ORIGINAL RESEARCH

The Plant Genome

Genomic selection for lentil breeding: Empirical evidence

Teketel A. Haile | **Taryn Heidecker** | **Derek Wright** | **Sandesh Neupane** |
Larissa Ramsay | **Albert Vandenberg** | **Kirstin E. Bett**

Dep. of Plant Sciences, Univ. of Saskatchewan, 51 Campus Drive, Saskatoon, SK, S7N 5A8, Canada

Correspondence
Kirstin E. Bett, Dep. of Plant Sciences, Univ. of Saskatchewan, 51 Campus Drive, Saskatoon, SK, S7N 5A8, Canada.
Email: k.bett@usask.ca

Abstract

Genomic selection (GS) is a marker-based selection initially suggested for livestock breeding and is being encouraged for crop breeding. Several statistical models are used to implement GS; however, none have been tested for use in lentil (*Lens culinaris* Medik.) breeding. This study was conducted to compare the accuracy of different GS models and prediction scenarios based on empirical data and to make recommendations for designing genomic selection strategies for lentil breeding. We evaluated nine single-trait (ST) models, two multiple-trait (MT) models, and a model that incorporates genotype × environment interaction (GEI) using populations from a lentil diversity panel and two recombinant inbred lines (RILs). The lines in all populations were phenotyped for five phenological traits and genotyped using a custom exome capture assay. Within-population, across-population, and across-environment genomic predictions were made. Prediction accuracy varied among the evaluated models, populations, prediction scenarios, and traits. Single-trait models showed similar accuracy in the absence of large effect quantitative trait loci (QTL) but BayesB outperformed all models when there were QTL with relatively large effects. Models that accounted for GEI and MT-GS models increased prediction accuracy for a low heritability trait by up to 66 and 14%, respectively. Moderate to high accuracies were obtained for within-population (range of .36–.85) and across-environment (range of .19–.89) predictions but across-population prediction accuracy was very low. Results suggest that GS can be implemented in lentil breeding to make predictions within populations and across environments, but across-population prediction should not be considered when the population size is small.

1 | INTRODUCTION

Lentil is a diploid ($2n = 14$), self-pollinating crop with a haploid genome size of approximately 4 Gb (Arumuganathan & Earle, 1991). It is an important pulse crop, which provides a dietary source of protein, carbohydrate, micronutrients, vitamins, and fiber. Lentil is used increasingly as an important source of protein for people whose diet is solely plant based. It has ecological advantages as a rotational crop in

cereal-based cropping systems by enabling better management of pests, herbicide residue, and soil nitrogen. In 2017, lentil was grown on more than 6.6 Mha across 52 countries, accounting for an annual production of 7.6 Mt (FAO, 2019). Until recently, lentil has received little attention in terms of genetic research. Advances in high-throughput genotyping technologies and a relative decline in sequencing costs have resulted in the availability of abundant single nucleotide polymorphisms (SNPs) covering the gene space of lentil through in-depth sequencing of the entire set of exons in the genome (Ogutcen, Ramsay, von Wettberg, & Bett, 2018). This development opened new opportunities to evaluate genomic tools such as GS for lentil improvement.

Genomic selection is a marker-based selection that uses a training population (TP) that is phenotyped and genotyped to train a statistical model. The model is then used to predict genomic estimated breeding values (GEBVs) of individuals in a breeding population that has been genotyped but not phenotyped (Meuwissen, Hayes, & Goddard, 2001). The GEBVs are then used to select superior lines rather than the actual phenotypes. Model prediction accuracy is tested using independent lines in a validation population for which phenotypic and genotypic data are available. This is commonly achieved by partitioning the same population into training and validation sets in which GEBVs are predicted for individuals in the validation set. The correlation between predicted values and the observed phenotypes is considered the prediction accuracy.

Unlike the standard marker-assisted selection, which uses a small number of markers associated with major QTL, GS uses a large set of genome-spanning markers to potentially capture all the QTL underlying a trait (Meuwissen et al., 2001). Fitting all markers simultaneously avoids multiple testing and the need to identify marker–trait associations based on an arbitrarily chosen significance threshold. Genomic selection was initially suggested for livestock breeding (Meuwissen et al., 2001) and was later evaluated and implemented in crop breeding, especially for wheat (*Triticum* sp.) and maize (*Zea mays* L.) (Bernardo, 2009; Bernardo & Yu, 2007; Beyene et al., 2015; Crossa et al., 2016; Crossa et al., 2014). The goal of GS is to enhance the genetic gain of quantitative traits by accelerating the breeding cycle and increasing selection intensity (Battenfield et al., 2016; Heffner, Lorenz, Jannink, & Sorrells, 2010). Thus far, there is no information on the utility of GS for lentil breeding.

The accuracy of GS is affected by several factors, including TP size, marker density, trait heritability, the genetic relationship between the training and breeding populations, linkage disequilibrium (LD), population structure, GEI, and the statistical model used (Burgueño, de los Campos, Weigel, & Crossa, 2012; Calus, 2010; Combs & Bernardo, 2013; de los Campos, Veturi, Vazquez, Lehermeier, & Pérez-Rodríguez, 2015; Heffner, Jannink, & Sorrells, 2011; Jarquín et al., 2014;

Core Ideas

- BayesB showed the highest prediction accuracy for traits controlled by few QTL with relatively large effects.
- Incorporating genotype × environment interactions improved prediction accuracy by up to 66%.
- Genomic selection can be implemented in lentil to make within-population and across-environment predictions.
- Across-population genomic prediction should not be considered in lentil when population size is small.

VanRaden et al., 2009; Wientjes, Veerkamp, & Calus, 2013). A large TP size, higher heritability, and higher marker density usually improve genomic prediction accuracy (Asoro, Newell, Beavis, Scott, & Jannink, 2011; Heffner et al., 2011; Liu et al., 2018; Meuwissen et al., 2001). However, uniform marker distribution across the genome is essential to tag important QTL and maintain GS accuracy (Bassi, Bentley, Charmet, Ortiz, & Crossa, 2016). The main assumption of GS is that each of the QTL is in LD with at least one nearby marker, and potentially all the genetic variance can be explained by the markers (Calus, 2010; Goddard & Hayes, 2007). The effect of population structure on prediction accuracy varies based on prediction strategies and the genetic architectures of traits and populations (Guo et al., 2014). It has also been shown that incorporating GEI into GS models can improve the accuracy of prediction (Crossa et al., 2015; de los Campos et al., 2015; Jarquín et al., 2017; Lopez-Cruz et al., 2015). Therefore, it is important to evaluate models that account for GEI in each target environment.

Genomic selection models fit a large number of markers on a small number of phenotypic observations, which may lead to overfitting when markers are treated as fixed effects (de los Campos, Hickey, Pong-Wong, Daetwyler, & Calus, 2013). Standard GS models treat markers as a random effect and reduce the dimension of the marker data by either selecting markers, shrinking marker effect estimates, or a combination of both (de los Campos et al., 2013). Several methods, including ridge regression, Bayesian regression, kernel-based approaches, and machine learning algorithms, have been developed that differ mainly in their assumptions about the variances of marker effects (Breiman, 2001; Gianola & van Kaam, 2008; Gianola, Fernando, & Stella, 2006; Habier, Fernando, Kizilkaya, & Garrick, 2011; Meuwissen et al., 2001; Park & Casella, 2008; Pérez, de los Campos, Crossa, & Gianola, 2010). Ridge regression assumes that all loci explain equal amounts of the genetic variance, while Bayesian models allow the variance to vary across loci

(Meuwissen et al., 2001). The assumption of common variance for all loci in a ridge regression is unrealistic and may underestimate the effects of known major QTL. Alternatively, models have been developed that treat markers associated with traits, markers tagging candidate genes, and previously discovered QTL as fixed effects (Spindel et al., 2016). Initially, GS models were developed for predicting a single trait from a single environment. In recent years, multivariate models and models that account for GEI have been developed (Burgueño et al., 2012; Jarquín et al., 2014; Jia & Jannink, 2012; Jiang et al., 2015; Lopez-Cruz et al., 2015). Multivariate and multi-environment models use the correlation of traits and environments to enhance GS accuracy (Jia & Jannink, 2012; Lopez-Cruz et al., 2015). To date, none of these models have been tested in lentil mainly due to the lack of dense genetic resources. The objectives of this study were to compare the accuracy of different GS models and prediction scenarios based on empirical data in lentil and to make recommendations for designing genomic selection strategies for lentil breeding.

