is the overall mean, G_i is the effect of i-th genotype, R_{jl} the effect of the j-th replicate in the l-th environment, B_{jkl} is the effect of the k-th block in the j-th replicate of the l-th environment, E_l is effect of the l-th environment, $(GE)_{il}$ is the interaction effect between the i-th genotype and the l-th environment, ε_{ijk} and ε_{ijkl} denote experimental errors. Genotypes were considered as a fixed effect and replication, and block nested in replication was considered random for the separate analysis of data in each environment. For the combined analysis of data across data sets and environments, replication nested in environment, block nested in replication and environment, environment (site-years), and genotype-by-environment interaction were considered as random effects. The Kenward–Roger degrees of freedom approximation method was used to compute the degrees of freedom for means and to control the Type I error (Kenward and Roger 1997; Littell et al. 2006).

Phenotypic data for the BP were analyzed separately in each environment and then combined across environments. The linear mixed model used for analysis of variance in individual and combined environments, respectively, was in the form

$$y_{ij} = \mu + G_i + B_j + \varepsilon_{ij} \quad (3)$$

and

$$y_{ijl} = \mu + G_i + B_{jl} + E_l + (\mathrm{GE})_{il} + \varepsilon_{ijl} \quad (4)$$

where y_{ij} and y_{ijl} denote the observed trait for i-th genotype in the j-th block of the l-th environment, μ is the overall mean, G_i is the effect of i-th genotype (entry and check cultivar), B_{jl} is the effect of the j-th block in the l-th environment, E_l is effect of the l-th environment, $(\mathrm{GE})_{il}$ is the interaction effect between the i-th genotype and the l-th environment, ε_{ij} and ε_{ijl} denote experimental errors. Genotypes (i.e. entries plus check cultivars) were considered a fixed effect and block was considered as a random effect for the separate analysis of data in each environment. For the combined analysis of data across environments, genotypes were considered as a fixed effect and environment, block nested in environment and genotype-by-environment interactions were considered as random effects. To control for block-to-block heterogeneity, trait values of entries were adjusted relative to the four check cultivars repeated in each block using the LSMEANS procedure in SAS (Wolfinger et al. 1997). The phenotypic data analyses included the 322 entries in the F_6, F_7, and F_8 generations, and the four check cultivars but only 304 entries that had marker data were included in the BP. Broad-sense heritability (H^2) on a plot basis was estimated for all traits in the TP using the equation $\sigma_g^2/(\sigma_g^2 + \sigma_{\mathrm{ge}}^2/e + \sigma_\varepsilon^2/\mathrm{re})$, where σ_g^2 is the genetic variance, σ_{ge}^2 is the genotype-by-environment interactions variance, σ_ε^2 is the residual variance, e is the number of environments, and r is the number of replications per environment. Variance components were estimated in SAS using restricted maximum likelihood method described in Holland et al. (2003), with the effect of genotype, environment, genotype-by-environment interactions, and replication considered as random effect. In the BP, a two-step analysis of the phenotypic data was followed to estimate heritability (Poland et al. 2012). First, adjusted entry means were derived for each environment using Eq. (3), and these were used in the second stage to estimate H^2 on a plot basis using the equation: $\sigma_g^2/(\sigma_g^2 + \sigma_\varepsilon^2)$ where σ_g^2 and σ_ε^2 are the estimated genetic and residual variance components, respectively. The variance components were calculated using the '*lmer*' function in the R package *lme4*, v1.1–7 (Bates et al. 2015).

Genome-wide association mapping in the training population

Prior to the assessment of GS, genome-wide association mapping was performed in the TP to understand the genetic architecture of traits and to fit significant markers as fixed effects in GS. Marker-trait association analyses were performed based on LS-means across environments using a mixed linear model that combined both population structure information and pairwise relatedness (kinship) as covariates using the software TASSEL, v3.0 (Bradbury et al. 2007). Population structure information was accounted for using five marker-based principal components (Price et al. 2006). Principal components and the kinship-matrix were computed from the marker data using TASSEL. Mixed linear model analyses were performed using the default settings of TASSEL (optimum compression level and PD3 variance

component estimation). To determine if each model adequately controlled population structure and pairwise relatedness, quantile–quantile (Q–Q) plots were generated for all traits based on the observed *P* values for all SNPs and the expected distribution of *P* values under the null hypothesis of no marker-trait association. Marker probabilities were declared significant relative to a false discovery rate (FDR) of 0.2 to control for multiple testing. The FDR was calculated for all SNPs based on the 'BH' method using the '*p. adjust*' function in R (Benjamini and Hochberg 1995).

QTL analysis in the breeding population

QTL analysis was performed using a reduced subset of 1219 evenly spaced SNPs (mean genetic distance of 2.9 cM between adjacent SNPs) selected from the total markers based on genetic distances using the software MapThin, v1.11 (Howey and Cordell 2012). The chromosomal positions of the SNPs were determined based on the hexaploid wheat consensus genetic linkage map (Wang et al. 2014a). Analyses of the additive effects at individual QTL (ICIM-ADD) were performed with the inclusive composite interval mapping (ICIM) procedure using QTL IciMapping v4.1 (Meng et al. 2015). ICIM was performed on the LS-means of each trait for individual environment and averaged (combined) across all environments. A critical logarithm of odds (LOD) threshold was estimated for each trait based on 1000 permutations at a significance level of 0.05. Mapping parameters of 1 cM walking distance and deletion of missing phenotypes were applied.

Genomic prediction models

We evaluated ridge regression best linear unbiased prediction (RR-BLUP), genomic best linear unbiased prediction (G-BLUP), BayesB, Bayesian Lasso (BL), Bayesian reproducing kernel Hilbert spaces (RKHS) regression, GS + de novo GWAS, and a reaction norm model for genomic prediction of six agronomic traits in wheat. These models were chosen because they have different assumptions that are appropriate for a range of trait genetic architectures. In RR-BLUP, all markers are included in the model and their effects are shrunken toward zero uniformly, assuming that every marker has equal contribution to the genetic variance (Meuwissen et al. 2001; Whittaker et al. 2000). G-BLUP is equivalent to RR-BLUP, but marker information is included in the model using genomic relationships between individuals computed from SNPs (VanRaden 2008). The genomic relationship matrix, which estimates the realized proportion of the genome that is shared by two individuals, was computed according to VanRaden (2008) and used to estimate GEBVs. BayesB is a variable selection model that assigns non-uniform variances to markers with different effect sizes (Meuwissen et al. 2001). BayesB uses a finite mixture of priors with a point of mass at zero and a scaled-t distribution (Meuwissen et al. 2001). The BL uses double-exponential (conditional Laplace) prior density and combines variable selection and shrinkage of estimates (de los Campos et al. 2009b; Park and Casella 2008). GS + de novo GWAS fits significant markers identified from fold-specific genome-wide association studies as fixed effects in GS (Bernardo 2014; Spindel et al. 2016). GS + de novo GWAS is identical to RR-BLUP if no marker is fitted as fixed effects (Spindel et al. 2016). In the TP, genome-wide association mapping was conducted in each fold of a five-fold cross-validation using the phenotypic and genotypic data of lines in the training set as described above. The *P* values were corrected for multiple testing and markers on each chromosome were binned into 2-cM distance based on the average LD decay in this population. Then, the marker with the lowest *P* value was extracted from each bin. This step was performed to avoid fitting markers tagging the same QTL as fixed effects. Similarly, QTL analysis was performed in each fold of a five-fold cross-validation in the BP using methods described above. Up to three most significant markers (FDR = 0.2) identified from genome-wide association mapping in the TP and markers nearest to the QTL peaks in the BP were fitted as fixed effects in GS + de novo GWAS (Online Resource 2). When no marker passed the FDR threshold, a marker with the lowest *P* value was fitted as fixed effect. The RKHS regression is a semi-parametric approach which captures both the additive and non-additive genetic effects among loci by creating a kernel matrix that includes interactions among markers (Gianola et al. 2006; Gianola and van Kaam 2008). We used the Gaussian kernel implemented in Pérez and de los Campos (2014), evaluated as the average squared-Euclidean distance between genotypes:

$$K_{(x_i,x_j)} = \exp\left\{-h \times \frac{\sum_{k=1}^{p}\left(x_{ik}-x_{jk}\right)^2}{p}\right\} \quad (5)$$

