

Strategies for Selecting Crosses Using Genomic Prediction in Two Wheat Breeding Programs

Bettina Lado, Sarah Battenfield, Carlos Guzmán, Martín Quincke, Ravi P. Singh, Susanne Dreisigacker, R. Javier Peña, Allan Fritz, Paula Silva, Jesse Poland,* and Lucía Gutiérrez*

Abstract

The single most important decision in plant breeding programs is the selection of appropriate crosses. The ideal cross would provide superior predicted progeny performance and enough diversity to maintain genetic gain. The aim of this study was to compare the best crosses predicted using combinations of mid-parent value and variance prediction accounting for linkage disequilibrium (V_{LD}) or assuming linkage equilibrium (V_{LE}). After predicting the mean and the variance of each cross, we selected crosses based on mid-parent value, the top 10% of the progeny, and weighted mean and variance within progenies for grain yield, grain protein content, mixing time, and loaf volume in two applied wheat (*Triticum aestivum* L.) breeding programs: Instituto Nacional de Investigación Agropecuaria (INIA) Uruguay and CIMMYT Mexico. Although the variance of the progeny is important to increase the chances of finding superior individuals from transgressive segregation, we observed that the mid-parent values of the crosses drove the genetic gain but the variance of the progeny had a small impact on genetic gain for grain yield. However, the relative importance of the variance of the progeny was larger for quality traits. Overall, the genomic resources and the statistical models are now available to plant breeders to predict both the performance of breeding lines per se as well as the value of progeny from any potential crosses.

Core Ideas

- Cross prediction strategies for grain yield and baking quality traits were compared.
- Crosses for all parent combinations were obtained via genomic prediction models.
- Mid-parent selection was similar to accounting for variance when selecting yield.
- The variance had a larger impact in cross predictions for quality traits.

THE MAIN OBJECTIVE of plant breeding is to increase the yield, productivity, adaptation, and quality of crops while optimizing resource use (Allard 1960). Genetic gain in plant breeding is accomplished through the selection of best genetic combinations between

B. Lado, L. Gutiérrez, Statistics Dep., Facultad de Agronomía, Univ. de la República, Garzón 780, Montevideo 12900, Uruguay; L. Gutiérrez, Dep. of Agronomy, Univ. of Wisconsin–Madison, 1575 Linden Dr, Madison, WI 53706; S. Battenfield, J. Poland, Wheat Genetics Resource Center, Dep. of Plant Pathology, 1712 Claflin Rd., Kansas State Univ., Manhattan, KS 66506; S. Battenfield (current address), AgriPro Wheat, Syngenta, 11783 Ascher Rd. Junction City, KS, 66441; C. Guzmán, R.P. Singh, S. Dreisigacker, R.J. Peña, CIMMYT, El Batán, Mexico, Mexico; A. Fritz, Dep. of Agronomy, Kansas State Univ., Manhattan, KS 66506; M. Quincke, P. Silva, Programa Nacional de Investigación Cultivos de Secano, Instituto Nacional de Investigación Agropecuaria, Est. Exp. La Estanzuela, Colonia 70000, Uruguay. B. Lado and S. Battenfield contributed equally. Received 14 Dec. 2016. Accepted 18 Mar. 2017.
*Corresponding authors (jpoland@ksu.edu; gutierrezzcha@wisc.edu). Assigned to Associate Editor Nicholas Tinker.

Published in Plant Genome
Volume 10. doi: 10.3835/plantgenome2016.12.0128

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

Abbreviations: BL, Bayes Lasso; GS, genomic selection; INIA, Instituto Nacional de Investigación Agropecuaria; RIL, recombinant inbred line; RR-BLUP, ridge regression best linear unbiased predictor; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; V_{LD} , variance accounting for linkage disequilibrium; V_{LE} , variance assuming linkage equilibrium.

genotypes (Fehr, 1987). The most important traits selected in breeding are generally complex quantitative traits that are influenced by many genes and the environment (Falconer and Mackay, 1996; Bernardo, 2010). Historically, strategies for selecting superior individuals have relied on phenotypic evaluation of individuals and pedigree information. However, at the end of the twentieth century, marker-assisted selection strategies were sought. To this end, quantitative trait loci (QTL) were detected to identify genes involved in the expression of complex traits through biparental QTL mapping (Lander and Botstein, 1989; Lande and Thompson, 1990), or genome-wide association studies (Jannink et al., 2001). These genetic markers were then used in marker-assisted selection strategies (Dekkers and Hospital, 2002). However, not all of the segregating QTL are identified in a given study and the ones that are identified are largely overestimated (Beavis, 1994). Therefore, methodologies for QTL identification have been insufficient to capture the effects of the many genes involved in common complex traits (Manolio et al., 2009) and the actual use of marker-assisted selection for identified QTL has been very limited in breeding programs (Bernardo, 2008).

Genomic prediction (Meuwissen et al., 2001) provides an alternative method of using genetic information in breeding decisions. Rather than using individual QTL information, genomic prediction involves the prediction of genotypic performance accounting for all genome-wide marker effects simultaneously (Meuwissen et al., 2001). When these predictions are used as a selection strategy, it is called genomic selection (GS). Genomic selection has been extensively researched in plants (reviewed in Heslot et al., 2015) while the implementation in plant breeding programs has been evaluated through optimization of the training population (Asoro et al., 2011; Zhao et al., 2013; Liu et al., 2015; Isidro et al., 2015) and genotype \times environment interaction (Burgueño et al., 2012; Heslot et al., 2014; Jarquín et al., 2014; Lado et al., 2016). Additionally, strategies of GS to select the best crosses by modeling the variance of a cross have been proposed but no study has assessed the consequences in cross selection (Bernardo, 2014; Mohammadi et al., 2015; Tiede et al., 2015).

Crossing is one of the main decisions of a breeding program. The number of possible crosses is orders of magnitude larger than the number of feasible crosses (Witcombe et al., 2013). Although breeders attempt to take into consideration all available information on the potential parents to determine which crosses to make, many crosses are discarded in subsequent years, as they do not deliver superior progeny (Koebner and Summers, 2003; Heslot et al., 2015). Methods to select optimum crosses accounting for the performance per se of each line (e.g., yield, adaptability, quality) were initially proposed on the basis of the progeny's mean and variance. The variance was predicted using morphological data, pedigree information, a few molecular markers, or a combination of these (Souza and Sorrells, 1991; Melchinger et al.,

1998; Utz et al., 2001; Bertan et al., 2007; Casassola and Brammer, 2013). These methods were then extended to genome-wide marker coverage (Endelman, 2011; Bernardo, 2014; Mohammadi et al., 2015; Tiede et al., 2015). The progeny mean was well-predicted from the mid-parent value (Bernardo, 2014). However, genetic variances were difficult to predict using either phenotypic, pedigree, or genetic distances (Souza and Sorrells, 1991; Melchinger et al., 1998; Utz et al., 2001; Hung et al., 2012) but could be adequately modeled by simulating progeny by using genomic prediction (Bernardo, 2014; Mohammadi et al., 2015; Tiede et al., 2015). Endelman (2011) proposed a strategy of estimating the mean and variance for the progeny of a cross using genomic data, where the progeny mean was estimated as the mid-parent value of the estimated parental performance. In this method, the variance was estimated from the marker effects, assuming linkage equilibrium (V_{LE}) between markers. Another strategy used to estimate progeny variance with large genomic data is to simulate the progeny, accounting for the true linkage structure and recombination, and then to predict the performance of the progeny (Bernardo, 2014; Mohammadi et al., 2015; Tiede et al., 2015).

The aim of this research was to evaluate the selection of crosses based on progeny variance, predicted either by assuming linkage equilibrium (V_{LE}) or by accounting for linkage disequilibrium (V_{LD}) between the parents. Specifically, we evaluated the impact of cross selection based only on mid-parent performance or mid-parent performance in a combination with progeny variance, where the variance was estimated either by assuming V_{LE} or V_{LD} . We analyzed data for grain yield in the INIA spring bread wheat breeding program in Uruguay, and grain protein content, mixograph optimal mixing time, and loaf volume quality traits from the CIMMYT spring bread wheat breeding program in Mexico. We evaluated multiple traits and breeding programs to limit possible trait or program bias in the prediction methodology results.