2 | MATERIALS AND METHODS

2.1 | Plant material and phenotyping

We used three populations to evaluate genomic prediction accuracy in lentil (Supplemental Table S1). The first population was a lentil diversity panel (LDP) composed of 324 accessions obtained from the gene banks of the International Center for Agricultural Research in the Dry Areas (ICARDA), the USDA, Plant Gene Resources of Canada (PGRC), and cultivars developed at the Crop Development Centre (CDC), University of Saskatchewan (https://knowpulse.usask.ca/Lentil-Diversity-Panel). This population was evaluated at Sutherland (52°9′ N, 106°30′ W) from 2016 to 2018 and Rosthern, SK (52°41′ N, 106°17′ W) in 2016 and 2017. The second population, LR-01, was composed of 110 RILs developed from a cross between ILL 1704 and 'CDC Robin'. CDC Robin is a red cotyledon lentil cultivar developed at the CDC, and ILL 1704 is a landrace from Ethiopia (Vandenberg et al., 2002). The third population, LR-11, included 120 RILs developed from a cross between 'ILL 8006-BM4' and 'CDC Milestone'. CDC Milestone is a yellow cotyledon lentil cultivar developed at the CDC (Vandenberg et al., 2001), and 'ILL 8006-BM4' was derived from 'Barimasur-4', a red cotyledon cultivar from Bangladesh (Sarker, Erskine, Hassan, Afzal, & Murshed, 1999). The LR-11 population was evaluated at Sutherland and Rosthern, SK, in 2017 and 2018. LR-01 was grown at the same two locations in 2018. For all populations, field trials were established in 1-m² three-row microplots arranged in a randomized complete block design with three replications in each site-year. Seeding rates of 60,

100, and 120 seeds m⁻² were used for the LDP, LR-01, and LR-11, respectively. Plots were seeded between late April and mid-May and harvested between mid-August and early September in each year.

The phenotypic traits measured in all populations included days to flowering (DTF), vegetative period (VEG), days to swollen pods (DTS), days to maturity (DTM), and reproductive period (REP). Days to flowering and DTS were recorded when 10% of the plants in a plot had at least one open flower and pods with fully swollen seeds, respectively. The trait DTM was recorded as the date when 10% of the plants displayed 50% pod maturity. Vegetative and reproductive periods were recorded as the number of days from emergence to flowering and from flowering to maturity, respectively. The phenotypic data were analyzed using analysis of variance (ANOVA) with SAS mixed models, version 9.4 (SAS Institute, 2015). For each population, the phenotypic data were analyzed separately for each environment (site-year) and combined across environments. Genotype was considered a fixed effect and replication was considered random for the data analysis of each environment. For combined data analyses, replication nested in environment, environment (site-years), and GEI were considered random effects. The Kenward–Roger degrees of freedom approximation method was used to compute the degrees of freedom for means (Kenward & Roger, 1997). Broad-sense heritability (H^2) on a plot basis was estimated for all traits in each population using the equation $\sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/e + \sigma_\varepsilon^2/re)$, where σ_g^2 is the genetic variance, σ_{ge}^2 is the GEI 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 the restricted maximum likelihood method described by Holland, Nyquist, and Cervantes-Martínez (2003), with genotype, environment, GEI, and replication considered as random effects.

2.2 | DNA extraction and genotyping

Genomic DNA was extracted from fresh leaves of 2- to 3-wk-old seedlings for all lines using a Qiagen MiniPrep Kit (Qiagen). Genotyping of all lines was performed using a custom exome capture assay as described by Ogutcen et al. (2018). Markers with more than 5% missing data and a minor allele frequency of less than 5% were removed prior to analysis (Tabangin, Woo, & Martin, 2009). A total of 9394, 24,395, and 39,297 markers that were common between the LDP and LR-11, the LDP and LR-01, and LR-01 and LR-11, respectively, were used for genomic prediction. Missing marker genotypes were imputed by replacing them with the population mean for that marker using the function A.mat in the R package rrBLUP, version 4.4 (Endelman, 2011).

2.3 | Statistical models and prediction scenarios

2.3.1 | Standard single-trait prediction models

Single-trait predictions were made using ridge regression best linear unbiased prediction (RR-BLUP), genomic best linear unbiased prediction (G-BLUP), BayesA, BayesB, BayesCπ, Bayesian Lasso, Bayesian ridge regression, and Bayesian reproducing kernel Hilbert spaces regression (RKHS). The RKHS regression is a semi-parametric approach that accounts for both additive and nonadditive genetic effects (de los Campos, Gianola, & Rosa, 2009), while all the other models are based only on the additive effect. The basic equation for the single trait models is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon \qquad (1)$$

where **y** is a vector of observed phenotypes, **X** is a design matrix for fixed effects, **β** is a vector of fixed effects, **Z** is a design matrix for random effects, **u** is a vector of random effects, and **ε** is a vector of residuals.

All models were fitted in R (R Core Team, 2016), using the Bayesian generalized linear regression (BGLR) package, version 1.0.8 (Pérez & de los Campos, 2014), and the rrBLUP package, version 4.4 (Endelman, 2011). The default settings of BGLR (five degrees of freedom and the scale parameter based on the sample variance of the phenotypes) were used (Pérez & de los Campos, 2014). For G-BLUP, the genomic relationship matrix was computed according to VanRaden (2008). For RKHS, we used a Gaussian kernel with a bandwidth parameter of $h = 1/M \times \{1/5, 1, 5\}$ as described by Pérez and de los Campos (2014), where M is the median squared Euclidean distance between all lines calculated using off-diagonals.

2.3.2 | Genomic selection model that combines genome-wide association studies

We used a GS+GWAS model to fit significant markers identified from fold-specific genome-wide association studies (GWAS) as fixed effects in GS (Spindel et al., 2016). GS+GWAS is equivalent to RR-BLUP when no marker is fitted as a fixed effect (Spindel et al., 2016). GS+GWAS is similar to the model in Equation (1) except that the allele dosages of the markers fitted as fixed effects were added as columns to the **X** matrix and then excluded from calculation of the genomic relationship matrix (Spindel et al., 2016). The accuracy of GS+GWAS was tested using a five-fold cross-validation design. In the LDP, GWAS was conducted in each fold based on the phenotypic and genotypic data of the TP using the compressed mixed linear model in GAPIT, version

3.0 (Lipka et al., 2012). Population structure and cryptic relatedness were accounted for using five marker-derived principal components and kinship as covariates. For LR-01 and LR-11, single marker regression was performed in each fold based on the phenotypic and genotypic data of the TP using the *lm* function in R. For all three populations, the *P* values were sorted from low to high, and multiple testing correction was performed based on a false discovery rate using the function p.adjust and the BH method in R (Benjamini & Hochberg, 1995). Then the SNPs on each chromosome were binned into 500-kb distance, and the SNP with the lowest *P* value was extracted from each bin. This step was performed to avoid fitting adjacent SNPs tagging the same QTL as fixed effects. Up to three most significant markers (false discovery rate of 0.1) were selected separately for each fold in the five-fold cross-validation method. When no marker met this threshold, only the most significant marker was selected. The selected markers were then included in the GS+GWAS model as fixed effects, while all the remaining markers from the full marker density were included as random effects. The GS+GWAS model was fitted in R using the kinship.BLUP function in the rrBLUP package (Endelman, 2011).

2.3.3 | A reaction norm model that accounts for genotype × environment interaction

The effect of modeling GEI on genomic prediction accuracy was assessed using a reaction norm model that incorporates the main effects and interactions of molecular markers and environments using covariance structures (EG-G × E) as described by Jarquín et al. (2014). The main effects of markers and environments were included in the model using techniques similar to the standard G-BLUP, while the interaction terms were modeled using a cell by cell product of two covariance structures, $[\mathbf{Z_g}\mathbf{G}\mathbf{Z'_g}] \circ [\mathbf{Z_E}\mathbf{Z'_E}]$, where $\mathbf{Z_g}$ and $\mathbf{Z_E}$ are the incidence matrices for lines and environments, respectively; **G** is a marker-derived genomic relationship matrix; and ∘ indicates the cell by cell (Hadamard) product between two matrices (Jarquín et al., 2014). In this model, phenotypes (y_{ijk}) are 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 using a marker-derived genomic relationship matrix ($\mathbf{g}_j$), plus an interaction term between genotypes and environments (gE_{ij}) and a residual term (ε_{ijk}):

$$y_{ijk} = \mu + E_i + g_j + gE_{ij} + \varepsilon_{ijk} \qquad (2)$$

with $E_i \sim N(0, \sigma_E^2)$, $g \sim N(0, \mathbf{G}\sigma_g^2)$, $gE \sim N(0, [\mathbf{Z_g}\mathbf{G}\mathbf{Z'_g}] \circ [\mathbf{Z_E}\mathbf{Z'_E}]\sigma_{gE}^2)$, and $\varepsilon_{ijk} \sim N(0, \sigma_g^2)$.