where x_i and x_j are the pairs of vectors of genotypes, p refers to the total number of markers, and h is a bandwidth parameter that controls how fast the (co)variance function drops as the distance between pairs of vector genotypes increases (de los Campos et al. 2009a; Pérez and de los Campos 2014). Kernel methods also allow the use of multiple kernels by evaluating the Gaussian kernel over a range of h values, which was termed kernel averaging (de los Campos et al. 2010). We followed recommendations by Pérez and de los Campos (2014) and used kernel averaging by estimating h as $h = 1/M \times \{1/5, 1, 5\}$, where M is $1/p$ of the median squared Euclidean distance between all lines calculated using off-diagonals only. The reaction norm model is equivalent to the standard G-BLUP model with the addition of a random environmental effect (Jarquín et al. 2014). In this model,

phenotypes (y_{ijk}) were described as the sum of an overall mean (μ) plus a random deviation due to the environment (E_i), which is a combination of site-years, plus marker covariates of the form: $g_j = \sum_{k=1}^{p} x_{jk} b_k$, where g_j represents an approximation of the true genetic value of the j-th line, x_{jk} is the genotype of the j-th line at the k-th marker, and b_k is the effect of the k-th marker, plus a residual term (ε_{ijk}). The regression equation is indicated using the following formula:

$$y_{ijk} = \mu + E_i + g_j + \varepsilon_{ijk} \quad (6)$$

with $E_i \sim N(0,\ \sigma_E^2)$, $g \sim N\left(0,\ G\sigma_g^2\right)$ and $\varepsilon_{ijk} \sim N(0,\ \sigma_\epsilon^2)$, where G is marker-derived genomic relationship matrix.

All statistical models were fitted in R (R Core Team 2016). The G-BLUP, BayesB, BL, RKHS, and reaction norm model were fitted using the *Bayesian generalized linear regression* (*BGLR*) package, v1.0.4 (Pérez and de los Campos 2014). RR-BLUP and GS + de novo GWAS were fitted using the '*mixed.solve*' and '*kinship.BLUP*' functions in the *rrBLUP* package, v4.4, respectively (Endelman 2011). The default settings of *BGLR* (five degrees of freedom and the scale parameter based on sample variance of the phenotypes) were used (Pérez and de los Campos 2014). Inferences for all Bayesian models were based on 50,000 iterations obtained after discarding 5000 samples as burn-in.

Cross-validation schemes

We tested different cross-validation schemes that simulate prediction scenarios that breeders may face when implementing GS. Genomic predictions were made for three different scenarios: within-population, across-population, and across-environment. In all three scenarios, predictions were made for days to heading, plant height, days to maturity, grain yield, test weight, and kernel weight. The first prediction scenario involved within-population genomic predictions using a five-fold cross-validation design. In each population, the lines were randomly divided into five mutually exclusive groups of approximately equal sizes. In each fold, the four groups were used as training set and the remaining one was used for validation. This was repeated five times until each group was used as a validation set. Assignment of genotypes to groups was repeated ten times resulting in 50 different cross-validation runs for each model. The overall means of the TP and BP across environments were used to make predictions. Within-population genomic predictions were made using RR-BLUP, G-BLUP, BayesB, BL, RKHS, and GS + de novo GWAS models. For a reliable comparison of these models, the same cross-validation folds were used in each model-trait combination. The total number of 17,887 and 16,115 polymorphic SNPs was used to make predictions in the TP and BP, respectively. To test significant differences in prediction accuracies among the evaluated models, the cross-validation results were analyzed using a one-way analysis of variance with the PROC MIXED procedure in SAS (SAS Institute 2015), using fold as a blocking factor. The LSMEANS procedure was used to determine differences between models.

The second prediction scenario involved across-population genomic predictions where GEBVs for the BP were predicted based on marker effects estimated from the TP (Online Resource 3). The effect of genetic relationship between the TP and the BP on model prediction performance was evaluated by excluding the two parents from the TP, including the two parents in the TP, and including the two parents along with 50 or 100 randomly selected lines from the BP in the TP. Moreover, genomic prediction accuracy was investigated by clustering the BP into two groups based on their genomic relationships to the parents. A kinship matrix was calculated using the EMMA algorithm within GAPIT (Lipka et al. 2012), to show the clusters among the BP and familial relatedness based on the marker genotypes. The first group was composed of 121 lines that were clustered with Pasteur or neither of the parents while the second group was composed of 183 lines that were clustered with CDC Utmost and CDC Plentiful. The second group was considered as closely related to the TP because CDC Utmost and CDC Plentiful were included in the TP, while the first group was considered as distantly related to the TP because Pasteur was not included in the TP. Genomic predictions were made for each group separately using the TP that included the parents. A total of 9187 polymorphic SNPs that were common between the TP and BP were used to make across-population genomic predictions using G-BLUP, BayesB, and GS + de novo GWAS models.

The third scenario was an across-environment prediction using three designs (Online Resource 4). The first design (CV1) involved predicting the performance of lines that have never been tested in any of the environments (newly developed lines), while the second design (CV2) involved predicting the performance of lines that were evaluated in some environments but not in others (i.e. in an incomplete field trials) (Burgueño et al. 2012). The third design involved forward prediction of future phenotypes across years. For this analysis, we performed different across-year predictions. The BP data from 2014 (F_4:F_6 generation) were used as TP to predict the phenotypes of the BP in 2015 (F_4:F_7 generation) and 2016 (F_4:F_8 generation), the BP data from 2015 were used as TP to predict phenotypes of BP in 2016, and the BP data from 2014 and 2015 combined were used as TP to predict phenotypes of the BP in 2016. The DP2 and BP were evaluated in similar environments in 2014 and 2015 and were combined and used as a TP to make similar across-year predictions in the BP. The DP2 and BP data from 2014 were used to predict the phenotypes of the BP in 2015 and 2016, the DP2 and BP from 2015 were used to

predict the phenotypes of the BP in 2016, and the DP2 and BP from 2014 and 2015 combined were used to predict the phenotypes of the BP in 2016. The 2014 DP2 data are LS-Means of two replications from one location, 2014 BP data are adjusted means from one location, while the DP2 and BP data in 2015 and 2016 are LS-Means from two locations (Saskatoon and Rosthern). Similar across-year genomic predictions were also made for each location separately. Phenotypic prediction accuracy (r_P) was the correlation of observed phenotypes of the BP in the environments used for model training and validation. We used 16,115 SNPs for predictions involving BP only and 9187 SNPs for predictions involving combined BP and DP2 data sets in the TP. Across-environment genomic predictions were made using the reaction norm model that incorporated the main effects of markers and environments as implemented in Jarquín et al. (2014). Model prediction accuracy was assessed based on the Pearson correlation between the predicted values and observed phenotypes of individuals in the validation set. In CV1 and CV2 designs, correlations between observed and predicted values were performed within the same environment. For the forward prediction, accuracy was computed as the correlation between the predicted values using the earlier year as training set with the observed phenotypes in subsequent years.

Results

Distributions of trait phenotypes

Phenotypic distributions of all traits in the TP and BP across years followed an approximately normal distribution (Online Resource 5 and 6). Broad-sense heritability was the highest for kernel weight (0.50 and 0.78) and days to heading (0.48 and 0.75) in the TP and BP, respectively. Moderate to high estimates of heritability were obtained for grain yield (0.55), plant height (0.55), days to maturity (0.64), and test weight (0.71) in the BP, but estimates were low to moderate for grain yield (0.28), days to maturity (0.32), plant height (0.33), and test weight (0.40) in the TP.

Marker-trait associations in the training population

Marker-trait associations were assessed for important agronomic traits in the TP. For all traits, the Q–Q plots were close to the diagonal line except for deviations toward the upper-right end of the diagonal indicating that models including population structure or kinship adequately controlled false positives (Online Resource 7). Twelve SNPs were significantly associated with plant height (Table 1). Nine of these SNPs were located on chromosome 4B (39.9–72.5 cM), two SNPs on chromosome 2A (109.5–110.1 cM), and one SNP on chromosome 5B (115.7 cM). These SNPs explained 6 to 15% of the phenotypic variance. The SNPs associated with plant height on 4B spanned a large genomic window which may suggest the presence of multiple QTL, but a single gene cannot be ruled out. Three SNPs (55.5–57.5 cM) explained the highest proportion of the phenotypic variance (Table 1). The two SNPs that were associated with days to heading were located on chromosomes 2D (19.3 cM) and 5B (110.56 cM) (Table 1). Each of these SNPs explained 8% of the phenotypic variance. No SNP passed the FDR threshold for grain yield, days to maturity, test weight, and kernel weight.

