Published November 22, 2019

Training Population Optimization for Genomic Selection

Inés Berro, Bettina Lado, Rafael S. Nalin, Martin Quincke, and Lucía Gutiérrez*

I. Berro, R.S. Nalin, L. Gutierrez, Dep. of Agronomy, Univ. of Wisconsin – Madison, 1575 Linden Dr., Madison, WI 53706; I. Berro, B. Lado, L. Gutiérrez, Statistics Dep., Facultad de Agronomía, Univ. de la República, Garzón 780, Montevideo 12900, Uruguay; R.S. Nalin, Dep. of Genetics, Escola Superior de Agricultura "Luiz de Queiroz", Univ. de São Paulo, Piracicaba, São Paulo, Brazil; M. Quincke, Instituto Nacional de Investigación Agropecuaria, Est. Exp. La Estanzuela, Ruta 50 km 11.5, 70006 Colonia, Uruguay.

ABSTRACT The effectiveness of genomic selection in breeding programs depends on the phenotypic quality and depth, the prediction model, the number and type of molecular markers, and the size and composition of the training population (TR). Furthermore, population structure and diversity have a key role in the composition of the optimal training sets. Our goal was to compare strategies for optimizing the TR for specific testing populations (TE). A total of 1353 wheat (*Triticum aestivum* L.) and 644 rice (*Oryza sativa* L.) advanced lines were evaluated for grain yield in multiple environments. Several within-TR optimization strategies were compared to identify groups of individuals with increased predictive ability. Additionally, optimization strategies to choose individuals from the TR with higher predictive ability for a specific TE were compared. There is a benefit in considering both the population structure and the relationship between the TR and the TE when designing an optimal TR for genomic selection. A weighted relationship matrix with stratified sampling is the best strategy for forward predictions of quantitative traits in populations several generations apart.

Abbreviations: AYT, advanced yield trial; BLUE, best linear unbiased estimate; CDmean, average coefficient of determination; EYT, elite yield trial; GBLUP, mixed model best linear unbiased prediction; GBS, genotyping-by-sequencing; GS, genomic selection; GY, grain yield; INIA, Instituto Nacional de Investigación Agropecuaria (National Agricultural Research Institute); IRBP, INIA rice breeding program; IWBP, INIA wheat breeding program; OTR, optimized training population; PEVmean, prediction error variance; PYT, preliminary yield trial; sBLUP, super best linear unbiased prediction; SNP, single nucleotide polymorphism; taBLUP, trait-specific relationship matrix best linear unbiased prediction; TE, testing population; TR, training population; W_C, weighted additive relationship matrix with a stratified sampling accounting for genetic cluster.

CORE IDEAS

- Training populations can be optimized for specific testing populations.
- Optimized training populations are smaller, more related, and more predictive.
- Stratified sampling with a relationship matrix weighted by marker effect is optimal.

Genomic selection (GS) consists of selecting individuals from a TE on the basis of genotypic values predicted from their genome-wide molecular marker scores and a statistical model adjusted with individuals that have phenotypic and genotypic information (Meuwissen et al., 2001). The group of individuals that were phenotyped and genotyped is called the TR (Heffner et al. 2009).

Genomic selection is preferred over marker-assisted selection approaches for complex traits (Habier et al., 2007; Lorenz et al., 2011) because it includes all molecular markers in the prediction model and because it considers the quantitative trait loci of both major and minor effects (Xu, 2003; Jannink et al., 2010; Poland and Rife, 2012; Smith et al., 2018). Simulated and empirical cross-validation studies in plants show that GS can accelerate progress in plant breeding compared with marker-assisted selection, resulting in higher genetic gains (Hayes et al., 2009; Crossa et al., 2010; Heffner et al., 2011), higher profit per unit of cost, and superior progenies (Bernardo and Yu, 2007; Heffner et al., 2009, 2011; Heslot et al.,

Citation: Berro, I., B. Lado, R.S. Nalin, M. Quincke, and L. Gutiérrez. 2019. Training population optimization for genomic selection. Plant Genome 12:190028. doi: 10.3835/plantgenome2019.04.0028

Received 1 Apr. 2019. Accepted 23 Sept. 2019.
*Corresponding author (gutierrezcha@wisc.edu).

2012; Poland and Rife, 2012; Lado et al., 2017). Genomic selection is also useful when phenotypic information is not available, is not reliable because of low heritability, is costly or labor-intensive (Lado et al., 2018), or when there is a strong genotype × environment interaction with the target environment.

The most widely used methods of estimating the genotypic values are ridge regression, which estimates marker effects by regressing the trait values on marker genotypes, and the mixed model best linear unbiased prediction (GBLUP), which uses the genomic-estimated relationship matrix to model the correlation among individuals (VanRaden, 2008). The genotypic values from ridge regression and GBLUP are equivalent (Habier et al., 2009; de los Campos et al., 2012).