Materials and Methods

Plant Material

INIA

A total of 1465 spring bread wheat lines from the INIA Wheat Breeding Program were used as the training population for GS for grain yield. The INIA lines consisted of all the lines from the preliminary yield trials (F_7) from 2010, 2011, and 2013, as well as the advanced (F_8) and elite (F_9) yield trials from 2010.

Grain yield evaluations were conducted in 35 environments in Uruguay including five locations evaluated in 4 yr and one location with four sowing dates. Locations used to evaluate the genotypes were Dolores (33°50'S, 58°14'W; 15 m asl), Durazno (33°33'S, 56°31'W; 91 m asl), La Estanzuela (34°20'S, 57°42'W; 81 m asl), Young (32°76'S, 57°57'W; 85 m asl), and Ruta2 (33°45'S, 57°90'W; 95 m asl). Four sowing dates were evaluated in

La Estanzuela (LE1, LE2, LE3, and LE4). The evaluation years were 2010 to 2014. The trials from 2012 and most of the locations of 2014 were not included in this study because of a strong genotype by environment interaction (see experimental details in Lado et al., 2016).

CIMMYT

In total, 6095 spring bread wheat lines from the CIMMYT Global Bread Wheat Breeding Program were evaluated for quality traits (Battenfield et al., 2016). The CIMMYT lines were materials in the preliminary yield trials between 2009 and 2015. All lines were grown in Ciudad Obregon, Sonora, Mexico (27°29'N, 109°56'W; 40 m asl). Three phenotypes were chosen to represent different stages of testing wheat as grain, dough, and final product. Grain protein content was determined by near-infrared spectroscopy, using NIR Systems 6500 (Foss, Denmark) using AACC International (2000a) Method 39–10, and reported at 12.5% moisture basis. Grain samples were tempered and milled using Brabender Quadrumat Senior (C. W. Brabender OHG, Germany). Mixograph optimal mixing time was assessed using the Swanson and Working Mixograph (National MFG Co., National Manufacturing Company, Lincoln, NE) according to AACC International Method 54–40A (AACC International, 2000b). Loaf volume was assessed on pup loaves baked according to AACC International (2000c) Method 10–09. Loaf volume and mixing time tests used the adjustment for optimal water absorption described in Guzmán et al. (2015).

Genotyping

Tissue was collected and the CTAB method (Saghai-Maroo et al., 1984) was used to isolate DNA for the genotyping-by-sequencing protocol as described in Poland et al. (2012a). The TASSEL-GBS pipeline (Glaubitz et al., 2014) was run with modifications for nonreference genomes in the INIA material (Poland et al., 2012b). The TASSEL Version 5.2 with alignment to the *T. aestivum* International Wheat Genome Sequencing Consortium genome assembly version 2.25 (Mayer et al., 2014) using Bowtie2 version 2.24 (Langmead and Salzberg, 2012) was used in the CIMMYT lines. Genotyping-by-sequencing tags with single nucleotide polymorphisms (SNPs) were mapped against the POPSEQ Synthetic W7984 by Opatá M85 reference to determine the map position (Chapman et al., 2015). Single nucleotide polymorphisms were filtered by setting maximum missing values of 20% for the INIA and 50% for the CIMMYT populations. Different thresholds for missing data were used to keep the balance between coverage and good SNP quality at each population. Low tolerance for missing values was used to increase the reliability of the imputed SNPs. Individuals with more than 50% missing information were discarded. The final number of individuals were 1465 and 5984 in the INIA and CIMMYT populations, respectively. SNP imputation was conducted using multivariate normal expectation maximization

imputation (Endelman, 2011; Poland et al., 2012a). This method assumes that marker genotypes follow a multivariate normal distribution. The imputation was done through multiple regression starting with the mean of the marker and was then updated using maximum likelihood to estimate the mean and covariance between lines with nonmissing data.

Phenotypic Data Analysis

Phenotypic best linear unbiased estimations were obtained for the grain yield of all genotypes present in each trial. Field analysis was conducted according to an experimental design that consisted of a series of smaller α -design trials connected through common checks. The following model was used to estimate INIA grain yield genotypic means:

$$y_{ijklm} = \mu + g_i + e_j + t_{k(j)} + r_{l(jk)} + b_{m(ljk)} + \varepsilon_{ijklm}, \quad [1]$$

where μ is the overall mean, g_i is the fixed effect of the i^{th} genotype; e_j is the fixed effect of j^{th} year–location combination; $t_{k(j)}$ is the random effect of the k^{th} trial nested within the j^{th} year–location; $r_{l(jk)}$ is the random effect of the l^{th} replicate nested within the k^{th} trial and j^{th} year–location; $b_{m(ljk)}$ is the random effect of the m^{th} incomplete block nested within the l^{th} replicate, k^{th} trial, and the j^{th} year–location; and ε_{ijklm} is the residual error for the i^{th} genotype in the m^{th} block within the l^{th} replicate in the k^{th} trial located in the j^{th} year–location; and $t_{k(j)}$, $r_{l(jk)}$, $b_{m(ljk)}$, and ε_{ijklm} are random variables: $t_{k(j)} \sim N(0, \sigma_t^2)$, $r_{l(jk)} \sim N(0, \sigma_r^2)$, $b_{m(ljk)} \sim N(0, \sigma_b^2)$, and $\varepsilon_{ijklm} \sim N(0, \sigma_e^2)$. The best linear unbiased estimations were estimated with the nlme package (Pinheiro et al., 2013) in R statistical software (R Development Core Team, 2016). Heritability was estimated via the model shown in Eq. [1] but with genotypes as the random effect, being $g_i \sim N(0, G\sigma_g^2)$. G is the genotypic variance–covariance matrix between lines estimated by using markers, as was proposed by VanRaden (2008), and those estimated by using the A.mat function in the rrBLUP package (Endelman, 2011). Genotypic variance (σ_g^2) and the average pairwise variance error of the best linear unbiased predictor (\bar{v}_{BLUP}) were estimated using the sommer R package (Covarrubias-Pazaran, 2016). Heritability for grain yield was estimated following Cullis et al. (2006):

$$H^2 = 1 - (\bar{v}_{BLUP} / 2\sigma_g^2). \quad [2]$$

For the CIMMYT population, quality data were not replicated within or across years and no statistical design was used within years, since only high-yielding selections from the yield trial were advanced to quality testing. The heritability of the quality traits was estimated as in Battenfield et al. (2016) for CIMMYT data from the ridge regression best linear unbiased predictor (RR-BLUP) model, which only models additive effects. Briefly,

$$H^2 = (\sigma_p^2 - \sigma_e^2) / \sigma_p^2, \quad [3]$$

where σ_p^2 is the phenotypic variance of the trait in the training population and σ_e^2 is the error variance of the RR-BLUP model. The σ_e^2 is assumed to include genotypic variance caused by dominance and epistatic effects, as well as any nongenetic errors. This leaves only additive genetic variance when σ_e^2 is subtracted from σ_p^2 .

Genomic Prediction Models

Two genomic prediction models were compared in order to estimate marker effects and the performance of each parent. The ridge regression model was evaluated via the rrBLUP package (Endelman, 2011) and the Bayesian Lasso (BL) model was calculated via the BGLR package (Pérez and de los Campos, 2010), both in R software (R Development Core Team, 2016). Genomic prediction models incorporated the genotypic matrix \mathbf{X} codified as $(-1, 0, 1)$, where +1 indicates that loci were homozygous to the most frequent allele at a given locus, -1 indicates that loci were homozygous to the least frequent allele at given locus, and 0 indicates heterozygous loci. The missing SNPs were imputed as fractions between -1 and 1, and then were rounded to the nearest whole number (to 1 if the SNP value was ≥ 0.5 , -1 if the SNP value was less than or below -0.5, and 0 if the SNP value was between -0.5 and 0.5). For BL, initial arbitrary values of trait heritability and error variance were used, then values were optimized via prediction accuracy to calculate the hyperparameter rate and shape from the double exponential distribution. Next, the marker effects and parent performance values were estimated from the posterior distribution. For the ridge regression model, marker effects were estimated by using mixed models with markers considered as random with $N(0, I\sigma_{\text{mrk}}^2)$. After that, to predict the performance of the parents, mixed models with lines (g_i) as random effects were used with $N(0, G\sigma_g^2)$, where G is the variance-covariance matrix estimated as in VanRaden (2008) via the A.mat function in the rrBLUP package (Endelman, 2011). The models were selected because of their use of different statistical approaches (mixed models and Bayesian, respectively) and their different assumptions about marker effects on predictions.