QTL identified in the breeding population

The number of markers used for QTL analysis varied from 23 on 4D to 96 on 5B, with an average of 58 markers per

Table 1 Markers significantly associated with plant height and days to heading in the training population

Trait	Marker	Chr	Position (cM)	*P* value	R^2 (%)
Plant height	*BS00022896_51*	2A	109.5	4.05E−04	6
	BS00012320_51	2A	110.1	1.46E−05	9
	RAC875_c12959_869	4B	39.9	7.42E−04	7
	Tdurum_contig64772_417	4B	50.8	1.02E−04	7
	BobWhite_rep_c49034_132	4B	55.5	2.13E−05	10
	Tdurum_contig33737_157	4B	56	2.70E−07	12
	IAAV971	4B	57.5	2.11E−08	15
	Excalibur_c56787_95	4B	58.1	3.84E−04	6
	Kukri_c11415_1074	4B	68.5	2.87E−04	6
	Kukri_c17224_278	4B	71.3	1.03E−04	8
	wsnp_Ra_c22026_31453420	4B	72.5	6.25E−05	7
	BS00022673_51	5B	115.7	2.62E−04	6
Days to heading	*wsnp_CAP12_c812_428290*	2D	19	3.81E−05	8
	BS00065128_51	5B	110.6	3.36E−05	8

chromosome. The total map length across the 21 chromosomes spanned 3,526 cM. The range of genetic distance between adjacent SNPs varied from 0.04 to 52.3 cM with a mean of 2.9 cM (Online Resource 8). A total of 16 QTL were detected for six agronomic traits in the BP (Table 2). Four QTL were identified for days to heading, three QTL each for test weight, days to maturity, and kernel weight, two QTL for plant height, and one QTL for grain yield (Table 2). Most of these QTL had only minor effects, with days to heading and maturity having the only stable large effect QTL. Chromosome 2D (18.4–23.4 cM) harbored the strongest QTL for days to heading (*QHd.usw-2D*), maturity (*QMat.usw-2D*), and plant height (*QHt.usw-2D*), which explained 19.2, 14.7, and 11.2% of the phenotypic variances, respectively (Table 2). *QHd.usw-2D* and *QMat.usw-2D* were detected in each of the five environments and accounted for 9.5 to 22.8 and 4.5 to 13.6% of the phenotypic variance, respectively (Online Resource 9). *QHt.usw-2D* was detected in all environments, except Rosthern 2016, and explained 4.9 to 11.2% of the phenotypic variance in each environment (Online Resource 9). Chromosome 7D (96.5–100.5 cM) also harbored QTL for days to heading (*QHd.usw-7D*) and maturity (*QMat.usw-7D.1*) that explained 11.8 and 10.7% of the phenotypic variance, respectively. *QHd.usw-7D* and *QMat.usw-7D.1* were also detected in four and three individual environments and explained 8.8 to 15.5 and 6.2 to 11.7% of the phenotypic variance, respectively (Online Resource 9). Two QTL were detected for days to maturity on chromosome 7D separated by 36 cM between the QTL peaks. Two stable QTL were identified each for days to heading and maturity on chromosomes 2D and 7D that together explained 31 and 25.4% of the phenotypic variance, respectively.

Three QTL were identified for test weight on 2B (*QTwt.usw-2B*), 4B (*QTwt.usw-4B*), and 7B (*QTwt.usw-7B*) (Table 2). Each of these QTL explained 3.3 to 7.9% of the phenotypic variance. *Twt.usw-2B* was detected in two environments and accounted for 4.2 and 5.7% of the phenotypic variance (Online Resource 9). *QTwt.usw-4B* was detected in four out of the five environments and explained 4.2 to 10.4% of the phenotypic variance, while *QTwt.usw-7B* was detected in three environments and accounted for 4 to 6.5% of the phenotypic variance in each environment (Online Resource 9). The three QTL for kernel weight were mapped on chromosome 2A (*QTkw.usw-2A*), 4A (*QTkw.usw-4A*), and 6A (*QTkw.usw-6A*) (Table 2). Each QTL explained 5.9 to 7.8% of the phenotypic variance. *QTkw.usw-2A* was detected at the same confidence interval in two of the individual environments and explained 6.4 and 9.1% of the phenotypic variance in each environment (Online Resource 9). *QTkw.usw-4A* was detected in three environments and accounted for 4.3 to 9.3% of the phenotypic variance. *QTkw.usw-6A* was detected in all environments except Kernen 2015 and explained from 8.5 to 11.4% of the phenotypic variance. The QTL for grain yield (*QYld.usw-2A*) was co-localized with *QTkw.usw-2A* on the distal end of 2AS. *QYld.usw-2A* explained 10.4% of the phenotypic variance in the combined data set and 18 and 5.3% of the phenotypic variance in two environments (Table 2; Online Resource 9).

Within-population genomic prediction accuracies

Prediction accuracies varied among the evaluated traits both in the TP and BP. The average prediction accuracies, based on 50 cross-validation runs in the TP, ranged from 0.56 to

Table 2 QTL detected using inclusive composite interval mapping in the breeding population

Trait	QTL name	Chr	Interval (cM)	Position (cM)	LOD	R^2 (%)	Add[a]
Days to heading	*QHd.usw-2D*	2D	19.4–22.4	19.9	32.9	19.2	1.1
	QHd.usw-4A	4A	105.1–107.1	106.6	9.2	4.0	−0.6
	QHd.usw-4B	4B	88.5–90.5	90	4.2	1.8	−0.3
	QHd.usw-7D	7D	96.5–97.5	97	23.1	11.8	1.0
Days to maturity	*QMat.usw-2D*	2D	18.4–21.4	18.9	17.6	14.7	0.9
	QMat.usw-7D.1	7D	96.5–100.5	98	14.7	10.7	0.9
	QMat.usw-7D.2	7D	133.5–134.5	134	4.6	3.4	−0.4
Test weight	*QTwt.usw-2B*	2B	119.1–122.1	120.6	4.8	3.3	0.3
	QTwt.usw-4B	4B	55.5–59.5	59	13.3	7.9	0.4
	QTwt.usw-7B	7B	69.5–71.5	70	8.5	5.1	−0.4
Kernel weight	*QTkw.usw-2A*	2A	0–2.5	1	6.8	7.3	−0.7
	QTkw.usw-4A	4A	99.1–102.1	100.6	7.8	5.9	−0.8
	QTkw.usw-6A	6A	81.4–83.4	81.9	3.8	7.8	0.8
Plant height	*QHt.usw-2D*	2D	18.4–23.4	19.9	11.5	11.2	2.0
	QHt.usw-6D	6D	80.5–83.5	82	6.6	6.1	1.7
Grain yield	*QYld.usw-2A*	2A	0–0.5	0	9.5	10.4	168.7

[a]Additive effect of the QTL

0.78 across different model-trait combinations (Table 3). In the BP, prediction accuracies based on 50 cross-validation runs ranged from 0.44 to 0.76 for all model-trait combinations (Table 3). There were significant differences ($P < 0.05$) among the evaluated models for all traits except grain yield in both populations and test weight in the TP (Table 3). BayesB and GS + de novo GWAS showed significantly higher accuracy than the other models for plant height in the TP. We applied the RKHS model to determine if accounting for non-additive genetic effects improves genomic prediction accuracy. The accuracy of RKHS was 5 to 13% higher than the models that are based on the additive effects (RR-BLUP, G-BLUP, BayesB, BL, and GS + de novo GWAS) for days to heading and maturity, but no improvement in accuracy was observed for the other traits. In the BP, GS + de novo GWAS showed higher accuracy than RR-BLUP, G-BLUP, BL, and RKHS for days to heading and maturity, but BayesB was the best performing model (Table 3). On the other hand, GS + de novo GWAS had lower accuracy than the other models for kernel weight in the TP and BP.

Across-population genomic prediction accuracies

Across-population genomic predictions were made for the BP using the marker effects estimated from the TP. When none of the parents of the BP were included in the TP, accuracy was very low (ranged from −0.16 to 0.33) (Table 4). We included two of the parents of the BP (CDC Utmost and CDC Plentiful) in the TP to improve the genetic relationship between the TP and BP, but it did not improve prediction accuracies for all traits and models (Table 4). To further investigate the importance of having related individuals in the TP, we included 50 or 100 randomly selected lines from the BP in the TP. Prediction accuracies increased when 50 lines from the BP were included in the TP (Table 4). The largest increase was observed for plant height, where accuracy increased from approximately 0 to 0.3 for all models. Increasing the number of BP lines in the TP to 100 further increased prediction accuracies for all traits except for plant height in G-BLUP (Table 4). GS + de novo GWAS gave substantially higher accuracy for days to heading for which a marker tagging the same QTL in the TP and BP was fitted as a fixed effect (Online Resource 2). Similarly, GS + de novo GWAS was the best performing model for all traits when the number of BP lines in the TP increased from 50 to 100.