The accuracy of the EG-G × E model was compared with a standard G-BLUP model plus a random environmental effect (E), yielding the EG model without the interaction term:

$$y_{ijk} = \mu + E_i + g_j + \varepsilon_{ijk} \quad (3)$$

with $E_i \sim N(0,\sigma_E^2)$, $g \sim N(0,\mathbf{G}\sigma_g^2)$, and $\varepsilon_{ijk} \sim N(0,\sigma_g^2)$.

Prediction accuracy of the reaction norm model was assessed using two cross-validation designs (CV1 and CV2) as implemented by Burgueño et al. (2012). The first cross-validation design (CV1) involved predicting phenotypes of lines that have never been tested in any of the environments (newly developed lines). The second design (CV2) involved predicting phenotypes of lines that were evaluated in some environments but not in others (i.e., incomplete field trials). CV1 assigns lines to folds, and therefore records of the same line are assigned to the same fold; in CV2, individual phenotypes of each line were assigned to folds, and thereby phenotypes of a line from different environments are potentially assigned to different folds (Jarquín et al., 2014; Pérez-Rodríguez et al., 2015). Both CV1 and CV2 were implemented in a five-fold design.

2.4 | Multiple-trait prediction models

We used MT-BayesA and MT-BayesA matrix models for joint prediction of multiple traits (Jiang et al., 2015). The equation for the multiple-trait models is

$$\mathbf{y}_i \mid \mu, \alpha, \Sigma \sim N\left(\mu + \sum_{j=1}^{p} Z_{ij}\alpha_j, \Sigma\right) \quad i = 1, \ldots, n \quad (4)$$

where y_i is an m-element vector of phenotypes for the ith individual, μ is an m-element vector for the overall population mean of m traits, Z_{ij} is the SNP genotype code for the ith individual at the jth marker, $\boldsymbol{\alpha}_j$ is an m-element vector for the effects of the jth SNP marker on all m traits, and $\boldsymbol{\Sigma}$ is an $m \times m$ covariance matrix of the residual effects.

MT-BayesA uses the genetic correlation between two or more traits to potentially improve the accuracy of prediction (Jia & Jannink, 2012). The MT-BayesA matrix is an antedependence-based model that uses correlations between traits as well as SNP effects simultaneously to improve prediction accuracy (Jiang et al., 2015). The standard GS models assume that marker effects are independently distributed; however, this is not always the case, especially when adjacent SNPs are in high LD with the same QTL (Jiang et al., 2015). The antedependence model considers the potential nonstationary correlations between SNP effects near a QTL (Yang & Tempelman, 2012). MT-BayesA and MT-BayesA matrix models were fitted using C language programs developed by Jiang et al. (2015).

2.5 | Genomic prediction scenarios

Three prediction scenarios were evaluated in this study: within population, across population, and across environment. Within-population genomic predictions were made in each population using a five-fold cross-validation design. In a five-fold cross-validation, each population was randomly divided into five groups of approximately equal size. In each fold, four groups were combined and used as the TP, and predictions were made for the remaining group. This was repeated five times until predictions were made for each of the five groups. Across-population genomic predictions were made by using the LDP as the TP to make predictions for the LR-01 and LR-11 populations. LR-01 was used as the TP to make prediction for LR-11 and vice versa. Across-environment predictions were made with the reaction norm model using the CV1 and CV2 approaches described above. Across-population predictions were made using 9394, 24,395, and 39,297 markers that were common between the LDP and LR-11, the LDP and LR-01, and LR-01 and LR-11, respectively. Within-population and across-environment genomic predictions were made using the 24,395 markers in the LDP and 39,297 markers in LR-01 and LR-11. The accuracy of all models was determined as Pearson's correlation (r) between the predicted values and observed phenotypes of lines in the validation set. For the reaction norm models, prediction accuracy was computed within each environment and fold. The same cross-validation folds were used for ST-GS, GS+GWAS, and MT-GS models for a reliable comparison of their accuracy. Inferences for all Bayesian models were based on 30,000 iterations obtained after discarding 10,000 samples as a burn-in.

3 | RESULTS

The distribution of all traits followed an approximately normal distribution in the LDP and LR-11, but it was slightly skewed toward higher values in LR-01 because some lines in LR-01 were susceptible to residue of a herbicide that had been applied the previous fall, which possibly led to delayed flowering and maturity (Supplemental Figure S1). There were strong positive correlations ($r \geq .72$, $P = .001$) among DTF, VEG, DTS, and DTM in the LDP and LR-11 populations (Supplemental Figures 1A and 1B). Reproductive period was moderately correlated with DTM but showed weak correlations with the other traits in the LDP and LR-11. In LR-01, VEG was strongly correlated with DTF ($r = .83$, $P = .001$) but weakly correlated with the other traits (Supplemental Figure 1c). There was also a strong correlation among DTS, DTM, and REP in LR-01 ($r \geq .76$, $P = .001$). A broad range of heritability estimates was obtained for all traits in each population (Table 1). The highest heritability was obtained for VEG in the

TABLE 1 Broad-sense heritability estimates for five traits based on data from the respective number of environments in the lentil diversity panel (LDP) and recombinant inbred lines LR-01 and LR-11 populations

	LDP		LR-01		LR-11	
Trait[a]	Environments	Heritability	Environments	Heritability	Environments	Heritability
DTF	5	.81	2	.56	4	.63
VEG	3	.82	2	.54	4	.68
DTS	5	.74	2	.80	2	.34
DTM	5	.62	2	.83	4	.42
REP	5	.29	2	.85	4	.17

[a]DTF, days to flowering; VEG, vegetative period; DTS, days to swollen pods; DTM, days to maturity; REP, reproductive period.

LDP and LR-11 and for REP in LR-01. Heritability of REP was the lowest in the LDP and LR-11.

3.1 | Within-population prediction accuracy

3.1.1 | Single-trait prediction accuracy

Single-trait prediction accuracy varied across traits, regardless of the model. Within-population prediction accuracies for all traits and models ranged from .65 to .85, .36 to .74, .41 to .79, and .30 to .59 in the LDP, LR-01, LR-11, and combined LR-01 + LR-11 populations, respectively (Figure 1). Prediction accuracy was lower for REP than for the other traits in the LDP. For each trait, all models showed similar accuracy except for REP, where the RKHS showed 3 to 5% higher accuracy than the other models in the LDP. In LR-11, the accuracy of BayesA, BayesB, and GS+GWAS was higher for DTF and VEG (Figure 1). GS+GWAS fitted three significant markers from fold-specific single-marker regression in the TP as fixed effects while all the remaining markers were fitted as random effects (Supplemental Table S2). GS+GWAS showed higher accuracy than the standard RR-BLUP, but its accuracy was either similar to or lower than that of BayesB (Figure 1). Similarly, BayesB showed slightly higher accuracy for DTF and VEG in LR-01 and the joint LR-01 + LR-11 populations. Prediction accuracy was high in the LDP compared with the biparental populations for all traits except for REP, for which the highest accuracy was obtained in LR-01. Combining the LR-01 and LR-11 populations resulted in either similar or lower accuracy than the accuracy within each population.

3.1.2 | Multiple-trait prediction accuracy

We compared the accuracies of MT-BayesA and MT-BayesA matrix models with ST-BayesA and the best ST-GS model with the highest accuracy. In each population, two MT predictions were made using different trait combinations based on their phenotypic correlation (Table 2). In the LDP and LR-11, the first prediction was made for DTF, VEG, DTS, and DTM ($r \geq .72$) while the second prediction was made

for DTM and REP, with correlations of $r = .40$ and .66 in the LDP and LR-11, respectively (Supplemental Figures 1A and 1B). In LR-01, the first prediction was made for DTF and VEG ($r = .83$), while the second prediction was made for DTS, DTM, and REP ($r \geq .76$) (Supplemental Figure 1C). The accuracy of the MT-GS models varied depending on the population and traits selected for joint prediction (Table 2). In the LDP, MT-BayesA and MT-BayesA matrix models showed 4 to 7% higher accuracy for REP than the accuracy of ST-BayesA, but accuracies remained the same for all the other traits. MT-BayesA performed similarly to the best ST-GS model (RKHS), but MT-BayesA matrix showed 3% higher accuracy than the accuracy of RKHS for REP (Figure 1, Table 2). In LR-01, the accuracy of the MT-GS models was either similar to or lower than the ST-BayesA except for VEG when it was predicted together with DTF. The accuracy of MT-BayesA was 10% higher than the accuracy of ST-BayesA for VEG but similar to the best ST-GS model (BayesB) in LR-01. When joint prediction was made for DTM and REP in LR-11, MT-BayesA matrix resulted in 15 and 14% higher accuracy for REP than ST-BayesA and the best ST-GS model (BayesB), respectively (Table 2). Similarly, the accuracy of the MT-GS models was higher than the accuracy of ST-BayesA for DTM, but they had lower accuracy than BayesB in LR-11.