The main factors affecting the predictive abilities of GS models are the number and quality of phenotypic observations of the TR (Heffner et al. 2009; Jannink et al. 2010; Cooper et al., 2014; Lado et al., 2018), the genome coverage of molecular markers in both the TR and the TE (Solberg et al., 2008; Poland and Rife, 2012), the size and composition of the TR (Jannink et al., 2010; Heffner et al., 2011; Wientjes et al., 2013), and the relationship between the TR and the TE (Pszczola et al., 2012; Wientjes et al., 2013; Crossa et al., 2014; Hickey et al., 2015). The amount of phenotypic information used in GS has increased over recent years with the development of high-throughput phenotyping methods and the integration of phenotypic information from different breeding programs in the same geographical region. However, obtaining more phenotypic information of better quality remains a challenge for improving the selection of superior individuals (Araus and Cairns, 2014; Cooper et al., 2014). Studies have shown that the higher the heritability, the higher the prediction accuracy (Combs and Bernardo, 2013; Isidro et al., 2015). In addition, the more complex the trait, the poorer the prediction accuracy regardless of the heritability (Combs and Bernardo, 2013). The number and genome-wide coverage of molecular markers affect the prediction accuracy, with more molecular markers providing higher predictive ability until a plateau is reached that depends on the population size, structure, and diversity (Lorenzana and Bernardo, 2009; Gorjanc et al., 2017). More molecular markers are generally needed to maintain the prediction accuracy over cycles of selection through a better representation of the linkage disequilibrium structure (Asoro et al., 2011; Habier et al., 2007; Heffner et al. 2011; Lorenzana and Bernardo, 2009; Norman et al., 2018). Additionally, a large number of molecular markers is required to capture the effect of all quantitative trait loci when the TR used is genetically diverse (Norman et al., 2018).

The diversity of the TR affects the accuracy of the predictions in many ways. Several studies have shown that the larger the TR, the higher the predictive ability (Habier et al., 2007; Hayes et al., 2009; Lorenzana and Bernardo, 2009; Asoro et al., 2011; Crossa et al., 2014). Furthermore, low prediction accuracies are obtained when the TR has narrow genetic diversity because it is not possible to accurately estimate all the genotypic effects that explain the variation in the phenotype (Norman et al., 2018). On the other hand, increasing the diversity of the TR by using individuals that are genetically distant from the TE decreases the prediction accuracy (Crossa et al., 2014).

Therefore, one of the most challenging aspects of effective GS is the design of an optimal TR. After a good marker system and high-quality phenotyping have been established, choosing an optimal TR is not trivial. On the one hand, increasing the TR population size increases the prediction accuracy (Lorenzana and Bernardo, 2009; Asoro et al., 2011). However, because GS models rely on linkage disequilibrium, some studies have found that smaller, more related populations might be optimal (Crossa et al., 2010), whereas other studies have argued that increasing the population size, even at the expense of genetic relatedness, might be optimal (Asoro et al., 2011) for predicting a new population (i.e. the TE). On the other hand, genetic diversity within the TR is fundamental for estimating marker effects appropriately (Norman et al., 2018). Finally, the region in the genome where individuals are more similar also affects genomic predictions, especially for oligogenic traits (Zhang et al., 2010; Wang et al., 2018). Some strategies have been designed to weight genetic relationship matrices on the basis of their marker effects regardless of their position or linkage disequilibrium [trait-specific relationship matrix best linear unbiased prediction (taBLUP)] (Zhang et al., 2010), or weighting only bin-selected quantitative trait nucleotides [super best linear unbiased prediction (sBLUP)] (Wang et al., 2014). Several attempts have been made to optimize the TR for a given TE based on maximizing the accuracy of the prediction. Rincent et al. (2012) proposed an iterative process of exchanging individuals to maximize a function derived from the generalized average coefficient of determination (CDmean), defined as the squared correlation between the true and the predicted contrast of genetic values (Laloë, 1993) that maximized prediction accuracy between the TR and a TE, and the use of the prediction error variance (PEVmean). Isidro et al. (2015) used these criteria and three different TR optimization methods based on stratified sampling and showed that population structure has a large effect on optimizing the TR and the best methods depend on it. When the population structure effect was small, the CDmean performed adequately. However, with a strong population structure, the best strategy for optimizing the TR was a stratified sampling based on the population structure (Isidro et al., 2015; Rincent et al., 2017). Because these methods rely on GBLUP theory, they work well with highly polygenic traits but low accuracy would be found in oligogenic traits determined by a few large-effect quantitative trait loci (Rincent et al., 2017; Wang et al., 2018). A method that would be superior across populations and traits could not be found (Rincent et al., 2012; Isidro et al., 2015). In summary, the composition of the TR should have a balance between the within-population genetic