To further investigate the effect of genetic relatedness on genomic prediction accuracy, the BP was divided into two groups based on the genomic relationship of the lines to the parents (Fig. 1). The first group (green and gold) was composed of 121 lines that were distantly related to the TP, while the second group (red and blue) was composed of 183 lines that were closely related to the TP (Fig. 1).

Genomic predictions were made for the two groups separately using the marker effects estimated from the TP that included two of the parents. Accuracy was slightly higher when predictions were made for lines closely related to the TP for all traits, except grain yield (Table 5). Overall, the prediction accuracy was very low even though the parents of the BP were included in the TP to enhance genomic relationships. This indicates that including few closely related parents in the TP of diverse lines may not ensure accurate prediction of GEBVs for progenies.

Table 3 Mean and standard deviations (from 50 cross-validation runs) of within population prediction accuracies in the training and breeding populations

Model	Days to heading	Plant height	Days to maturity	Grain yield	Test weight	Kernel weight
Training population						
RR-BLUP	0.62 (0.01)b	0.57 (0.01)b	0.59 (0.01)bc	0.56 (0.01)a	0.64 (0.01)a	0.78 (0.01)a
G-BLUP	0.63 (0.01)b	0.57 (0.01)b	0.59 (0.01)b	0.57 (0.01)a	0.65 (0.01)a	0.78 (0.01)a
BayesB	0.63 (0.01)b	0.62 (0.01)a	0.59 (0.01)bc	0.57 (0.01)a	0.64 (0.01)a	0.78 (0.01)a
BL	0.63 (0.01)b	0.57 (0.01)b	0.59 (0.01)b	0.56 (0.01)a	0.65 (0.01)a	0.78 (0.01)a
RKHS	0.66 (0.01)a	0.58 (0.01)b	0.63 (0.01)a	0.57 (0.01)a	0.66 (0.01)a	0.77 (0.01)ab
GS + de novo GWAS	0.62 (0.01)b	0.63 (0.01)a	0.56 (0.01)c	0.59 (0.01)a	0.63 (0.01)a	0.75 (0.01)b
Breeding population						
RR-BLUP	0.55 (0.01)c	0.44 (0.01)b	0.53 (0.01)c	0.46 (0.01)a	0.58 (0.01)b	0.66 (0.01)a
G-BLUP	0.54 (0.01)c	0.45 (0.01)b	0.53 (0.01)c	0.46 (0.01)a	0.59 (0.01)b	0.66 (0.01)a
BayesB	0.76 (0.01)a	0.48 (0.01)a	0.73 (0.01)a	0.46 (0.01)a	0.63 (0.01)a	0.66 (0.01)a
BL	0.55 (0.01)c	0.44 (0.01)b	0.53 (0.01)c	0.46 (0.01)a	0.59 (0.01)b	0.66 (0.01)a
RKHS	0.49 (0.01)d	0.46 (0.01)ab	0.49 (0.01)d	0.49 (0.01)a	0.58 (0.01)b	0.68 (0.01)a
GS + de novo GWAS	0.68 (0.01)b	0.48 (0.01)a	0.60 (0.01)b	0.47 (0.01)a	0.58 (0.01)b	0.62 (0.01)b

Within each population and trait, accuracies followed by the same letter were not significantly different at $\alpha = 0.05$

Table 4 Prediction accuracy based on the Pearson correlations between GEBVs calculated with different supplements to the training population and phenotypes of lines in the validation set

Training population[a]	Model	Days to heading	Plant height	Days to maturity	Grain yield	Test weight	Kernel weight
TP without parents	G-BLUP	0.153	0.052	0.146	0.052	0.194	0.243
	BayesB	0.216	−0.090	0.266	0.079	0.182	0.276
	GS + de novo GWAS	0.330	−0.157	0.263	0.071	0.174	0.248
TP with two parents	G-BLUP	0.156	0.072	0.162	0.069	0.208	0.209
	BayesB	0.194	−0.070	0.208	0.078	0.192	0.228
	GS + de novo GWAS	0.327	−0.171	0.261	0.079	0.174	0.220
TP with two parents plus 50 lines from the BP	G-BLUP	0.199	0.311	0.296	0.162	0.355	0.478
	BayesB	0.242	0.308	0.340	0.181	0.359	0.501
	GS + de novo GWAS	0.389	0.330	0.396	0.158	0.339	0.501
TP with two parents plus 100 lines from the BP	G-BLUP	0.287	0.309	0.351	0.324	0.467	0.590
	BayesB	0.594	0.323	0.519	0.332	0.480	0.602
	GS + de novo GWAS	0.620	0.363	0.590	0.374	0.501	0.609

Predictions were made for six traits using G-BLUP, BayesB, and GS + de novo GWAS models

[a]TP, training population; BP, breeding population

Across-environment genomic prediction accuracies

Across-environment prediction accuracies varied based on traits and environments. In CV1, which involved predicting the performance of lines that were not tested in any of the environments, prediction accuracy ranged from 0.28 to 0.56 for all traits (Table 6). The CV2 strategy mimics prediction in incomplete field trials where some lines were evaluated in some environments but not in others. The accuracies obtained in CV2 were much higher (ranged from 0.40 to 0.89) than those observed in CV1 because in CV2 prediction of performance for a line can benefit from records of the same line collected in other environments (Table 6).

Moderate to high accuracies were obtained for all traits when forward predictions were made across years (Table 7). When the BP in 2014 (F_4:F_6 generation) was used to predict phenotypes of the BP in 2015 (F_4:F_7 generation) and 2016 (F_4:F_8 generation), accuracies for all traits ranged from 0.56 to 0.76 and 0.65 to 0.84, respectively (Table 7). When the BP in 2015 was used to predict phenotypes of the BP in 2016, accuracies ranged from 0.62 to 0.85. Combining data from 2014 and 2015 resulted in higher accuracy (ranged from 0.68 to 0.90) for all traits except plant height. Prediction accuracies were slightly lower or similar for all traits when the DP2 and BP data sets were combined to make across-year predictions for the BP (Table 7).

Phenotypic prediction accuracies are correlations between the phenotypes of the BP from the respective environments used to train and validate GS models. These accuracies ranged from 0.76 to 0.88 (days to heading), 0.66 to 0.72 (plant height), 0.70 to 0.82 (days to maturity), 0.62 to 0.76 (grain yield), 0.73 to 0.81 (test weight), and 0.81 to 0.91 (kernel weight) (Table 7). The deviation of these numbers from one suggests the presence of genotype-by-environment interaction. The ratio of genomic to phenotypic prediction accuracies across years (r_{GS}/r_P) ranged from 0.86 (test weight) to 1.03 (days to maturity) with a mean ratio of 0.96, indicating that accuracies obtained from genomic and phenotypic predictions across years are highly comparable. We also evaluated across-year prediction accuracies for each location separately (Online Resource 10). Overall, across-year genomic and phenotypic prediction accuracies were higher when data from multiple locations or years were combined as opposed to using data from a single location or year. The ratio of genomic to phenotypic prediction accuracy (r_{GS}/r_P) at Kernen ranged from 0.86 (test weight) to 1.10 (grain yield) with a mean ratio of 0.99, while the ratio ranged from 0.90 (grain yield) to 1.03 (days to heading) with a mean ratio of 0.96 at Rosthern (Online Resource 10).