3.2 | Across-population prediction accuracy

Across-population prediction accuracies were very low for all traits relative to within-population prediction accuracy except for REP in LR-01 (Figures 1 and 2). When the LDP was used as the TP to make prediction in LR-01 and LR-11, accuracies for all traits ranged from .21 to .57 and from .09 to .41, respectively (Figures 2a and 2b). RKHS consistently resulted in higher across-population prediction accuracies for all traits when the LDP was used to make prediction in LR-01 and LR-11 (Figures 2a and 2b). In LR-01, prediction accuracy was the highest for REP (ranged from .54 to .56) which also had the highest heritability (Figure 2a). In LR-11, accuracies were close to zero for DTF but relatively higher accuracies (ranging from .25 to .41) were obtained for the
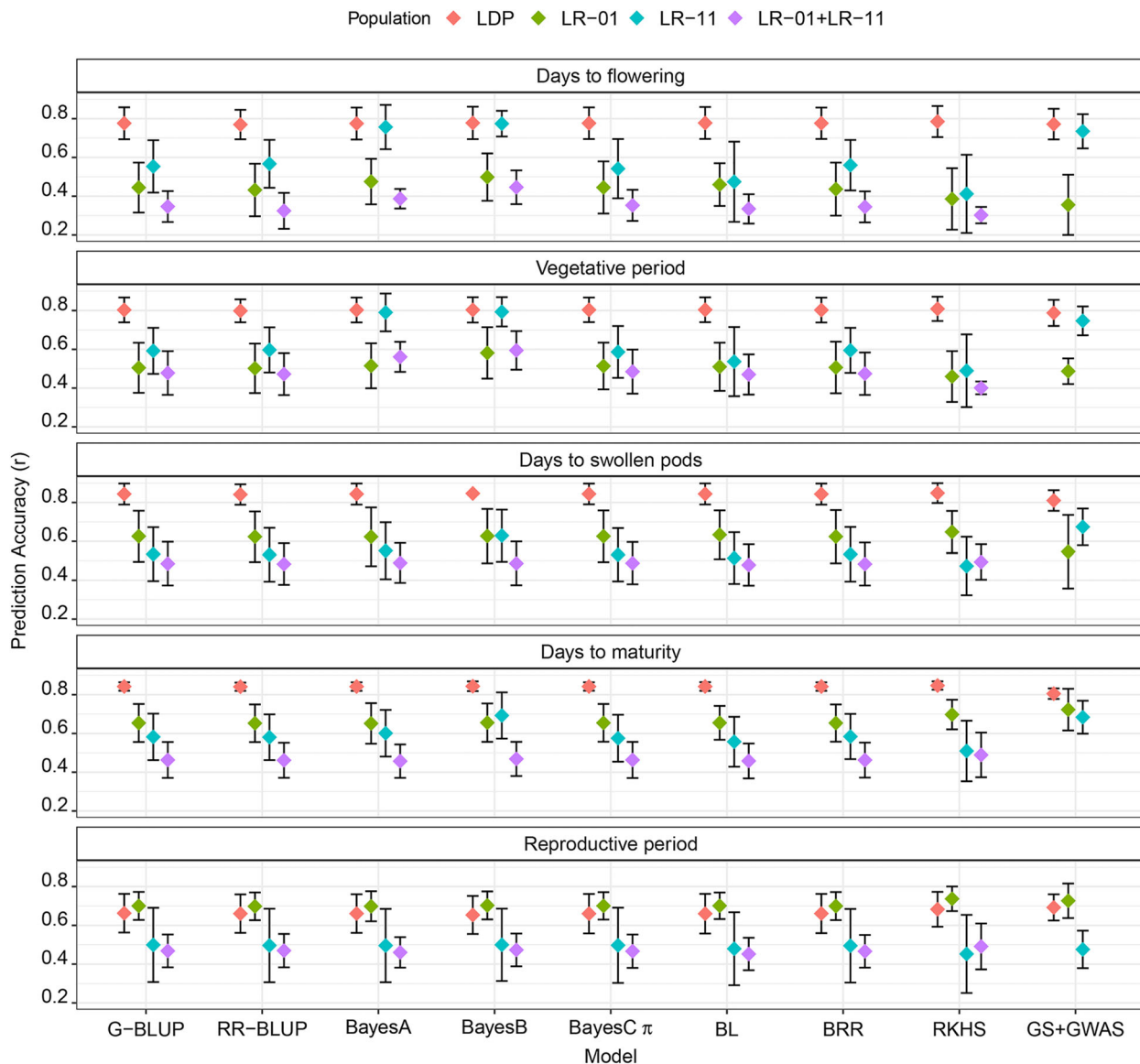
FIGURE 1 Average within-population prediction accuracy (based on five-fold cross-validation) of nine genomic selection models for five traits measured in the lentil diversity panel (LDP), recombinant inbred lines LR-01 and LR-11, and a joint recombinant inbred line population (LR-01+LR-11). Vertical bars show standard deviation of the mean. G-BLUP, genomic best linear unbiased prediction; RR-BLUP, ridge regression best linear unbiased prediction; BL, Bayesian Lasso; BRR, Bayesian ridge regression; RKHS, Bayesian reproducing kernel Hilbert spaces regression; GS+GWAS, genomic selection model that combines genome-wide association studies

other traits (Figure 2b). When one biparental population was used to make prediction for the other biparental population, accuracies were close to zero or negative for DTF and VEG in all models except in BayesA and BayesB (Figures 2c and 2d). In contrast, slightly higher accuracies (ranging from .26 to .38) were obtained for DTS and DTM in LR-01 using LR-11 as the training set (Figure 2c).

3.3 | Across-environment prediction accuracy

Across-environment predictions were made using two cross-validation designs that simulate prediction for newly devel-

oped lines (CV1) and incomplete field trials (CV2). We compared the prediction accuracy of the EG-G × E model that incorporates GEI with the EG model that included only the main effects of environments and markers without the interaction term. Prediction accuracies varied depending on the cross-validation design and population. For all traits in the LDP, prediction accuracies of EG-G × E ranged from .52 to .82 in CV1 and from .63 to .95 in CV2 (Table 3). The lowest accuracies were obtained for REP in both cross-validation designs. For REP, modeling GEI resulted in 7–30 and 4–35% higher accuracy than EG in CV1 and CV2, respectively (Table 3). Similarly, there was a 4–6% increase in prediction accuracy for DTM when modeling GEI in Sutherland 2016.

TABLE 2 Average prediction accuracy (from five-fold cross-validation) of multiple-trait (MT)-BayesA, MT-BayesA matrix, single-trait (ST)-BayesA, and the best ST model in the lentil diversity panel (LDP) and recombinant inbred lines LR-01 and LR-11 populations

Population	Prediction[a]	Trait[b]	MT-BayesA	MT-BayesA matrix	ST-BayesA	Best ST model
LDP	1 ($r \geq .72$)	DTF	.775	.776	.775	.785
		VEG	.800	.799	.803	.809
		DTS	.837	.838	.844	.849
		DTM	.835	.837	.842	.847
	2 ($r = .40$)	DTM	.836	.834	.842	.847
		REP	.686	.706	.661	.683
LR-01	1 ($r = .83$)	DTF	.498	.428	.475	.499
		VEG	.567	.489	.515	.582
	2 ($r \geq .76$)	DTS	.544	.501	.623	.648
		DTM	.641	.586	.680	.697
		REP	.682	.647	.699	.737
LR-11	1 ($r \geq .72$)	DTF	.695	.726	.757	.774
		VEG	.705	.735	.790	.794
		DTS	.596	.629	.552	.629
		DTM	.620	.643	.601	.693
	2 ($r = .66$)	DTM	.651	.642	.601	.693
		REP	.499	.572	.496	.500

[a]Joint predictions were made for correlated traits.

[b]DTF, days to flowering; VEG, vegetative period; DTS, days to swollen pods; DTM, days to maturity; REP, reproductive period.