Random cross-validation was conducted by using 60% of the lines to train the model, predicting the remaining 40% of entries for both the INIA and CIMMYT datasets separately. Each model was iterated 100 times and then, the model with the best predictive ability was used for subsequent analysis. All genotyped parents were included in the training population to estimate the marker effects and the parents' performance for grain yield ($n = 1465$), grain protein content, mixing time, and loaf volume ($n = 5984$), via the model with the best predictive ability. However, as there were no differences in the accuracy of predictions between the models, the RR-BLUP model was used to estimate marker effects for all traits for the remaining analyses.

Cross Selection Based on the Performance of Simulated Progeny

Progeny of all possible pairwise cross combinations were simulated and their performance was predicted for grain yield in the INIA population. However, in the CIMMYT population, only the lines evaluated in 2015 ($n = 1425$) were used as parents for progeny simulation and prediction. Reciprocal crosses were not predicted, as negligible maternal effects were assumed to be present (McNeal et al., 1968; Jumbo and Carena, 2008).

The V_{LE} of the progeny within each cross ($V_{LE(i)}$) was predicted, taking into account allelic differences between the parents and markers effects for each trait, following Endelman (2011):

$$V_{LE(i)} = \sum_{k=1}^{k=M} (1 - (p_{k+} - p_{k-})^2) \times \hat{u}[k]^2, \quad [4]$$

where p_{k+} is the frequency of the k^{th} SNP of biallele +1 for parents of progeny in i^{th} cross and p_{k-} is the same for the biallele -1, and \mathbf{u} is the vector of marker effects with a length M (the number of SNPs). Marker effects were estimated via RR-BLUP. For example, if both parents had the allele +1, the frequencies were $p_{k+} = 1$ and $p_{k-} = 0$; if one parent had the allele +1 and the other had -1, the frequencies were $p_{k+} = 0.5$ and $p_{k-} = 0.5$; if both parents had the alleles 0, $p_{k+} = 0$ and $p_{k-} = 0$, and so on. Next, the distribution of predicted performance values for the population of progeny in each cross was simulated, assuming a normal distribution for the progeny. The parameters used in the normal distribution were the mean (equal to the mid-parent value) and the variance (equal to the progeny variance estimated as in Eq. [4]). The mean of the top 10% of the progeny was estimated for each cross.

To predict V_{LD} , we used the PopVar package described in Mohammadi et al. (2015) and implemented in R software (R Development Core Team, 2016). Within PopVar, recombinant inbred lines (RILs) were simulated via the Rqtl package (Broman et al., 2003), which requires ordered markers with genetic distances to account for recombination. The function simulates recombination points along the chromosome by using independent crossovers, following the Stahl model (Stahl, 1979). Using this package, we simulated 1000 RILs for each cross and then estimated the predicted performance of each individual simulated RIL. The mean performance of the 1000 RILs was used to estimate the progeny mean performance of each cross. The variance of the predicted performance of the 1000 RILs was used to estimate the progeny variance performance of each cross. The mean of the top 10% of the progeny was estimated for each cross.

Finally, the best 1000 crosses were selected for grain yield, grain protein, and loaf volume on the basis of either their mid-parent performance value or the predicted performance of the top 10% of their progeny, assuming V_{LD} or V_{LE} . For mixing time, where an intermediate optimal time is desired, the best crosses were selected on the basis of having the lowest variance within a window of targeted predicted progeny performance values (i.e., 2.5–4.0 min.).

Weighting Progeny Variance in Cross Selection

Five values of performance were established as thresholds to illustrate the trade-offs of selection on the basis of maximizing the expected mean or accounting for the variance. Therefore, five groups with different weights on the relative importance of the variance were established. The maximum parental predicted performance minus 20, 40, 60, 80 or 100% of the difference between the maximum and the mean predicted performance was used to define the groups G1 (20%), G2 (40%), G3 (60%), G4 (80%), and G5 (100%). One hundred crosses were selected within each group, taking into account the mean predicted progeny performance above the threshold and the maximum variance for three traits (mixing time was excluded from this analysis because the selection criteria for this trait are different). Finally, the mean of all progenies and the mean of the top 10% of the progeny for the 100 selected progenies were calculated.

Results and Discussion

Genotype Data

For INIA data, a total of 81,044 SNPs were identified after the TASSEL pipeline was run and duplicate markers were removed. Of these, 23,771 were mapped using the POPSEQ genetic framework. Finally, we kept 3884 SNPs with less than 20% missing data and 1465 individuals with less than 50% missing data. For CIMMYT data, a total of 31,166 SNPs were identified after the TASSEL pipeline was run and duplicate markers were removed. Of these, 8443 SNPs were mapped using POPSEQ. Finally, we kept 1164 SNPs with less than 50% missing data and 5984 individuals with less than 50% missing data. The SNPs were distributed in all chromosomes with a median of 0 and 0 cM and maximum gap of 61 cM and 94 cM, for the INIA and CIMMYT genetic maps, respectively (Supplemental Fig. S1).

Phenotype Data

Mean grain yield for the INIA dataset was 5160 kg ha⁻¹ (SE = 853) and the heritability was 0.67. The means of the three end-use quality traits from CIMMYT were 11.97% (SE = 0.77) for grain protein content, 3.13 min (SE = 0.80) for mixing time, and 802 cm³ (SE = 59) for loaf volume. The heritability of the CIMMYT quality traits were 0.67, 0.71, and 0.68 for grain protein content, mixing time, and loaf volume, respectively.

Genomic Selection Models

High correlations between observed and predicted breeding values were found for grain yield in the INIA dataset ($r_{(y,\hat{y})} = 0.42$) and for grain protein content ($r_{(y,\hat{y})} = 0.63$), mixing time ($r_{(y,\hat{y})} = 0.64$), and loaf volume ($r_{(y,\hat{y})} = 0.59$) in the CIMMYT dataset. The predictive ability using both models, BL and RR-BLUP, were similar. As expected, the correlation of marker effects between models decreased at the extreme values for each trait (Supplemental Fig. S2). The BL method induces less shrinkage

than RR-BLUP in marker effects estimation as it sets some of the marker effects to zero and thus BL estimates fewer but larger effects than RR-BLUP (Park and Casella, 2008). Mixing time shows the largest shrinkage effect for RR-BLUP (Supplemental Fig. S2). This was expected, because mixing time is influenced by a few genes with a large effect (Payne et al., 1987). However, when the marker effect values were combined to predict performance of the parents, there was no difference between GS models (Supplemental Fig. S2). It has been documented elsewhere that different GS models had similar prediction ability in plants (Heslot et al., 2012). Thus because of the similarity in accuracies between RR-BLUP and BL predictions, all further analyses were conducted using RR-BLUP model.

Progeny Prediction

The predicted mean performance of the progeny was perfectly correlated between the mid-parent value and mean of the simulated progenies for all traits (data not shown). This result was expected because in an additive model, the expected value of the mean of the progeny calculated as the mid-parent value is the same as the mean of the RILs' performance. The same results were found in maize (*Zea mays* L.) for silking date and protein (Bernardo, 2014) and in barley (*Hordeum vulgare* L.) for yield and deoxynivalenol content (Mohammadi et al., 2015).

The correlation between V_{LD} and V_{LE} was different for each trait and was probably associated with trait complexity (Fig. 1). The order of the variance correlation between estimations was grain yield < grain protein content < loaf volume < mixing time. Variance predictions of more complex traits are more inconsistent (Bernardo, 2008) and thus modeling variances for these traits has less impact in progeny predictions. This is probably because the error in the estimates of marker effects has a stronger influence on the variance than on the mean predictions of a cross (Zhong and Jannink, 2007). The higher variances correlation for mixing time found in our study can be explained by the fact that mixing time is a trait that is influenced by a few major genes (Payne et al., 1987; Simons et al., 2012).