Discussion

Marker-trait associations were performed in the TP and BP to understand the genetic architecture of traits and to fit significant markers as fixed effects in GS. In the TP, twelve SNPs on chromosome 4B, 2A, and 5B were significantly associated with plant height and explained 6 to 15% of the phenotypic variance (Table 1). Nine of these SNPs were mapped to chromosome 4B. These results are consistent with previous studies that reported major plant height QTL on 4B in wheat (Gao et al. 2015; Guo et al. 2017; N'Diaye et al. 2018). N'Diaye et al. (2018) also reported a haplotype locus controlling plant height in wheat on chromosome 2A that explained 15% of the phenotypic variance. Similarly, McCartney et al. (2005) reported a prominent QTL for plant height

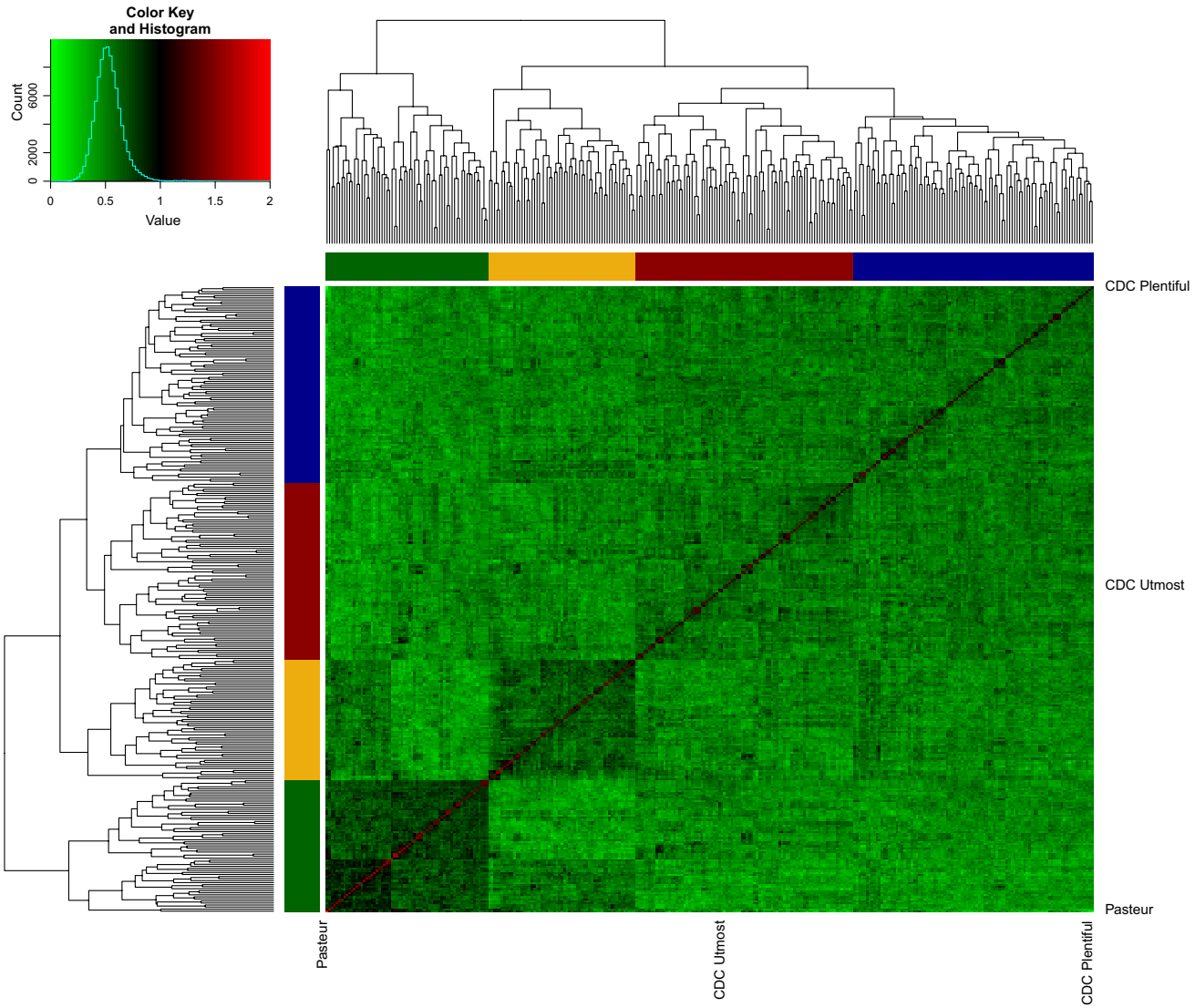

Fig. 1 Heat map and dendrogram of a genomic relationship matrix estimated from 16 K SNPs among the 304 wheat lines and three parents. Color codes show groups of lines based on their genomic relationships. Both rows and columns represent the lines

Table 5 Prediction accuracy based on the Pearson correlations between GEBVs and observed phenotypes

Model	Days to heading	Plant height	Days to maturity	Grain yield	Test weight	Kernel weight
Lines distantly related to the training population						
G-BLUP	0.142	−0.071	0.083	0.127	0.079	0.164
BayesB	0.197	−0.124	0.136	0.141	0.056	0.188
GS + de novo GWAS	0.277	−0.179	0.225	0.182	0.080	0.174
Lines closely related to the training population						
G-BLUP	0.164	0.143	0.198	−0.002	0.302	0.293
BayesB	0.223	−0.026	0.289	−0.003	0.260	0.327
GS + de novo GWAS	0.368	−0.108	0.299	−0.004	0.257	0.271

Table 6 Average and standard deviation of across-environment prediction accuracies (from five-fold design) based on correlations between observed and predicted phenotypes of six traits using the reaction norm model for cross-validation CV1 (newly developed lines) and CV2 (prediction in incomplete field trials)

Environment[a]	Days to heading	Plant height	Days to maturity	Grain yield	Test weight	Kernel weight
CV1						
Kernen 2014	0.46 (0.08)	0.45 (0.03)	0.52 (0.08)	0.42 (0.07)	0.43 (0.22)	0.56 (0.11)
Kernen 2015	0.46 (0.08)	0.37 (0.19)	0.47 (0.08)	0.32 (0.09)	0.40 (0.15)	0.49 (0.10)
Kernen 2016	0.50 (0.06)	0.46 (0.11)	0.53 (0.04)	0.45 (0.11)	0.54 (0.13)	0.55 (0.06)
Rosthern 2015	0.42 (0.07)	0.40 (0.05)	0.46 (0.10)	0.32 (0.08)	0.41 (0.21)	0.50 (0.10)
Rosthern 2016	0.49 (0.04)	0.28 (0.05)	0.48 (0.08)	0.41 (0.10)	0.46 (0.16)	0.52 (0.11)
CV2						
Kernen 2014	0.79 (0.05)	0.77 (0.04)	0.75 (0.06)	0.68 (0.10)	0.43 (0.22)	0.89 (0.02)
Kernen 2015	0.84 (0.04)	0.62 (0.10)	0.79 (0.05)	0.59 (0.07)	0.40 (0.15)	0.82 (0.04)
Kernen 2016	0.89 (0.03)	0.68 (0.07)	0.76 (0.10)	0.64 (0.06)	0.54 (0.13)	0.84 (0.04)
Rosthern 2015	0.82 (0.04)	0.72 (0.10)	0.78 (0.03)	0.60 (0.07)	0.41 (0.21)	0.81 (0.08)
Rosthern 2016	0.83 (0.04)	0.49 (0.14)	0.73 (0.09)	0.70 (0.07)	0.46 (0.16)	0.87 (0.02)

[a]Site-years where the breeding population were grown

Table 7 Across-year genomic and phenotypic prediction accuracies based on combined data from two sites

Training set[a]	Validation set[a]	Days to heading	Plant height	Days to maturity	Grain yield	Test weight	Kernel weight
Genomic prediction accuracy (r_{GS})[b]							
BP 2014	BP 2015	0.751	0.700	0.699	0.562	0.689	0.761
BP 2014	BP 2016	0.772	0.647	0.713	0.665	0.705	0.843
BP 2015	BP 2016	0.853	0.618	0.795	0.644	0.756	0.835
BP 2014 and 2015	BP 2016	0.870	0.683	0.800	0.715	0.807	0.899
BP + DP2 2014	BP 2015	0.687	0.684	0.661	0.553	–	0.758
BP + DP2 2014	BP 2016	0.699	0.638	0.669	0.651	–	0.839
BP + DP2 2015	BP 2016	0.807	0.583	0.759	0.632	0.735	0.819
BP + DP2 2014 and 2015	BP 2016	0.836	0.662	0.781	0.690	–	0.893
Phenotypic prediction accuracy (r_P)[c]							
BP 2014	BP 2015	0.764	0.723	0.709	0.620	0.804	0.813
BP 2014	BP 2016	0.775	0.673	0.695	0.698	0.738	0.888
BP 2015	BP 2016	0.859	0.661	0.799	0.685	0.786	0.860
BP 2014 and 2015	BP 2016	0.881	0.717	0.818	0.767	0.807	0.912
r_{GS}/r_P[d]							
BP 2014	BP 2015	0.983	0.968	0.986	0.906	0.857	0.936
BP 2014	BP 2016	0.996	0.961	1.026	0.953	0.955	0.949
BP 2015	BP 2016	0.993	0.935	0.995	0.940	0.962	0.971
BP 2014 and 2015	BP 2016	0.988	0.953	0.978	0.932	1.00	0.986