For all other traits, however, modeling the interaction term did not improve the accuracy of prediction. In LR-01, accuracies for all traits ranged from .41 to .66 and .63 to .88 in CV1 and CV2, respectively (Table 3). Both models showed similar accuracy for all traits in LR-01, indicating that there was no benefit of modeling GEI when trait heritability is high. Similarly, prediction accuracies of EG-G × E in LR-11 ranged from .19 to .60 and .36 to .89 for all traits in CV1 and CV2, respectively. The lowest accuracies were obtained for REP, which also had low heritability ($h^2 = .17$). Similar accuracies were obtained between the two models for all traits in LR-11 except for REP in CV2, where incorporating the interaction term resulted in 18–66% higher accuracy. This shows that modeling GEI improved the accuracy of prediction for a low-heritability trait but modeling only the main effects was enough for traits of high heritability. Overall, CV2 consistently resulted in higher accuracy than CV1 for all traits and populations (Table 3).

4 | DISCUSSION

This study evaluated the accuracy of ST-GS models, MT-GS models, a model that included significant markers from GWAS as fixed effects, and a model that accounted for GEI using a lentil diversity panel and two RIL populations. Moreover, within-population, across-population, and across-environment predictions were tested to simulate scenarios

that breeders may face when implementing GS. Highly variable prediction accuracies were obtained depending on the trait, population, prediction scenario, and statistical model used.

4.1 | Within-population genomic prediction

Nine ST-GS and two MT-GS models were used to make within-population genomic predictions based on a five-fold cross-validation design. Prediction accuracies ranged from .65 to .85, .36 to .74, .41 to .79, and .30 to .59 for all traits in the LDP, LR-01, LR-11, and combined LR-01 + LR-11 populations, respectively (Figure 1). The degree of genetic relationship between the training and validation sets is an important factor that affects the accuracy of GS prediction, with the accuracy being higher for lines that are closely related (Habier, Fernando, & Dekkers, 2007; Riedelsheimer et al., 2013). Lines in the biparental populations are closely related compared with the lines in the LDP, but accuracies were generally higher in the LDP than the biparental populations. This could be because the number of lines in the LDP is nearly three times that of the biparental populations. Training population size is another important factor that affects the accuracy of genomic predictions. Increasing the TP size increases the accuracy of prediction because it provides more data to estimate marker effects (Asoro et al., 2011; Saatchi, Miraei-Ashtiani, Javaremi, Moradi-Shahrebabak, &
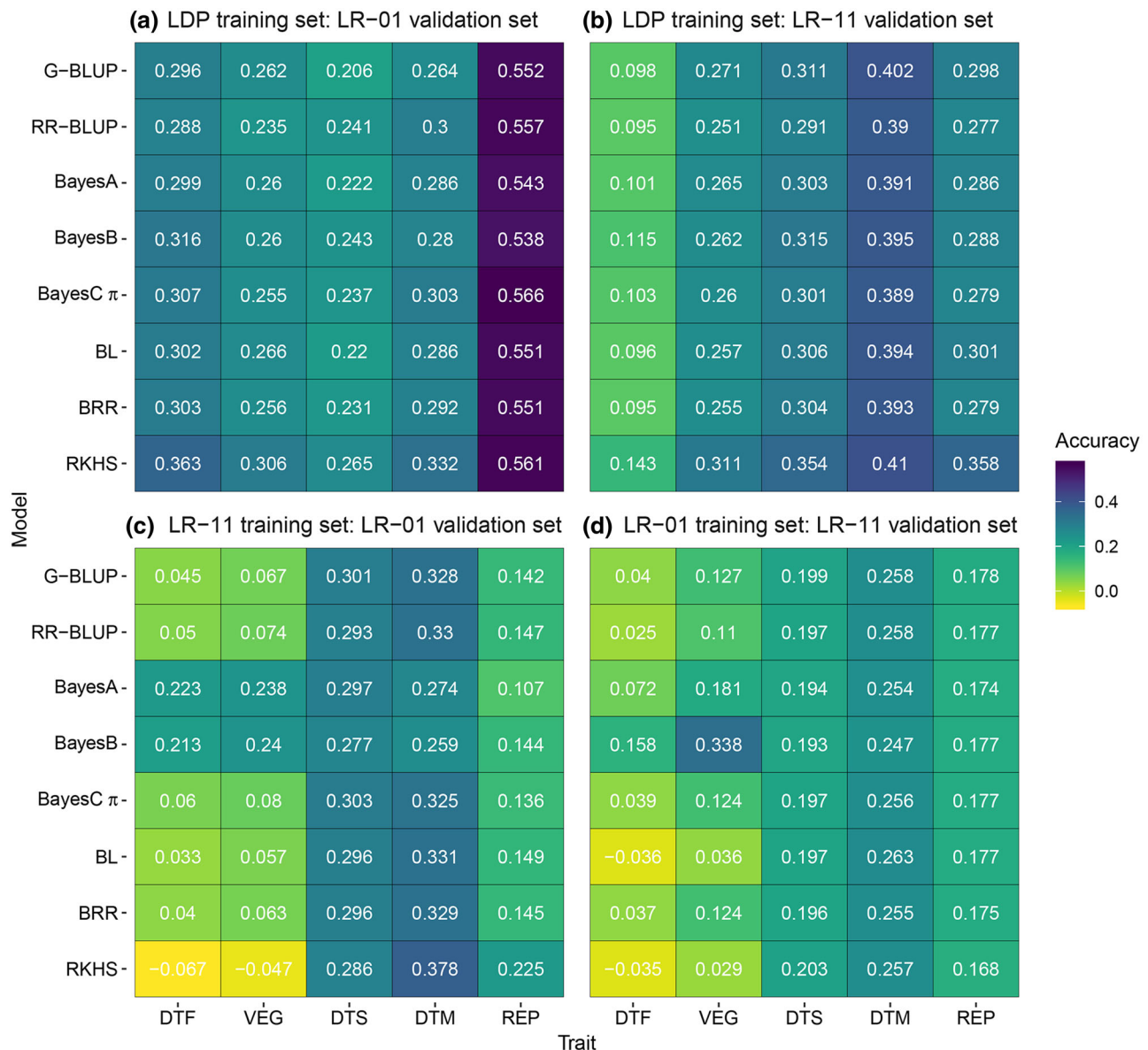
(a) LDP training set: LR−01 validation set

Model	DTF	VEG	DTS	DTM	REP
G−BLUP	0.296	0.262	0.206	0.264	0.552
RR−BLUP	0.288	0.235	0.241	0.3	0.557
BayesA	0.299	0.26	0.222	0.286	0.543
BayesB	0.316	0.26	0.243	0.28	0.538
BayesC π	0.307	0.255	0.237	0.303	0.566
BL	0.302	0.266	0.22	0.286	0.551
BRR	0.303	0.256	0.231	0.292	0.551
RKHS	0.363	0.306	0.265	0.332	0.561

(b) LDP training set: LR−11 validation set

Model	DTF	VEG	DTS	DTM	REP
G−BLUP	0.098	0.271	0.311	0.402	0.298
RR−BLUP	0.095	0.251	0.291	0.39	0.277
BayesA	0.101	0.265	0.303	0.391	0.286
BayesB	0.115	0.262	0.315	0.395	0.288
BayesC π	0.103	0.26	0.301	0.389	0.279
BL	0.096	0.257	0.306	0.394	0.301
BRR	0.095	0.255	0.304	0.393	0.279
RKHS	0.143	0.311	0.354	0.41	0.358

(c) LR−11 training set: LR−01 validation set

Model	DTF	VEG	DTS	DTM	REP
G−BLUP	0.045	0.067	0.301	0.328	0.142
RR−BLUP	0.05	0.074	0.293	0.33	0.147
BayesA	0.223	0.238	0.297	0.274	0.107
BayesB	0.213	0.24	0.277	0.259	0.144
BayesC π	0.06	0.08	0.303	0.325	0.136
BL	0.033	0.057	0.296	0.331	0.149
BRR	0.04	0.063	0.296	0.329	0.145
RKHS	−0.067	−0.047	0.286	0.378	0.225

(d) LR−01 training set: LR−11 validation set

Model	DTF	VEG	DTS	DTM	REP
G−BLUP	0.04	0.127	0.199	0.258	0.178
RR−BLUP	0.025	0.11	0.197	0.258	0.177
BayesA	0.072	0.181	0.194	0.254	0.174
BayesB	0.158	0.338	0.193	0.247	0.177
BayesC π	0.039	0.124	0.197	0.256	0.177
BL	−0.036	0.036	0.197	0.263	0.177
BRR	0.037	0.124	0.196	0.255	0.175
RKHS	−0.035	0.029	0.203	0.257	0.168

Accuracy: 0.4 / 0.2 / 0.0

FIGURE 2 Across-population prediction accuracy of eight models for days to flowering (DTF), vegetative period (VEG), days to swollen pods (DTS), days to maturity (DTM), and reproductive period (REP): (a) the lentil diversity panel (LDP) was used as a training set and recombinant inbred line LR-01 was a validation set; (b) the LDP was used as a training set and recombinant inbred line LR-11 was a validation set; (c) LR-11 was used as training set and LR-01 was a validation set; and (d) LR-01 was used as training set and LR-11 was a validation set. G-BLUP, genomic best linear unbiased prediction; RR-BLUP, ridge regression best linear unbiased prediction; BL, Bayesian Lasso; BRR, Bayesian ridge regression; RKHS, Bayesian reproducing kernel Hilbert spaces regression

Mehrabani-Yeghaneh, 2010; VanRaden et al., 2009). All models showed similar accuracy in the LDP except for RKHS, which showed 3–5% higher accuracy than the other models for REP. BayesA, BayesB, and GS+GWAS resulted in higher accuracy for DTF and VEG in LR-11 (Figure 1). In theory, RR-BLUP is expected to have higher accuracy for traits controlled by many QTL with small effects, while Bayesian models have higher accuracy when few QTL with large effects control most of the phenotypic variance (Lorenz et al., 2011). The higher accuracy of BayesA and BayesB in LR-11 could

be due to the presence of large-effect QTL controlling DTF and VEG.