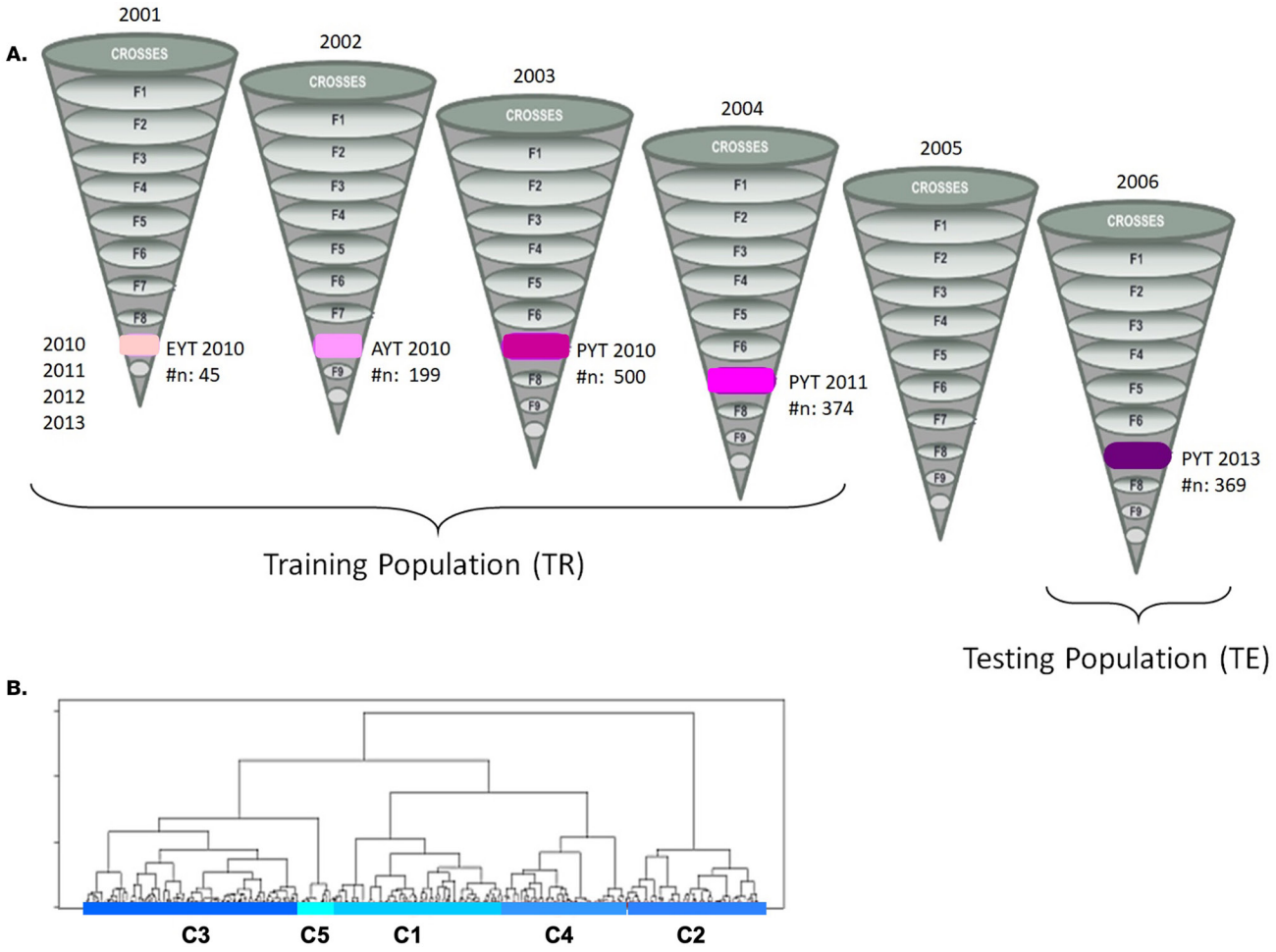

Fig. 1. Structure of the advanced wheat inbred lines from the Instituto Nacional de Investigación Agropecuaria Wheat Breeding Program and their genetic relationships. (A) The composition of the training population (TR) and the testing population (TE), with elite yield trials (EYT), advanced yield trials (AYT) and preliminary (PYT) yield trials from 2010 and 2011 comprising the TR. The PYT from 2013 was the forward TE. (B) Dendrogram of the training population constructed with the additive relationship matrix (**K**) where five groups were identified.

diversity, the relatedness with the TE, and the regions of the genome where individuals are more similar.

The goal of this study was to compare strategies for optimizing the TR for genomic prediction models in a rice and a wheat breeding program. First, we evaluated how the size, genetic relationships among individuals, and population structure affect the prediction ability within the TR. Second, we proposed some strategies for optimizing the TR based on different methods that account for the genetic relationship between the TR and the TE and we compared these to other methods proposed in the literature.

MATERIALS AND METHODS

Wheat Population

Plant Material

A total of 1353 spring bread wheat advanced inbred lines from the Wheat Breeding Program of the Instituto Nacional de Investigación Agropecuaria (INIA, National Agricultural Research Institute) of Uruguay (IWBP) were used. The IWBP lines consisted of all the advanced inbred lines from the preliminary yield trials (PYT) from 2010, 2011, and 2013, as well as the lines from the advanced yield trials (AYT) and elite yield trials (EYT) from 2010 (Fig. 1).

Phenotyping

Grain yield (GY) evaluations were conducted in five locations in Uruguay from 2010 to 2014, including one location with four sowing dates. Locations used to evaluate the genotypes were Dolores (33°50′S, 58°14′W; 15 m a.s.l.), Durazno (33°33′S, 56°31′W; 91 m a.s.l.), La Estanzuela (34°20′S, 57°42′W; 81 m a.s.l.), Young (32°76′S, 57°57′W; 85 m a.s.l.), and Ruta2 (33°45′S, 57°90′W; 95 m a.s.l.). For a full description of the number of lines evaluated in each location and years, see Lado et al. (2016).