–, Data not available

[a]*BP* breeding population, *DP2* diversity panel two

[b]Pearson correlations between observed and predicted phenotypes

[c]Pearson correlations between phenotypes of the breeding population across years

[d]r_{GS}/r_P is the ratio of genomic prediction accuracy to phenotypic prediction accuracy

on the long arm of chromosome 5B in spring bread wheat. Two SNPs on chromosomes 2D and 5B were associated with days to heading in the TP (Table 1). Each of these SNPs explained 8% of the phenotypic variance. Similarly, QTL for days to heading (*QHd.usw-2D*) and maturity (*QMat.usw-2D*) were identified in the BP on chromosome 2D which explained 19.2 and 14.7% of the phenotypic variance, respectively. Chromosome 7D also harbored QTL for days

to heading (*QHd.usw-7D*) and maturity (*QMat.usw-7D.1*) that explained 11.8 and 10.7% of the phenotypic variance in the BP, respectively. Previous studies reported QTL for days to heading and maturity on the short arm of chromosome 2D which harbors a major photoperiod response locus in wheat (Beales et al. 2007; Carter et al. 2011). Cuthbert et al. (2008) also reported QTL for days to heading and maturity on chromosome 7D, less than 5 cM distance from the days to heading (*QHd.usw-7D*) and maturity (*QMat.usw-7D.1*) QTL identified in this study, suggesting that these QTL may be the same. Similarly, McCartney et al. (2005) reported a QTL for maturity that explained 25.7% of the phenotypic variance on the short arm of chromosome 7D based on a Canadian spring wheat mapping population. Overall, stable QTL with relatively large effects were identified only for days to heading and maturity in the BP but relatively small genetic effects for most of the individual QTL identified for the other traits.

Within-population prediction

Moderate to high accuracies were obtained when predictions were made based on five-fold cross-validation within each population. Average prediction accuracies ranged from 0.56 to 0.78 and 0.44 to 0.76 across different model-trait combinations in the TP and BP, respectively (Table 3). There were significant differences in prediction accuracy among the models (Table 3). In the TP, GS + de novo GWAS resulted in 9 to 11% higher accuracy than RR-BLUP, G-BLUP, BL, and RKHS, but its accuracy was similar to BayesB for plant height. Significant marker-trait associations were detected only for days to heading and plant height, but no SNP passed the FDR threshold for all other traits in the TP (Table 1). Three highly significant markers were fitted as fixed effects in 80% of the cross-validation runs for plant height but for the other traits only one marker was fitted as fixed effect in most cross-validation runs (Online Resource 2). In the BP, the accuracy of BayesB was significantly higher, 7 to 55% higher than the other models, for days to heading, days to maturity, and test weight (Table 3). GS + de novo GWAS also showed 11 to 41% higher accuracy than RR-BLUP, G-BLUP, BL, and RKHS, but its accuracy was significantly lower compared to BayesB. There were stable QTL with relatively large effects underlying days to heading and maturity, but effects of the QTL identified for the other traits were relatively small (Table 2; Online Resource 2). Spindel et al. (2016) reported up to 30% gains in prediction accuracy using GS + de novo GWAS over the standard RR-BLUP model for flowering time in rice, which had a large GWAS peak. Rice and Lipka (2019) also evaluated the performance of GS + de novo GWAS using 216 simulated traits with a range of genetic architectures in maize and sorghum and reported improved accuracies relative to RR-BLUP for 60 traits. Conversely, no improvement or a decrease in accuracy was observed for the remaining traits.

The observed differences in the accuracy of GS models are related to how they account for the variances of marker effects. Infinitesimal models such as RR-BLUP assign uniform variance to all markers, while Bayesian models allow non-uniform marker variances (Asoro et al. 2011; Habier et al. 2007; Meuwissen et al. 2001). Treating markers as random variables and shrinking their effects uniformly as performed in RR-BLUP do not explicitly model the effects of major QTL versus unknown small effects QTL (Bernardo 2014). Bernardo (2014) suggested that when a few major genes each accounting for more than 10% of the genetic variance are present for a quantitative trait, these major genes should be fitted as fixed effects instead of random effects in GS models. In this study, BayesB outperformed GS + de novo GWAS in the presence of QTL with relatively large effects. Most GS models have similar performance for polygenic traits, but variable selection models such as BayesB did have advantages for traits controlled by fewer large effects QTL (Clark et al. 2011; Daetwyler et al. 2013). Meuwissen et al. (2001) showed that BayesB was able to predict the position of large QTL that accounted for more than 10% of the total genetic variance, but it failed to identify smaller QTL and treated their contribution in the same way as in BLUP. Studies that incorporated fixed-effect marker covariates in infinitesimal models reported improved prediction accuracies relative to models that treat all QTL or markers equally (Arruda et al. 2016; Bernardo 2014; Rice and Lipka 2019; Spindel et al. 2016). However, none of these studies included BayesB, which is recommended for prediction of traits controlled by few QTL with large effects. Spindel et al. (2016) reported that adding fixed-effect marker covariates identified from GWAS never decreased accuracy compared to RR-BLUP. But in this study, GS + de novo GWAS gave lower accuracy for kernel weight in the TP and BP (Table 3). In the TP, no marker passed the FDR threshold for kernel weight and only one marker with the lowest *P*-value was fitted as fixed effect in all cross-validation runs (Online Resource 2). These markers may not necessarily be associated with the QTL underlying kernel weight. Arruda et al. (2016) reported reduced accuracy when randomly selected markers were treated as fixed effects in RR-BLUP. Rice and Lipka (2019) also showed that incorporation of at least one fixed-effect covariate in RR-BLUP can increase the bias of predicted GEBVs. Similarly, most of the QTL for kernel weight explained less than 10% of the variance in the BP. Bernardo (2014) showed that having a single gene treated as fixed effect in RR-BLUP was never disadvantageous, except when the variance explained by the major gene was less than 10%. Overall, this study showed that variable selection models such as BayesB have advantages over GS + de novo GWAS when few QTL with large effect underlie traits,

but BayesB also gave accuracies comparable to the other models in the absence of detectable large effect QTL. Thus, the additional step of identifying marker-trait associations is unnecessary in GS and genomic predictions can be streamlined using BayesB that has acceptable performance across a range of trait genetic architectures.

Across-population prediction

We obtained moderate to high prediction accuracies when performing within-population predictions that partitioned individuals of the same population into training and validation sets. However, this approach may have limited application in a breeding program because inferences are made on known populations that have already been phenotyped. It would be more useful to plant breeders to use data that is routinely generated in a breeding program to make predictions for independent populations; therefore, we assessed the ability of GS models to make across-population genomic predictions. Across-population prediction accuracies were very low (ranged from −0.16 to 0.33) compared to accuracies based on five-fold cross-validation in each population (Tables 3 and 4). Using two doubled haploid wheat populations, Thavamanikumar et al. (2015) also showed that prediction accuracies based on tenfold cross-validation in each population were generally higher than those obtained when marker effects from one population were used to predict traits in the other population. Daetwyler et al. (2014) reported a higher prediction accuracy when each line in the validation set had at least one close relationship to the training lines. In this study, there was no improvement in accuracy when parents of the BP were included in the TP. However, including 50 or 100 lines from the BP in the TP resulted up to three-fold increase in the accuracy of prediction for the BP. The low across-population prediction accuracy when none of the lines from the BP were included in the TP could be because of the distant relationship between the TP and BP. When the populations are distantly related, marker effect estimates can be inconsistent because of differences in alleles, allele frequencies, and linkage phases between the TP and BP (Bassi et al. 2016). When the data from both populations are combined in the TP, markers that are in LD with the QTL that are shared between populations can be used to make more accurate predictions.

The accuracy of predicting GEBVs depends on both genetic relationships among individuals as well as LD between markers and QTL (Habier et al. 2007). In the absence of close relationships between the training and test populations, prediction accuracy is driven by distant relationships that will be captured when there is strong LD (Clark et al. 2012). Moreover, estimation of marker effects across populations requires not only strong LD but the same linkage phase between the marker and the QTL in each population (Goddard and Hayes 2007). For days to heading and maturity, GS + de novo GWAS gave higher accuracy when markers tagging the same QTL in the TP and BP were treated as having fixed effects (Table 4; Online Resource 2). However, for the other traits where different QTL were involved, the accuracy of GS + de novo GWAS was either similar or lower compared to BayesB (Table 4). Overall, our results indicate that across-population predictions have low accuracy. These findings agree with what has been observed in other studies (Charmet et al. 2014; Crossa et al. 2014; Riedelsheimer et al. 2013; Wang et al. 2014b; Windhausen et al. 2012). Windhausen et al. (2012) reported accuracies close to zero for prediction of testcross performance in F_2-derived lines using marker effects estimated from a maize diversity panel that included crosses of lines. Based on five biparental doubled-haploid maize populations developed from crosses involving four parents, Riedelsheimer et al. (2013) reported mean accuracies of zero or negative values when prediction was made for individuals in biparental families using a model trained based on data from unrelated biparental families. However, using half-sib and full-sib families in the training set improved prediction accuracy to 0.25 and 0.59, respectively. Using two bi-parental hybrid rye populations that share one common parent, Wang et al. (2014b) also reported substantially low accuracy when one population was used as TP to estimate GEBVs of another population, but accuracy increased when both populations contributed to training and validation sets. This indicates that the inability for GS models to make accurate across-population prediction is a common trend and will limit the potential application of GS in wheat breeding.