The accuracy of RKHS was slightly higher for REP in the LDP but it resulted in similar or lower accuracy than the other models for all traits (Figure 1). This suggests that the contribution of nonadditive genetic effects to the total genetic variance is negligible in the populations used for this study. Previous studies in wheat and maize also reported no benefit of models accounting for nonadditive effects over simple additive models (Lorenzana & Bernardo, 2009; Sallam,

TABLE 3 Across-environment prediction accuracy (from five-fold design) for four traits predicted using the EG and EG-G × E models for cross-validation CV1 and CV2

Environment	Model EG				Model EG-G × E			
	DTF	DTM	DTS	REP	DTF	DTM	DTS	REP
	Lentil diversity panel							
	CV1[a]							
Rosthern, 2016	.749 (.039)	.766 (.063)	.797 (.036)	.488 (.085)	.734 (.053)	.761 (.038)	.802 (.042)	.524 (.078)
Rosthern, 2017	.802 (.073)	.738 (.061)	.794 (.070)	.558 (.100)	.803 (.074)	.752 (.053)	.798 (.076)	.597 (.104)
Sutherland, 2016	.798 (.074)	.746 (.096)	.807 (.063)	.529 (.099)	.796 (.078)	.777 (.071)	.811 (.059)	.609 (.032)
Sutherland, 2017	.812 (.066)	.769 (.074)	.822 (.078)	.490 (.186)	.819 (.062)	.779 (.034)	.822 (.078)	.638 (.121)
Sutherland, 2018	.787 (.069)	.729 (.088)	.806 (.055)	.585 (.056)	.785 (.077)	.746 (.060)	.821 (.057)	.699 (.058)
	CV2							
Rosthern, 2016	.896 (.027)	.856 (.032)	.875 (.026)	.595 (.086)	.901 (.029)	.861 (.037)	.877 (.042)	.632 (.112)
Rosthern, 2017	.943 (.013)	.826 (.033)	.890 (.022)	.625 (.093)	.947 (.011)	.840 (.036)	.915 (.017)	.649 (.083)
Sutherland, 2016	.950 (.018)	.811 (.033)	.922 (.011)	.579 (.100)	.953 (.013)	.857 (.025)	.921 (.019)	.698 (.074)
Sutherland, 2017	.958 (.006)	.864 (.025)	.933 (.018)	.558 (.080)	.950 (.006)	.878 (.023)	.936 (.006)	.720 (.047)
Sutherland, 2018	.923 (.018)	.807 (.055)	.903 (.018)	.526 (.100)	.931 (.032)	.812 (.054)	.925 (.013)	.708 (.047)
	Recombinant inbred line LR-01							
	CV1							
Rosthern, 2018	.484 (.156)	.573 (.086)	.543 (.140)	.608 (.053)	.485 (.149)	.576 (.086)	.548 (.139)	.611 (.054)
Sutherland, 2018	.408 (.147)	.613 (.074)	.552 (.149)	.654 (.066)	.406 (.150)	.615 (.073)	.559 (.141)	.656 (.065)
	CV2							
Rosthern, 2018	.637 (.187)	.831 (.021)	.836 (.082)	.849 (.033)	.638 (.187)	.850 (.056)	.837 (.081)	.854 (.062)
Sutherland, 2018	.631 (.075)	.870 (.033)	.765 (.102)	.882 (.033)	.630 (.074)	.862 (.031)	.767 (.102)	.867 (.040)
	Recombinant inbred line LR-11							
	CV1							
Rosthern, 2017	.587 (.120)	.499 (.127)	—[b]	.295 (.163)	.585 (.123)	.499 (.128)	–	.299 (.162)
Rosthern, 2018	.457 (.138)	.331 (.052)	.428 (.102)	.296 (.162)	.456 (.139)	.331 (.047)	.430 (.103)	.299 (.157)
Sutherland, 2017	.595 (.153)	.575 (.110)	–	.194 (.165)	.595 (.157)	.574 (.106)	–	.192 (.166)
Sutherland, 2018	.512 (.211)	.507 (.161)	.440 (.154)	.393 (.146)	.514 (.210)	.509 (.155)	.442 (.153)	.395 (.145)
	CV2							
Rosthern, 2017	.891 (.044)	.738 (.050)	–	.370 (.327)	.890 (.045)	.736 (.049)	–	.521 (.154)
Rosthern, 2018	.736 (.060)	.545 (.163)	.597 (.146)	.383 (.252)	.736 (.060)	.546 (.162)	.597 (.147)	.453 (.135)
Sutherland, 2017	.889 (.042)	.794 (.087)	–	.220 (.304)	.889 (.042)	.792 (.090)	–	.364 (.212)
Sutherland, 2018	.892 (.053)	.770 (.114)	.640 (.126)	.326 (.173)	.892 (.054)	.769 (.14)	.639 (.126)	.442 (.131)

Notes: Average and standard deviation (in parentheses). DTF, days to flowering; VEG, vegetative period; DTS, days to swollen pods; DTM, days to maturity; REP, reproductive period.

[a]CV1, prediction for lines that were not evaluated in any of the environments; CV2, prediction for lines that were evaluated in some of the environments.

[b]Data not available.

Endelman, Jannink, & Smith, 2015; Zhao, Zeng, Fernando, & Reif, 2013). In contrast, improved prediction accuracies were reported using RKHS and models that incorporate epistasis (Crossa et al., 2010; Pérez-Rodríguez et al., 2012). This suggests that the benefit of accounting for nonadditive effects in GS may vary depending on the trait and population type.

GS+GWAS fitted up to three markers that are significantly associated with traits as fixed effects (Supplemental Table S2). There was no benefit of fitting significant markers as fixed effects in the LDP because all markers explained less

than 5% of the phenotypic variance (Supplemental Table S2). GS+GWAS works best for traits with one or more medium to large effect QTL segregating in the population but it has no advantage if a trait has no significant GWAS peaks (Spindel et al., 2016). Bernardo (2014), suggested that major QTL should be fitted as having fixed effects in GS, especially if a few major QTL are present and if each QTL explains more than 10% of the genetic variance. Including up to three significant markers as a fixed effect in GS+GWAS resulted in 18 to 30% higher accuracy than the standard RR-BLUP for

DTF, VEG, DTS ,and DTM in LR-11 (Figure 1). Similarly, 4 and 11% higher accuracy was obtained for REP and DTM, respectively, in LR-01. However, no improvement in accuracy was observed for the other traits in LR-01 and LR-11. Spindel et al. (2016) showed that GS+GWAS performed better than the standard RR-BLUP in all cases, and up to 30% higher accuracy was obtained for flowering time of rice (*Oryza sativa* L.), which had a large GWAS peak. Although GS+GWAS gave higher accuracies than the standard RR-BLUP in LR-01 and LR-11, its accuracy was either similar to or lower than BayesB, indicating that variable selection models such as BayesB can be equally effective to account for large-effect QTL in GS.

Combining LR-01 and LR-11 to increase the TP size did not improve the accuracy of prediction for all traits (Figure 1). Previous studies in wheat and barley (*Hordeum vulgare* L.) also indicated that prediction accuracies did not improve when unrelated populations from different breeding programs were merged to increase TP size (Charmet et al., 2014; Lorenz, Smith, & Jannink, 2012). Unrelated populations may have different LD phases, and combining them into a single TP reduces the overall LD (Goddard, 2012). Moreover, combining multiple populations may create strong population structure and allele frequency differences between subpopulations, which reduces GS accuracy (Riedelsheimer et al., 2013). The standard GS models assume that marker effects are constant across subpopulations and fail to account for differences in allele frequency due to population structure. Therefore, models that account for population structure need to be tested when unrelated populations are combined to increase TP size.