Genotyping

Genotyping-by-sequencing (GBS) data were obtained for all 1353 IWBP lines. Tissue was collected from plants grown in either the field or the greenhouse. The cetyl trimethylammonium bromide method (Saghai-Maroof et al., 1984) was used to isolate DNA for the GBS protocol as in

Poland and Rife (2012). The TASSEL-GBS pipeline (Glaubitz et al., 2014) was run with a modification for nonreference genomes (Poland and Rife, 2012). Briefly, markers with a minor allele frequency below 1% or with more than 80% missing data were discarded. Marker–data imputation was conducted via the realized relationship matrix through the multivariate normal expectation maximization method of the *rrBLUP* package (Endelman, 2011) in R software (R Development Core Team, 2018). We identified 81,999 single nucleotide polymorphisms (SNPs).

Phenotypic Data Analysis

Best linear unbiased estimates (BLUEs) of GY were obtained for all genotypes present in each trial by the *nlme* package (Pinheiro et al., 2007) in R software (R Development Core Team, 2018). Field analysis was conducted according to the experimental design. Since the PYT consisted of a series of smaller alpha-design trials grouped by heading date and connected through common checks, the following model [Eq. 1] was used to estimate genotypic means for each heading date group in each environment (i.e. combination of location and year):

$$y_{ijkl} = \mu + \alpha_i + \lambda_j + \gamma_{k(j)} + \beta_{l(kj)} + \varepsilon_{ijkl}, \quad [1]$$

where μ is the overall mean, α_i is the effect of the i^{th} genotype, λ_j is the effect of the j^{th} trial, $\gamma_{k(j)}$ is the effect of the k^{th} replication within the j^{th} trial, $\beta_{l(kj)}$ is the effect of the l^{th} incomplete block within the k^{th} replication and the j^{th} trial, and ε_{ijkl} is the residual error from the i^{th} genotype in the l^{th} block within the k^{th} replication in the j^{th} trial, with λ_j, $\beta_{l(jk)}$, and ε_{ijkl} as the random variables $\lambda_j \sim N(0,\sigma_\lambda^2)$, $\beta_{l(jk)} \sim N(0,\sigma_\beta^2)$, and $\varepsilon_{ijkl} \sim N(0, \sigma_\varepsilon^2)$, all of which are independent. The AYT and EYT consisted of alpha-designs grouped by heading date; therefore, the following model [Eq. 2] was used to estimate genotypic means for each heading date group in each environment (i.e., combination of location and year):

$$y_{ijk} = \mu + \alpha_i + \gamma_j + \beta_{k(j)} + \varepsilon_{ijk}, \quad [2]$$

where μ, α_i, γ_j, $\beta_{k(j)}$, and ε_{ijk}, were defined as in Eq [1], with $\beta_{k(j)}$ and ε_{ijk} as the random variables $\beta_{k(j)} \sim N(0,\sigma_\beta^2)$ and $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$, both of which are independent.

To decrease the genotype × environment interactions, Mega-Environments 1 and 2, as identified in Lado et al. (2016) were used for this study. We did not use Mega-Environment 3 because a large genotype × environment interaction exists between Mega-Environment 3 and Mega-Environments 1 and 2. The mega-environments were constructed as groups of environments with low within genotype × environment interaction and high among genotype × environment interaction. Mega-Environments 1 and 2 include all locations from 2010, 2011, and 2013 and one of the La Estanzuela environments from 2014. They have a high correlation between environments. The final model used to obtain genotypic means across environments was:

$$y_{ijk} = \mu + \alpha_i + \delta_j + \gamma_{k(j)} + \varepsilon_{ijk}, \quad [3]$$

where μ, α_i, and $\gamma_{k(j)}$, were defined as in Eq. [1]; δ_j is the j^{th} environment (i.e., location and year) within a mega-environments, with $\gamma_{k(j)}$ and ε_{ijk} being the random variables $\gamma_{k(j)} \sim N(0,\sigma^2_\gamma)$ and $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$, both of which are independent. The BLUEs were estimated for each genotype by the *nlme* package (Pinheiro et al., 2007) in R software (R Development Core Team, 2018).

Rice Population

Plant Material

A total of 644 lines from the INIA Rice Breeding Program of Uruguay (IRBP) were used. The population consisted of 325 lines from the *O. sativa* ssp. *indica*, 314 lines from *O. sativa* ssp. *japonica* (*tropical japonica*), two *indica* cultivars [El Paso 144 (Yan et al., 2007) and INIA Olimar (Blanco et al., 1993; Instituto Nacional de Semillas, 2017) that are the most widely grown *indica* cultivars in Uruguay, and three *tropical japonica* cultivars [INIA Parao (Molina et al., 2011), INIA Tacuarí (Blanco et al.,1993), and INIA Caraguatá (Blanco et al., 1993)]. All cultivars were used as checks in all phenotyping experiments.

Phenotyping

Rice lines were evaluated for GY in the Experimental Unit of Paso de la Laguna (33°16′S, 54°10′W), Treinta y Tres, Uruguay, during three growing seasons (October–March): 2010–2011, 2011–2012, and 2012–2013. For a full description, see Monteverde et al. (2018).