The variance of predicted performance from 1000 progeny was larger for V_{LD} than for V_{LE} (Fig. 2 and Fig. 3). A potential explanation for this is as follows. The genetic variance for each pair of markers is the sum of the variance of each marker, plus twice the covariance between the two markers (Falconer and Mackay, 1996). When we assume linkage equilibrium, the markers were independent and therefore, the covariance between the markers was zero. If markers are in the coupling phase, the covariance is positive and therefore V_{LE} is smaller than V_{LD} (Lynch and Walsh, 1998; Bernardo, 2010). However, if the markers are in the repulsion phase, the covariance between markers becomes negative and therefore, V_{LE} is larger than V_{LD} (Lynch and Walsh, 1998; Bernardo, 2010). Hence, the relative magnitude of the variances (i.e., V_{LE} and V_{LD}) will depend on the marker

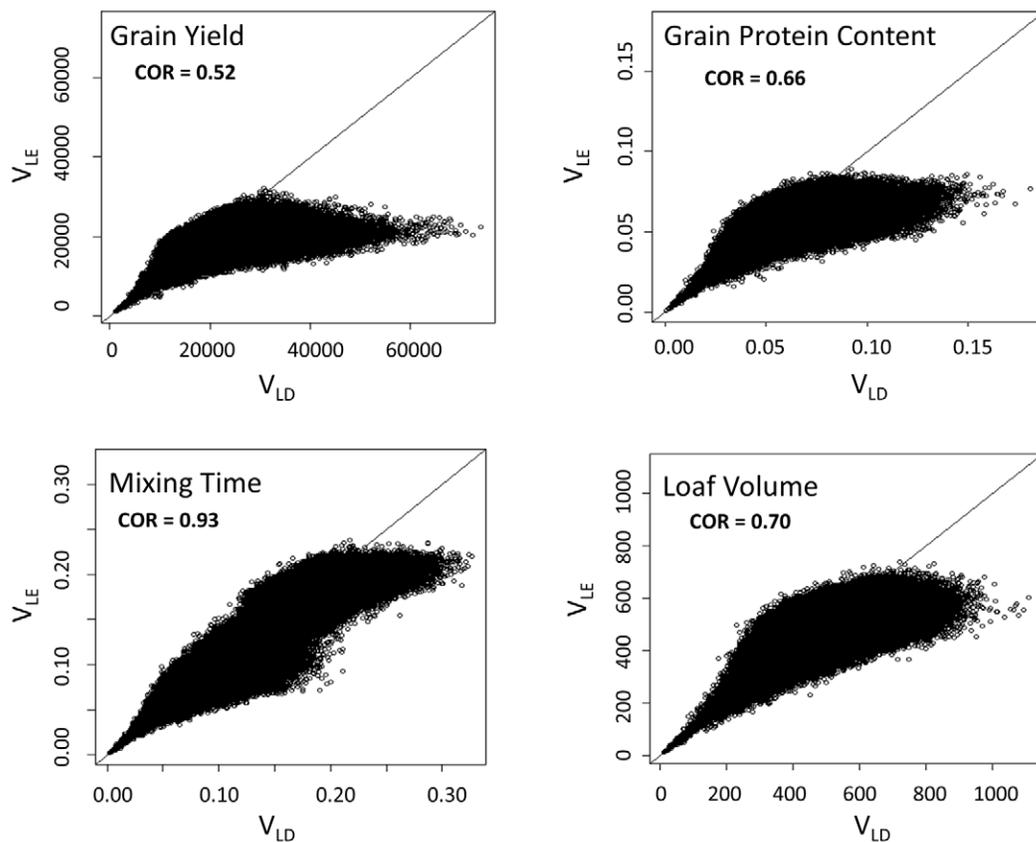


Fig. 1. Correlation between the variance of predicted progeny performance accounting for linkage disequilibrium (V_{LD}) or assuming linkage equilibrium (V_{LE}), for wheat grain yield, grain protein content, mixing time, and loaf volume.

phasing. In our case, we believe we have coupling phases in several situations. For example, one of the crosses that exhibits the largest differences in variance estimated by V_{LD} and V_{LE} showed a predominance of coupling phases for most chromosomes (Fig. 4).

We found a triangular relationship between variance and the mean of predicted progeny performance for the crosses in all traits when modeling V_{LD} , but not when modeling V_{LE} (Fig. 2 and Fig. 3). This relationship was stronger for grain yield than for the other traits (Fig. 2 and Fig. 3). This triangular relationship is generated when lines with similar trait values at the end of their distribution (i.e., low \times low or high \times high) are crossed, resulting in progeny with extreme values for mean and low variance (Bernardo, 2014). On the other hand, when lines from opposite extremes are crossed (i.e., low \times high), the progeny have intermediate means with higher variance. This is supported by previous findings by Bernardo (2014) and Mohammadi et al. (2015). Furthermore, as in Bernardo (2014), we found a linear relationship between the mean and variance of crosses from the same parent prediction when using V_{LD} (Fig. 4), which caused the triangular shape. Recombination between extreme parents would account for differences arising from coupling and repulsion linkage. These differences may not have an immediate impact but could become relevant for longer-term selection (Bernardo, 2014).

Cross Selection

When selecting crosses based on mid-parent values or on the basis of the mean of the top 10% of the progeny via two variance estimates (V_{LD} and V_{LE}), we observed that many of the selected crosses and parents were the same for all three selection methods (Fig. 2 and Fig. 3). Of the 1000 best crosses based on the two variance methods, there were 879 crosses in common for grain yield, 757 for grain protein content, 769 for mixing time, and 738 for loaf volume (Fig. 2B,E and Fig. 3B,E). In addition, 837 of the common crosses selected for grain yield were the same ones with the highest mid-parent values. This was also confirmed by the high correlation ($r = 0.94$, $p < 0.001$) between total progeny mean and the mean of the top 10% of the predicted progeny performance for the 1000 best crosses (selected by top 10%). These results indicate that the predicted mean progeny performance is the strongest driver for selecting superior crosses for grain yield. Thus regardless of the approach for predicting progeny variance, the crosses with the highest mid-parent values also provided the best top 10% of the progeny. This was largely independent of the genetic distance between parents. Mid-parent strategies are superior for selecting crosses than a combination of mid-parent and progeny variance when there is more variation in the progeny means than in the progeny variance (Zhong and Jannink, 2007).