Similar accuracies were obtained between MT-GS and ST-GS models when joint predictions were made for correlated high heritability traits (Tables 1 and 2). However, the accuracy of MT-BayesA matrix increased by 3 and 14% for REP when it was predicted with DTM in the LDP and LR-11, respectively (Table 2). Jiang et al. (2015) also showed that the prediction accuracy of the simplified version of the MT-BayesA matrix model was 2.8–8.6% higher than that of the best ST-GS model. Previous studies based on simulated data showed that MT-GS models have higher accuracy for a low heritability trait ($h^2 = .1$) jointly predicted with a correlated high heritability trait ($h^2 \geq .5$) (Hayashi & Iwata, 2013; Jia & Jannink, 2012; Jiang et al., 2015). However, no improvement in accuracy was observed compared with ST-GS models for traits of high heritability and in the absence of genetic correlation between traits (Hayashi & Iwata, 2013; Jia & Jannink, 2012; Jiang et al., 2015).

4.2 | Across-population genomic prediction

Variable across-population prediction accuracies were obtained depending on the trait and population used as TP.

Overall, across-population prediction accuracies were low relative to within-population prediction accuracy for all traits except for REP in LR-01 (Figures 1 and 2). The prediction accuracy was the highest (range .54–.56) for REP, which indicates that moderate across-population prediction can be made for highly heritable traits (Figure 2a). Prediction in LR-01 and LR-11 ranged from .21 to .57 and .09 to .41, respectively, when the LDP was used as the TP (Figures 2a and 2b). When one biparental population was used to predict the other biparental population, accuracies were close to zero or negative for DTF and VEG in all models except in BayesA and BayesB (Figures 2c and 2d). Previous studies in wheat and maize also reported mean accuracies of zero or negative values when an unrelated TP was used to make prediction for an independent population (Charmet et al., 2014; Crossa et al., 2014; Riedelsheimer et al., 2013; Windhausen et al., 2012). Genomic selection models use both the genetic relationships among individuals and LD between markers and QTL to improve prediction accuracy (Habier et al., 2007). Across-population prediction not only requires strong LD but similar linkage phases in each population (Goddard & Hayes, 2007). When the TP is unrelated to the breeding population, marker effects estimated in one population cannot be transferred to the other population due to differences in allele frequency and LD phase resulting in low prediction accuracy (Bassi et al., 2016; Windhausen et al., 2012). In plant breeding, early generation nurseries are comprised of different families from many crosses, and the most attractive application of GS is to estimate GEBVs of lines in these families based on marker effects estimated from an independent population. However, the results of this study, as well as previous studies, showed that GS has very low accuracy for such applications. Based on a simulation study, Meuwissen (2009) suggested that a substantially higher marker density and TP size is required for accurate prediction of GEBVs in unrelated individuals. Therefore, across-population prediction should not be considered unless the two populations are closely related or the size of the TP is very large.

4.3 | Across-environment genomic prediction

Across-environment predictions were made using reaction norm models that incorporated the main and interaction terms of environments and markers (Jarquín et al., 2014). Genotype × environment interaction is an important factor that affects both phenotypic and genomic selection accuracy. Prediction accuracies in CV1 ranged from .49 to .82, .41 to .66, and .19 to .60 in the LDP, LR-01, and LR-11, respectively (Table 3). Prediction accuracies in CV2 were consistently higher than those in CV1 and ranged from .53 to .96, .63 to .88, and .22 to .89 in the LDP, LR-01, and LR-11, respectively. Similar results

were obtained in previous studies that used the same cross-validation designs in wheat and cotton (*Gossypium hirsutum* L.) (Burgueño et al., 2012; Crossa et al., 2015; Jarquín et al., 2014; Jarquín et al., 2017; Lopez-Cruz et al., 2015; Pérez-Rodríguez et al., 2015). This is because CV2 allows borrowing of information for the same line across environments. Modeling the interaction term improved the accuracy of prediction for REP by 4–35 and 18–66% in the LDP and LR-11, respectively (Table 3). But for all other traits, modeling GEI resulted in similar accuracies to the model that included only the main effects of markers and environments. Previous studies in wheat and maize showed that modeling GEI improves the accuracy of genomic predictions (Crossa et al., 2015; Cuevas et al., 2017; Jarquín et al., 2014; Jarquín et al., 2017; Lopez-Cruz et al., 2015). Jarquín et al. (2014) used the reaction norm model to predict the grain yield of wheat and reported a 35% increase in accuracy when adding interaction terms between markers and environments. Recently, Jarquín et al. (2017) showed that modeling GEI resulted in 16 to 82% higher accuracy than a baseline model that did not include the GEI term. Other studies in wheat and cotton that used the reaction norm model with pedigrees instead of molecular markers also obtained the highest prediction accuracy when a GEI term was included in the model (Pérez-Rodríguez et al., 2015; Sukumaran, Crossa, Jarquín, & Reynolds, 2017).

5 | CONCLUSIONS

This research provides the first empirical evidence of GS in lentil. Comparison of different models and approaches showed that most ST-GS models have similar accuracy in the absence of large effect QTL underlying the traits, but BayesB is the best model when QTL with relatively large effects are present. Genomic selection models that accounted for GEI and MT-GS models improved the accuracy of prediction for a low heritability trait, but they have no advantage when trait heritability is high. Within-population and across-environment genomic predictions resulted in moderate to high accuracies, but across-population prediction accuracies were very low. This suggests that GS can be implemented in lentil to make within-population or across-environment predictions, but across-population prediction should not be considered when the population size is small.

AUTHOR CONTRIBUTIONS

TAH generated phenotypic data, performed all statistical analyses, and wrote the manuscript. TH, DW, SN generated phenotypic data, and edited the manuscript. LR generated the SNP data set. AV designed the experiment and provided germplasm materials. KEB designed the experiment, supervised the project, and edited the manuscript. All authors read and approved the submitted version.

DATA AVAILABILITY
The data supporting this study are available at https://knowpulse.usask.ca/Lentil-AGILE-GenomicSelection or from the authors upon request.

CONFLICT OF INTEREST
The authors declare that they have no conflict of interest.

ORCID
Teketel A. Haile https://orcid.org/0000-0002-6212-9489
Kirstin E. Bett https://orcid.org/0000-0001-7959-6959

REFERENCES

Arumuganathan, K., & Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, *9*, 208–218. https://doi.org/10.1007/bf02672069.

Asoro, F. G., Newell, M. A., Beavis, W. D., Scott, M. P., & 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, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science*, *242*, 23–36. https://doi.org/10.1016/j.plantsci.2015.08.021.

Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., … Poland, J. A. (2016). Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *Plant Genome*, *9*, 1–12. https://doi.org/10.3835/plantgenome2016.01.0005.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*, 289–300.

Bernardo, R. (2009). Genomewide selection for rapid introgression of exotic germplasm in maize. *Crop Science*, *49*, 419–425. https://doi.org/10.2135/cropsci2008.08.0452.

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

Bernardo, R., & Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Science*, *47*, 1082–1090. https://doi.org/10.2135/cropsci2006.11.0690.

Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B., … Crossa, J. (2015). Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Science*, *55*, 154–163. https://doi.org/10.2135/cropsci2014.07.0460.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/a:1010933404324.

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 Science*, *52*, 707–719. https://doi.org/10.2135/cropsci2011.06.0299.

Calus, M. P. L. (2010). Genomic breeding value prediction: Methods and procedures. *Animal*, *4*, 157–164. https://doi.org/10.1017/S1751731109991352.

Charmet, G., Storlie, E., Oury, F. X., Laurent, V., Beghin, D., Chevarin, L., … Robert, O. (2014). Genome-wide prediction of three important traits in bread wheat. *Molecular Breeding*, *34*, 1843–1852. https://doi.org/10.1007/s11032-014-0143-y.

Combs, E., & Bernardo, R. (2013). Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome*, *6*, 1–7. https://doi.org/10.3835/plantgenome2012.11.0030.

Crossa, J., de los Campos, G., Perez, P., Gianola, D., Burgueno, J., Araus, J. L., … Braun, H.-J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, *186*, 713–724. https://doi.org/10.1534/genetics.110.118521.

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 Science*, *56*, 1–17. https://doi.org/10.2135/cropsci2015.04.0260

Crossa, J., Jarquin, D., Franco, J., Perez-Rodriguez, P., Burgueno, J., Saint-Pierre, C., … Singh, S. (2016). Genomic prediction of gene bank wheat landraces. *G3: Genes, Genomes, Genetics*, *6*, 1819–1834. https://doi.org/10.1534/g3.116.029637.