Across-environment prediction

Genomic selection can be applied in plant breeding to predict the performance of lines in different environments. The CV1 approach involves prediction of performance for lines that were not evaluated in any of the tested environments. In this case, prediction of performance is entirely based on the phenotypes of other lines and accuracy depends on the genetic relationships between the lines used in the TP and BP. In CV2, prediction of performance for unobserved lines in a target environment is based on the phenotypes of these lines and others from different environments plus phenotypes of the other lines from the target environment. The two methods tested for cross-validation differed in their prediction accuracies. The accuracy of prediction in CV1 ranged from 0.28 to 0.56, while the CV2 prediction accuracies ranged from 0.40 to 0.89 for all traits (Table 6). CV2 consistently gave higher prediction accuracy compared to CV1 because prediction of performance for a line can benefit from records of the same line from other environments. This has been observed in several studies that used similar

cross-validation designs (Burgueño et al. 2012; Crossa et al. 2015; Jarquín et al. 2014, 2017; Lopez-Cruz et al. 2015; Pérez-Rodríguez et al. 2015). This suggests that including information of the same line from correlated environments is important to improve prediction accuracy in the target environment. The CV1 and CV2 designs can be used to make prediction in early generation of a breeding material. Different subsets of families could be tested in different sets of environments by implementing incomplete designs and predictions can be made using the CV2 approach, subsequently increasing the number of lines evaluated. Similarly, the performance of newly developed lines can be predicted for a target environment based on the records of other related lines from that environment using the CV1 approach. These strategies would help to reduce resource requirements for field testing of early generation breeding material and allow evaluation of large number of crosses and lines from each cross.

Moderate to high accuracies (ranged from 0.56 to 0.90) were obtained for all traits when forward predictions were made across years (Table 7). Combining data from 2014 and 2015 in the TP resulted in the highest accuracy for all traits except plant height. Our results agree with Wang et al. (2014b), who reported that limiting the number of locations or years in field testing for the TP reduced the accuracy of GS predictions. On the other hand, combining the DP2 and BP data sets to increase the TP size resulted in slightly lower or similar across-year prediction accuracies for all traits in the BP (Table 6). This suggests that combining different populations to increase the size of the TP may not be advantageous. Combining different populations may reduce LD since the phase of LD varies across populations (Goddard 2012). Moreover, combining multiple related or unrelated populations into one TP may reduce prediction accuracy because of intense population structures in the TP (Riedelsheimer et al. 2013). Lack of improvement in accuracy when combining different populations in the TP could be because only LD that is persistent across those populations is utilized in the model (Calus 2010). Charmet et al. (2014) also reported that prediction accuracies did not improve when unrelated populations from different breeding programs were merged to increase TP size. Overall, GS can make accurate predictions across-years and including more sample years improves accuracies.

Across-year genomic and phenotypic prediction accuracies are highly comparable. The ratio of genomic to phenotypic prediction accuracies across years (r_{GS}/r_P) ranged from 0.86 (test weight) to 1.03 (days to maturity) with a mean ratio of 0.96, indicating that accuracies obtained from genomic and phenotypic predictions across years are highly comparable. Heffner et al. (2011) also reported similar ratios ranging from 0.84 (days to heading) to 1.09 (grain yield) with a mean ratio of 0.95 when across-year predictions were made in winter wheat. Similarly, Zhong et al. (2009) reported comparable accuracies between genomic and phenotypic selection methods. These results indicate that data from earlier years can be used to make accurate predictions in subsequent years and that GS can be applied to predict the performance of a breeding material in future years.

Conclusions

This study compared different prediction scenarios using seven statistical methods and three cross-validation schemes. Prediction accuracy was highly variable based on the cross-validation design, which suggests the importance to use a design that resembles the variation within a breeding program. Across-population prediction accuracies were very low even when few closely related lines were included in the TP. On the other hand, within-population and across-environment predictions resulted in moderate to high accuracies. Combining data across locations or years in the TP resulted in higher prediction accuracy indicating the importance of evaluating the training set in more than one environment to achieve higher prediction accuracy. However, combining data from two unrelated populations to increase the TP size did not improve accuracy. Comparison of different statistical methods based on within-population prediction indicated that BayesB is superior to RR-BLUP, G-BLUP, BL, RKHS, and GS + de novo GWAS when there are QTL with relatively large effects. But including fixed-effect marker covariates in RR-BLUP was advantageous for an across-population prediction when the same QTL underlie traits in both populations. The reaction norm model that included a random environmental effect also gave accurate prediction of phenotypes across environments. Based on the findings from this study, BayesB and the reaction norm model are being deployed in our wheat breeding program to make GS predictions within populations and across locations and years.

Acknowledgements This study was conducted as part of the Canadian Triticum Applied Genomics ($CTAG^2$) project and funded by Genome Canada (Grant number: 8310), Saskatchewan Ministry of Agriculture, Western Grains Research Foundation, Saskatchewan Wheat Development Commission, and Government of Saskatchewan. We would like to acknowledge the technical assistance from the Durum Wheat Breeding and Genetics field and molecular laboratory staff at the University of Saskatchewan.

Author contributions statement TAH designed the experiment, generated phenotypic and marker data, performed all analyses, and wrote the manuscript. SW, AN, and JMC edited the manuscript. PJH designed the experiment, developed and maintained early generation of the breeding population, and edited the manuscript. RDC and REK collected phenotypic data, and edited the manuscript. CJP acquired funding, designed the experiment, supervised the project, collected phenotypic data and edited the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Arruda MP, Lipka AE, Brown PJ, Krill AM, Thurber C, Brown-Guedira G, Dong Y, Foresman BJ, Kolb FL (2016) Comparing genomic selection and marker-assisted selection for *Fusarium* head blight resistance in wheat (*Triticum aestivum* L.). Mol Breed 36:84. https://doi.org/10.1007/s11032-016-0508-5

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L (2011) Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome 4:132–144. https://doi.org/10.3835/plantgenome2011.02.0007

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). Plant Sci 242:23–36. https://doi.org/10.1016/j.plantsci.2015.08.021

Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/jss.v067.i01

Beales J, Turner A, Griffiths S, Snape JW, Laurie DA (2007) A *pseudo-response regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). Theor Appl Genet 115:721–733. https://doi.org/10.1007/s00122-007-0603-4

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Bernardo R (2014) Genomewide selection when major genes are known. Crop Sci 54:68–75. https://doi.org/10.2135/cropsci2013.05.0315

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Breiman A, Graur D (1995) Wheat evolution. Isr J Plant Sci 43:85–98. https://doi.org/10.1080/07929978.1995.10676595

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci 52:707–719. https://doi.org/10.2135/cropsci2011.06.0299

Calus MPL (2010) Genomic breeding value prediction: methods and procedures. Animal 4:157–164. https://doi.org/10.1017/S1751731109991352

Carter AH, Garland-Campbell K, Kidwell KK (2011) Genetic mapping of quantitative trait loci associated with important agronomic traits in the spring wheat (*Triticum aestivum* L.) cross 'Louise' × 'Penawawa'. Crop Sci 51:84–95. https://doi.org/10.2135/cropsci2010.03.0185

Cavanagh CR, Chao S, Wang S et al (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci USA 110:8057–8062. https://doi.org/10.1073/pnas.1217133110

Charmet G, Storlie E, Oury FX et al (2014) Genome-wide prediction of three important traits in bread wheat. Mol Breeding 34:1843–1852. https://doi.org/10.1007/s11032-014-0143-y

CIMMYT (2005) Laboratory protocols: CIMMYT Applied Molecular Genetics Laboratory, 3rd edn. Mexico, D.F

Clark SA, Hickey JM, Daetwyler HD, van der Werf JHJ (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. https://doi.org/10.1186/1297-9686-44-4