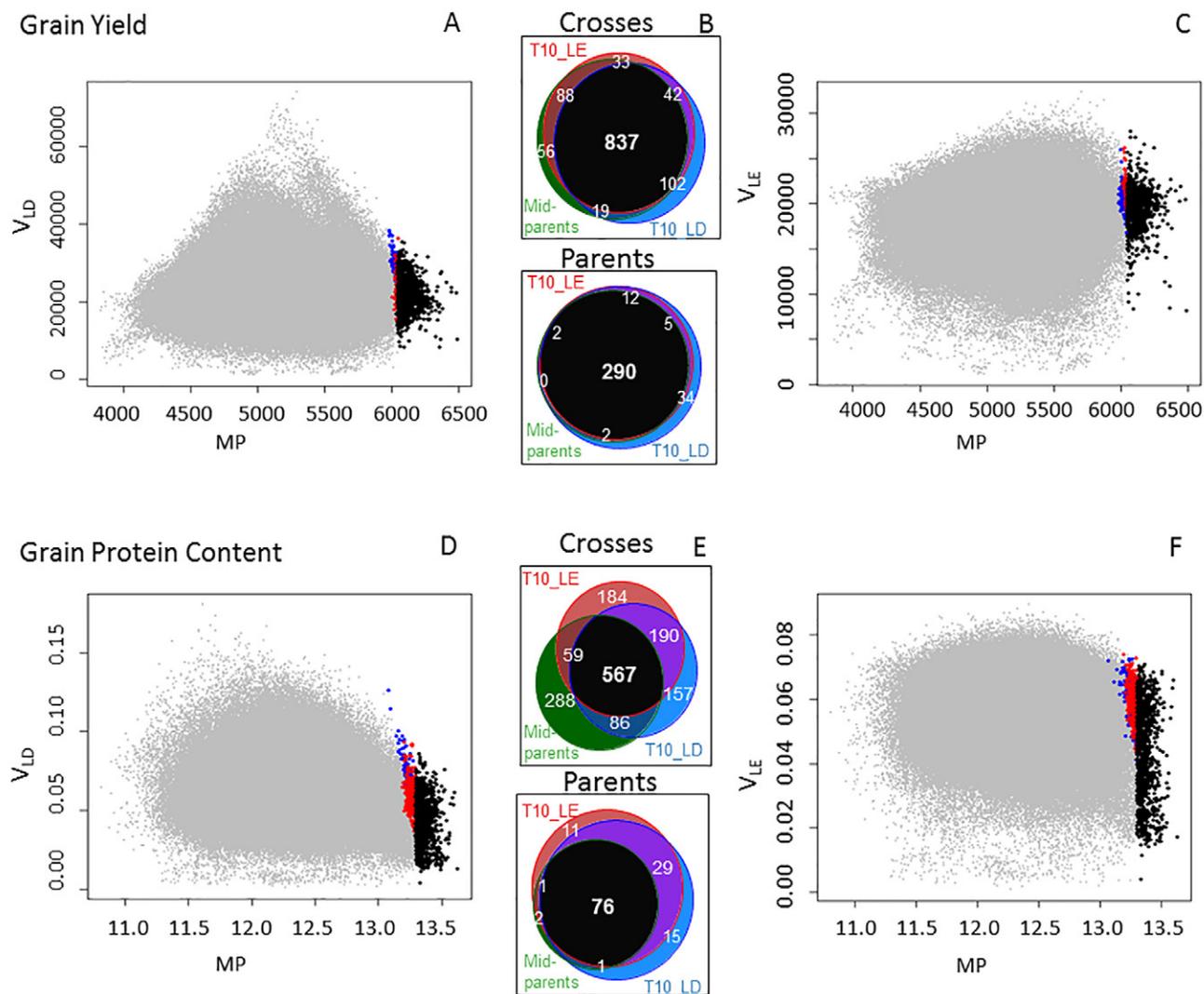


Fig. 2. Cross performance characterization. Expected mid-parent performance (MP) versus the predicted progeny variance accounting for linkage disequilibrium (V_{LD}) for yield (A) and grain protein content (D). Expected MP versus the predicted progeny variance assuming linkage equilibrium (V_{LE}) for yield (C) and grain protein content (F). The top 1000 crosses selected by MP (green), V_{LD} (blue), V_{LE} (red), and crosses in common among the two groups (black) are highlighted in each plot (A, C, D, F). Venn diagrams represent the 1000 crosses; the parents of those crosses selected by using the MP (green), V_{LD} (blue), V_{LE} (red); or common crosses among the three groups (black) for grain yield (B) and grain protein content (E).

For quality traits, the percentage of crosses in common between the three cross prediction strategies were 56.7% for grain protein content and 59.8% loaf volume (Fig. 2 and Fig. 3). In addition, the correlations between the total progeny mean and the mean of the top 10% of the predicted progeny performance for the best 1000 crosses (selected by top 10%) were, 0.67 ($p < 0.001$) for grain protein content and 0.80 ($p < 0.001$) for loaf volume. These results indicate that for quality traits, the influence of variance was a more relevant factor in determining cross selection. The results found by Bernardo (2014) for silking date and protein concentration in maize are more similar to those found for wheat quality traits in which crosses with an undesirable mean value but high variance could reach the same mean as the top 10% of the progeny.

Accounting for more variance (i.e., by having lower thresholds) decreased the number of shared crosses

between V_{LD} and V_{LE} (Fig. 5, left and right plots). Additionally, the mean of the top 10% of the progeny increased as the threshold value increased, regardless of cross variance (Fig. 5, middle plots). This effect was more pronounced for grain yield, which had the highest ratio between the variance of mean predicted progeny performance values and the variance of progeny SD among crosses (Table 1).

Although the differences between the variance predicted by the strategies considering V_{LD} or V_{LE} are high for some traits, the parents and the crosses selected on the basis of the top 10% performance of the progeny are similar. Therefore, the use of V_{LE} as a substitute for V_{LD} is a reasonable compromise if computation time or the lack of a linkage map are important considerations. The use of V_{LE} could be favored when one considers only one generation of improvement. For grain yield, where all three

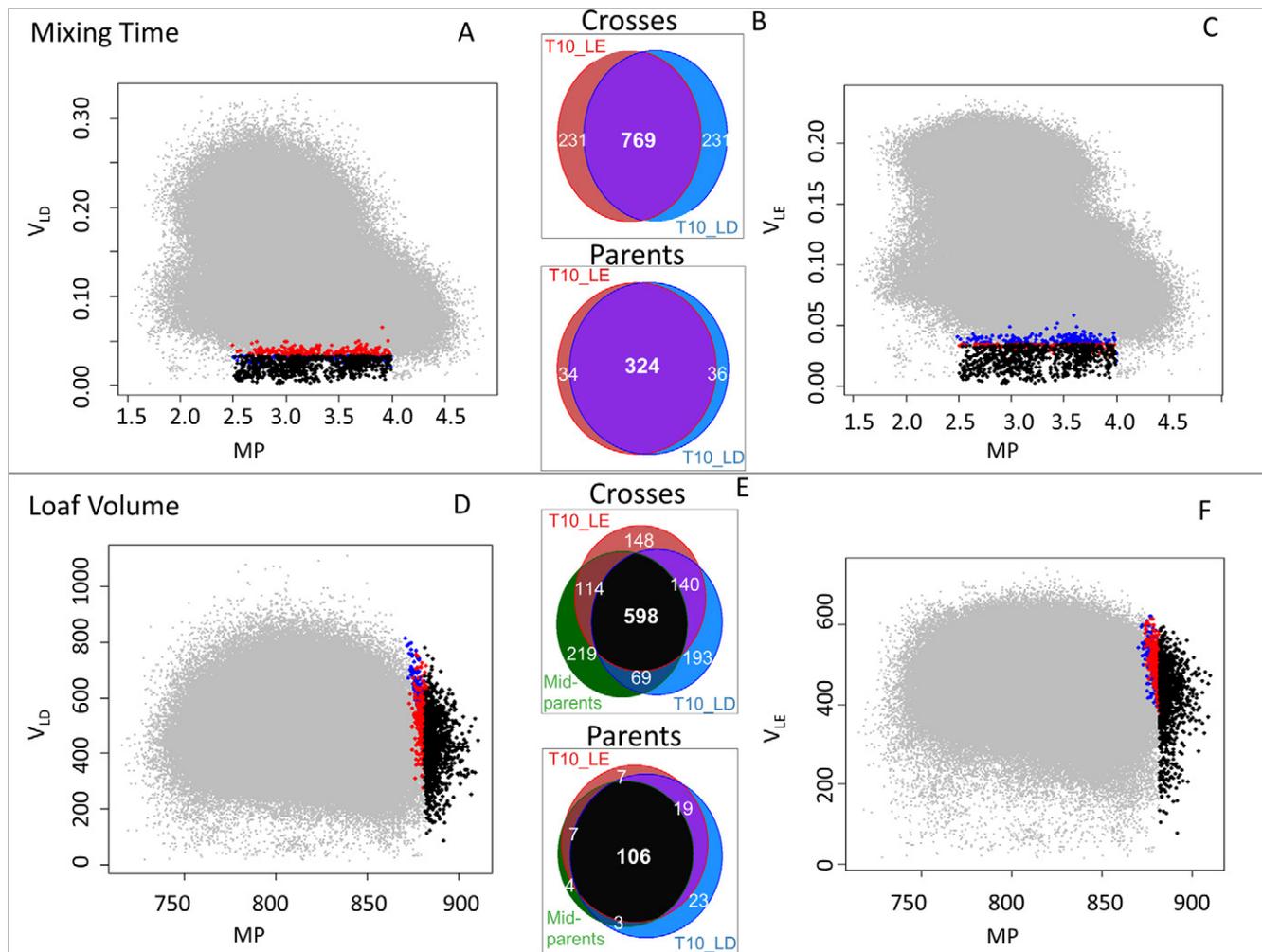


Fig. 3. Cross performance characterization. Expected mid-parent performance (MP) versus the predicted progeny variance accounting for linkage disequilibrium (V_{LD}) for mixing time (A) and loaf volume (D). Expected MP versus the predicted progeny variance assuming linkage equilibrium (V_{LE}) for mixing time (C) and loaf volume (F). The top 1000 crosses selected by using MP (green), V_{LD} (blue), V_{LE} (red), and crosses in common among the two groups (black) are highlighted in each plot (A, C, D, F). Venn diagrams represent the 1000 crosses or the parents of those crosses selected only by using V_{LD} (blue), V_{LE} (red) for mixing time (B), and MP (green), alongside common crosses among the three groups (black) for loaf volume (E).

methods gave similar results, the use of the mid-parent value would be the simplest method of cross prediction.