Genotyping

Genotyping-by-sequencing data were obtained for the 644 advanced inbred lines and cultivars from the IRBP. DNA was extracted from young leaf tissue from plants grown at the Biotechnology Unit in Las Brujas, Canelones, Uruguay. The extraction was done with the Qiagen Dneasy kit (www.qiagen.com, accessed 19 Oct. 2019). The GBS libraries and sequencing were done at the Biotechnology Resource Center in the Genomic Diversity Facility at Cornell University in Ithaca, New York. Libraries were prepared according to the protocol of Elshire et al. (2011). Because of the strong population structure present within the lines, three datasets were obtained: an *indica* set, a *tropical japonica* set, and a combined *indica* and *tropical japonica* set. For both subspecies, *indica* and *tropical japonica*, SNPs were called from fastq files via the TASSEL version 3.0 GBS pipeline (Bradbury et al., 2007) as described in Spindel et al. (2013). Alignment to the Michigan State University Nipponbare rice reference genome version 7.0 was performed with Bowtie 2 (Langmead and Salzberg, 2012). Imputation of missing data for each of the *indica* (SNP_I) and *tropical japonica* (SNP_{TJ}) genotypes was performed with the FILLIN algorithm implemented in TASSEL version 5.0 (Bradbury et al., 2007). The average imputation accuracy was approximately 94% for both the *indica* and *tropical japonica* datasets. Single nucleotide polymorphism markers that had more than 50% missing data after the imputation along with

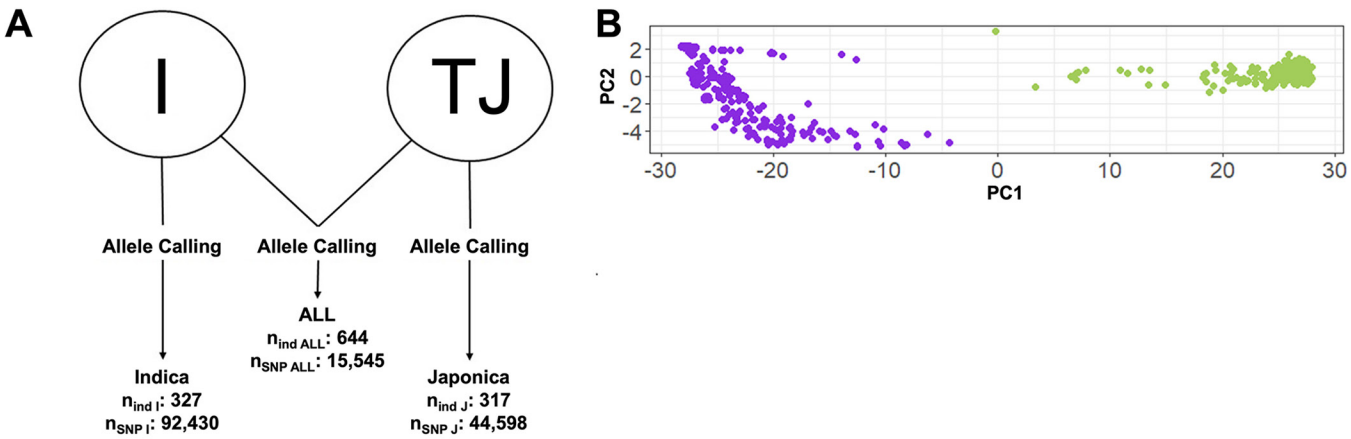

Fig. 2. Structure of the advanced rice inbred lines from the Instituto Nacional de Investigación Agropecuaria Rice Breeding Program and their genetic relationships. (A) Three datasets were used: *indica* with 92,430 single nucleotide polymorphisms (SNPs) and 327 advanced inbred lines, *tropical japonica* with 44,598 SNPs and 317 advanced inbred lines, and a combined *indica* and *tropical japonica* set with 15,545 SNPs and 644 individuals. (B) The first two principal components discriminating *indica* (purple, n = 327) and *tropical japonica* (green, n = 317) individuals in the combined data set.

monomorphic SNPs and SNPs with a minor allele frequency below 5% were removed from the datasets, as reported in Quero et al. (2018). The remaining missing data were imputed by the mean to perform principal component analysis. The final *indica* dataset contained 92,430 markers and the *tropical japonica* dataset had 44,598 markers. For the combined *indica* and *tropical japonica* set (SNP_{ALL}), SNPs were called from fastq files via the TASSEL version 3.0 GBS pipeline (Bradbury et al. 2007); the final dataset had 15,545 markers (Fig. 2).

Phenotypic Data Analysis

The BLUEs of GY were obtained for all genotypes present in each trial via the *lme4* package (Bates and Sarkar, 2010) in R software (R Development Core Team, 2018). Field analyses were conducted according to experimental design, which consisted of a series of smaller trials with randomized complete block designs connected through common checks. The following model, which used spatial correction for rows and columns, was used to estimate GY genotypic means for each environment (i.e., year) and subspecies (i.e., *indica* and *tropical japonica*):

$$y_{ijklm} = \mu + \alpha_i + \lambda_j + \gamma_{k(j)} + \eta_{l(j)} + \kappa_{m(j)} + \varepsilon_{ijklm}, \quad [4]$$