Clark SA, Hickey JM, van der Werf JHJ (2011) Different models of genetic variation and their effect on genomic evaluation. Genet Sel Evol 43:18. https://doi.org/10.1186/1297-9686-43-18

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos T Roy Soc B 363:557–572. https://doi.org/10.1098/rstb.2007.2170

Crossa J, de los Campos G, Maccaferri M, Tuberosa R, Burgueño J, Pérez-Rodríguez P (2015) Extending the marker × environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. Crop Sci 56:1–17. https://doi.org/10.2135/cropsci2015.04.0260

Crossa J, Pérez P, Hickey J et al (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112:48–60. https://doi.org/10.1038/hdy.2013.16

Cuthbert JL, Somers DJ, Brûlé-Babel AL, Brown PD, Crow GH (2008) Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). Theor Appl Genet 117:595–608. https://doi.org/10.1007/s00122-008-0804-5

Daetwyler HD, Bansal UK, Bariana HS, Hayden MJ, Hayes BJ (2014) Genomic prediction for rust resistance in diverse wheat landraces. Theor Appl Genet 127:1795–1803. https://doi.org/10.1007/s00122-014-2341-8

Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193:347–365. https://doi.org/10.1534/genetics.112.147983

de los Campos G, Gianola D, Rosa GJM (2009a) Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. J Anim Sci 87:1883–1887. https://doi.org/10.2527/jas.2008-1259

de los Campos G, Naya H, Gianola D et al (2009b) Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182:375–385. https://doi.org/10.1534/genetics.109.101501

de los Campos G, Gianola D, Rosa GJM, Weigel KA, Crossa J (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet Res 92:295–308. https://doi.org/10.1017/S0016672310000285

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package *rrBLUP*. Plant Genome 4:250–255. https://doi.org/10.3835/plantgenome2011.08.0024

FAO (2020) Crop statistics. https://www.fao.org/faostat/en/#data/QC. Accessed 05 Aug 2020

Federer WT (1961) Augmented designs with one-way elimination of heterogeneity. Biometrics 17:447–473

Gao F, Wen W, Liu J et al (2015) Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese Spring. Front Plant Sci 6:1099. https://doi.org/10.3389/fpls.2015.01099

Gianola D, Fernando RL, Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173:1761–1776. https://doi.org/10.1534/genetics.105.049510

Gianola D, van Kaam JBCHM (2008) Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics 178:2289–2303. https://doi.org/10.1534/genetics.107.084285

Goddard ME (2012) Uses of genomics in livestock agriculture. Anim Prod Sci 52:73–77. https://doi.org/10.1071/AN11180

Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet 124:323–330. https://doi.org/10.1111/j.1439-0388.2007.00702.x

Guo Y, Du Z, Chen J, Zhang Z (2017) QTL mapping of wheat plant architectural characteristics and their genetic relationship with

seven QTLs conferring resistance to sheath blight. PLoS ONE 12:e0174939. https://doi.org/10.1371/journal.pone.0174939
Habier D, Fernando RL, Dekkers JC (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397. https://doi.org/10.1534/genetics.107.081190
Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. BMC Bioinform 12:186. https://doi.org/10.1186/1471-2105-12-186
Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol 42:5
Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME (2009) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol 41:1. https://doi.org/10.1186/1297-9686-41-51
Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65–75. https://doi.org/10.3835/plantgenome2010.12.0029
Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1–12. https://doi.org/10.1016/j.cj.2018.03.001
Holland JB, Nyquist WE, Cervantes-Martínez CT (2003) Estimating and interpreting heritability for plant breeding: an update. In: Janick J (ed) Plant breeding reviews. John Wiley & Sons, New Jersey, pp 9–112
Howey R, Cordell HJ (2012) MapThin. https://www.staff.ncl.ac.uk/richard.howey/mapthin/. Accessed 17 Mar 2017
Jarquín D, Crossa J, Lacaze X et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607. https://doi.org/10.1007/s00122-013-2243-1
Jarquín D, Lemes da Silva C, Gaynor RC et al (2017) Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. Plant Genome. https://doi.org/10.3835/plantgenome2016.12.0130
Jia Y, Jannink J-L (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522. https://doi.org/10.1534/genetics.112.144246
Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997. https://doi.org/10.2307/2533558
Lipka AE, Tian F, Wang Q et al (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399. https://doi.org/10.1093/bioinformatics/bts444
Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) SAS® system for mixed models, 2nd edn. SAS Institute Inc., Cary
Lopez-Cruz M, Crossa J, Bonnett D et al (2015) Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 (Bethesda) 5:569–582. https://doi.org/10.1534/g3.114.016097
Maccaferri M, Zhang J, Bulli P et al (2015) A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). G3 (Bethesda) 5:449–465. https://doi.org/10.1534/g3.114.014563
McCallum BD, DePauw RM (2008) A review of wheat cultivars grown in the Canadian prairies. Can J Plant Sci 88:649–677. https://doi.org/10.4141/CJPS07159
McCartney CA, Somers DJ, Humphreys DG et al (2005) Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452×'AC domain'. Genome 48:870–883. https://doi.org/10.1139/g05-055
Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J 3:269–283. https://doi.org/10.1016/j.cj.2015.01.001
Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829
Meuwissen THE (2009) Accuracy of breeding values of "unrelated" individuals predicted by dense SNP genotyping. Genet Sel Evol 41:35. https://doi.org/10.1186/1297-9686-41-35
N'Diaye A, Haile JK, Nilsen K et al (2018) Haplotype loci under selection in Canadian durum wheat germplasm over 60 years of breeding: association with grain yield, quality traits, protein loss and plant height. Front Plant Sci 9:1589. https://doi.org/10.3389/fpls.2018.01589
Park T, Casella G (2008) The Bayesian Lasso. J Am Stat Assoc 103:681–686. https://doi.org/10.1198/016214508000000337
Pérez-Rodríguez P, Crossa J, Bondalapati K, De Meyer G, Pita F, Gdl C (2015) A pedigree-based reaction norm model for prediction of cotton yield in multienvironment trials. Crop Sci 55:1143–1151. https://doi.org/10.2135/cropsci2014.08.0577
Pérez P, de los Campos G (2014) Genome-wide regression and prediction with the *BGLR* statistical package. Genetics 198:483–495. https://doi.org/10.1534/genetics.114.164442
Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909. https://doi.org/10.1038/ng1847
Poland J, Endelman J, Dawson J et al (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103–113. https://doi.org/10.3835/plantgenome2012.06.0006
R Core Team (2016) R: A language and environment for statistical computing. Vienna, Austria. https://www.R-project.org/
Randhawa HS, Asif M, Pozniak C et al (2013) Application of molecular markers to wheat breeding in Canada. Plant Breed 132:458–471. https://doi.org/10.1111/pbr.12057
Rice B, Lipka AE (2019) Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. Plant Genome 12:180052. https://doi.org/10.3835/plantgenome2018.07.0052
Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194:493–503. https://doi.org/10.1534/genetics.113.150227
SAS Institute (2015) The SAS system for windows, 9.4 edn., Cary, North Carolina
Spindel JE, Begum H, Akdemir D, Collard B, Redoña E, Jannink JL, McCouch S (2016) Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity 116:395–408. https://doi.org/10.1038/hdy.2015.113
Thavamanikumar S, Dolferus R, Thumma BR (2015) Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. G3 (Bethesda) 5:1991–1998. https://doi.org/10.1534/g3.115.019745
VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423. https://doi.org/10.3168/jds.2007-0980
Wang S, Wong D, Forrest K et al (2014a) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796. https://doi.org/10.1111/pbi.12183
Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014b) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field

evaluation locations and test years. BMC Genom 15:556–556. https://doi.org/10.1186/1471-2164-15-556

Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. Genet Res (Camb) 75:249–252. https://doi.org/10.1017/S0016672399004462ER

Windhausen VS, Atlin GN, Hickey JM et al (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 (Bethesda) 2:1427–1436. https://doi.org/10.1534/g3.112.003699

Winfield MO, Allen AM, Burridge AJ et al (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnol J 14:1195–1206. https://doi.org/10.1111/pbi.12485

Wolfinger R, Federer WT, Cordero-Brana O (1997) Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED. Agron J 89:856–859. https://doi.org/10.2134/agronj1997.00021962008900060002x

Yang W, Tempelman RJ (2012) A Bayesian antedependence model for whole genome prediction. Genetics 190:1491–1501. https://doi.org/10.1534/genetics.111.131540

Zhong S, Dekkers JC, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355–364. https://doi.org/10.1534/genetics.108.098277