On the basis of our results, the best strategy of maintaining diversity is not clear. Fu (2015) discussed the difficulty of measuring and accounting for the impacts of plant breeding on crop genetic diversity. Jacobson et al. (2015) and Rutkoski et al. (2015) found that the losses of genetic diversity incurred when using genome-wide selection is higher than that for phenotypic selection, although there are situations in which some losses of diversity are necessary to obtain increased genetic gains. However, to sustain genetic gain in wheat breeding programs, genetic diversity is essential because it allows effective recombination, which, in turn, creates more diversity (Fehr, 1987). This diversity is fundamental when new challenges are encountered, such as changing races in biotic stressors, or changing environmental factors and abiotic stressors (Dreisigacker et al., 2005; Borlaug, 2007; Babiker et al.,

2015). Thus, Fu (2015) emphasized the importance of maintaining allele variability to have the possibility to improve under unfavorable conditions.

Conclusions

Mid-parent value has historically been used for cross selection in plant breeding. Here, progeny mean was estimated using genomic predicted mid-parent value, which was identical to the mean of the simulated progeny. Modeling the variance to predict the best crosses had a small impact for complex traits like grain yield. However, for quality traits, the influence of variance was a large factor in determining cross selection. Differences in selection assuming either V_{LD} or V_{LE} were small for all traits and the assumption of V_{LE} was less computationally intensive. Although selection of parental combinations based on the expected progeny variance for yield

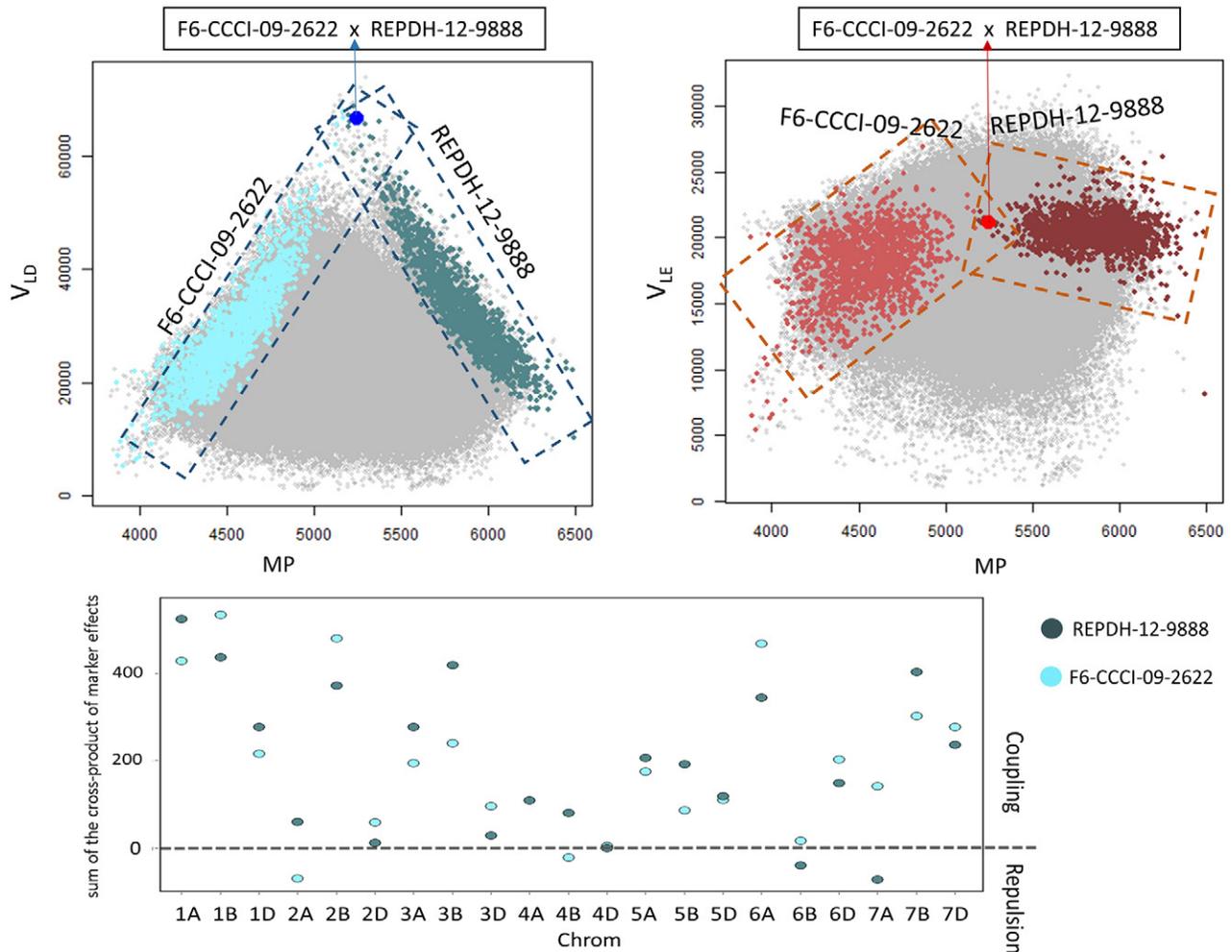


Fig. 4. Relationship between mid-parent performance (MP) and variance for all possible crosses predicted considering linkage disequilibrium (V_{LD} ; left) or linkage equilibrium (V_{LE} ; right). An extreme cross between a high-performing parent and a low-performing parent is highlighted, as well as the progeny of each of those parents (i.e., dark colors for the high-performing parent and light colors for the low-performing parent). The bottom panel of the figure represents the average allelic phasing for each chromosome calculated for adjacent close markers.

had little impact on improving the overall cross performance, maintaining genetic diversity should also be considered. In reality, this may be accomplished through the choice of parents that achieve high average yield through slightly different strategies or yield components.

Supplemental Information

Supplemental Fig. S1. Single nucleotide polymorphism (SNP) map with marker distribution across the wheat genome in the INIA (on the left) and CIMMYT (on the right) populations. Single nucleotide polymorphisms were mapped against the POPSEQ Synthetic W7984 by Opata M85 reference.

Supplemental Fig. S2. RR-BLUP vs BL scatterplots for marker effects (upper row), mid-parent value (MP, middle row), and progeny variance assuming linkage equilibrium (V_{LE} , bottom row) for grain yield, grain protein content, mixing time and loaf volume (traits by column).

Conflict of Interest Disclosure

The authors declare no conflict of interest.

Acknowledgments

Support for grain yield phenotyping was provided by INIA. We express our appreciation for the effort of the technical personnel of the INIA wheat breeding program: Leonardo Hernández, María Ferreira, Dumas Laun, Abel López, and José Flores. Support for phenotyping of the CIMMYT quality traits was provided by CGIAR Comite Regional Permanente (Permanent Regional Committee) WHEAT, Durable Rust Resistance Project, and Fondo Sectorial Secretaria de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (Secretary of Agriculture, Livestock, Rural Development, Fishery and Food, of Mexico) Consejo Nacional de Ciencia y Tecnología (National Council on Science and Technology, of Mexico) (No. 146788: “Sistema de mejoramiento genético para generar variedades resistentes a royas, de alto rendimiento y alta calidad para una producción sustentable en México de trigo”) of the Mexican government. Funding for CIMMYT GS was provided by the US Agency for International Development (USAID) Feed the Future Initiative (USAID Cooperative Agreement No. AID-OAA-A-13-0005) and the Bill & Melinda Gates Foundation through a grant to Cornell University for “Genomic Selection: The next frontier for rapid gains in maize and wheat improvement”. The opinions expressed herein are those of the author(s) and do