where μ, α_i, λ_j, and $\gamma_{k(j)}$ were defined as in Eq. [1]; $\eta_{l(j)}$ is the random effect associated with the l^{th} row in the j^{th} trial; $\kappa_{l(j)}$ is the random effect associated with the l^{th} column in the j^{th} trial, with λ_j, $\eta_{l(j)}$, $\kappa_{m(j)}$, and ε_{ijklm} as the random variables $\lambda_j \sim N(0,\sigma_\lambda^2)$, $\eta_{l(j)} \sim N(0,\sigma_\eta^2)$, $\kappa_{m(j)} \sim N(0,\sigma_\kappa^2)$, and $\varepsilon_{ijklm} \sim N(0, \sigma_\varepsilon^2)$, all of which are independent. The GY in *tropical japonica* in 2011 was not improved by the spatial corrections based on Akaike information criterion values and therefore we used Eq. [4] without $\eta_{l(j)}$ and $\kappa_{m(j)}$ for this trait. A final model with all the environments (i.e., years) was used to obtain overall genotypic means via the *lm* function in R software (R Development Core Team, 2018):

$$y_{ij} = \mu + \alpha_i + \delta_j + \varepsilon_{ij}, \quad [5]$$

where μ is the overall mean, α_i is the effect of the i^{th} genotype, δ_j is the j^{th} environment (i.e., year), and ε_{ij} is the residual error, $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$.

Overall GBLUP Model

Genotypic values were predicted via an overall GBLUP model following de los Campos and Pérez (2010):

$$y = 1\mu + u + \varepsilon, \quad [6]$$

where $y_{(N\times 1)}$ is the vector of mean yield for each genotype in all environments (i.e., the BLUEs from a model accounting for field design and environment) of length N (N = population size or number of genotypes in the set), $1_{(N\times 1)}$ is a vector of ones of length N, $\boldsymbol{\mu}$ is the overall mean, and $u_{(N\times 1)}$ is a random vector of genotypic predictors with $u \sim N(0, \mathbf{K}_{(N\times N)}\sigma^2_g)$, where $\mathbf{K}$ is the realized additive relationship matrix estimated as the cross-product of the centered and standardized marker states divided by the number of markers and estimated with the *rrBLUP* package (Endelman, 2011) in R software (R Development Core Team, 2018); ε is the vector of residuals with $\varepsilon \sim N(0, \mathbf{I}_{(N\times N)}\sigma^2_\varepsilon)$ where $\mathbf{I}$ is an identity matrix of size N.

Structured GBLUP Model

We used a GBLUP approach where subgroup-specific marker effects were estimated. The marker effects are described as the sum of common effects for all groups and group-specific random deviation. The structured GBLUP model can be represented as:

$$y = 1\mu + u_0 + u_1 + \varepsilon, \quad [7]$$

where $y_{(N\times 1)}$ is the vector of mean yield for each genotype in all the environments (i.e., the BLUEs from a model accounting for field design and environment) of length N (N = population size or number of genotypes in the set); $1_{(N\times 1)}$ is a vector of ones of length N; $\boldsymbol{\mu}$ is the overall mean; $u_{0(N\times 1)}$ is a random vector of genotypic predictors with $u_0 \sim N(0, \mathbf{K}\sigma^2_{u0})$, where $\mathbf{K}_{(N\times N)}$ is the realized

additive relationship matrix estimated as in the overall GBLUP model; $u_{1(N\times 1)}$ is a random vector for the group-specific genotypic predictors; and $u_1 \sim N(0, \mathbf{K}_1)$, where $\mathbf{K}_1$ is a block diagonal matrix, with each block being the realized additive relationship matrix estimated for each specific group with a group variance and off-diagonal zeros. Assuming two groups, $\mathbf{K}_1$ is:

$$K_1 = \frac{\begin{pmatrix} \sigma^2_{u_1} X_1 X_1' & 0 \\ 0 & \sigma^2_{u_2} X_2 X_2' \end{pmatrix}}{p}, \qquad [8]$$

where $\sigma^2_{u_1}$ and $\sigma^2_{u_2}$ are the genetic variance specific to Groups 1 and 2, respectively, and p is the number of markers. In this model, u_0 allows information to be shared between groups, whereas u_1 captures group-specific effects. For this model, we used the *BGLR* package (de los Campos and Pérez, 2010) in R software (R Development Core Team, 2018).

Within-Population Optimization

Three strategies were used to optimize the TR. For each strategy, a GBLUP model was fitted by the *rrBLUP* package (Endelman, 2011) in R software (R Development Core Team, 2018). The predictive ability was estimated as the correlation between predicted and observed genotypic values obtained via cross-validation (Burgueño et al., 2012; De Leon et al., 2016). Models were compared in terms of their predictive ability to random samples of the same size and to using all the individuals in the TR.

Strategy 1: Grouping Based on Genetic Relationship

Wheat Population. A clustering algorithm with the realized additive relationship matrix was used to group individuals in the IWBP TR population with the *cluster* package (Kaufman and Rousseeuw, 1990) in R software (R Development Core Team, 2018). The Ward hierarchical agglomerative method with the pseudo-*F* statistic was used to group similar individuals. The overall GBLUP model and the structured GBLUP model with groups as classes were used to predict genotypic values in the whole population.

Rice Population. The IRBP was grouped by subspecies: *indica* or *tropical japonica*. Given the high level of allelic phasing in the rice population, several dataset models were compared: (i) SNP_{ALL} to predict all individuals, (ii) SNP_{ALL} to predict *indica* individuals, (iii) SNP_{ALL} to predict *tropical japonica* individuals, (iv) SNP_{TJ} to predict *tropical japonica* individuals, and (v) SNP_{I} to predict *indica* individuals. The overall GBLUP model was used to predict genotypic values in the rice population.