Crossa, J., Pérez, P., Hickey, J., Burgueño, J., Ornella, L., Cerón-Rojas, J., … Mathews, K. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity*, *112*, 48–60. https://doi.org/10.1038/hdy.2013.16.

Cuevas, J., Crossa, J., Montesinos-Lopez, O. A., Burgueno, J., Perez-Rodriguez, P., & de los Campos, G. (2017). Bayesian genomic prediction with genotype × environment interaction kernel models. *G3: Genes, Genomes, Genetics*, *7*, 41–53. https://doi.org/10.1534/g3.116.035584.

de los Campos, G., Gianola, D., & Rosa, G. J. M. (2009). Reproducing kernel Hilbert spaces regression: A general framework for genetic evaluation. *Journal of Animal Science*, *87*, 1883–1887. https://doi.org/10.2527/jas.2008-1259.

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, *193*, 327–345. https://doi.org/10.1534/genetics.112.143313.

de los Campos, G., Veturi, Y., Vazquez, A. I., Lehermeier, C., & Pérez-Rodríguez, P. (2015). Incorporating genetic heterogeneity in whole-genome regressions using interactions. *Journal of Agricultural, Biological, and Environmental Statistics*, *20*, 467–490. https://doi.org/10.1007/s13253-015-0222-5.

Endelman, J. B. (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. (2019). *Crops*. FAOSTAT. Retrieved from http://www.fao.org/faostat/en/#data/QC (accessed 4 June 2019).

Gianola, D., Fernando, R. L., & 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, J. B. C. H.M. (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, M. E. (2012). Uses of genomics in livestock agriculture. *Animal Production Science*, *52*, 73–77. https://doi.org/10.1071/AN11180.

Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, *124*, 323–330. https://doi.org/10.1111/j.1439-0388.2007.00702.x.

Guo, Z., Tucker, D. M., Basten, C. J., Gandhi, H., Ersoz, E., Guo, B., … Gay, G. (2014). The impact of population structure on genomic prediction in stratified populations. *Theoretical and Applied Genetics*, *127*, 749–762. https://doi.org/10.1007/s00122-013-2255-x.

Habier, D., Fernando, R. L., & Dekkers, J. C. (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, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, *12*, 186. https://doi.org/10.1186/1471-2105-12-186.

Hayashi, T., & Iwata, H. (2013). A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics*, *14*, 34. https://doi.org/10.1186/1471-2105-14-34.

Heffner, E. L., Jannink, J.-L., & Sorrells, M. E. (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, E. L., Lorenz, A. J., Jannink, J.-L., & Sorrells, M. E. (2010). Plant breeding with genomic selection: Gain per unit time and cost. *Crop Science*, *50*, 1681–1690. https://doi.org/10.2135/cropsci2009.11.0662.

Holland, J. B., Nyquist, W. E., & Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: An update. *Plant Breeding Reviews*, *22*, 9–112. https://doi.org/10.1002/9780470650202.

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., … de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, *127*, 595–607. https://doi.org/10.1007/s00122-013-2243-1.

Jarquín, D., Lemes da Silva, C., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., … Crossa, J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. *Plant Genome*, *10*, 1–15. 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.

Jiang, J., Zhang, Q., Ma, L., Li, J., Wang, Z., & Liu, J. F. (2015). Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity*, *115*, 29–36. https://doi.org/10.1038/hdy.2015.9.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*, 983–997. https://doi.org/10.2307/2533558.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., … Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, *28*, 2397–2399. https://doi.org/10.1093/bioinformatics/bts444.

Liu, X., Wang, H., Wang, H., Guo, Z., Xu, X., Liu, J., … Xu, Y. (2018). Factors affecting genomic selection revealed by empirical evidence in maize. *Crop Journal*, *6*, 341–352. https://doi.org/10.1016/j.cj.2018.03.005.

Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J.-L., … de los Campos, G. (2015). Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3: Genes, Genomes, Genetics*, *5*, 569–582. https://doi.org/10.1534/g3.114.016097.

Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., … Jannink, J.-L. (2011). Genomic selection in plant breeding: Knowledge and prospects. *Advances in Agronomy*, *110*, 77–123. https://doi.org/10.1016/B978-0-12-385531-2.00002-5.

Lorenz, A. J., Smith, K. P., & Jannink, J.-L. (2012). Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. *Crop Science*, *52*, 1609–1621. https://doi.org/10.2135/cropsci2011.09.0503.

Lorenzana, R. E., & Bernardo, R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics*, *120*, 151–161. https://doi.org/10.1007/s00122-009-1166-3.

Meuwissen, T. H. E. (2009). Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics, Selection, Evolution*, *41*, 35. https://doi.org/10.1186/1297-9686-41-35.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*, 1819–1829.

Ogutcen, E., Ramsay, L., von Wettberg, E. B., & Bett, K. E. (2018). Capturing variation in *Lens* (Fabaceae): Development and utility of an exome capture array for lentil. *Applications in Plant Sciences*, *6*, e1165. https://doi.org/10.1002/aps3.1165.

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*, 681–686. https://doi.org/10.1198/016214508000000337.

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.

Pérez, P., de los Campos, G., Crossa, J., & Gianola, D. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome*, *3*, 106–116. https://doi.org/10.3835/plantgenome2010.04.0005.

Pérez-Rodríguez, P., Crossa, J., Bondalapati, K., De Meyer, G., Pita, F., & de los Campos, G. (2015). A pedigree-based reaction norm model for prediction of cotton yield in multienvironment trials. *Crop Science*, *55*, 1143–1151. https://doi.org/10.2135/cropsci2014.08.0577.

Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y., & Dreisigacker, S. (2012). Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3: Genes, Genomes, Genetics*, *2*, 1595–1605. https://doi.org/10.1534/g3.112.003665.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Riedelsheimer, C., Endelman, J. B., Stange, M., Sorrells, M. E., Jannink, J. L., & Melchinger, A. E. (2013). Genomic predictability of interconnected biparental maize populations. *Genetics*, *194*, 493–503. https://doi.org/10.1534/genetics.113.150227.

Saatchi, M., Miraei-Ashtiani, S. R., Javaremi, A. N., Moradi-Shahrebabak, M., & Mehrabani-Yeghaneh, H. (2010). The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. *African Journal of Biotechnology*, *9*, 438–442.

Sallam, A. H., Endelman, J. B., Jannink, J. L., & Smith, K. P. (2015). Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *Plant Genome*, *8*, 1–15. https://doi.org/10.3835/plantgenome2014.05.0020.

Sarker, A., Erskine, W., Hassan, M. S., Afzal, M. A., & Murshed, A. N. M. M. (1999). Registration of 'Barimasur-4' lentil. *Crop Science*, *39*, 876.

SAS Institute. (2015). *The SAS system for Windows. Version 9.4*. Cary, NC: SAS Institute.

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J. L., & 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.

Sukumaran, S., Crossa, J., Jarquín, D., & Reynolds, M. (2017). Pedigree-based prediction models with genotype × environment interaction in multienvironment trials of CIMMYT wheat. *Crop Science*, *17*, 1865–1880. https://doi.org/10.2135/cropsci2016.06.0558.

Tabangin, M. E., Woo, J. G., & Martin, L. J. (2009). The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC Proceedings*, *3*(Supplement 7), S41. https://doi.org/10.1186/1753-6561-3-S7-S41.

Vandenberg, A., Kiehn, F. A., Vera, C., Gaudiel, R., Buchwaldt, L., Dueck, S., … Slinkard, A. E. (2002). CDC Robin lentil. *Canadian Journal of Plant Science*, *82*, 111–112. https://doi.org/10.4141/P01-003.

Vandenberg, A., Kiehn, F. A., Vera, C., Gaudiel, R., Buchwaldt, L., Kirkland, K. J., … Slinkard, A. E. (2001). CDC Milestone lentil. *Canadian Journal of Plant Science*, *81*, 113–114. https://doi.org/10.4141/P00-043.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, *92*, 16–24. https://doi.org/10.3168/jds.2008-1514.

Wientjes, Y. C., Veerkamp, R. F., & Calus, M. P. (2013). The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics*, *193*, 621–631. https://doi.org/10.1534/genetics.112.146290.

Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J. L., Sorrells, M. E., … Melchinger, A. E. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3: Genes, Genomes, Genetics*, *2*, 1427–1436. https://doi.org/10.1534/g3.112.003699.

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

Zhao, Y., Zeng, J., Fernando, R., & Reif, J. C. (2013). Genomic prediction of hybrid wheat performance. *Crop Science*, *53*, 802–810. https://doi.org/10.2135/cropsci2012.08.0463.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.