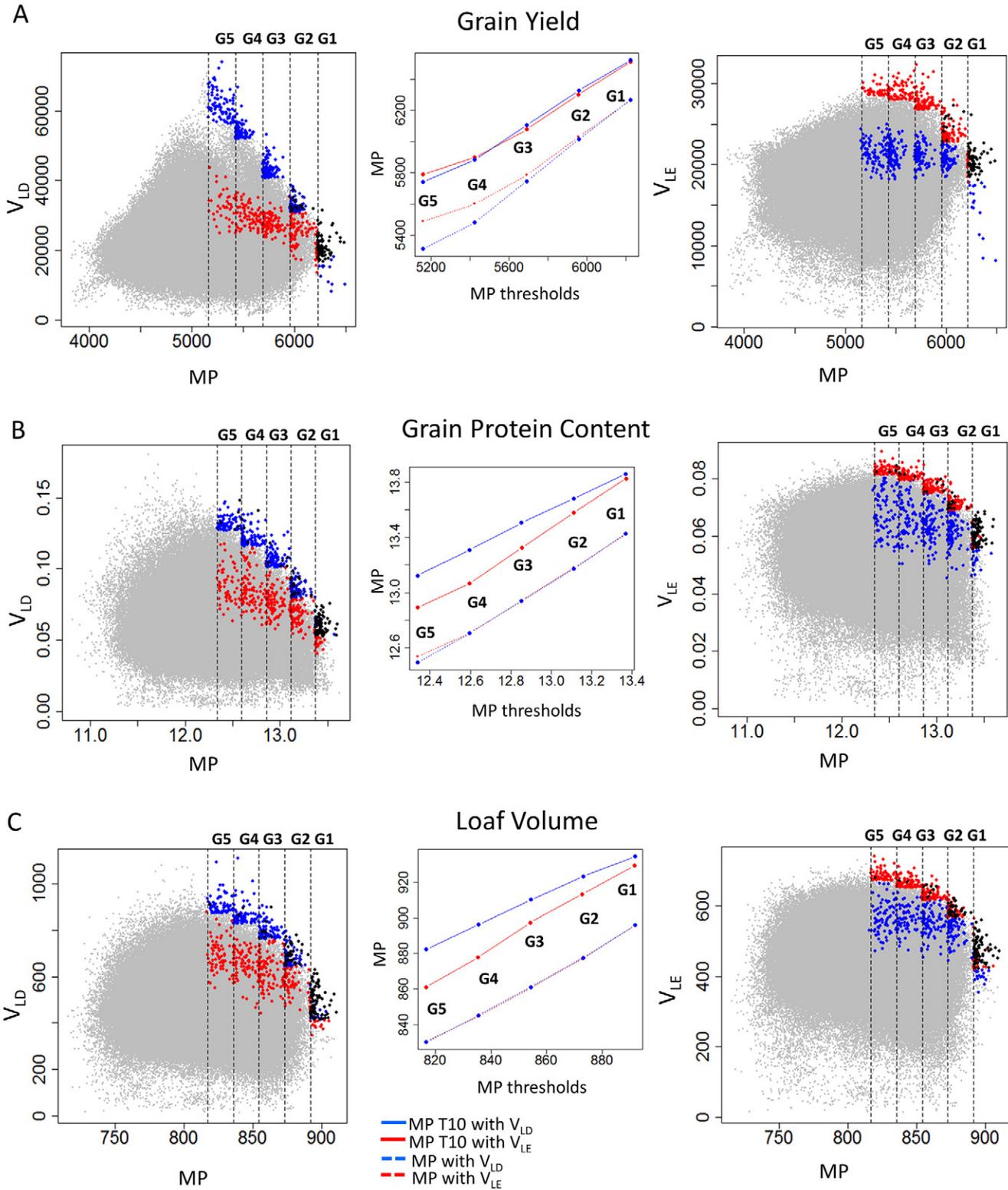


Fig. 5. The best 100 crosses selected using five progeny mean predicted performance thresholds [G1 (20%) > G2 (40%) > G3 (60%) > G4 (80%) > G5(100%)] and maximum variance are indicated in two correlation plots within each trait. Plots on the left are the expected mid-parent performance (MP) versus the variance accounting for linkage disequilibrium (V_{LD}). Plots on the right are the MP versus variance assuming linkage equilibrium (V_{LE}). Blue and red points indicate the best 100 crosses selected by accounting for linkage disequilibrium or assuming linkage equilibrium, respectively. Black points indicate the common crosses between V_{LD} and V_{LE} . The mean predicted progeny performance values of 100% (MP; broken line) or 10% (MP T10; continuous line) of the top performing progeny considering linkage disequilibrium (V_{LD} ; blue) or linkage equilibrium (V_{LE} ; red) for each selection group (G1–G5) are shown in the middle plots.

Table 1. Variance of mean predicted progeny performance ($Var(MP)$) among crosses and variance of the SD of predicted progeny performance ($\sqrt{V_{ID}}$) and the ratio between them [$Var(\sqrt{V_{ID}})/Var(MP)$]. Predictions were obtained by considering linkage disequilibrium for four traits: grain yield, grain protein content, mixing time, and loaf volume.

Trait	$Var(MP)$	$Var(\sqrt{V_{ID}})$	Ratio
Grain yield (kg ha ⁻¹)	90,969	28,400,576	312
Grain protein content (%)	0.056	0.022	2.5
Mixing time (min.)	0.070	0.012	5.8
Loaf volume (cm ³)	0.093	0.011	8.5

not necessarily reflect the views of USAID. Support for the doctoral work of SB was provided by Monsanto's Beachell-Borlaug International Scholar's Program. Support for the doctoral work of BL was provided by Agencia Nacional de Investigación e Innovación-Uruguay, through Grant POS_NAC_2013_1_11261 and by Comisión Sectorial de Investigación Científica-Uruguay, through grants in the program internships abroad.

References

- AACC International. 2000a. AACCI Method 39-10.01. Near-infrared reflectance method for protein determination in small grains. AACC International. <http://methods.aaccnet.org/summaries/39-10-01.aspx> (accessed 10 May 2017).
- AACC International. 2000b. AACCI Method 54-40.02. Physical dough tests. Mixograph method. AACC International. <http://methods.aaccnet.org/summaries/54-40-02.aspx> (accessed 10 May 2017).
- AACC International. 2000c. AACCI Method 10-09.01. Baking quality. Basic straight-dough bread-baking method—long fermentation. AACC International. <http://methods.aaccnet.org/summaries/10-09-01.aspx> (accessed 10 May 2017).
- Allard, R.W. 1960. Principles of plant breeding. John Wiley and Sons Inc., New York.
- Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome* 4(2):132–144. doi:10.3835/plantgenome2011.02.0007
- Babiker, E.M., T.C. Gordon, S. Chao, M. Newcomb, M.N. Rouse, Y. Jin, et al. 2015. Mapping resistance to the Ug99 race group of the stem rust pathogen in a spring wheat landrace. *Theor. Appl. Genet.* 128(4):605–612. doi:10.1007/s00122-015-2456-6
- Battenfield, S.D., C. Guzmán, R.C. Gaynor, R.P. Singh, R.J. Peña, S. Dreisigacker, et al. 2016. Genomic selection for processing and end-use quality traits in the CIMMYT Spring Bread Wheat Breeding Program. *Plant Genome* 9(2):1–12. doi:10.3835/plantgenome2016.01.0005
- Beavis, W.D. 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. In D.B. Wilkinson, editor. Proceedings of the 49th Annual Corn and Sorghum Research Conference. American Seed Trade Association, Washington, DC. p. 250–265
- Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Sci.* 48(5):1649–1664. doi:10.2135/cropsci2008.03.0131
- Bernardo, R. 2010. Breeding for quantitative traits in plants, 2nd ed.: Stemma Press, Woodbury, MN.
- Bernardo, R. 2014. Genomewide selection of parental inbreds: Classes of loci and virtual biparental populations. *Crop Sci.* 54(1):1–33.
- Bertan, I., F.I.F. De Carvalho, and A.A.C. De Oliveira. 2007. Parental selection strategies in plant breeding programs. *J. Crop Sci. Biotechnol.* 10(4):211–222.
- Borlaug, N.E. 2007. Sixty-two years of fighting hunger: Personal recollections. *Euphytica* 157(3):287–297. doi:10.1007/s10681-007-9480-9
- Broman, K.W., H. Wu, S. Sen, and G.A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19(7):889–890. doi:10.1093/bioinformatics/btg112
- Burguño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52(2):707–719. doi:10.2135/cropsci2011.06.0299
- Casassola, A., and S. Brammer. 2013. Parental selection of wheat lines based on phenotypic characterization and genetic diversity. *Crop Breed. Appl. Biotechnol.* (13):49–58. doi:10.1590/S1984-70332013000100006
- Chapman, J.A., M. Mascher, A. Buluç, K. Barry, E. Georganas, A. Session, et al. 2015. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.* 16(1):26. doi:10.1186/s13059-015-0582-8
- Covarrubias-Pazaran, G. 2016. Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS ONE* 11(6):E0156744. doi:10.1371/journal.pone.0156744
- Cullis, B.R., B. Smith, and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* 11(4):381–393. doi:10.1198/108571106X154443
- Dekkers, J.C.M., and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3(1):22–32. doi:10.1038/nrg701
- Dreisigacker, S., P. Zhang, M.L. Warburton, B. Skovmand, D. Hoisington, and E. Melchinger. 2005. Genetic diversity among and within CIMMYT wheat landrace accessions investigated with SSRs and implications for plant genetic resources management. *Crop Sci.* 45(2):653–661. doi:10.2135/cropsci2005.0653
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4(3):250–255. doi:10.3835/plantgenome2011.08.0024
- Falconer, D.S., and T.F. Mackay. 1996. Introduction to quantitative genetics. Prentice Hall, London.
- Fehr, W.R. 1987. Principles of cultivar development. Macmillan Inc., New York.
- Fu, Y.B. 2015. Understanding crop genetic diversity under modern plant breeding. *Theor. Appl. Genet.* 128(11):2131–2142. doi:10.1007/s00122-015-2585-y
- Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9(2):E90346. doi:10.1371/journal.pone.0090346
- Guzmán, C., G. Posadas-Romano, N. Hernández-Espinosa, A. Morales-Dorantes, and R.J. Peña. 2015. A new standard water absorption criteria based on solvent retention capacity (SRC) to determine dough mixing properties, viscoelasticity, and bread-making quality. *J. Cereal Sci.* 66:59–65. doi:10.1016/j.jcs.2015.10.009
- Heslot, N., D. Akdemir, M.E. Sorrells, and J.-L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127(2):463–480. doi:10.1007/s00122-013-2231-5
- Heslot, N., J.-L. Jannink, and M.E. Sorrells. 2015. Perspectives for genomic selection applications and research in plants. *Crop Sci.* 55: 1–12. doi:10.2135/cropsci2014.03.0249
- Heslot, N., H.-P. Yang, M.E. Sorrells, and J.-L. Jannink. 2012. Genomic selection in plant breeding: A comparison of models. *Crop Sci.* 52(1): 146–160. doi:10.2135/cropsci2011.06.0297
- Hung, H.-Y., C. Browne, K. Guill, N. Coles, M. Eller, A. Garcia, et al. 2012. The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. *Heredity* 108(5):490–499. doi:10.1038/hdy.2011.103
- Isidro, J., J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M.E. Sorrells. 2015. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* 128:145–158. doi:10.1007/s00122-014-2418-4

- Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2015. Minimal loss of genetic diversity after genomewide selection within biparental maize populations. *Crop Sci.* 55(2):783–789. doi:10.2135/cropsci2014.10.0744
- Jannink, J.-L., M.C.A.M. Bink, and R.C. Jansen. 2001. Using complex plant pedigrees to map valuable genes. *Trends Plant Sci.* 6:337–342. doi:10.1016/S1360-1385(01)02017-9
- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, et al. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3):595–607. doi:10.1007/s00122-013-2243-1
- Jumbo, M.B., and M.J. Carena. 2008. Combining ability, maternal, and reciprocal effects of elite early-maturing maize population hybrids. *Euphytica* 16:325–333. doi:10.1007/s10681-007-9618-9
- Koebner, R.M.D., and R.W. Summers. 2003. 21st Century wheat breeding: Plot selection or plate detection? *Trends Biotechnol.* 21(2):59–63. doi:10.1016/S0167-7799(02)00036-7
- Lado, B., P.G. Barrios, M. Quincke, P. Silva, and L. Gutiérrez. 2016. Modeling genotype \times environment interaction for genomic selection with unbalanced data from a wheat breeding program. *Crop Sci.* 56:1–15. doi:10.2135/cropsci2015.04.0207
- Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124(3):743–756.
- Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121(1):185–199.
- Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9(4):357–359. doi:10.1038/nmeth.1923
- Liu, H., H. Zhou, Y. Wu, X. Li, J. Zhao, T. Zuo, et al. 2015. The impact of genetic relationship and linkage disequilibrium on genomic selection. *PLoS ONE* 10(7):1–13.
- Lynch, M., and B. Walsh. 1998. *Genetics and analysis of quantitative traits*, Sinauer Associates Inc., Sunderland, MA.
- Manolio, T., F. Collins, and N. Cox. 2009. Finding the missing heritability of complex diseases. *Nature* 461(7265):747–753. doi:10.1038/nature08494
- Mayer, K.F.X., J. Rogers, J. Dole, C. Pozniak, K. Eversole, C. Feuillet, et al. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345(6194):1251788–1251788.
- McNeal, F.H., M.A. Berg, and C.A. Watson. 1968. Reciprocal crosses and their influence on wheat quality. *Crop Sci.* 8:485–487. doi:10.2135/cropsci1968.0011183X000800040027x
- Melchinger, A.E., R.K. Gumber, R.B. Leipter, M. Vuylsteke, and M. Kuiper. 1998. Prediction of testcross means and variances among F_3 progenies of F_1 crosses from testcross means and genetic distances of their parents in maize. *Theor. Appl. Genet.* 96(3–4):503–512. doi:10.1007/s001220050767
- Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829.
- Mohammadi, M., T. Tiede, and K.P. Smith. 2015. PopVar: A genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. *Crop Sci.* 55(5):2068–2077. doi:10.2135/cropsci2015.01.0030
- Park, T., and G. Casella. 2008. The Bayesian Lasso. *J. Am. Stat. Assoc.* 103(482):681–686. doi:10.1198/016214508000000337
- Payne, P.I., M.A. Nightingale, A.F. Krattiger, and L.M. Holt. 1987. The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. *J. Sci. Food Agric.* 40(1):51–65. doi:10.1002/jsfa.2740400108
- Pérez, P., and G. de los Campos. 2010. Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome* 3(2):106–116. doi:10.3835/plantgenome2010.04.0005
- Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team. 2013. *Linear and nonlinear mixed effects models*. Springer-Verlag: New York.
- Poland, J., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012a. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7(2):E32253. doi:10.1371/journal.pone.0032253
- Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, et al. 2012b. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5(3):103–113. doi:10.3835/plantgenome2012.06.0006
- R Development Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org> (accessed 11 May 2017).
- Rutkoski, J., R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.L. Jannink, et al. 2015. Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. *Plant Genome* 8(2):1–10. doi:10.3835/plantgenome2014.10.0074
- Saghai-Maroo, M., K.M. Soliman, R. Jorgensen, and R.W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA* 81(24):8014–8018. doi:10.1073/pnas.81.24.8014
- Simons, K., J.A. Anderson, M. Mergoum, J.D. Faris, D.L. Klindworth, S.S. Xu, et al. 2012. Genetic mapping analysis of bread-making quality traits in spring wheat. *Crop Sci.* 52(5):2182–2197. doi:10.2135/cropsci2012.03.0175
- Souza, E., and M.E. Sorrells. 1991. Prediction of progeny variation in oat from parental genetic relationships. *Theor. Appl. Genet.* 82(2):233–241. doi:10.1007/BF00226219
- Stahl, F.W. 1979. Special sites in generalized recombination. *Annu. Rev. Genet.* 13(5):7–24. doi:10.1146/annurev.gen.13.120179.000255
- Tiede, T., L. Kumar, M. Mohammadi, and K.P. Smith. 2015. Predicting genetic variance in bi-parental breeding populations is more accurate when explicitly modeling the segregation of informative genomewide markers. *Mol. Breed.* 35(10):1–13. doi:10.1007/s11032-015-0390-6
- Utz, H.F., M. Bohn, and A.E. Melchinger. 2001. Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents. *Crop Sci.* 41(5):1470–1478. doi:10.2135/cropsci2001.4151470x
- VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91(11):4414–4423. doi:10.3168/jds.2007-0980
- Witcombe, J.R., S. Gyawali, M. Subedi, D.S. Virk, and K.D. Joshi. 2013. Plant breeding can be made more efficient by having fewer, better crosses. *BMC Plant Biol.* 13(1):22. doi:10.1186/1471-2229-13-22
- Zhao, Y., J. Zeng, R. Fernando, and J.C. Reif. 2013. Genomic prediction of hybrid wheat performance. *Crop Sci.* 53(3):802–810. doi:10.2135/cropsci2012.08.0463
- Zhong, S., and J.L. Jannink. 2007. Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. *Genetics* 177(1):567–576. doi:10.1534/genetics.107.075358