Strategy 2: Grouping Based on Trials (Wheat Only)

The wheat population was evaluated in multiple trials, each having different phenotypic data quality. The overall GBLUP model and the structured GBLUP model that had field trials (EYT, AYT, and PYT) as classes were compared to predict genotypic values in the whole population (Fig. 1A).

Strategy 3: Grouping Based on Maturity (Wheat Only)

The advanced inbred lines were routinely grouped into early or late maturity lines on the basis of their heading date. Different breeding objectives for GY were then pursued for each group: early-maturity lines (short-cycle lines) were selected for high yield based on a large number of grains per spike, whereas late-maturity lines (long-cycle lines) were selected for high yield based on a large number of spikes. Since different selection pressures have been imposed on the two groups, the overall GBLUP model and the structured GBLUP model that used maturity as classes were compared to predict breeding values in the whole population.

Model Validation

Predictive ability was estimated as the correlation between the predicted and observed genotypic values via Type 1 cross-validation (Legarra et al., 2008; Burgeño et al., 2012; De Leon et al., 2016). A *k*-fold cross validation within TR was used, following Burgueño et al. (2012). Briefly, the observations were randomly divided into *k* nonoverlapping subsets. Next, *k* –1 groups were used as training sets and the remaining group was used as the validation set (i.e., genotypic values were obtained for each individual in the validation set). This procedure was followed until the genotypic values of individuals in all *k* subsets had been predicted. One hundred replications of the cross-validation with $k = 7$ were performed and the correlation between the predicted genotypic values and observed genotypic means was used to estimate the predictive ability.

Optimization of the TR to a Specific TE

For optimization of the TR to predict a specific TE, we used 984 wheat lines of the EYT, AYT, PYT 2010, and PYT 2011 as the TR and 369 PYT 2013 lines as the TE (Fig. 1A). Four strategies were evaluated for the selection of the training population; random selection, the genetic relationships between the TR and the TE, or two optimization criteria: the CDmean proposed by Rincent et al. (2012) and PEVmean.

Strategy 4: Selection Based on the Estimated Additive Genetic Relationship

The optimized TR (OTR) was constructed by choosing individuals from the TR on the basis of their high similarity to the TE. Two criteria were used to define similarity for each individual: the average relationship with the TE ($\bar{K}$) and the median relationship with the TE ($K_{0.5}$). The predictive ability of the OTR was evaluated for selecting the top 15, 20, …, 100% individuals from the TR.

Strategy 5: Selection Based on the Weighted Estimated Additive Genetic Relationship

Similar to Strategy 4, the OTR was constructed by choosing individuals from the TR on the basis of their similarity to the TE, but the weighted additive relationship matrix (**W**) was used instead of the kinship matrix. The weighted relationship matrix is used in a similar manner

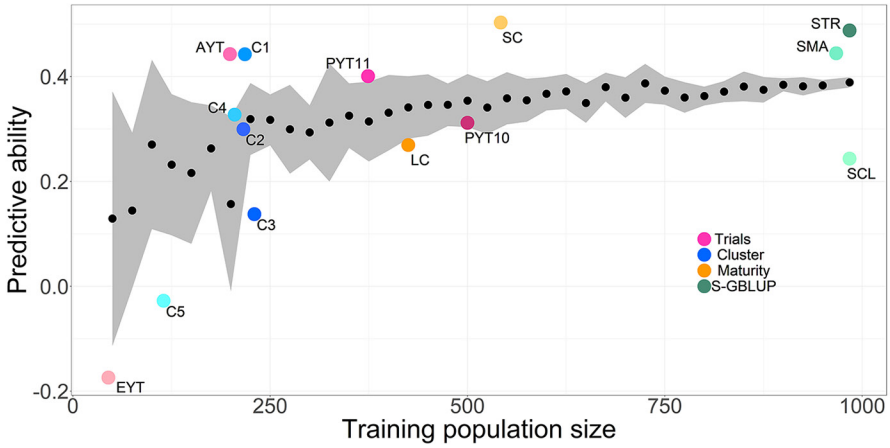

Fig. 3. Predictive ability of three clustering strategies [genetic groups based on the realized additive relationship matrix: Group to Group 5 (C1–C5); field trials: elite yield trials (yet), advanced yield trials (AYT), preliminary yield trials from 2010 (PYT10), and preliminary yield trials from 2011 (PYT11); and maturity: short-cycle lines (SC) and long-cycle lines (LC)], and the structured mixed model best linear unbiased prediction (GBLUP) model used for the within-population optimization for the wheat dataset.

to the taBLUP of Zhang et al. (2010) or the sBLUP of Wang et al. (2018), with the difference being that we are using this matrix to identify the most related individuals, but not in the prediction model. The genetic distance between the TR lines and the TE lines was calculated as the mean marker distance weighted by the estimated marker effects with a modification of the variance estimate of Endelman (2011):

$$Wij = \sqrt{\sum_{k=1}^{k=M}\left(\frac{\left|x_{\cdot TR[i,k]} - x_{\cdot TE[j,k]}\right|}{2}\right)\times \hat{u}[k]^2}, \qquad [9]$$

where i is the i^{th} line of the TR, j is the j^{th} line of the TE, k is the k^{th} marker, $X_{TR(n.TR'm)}$ and $X_{TE(n.TE'm)}$ are the genotypic matrices with the marker state (–1, 1), M is the number of markers, and u is the vector of the estimated marker effects. The *mixed.solve* function of the *rrBLUP* package (Endelman, 2011) in R (R Development Core Team, 2018) was used to estimate u. Three criteria were used to define similarity: (i) the average relationship to the TE ($\bar{W}$), (ii) the median relationship to the TE ($W_{0.5}$), and (iii) a proportional stratified sampling based on the average relationship to the TE and the groups defined in Strategy 1 ($\mathbf{W}_C$). The predictive ability of the OTR was evaluated for selecting the top 15, 20, and 100% individuals from the TR via the GBLUP models.

Strategy 6: Selection Based on the CDmean and PEVmean Genetic Algorithms

The OTR was constructed by choosing individuals from the TR according to a genetic algorithm that maximized the precision of the prediction of the difference between the values of each nonphenotyped individual (TE) and the mean of the population of candidate individuals (TR). We used the CDmean, defined as the squared correlation between the true and the predicted contrast of genetic values, and the PEVmean, defined as the variance of the distance between the true and the predicted contrast of genetic values, as optimization criteria. The optimization algorithm code was adapted from code provided by R. Rincent (pers. comm., 29 May 2017) and was implemented in R (R Development Core Team, 2018). The code is available as Supplemental File S1. For each sample size and each criterion, 50 repetitions of the algorithm with 800 iterations were used. The predictive ability of the OTR was evaluated for selecting the top 15, 20, and 100% individuals from the TR.

Random Selection

Finally, a random selection of TR lines was used to predict the TE by taking subgroups of different sizes. For subgroups with 15, 20, and 100% of the lines, GBLUP models were trained and used to predict the TE.

RESULTS

Within-Population TR Optimization

Wheat Population

The predictive ability increased with the number of individuals used in the TR (Fig. 3); however, after 600 individuals, the increase in predictive ability became marginal. Some groups had higher predictive ability than those obtained with either a random sample of the same size or the entire TR (Fig. 3). Group 1 obtained from the cluster analysis was better predicted than the other cluster groups, AYT was better predicted than the other trials groups, and the short-cycle lines were better than the long-cycle lines. The highly predictive groups had high within-population structure (Fig. 4) that was not associated with yield (data not shown). The model that includes the structure of the trials or cycle groupings showed higher predictive ability than random samples of the same size (Fig. 3).

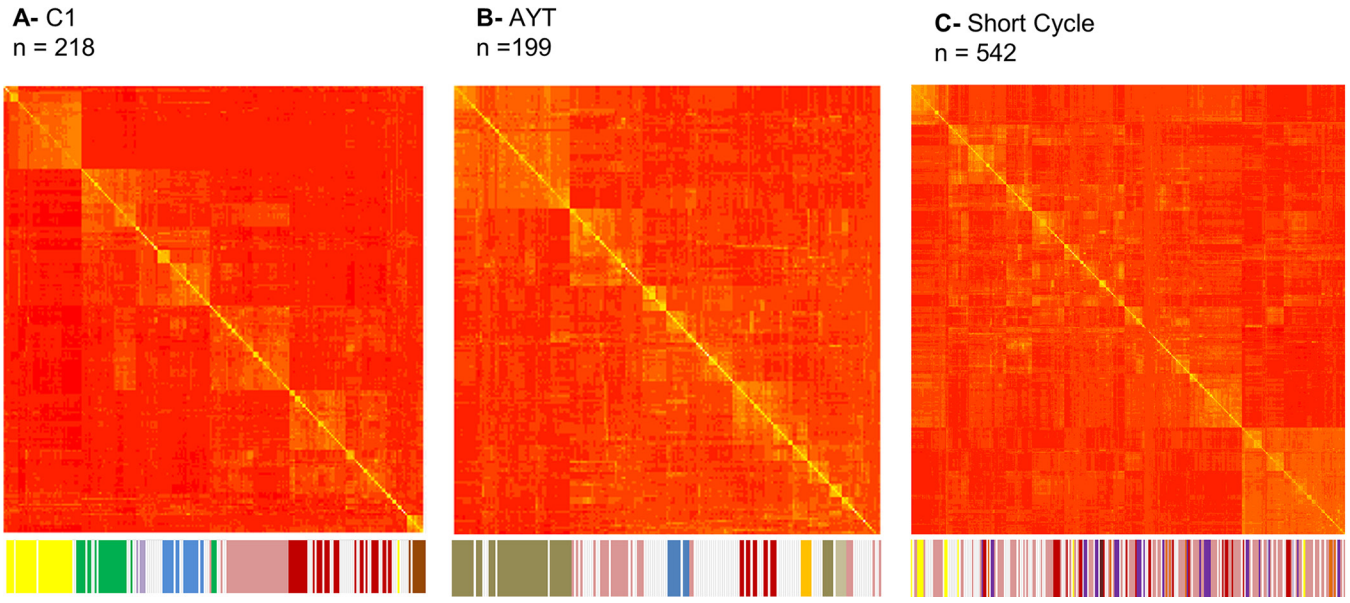

Fig. 4. Within-group population structure for the highly predictive groups: (A) Group 1, (B) advanced yield trials (AYT), and (C) short-cycle lines for the wheat dataset represented by heatmaps of the realized additive relationship matrix, with the parental contribution given below (i.e., each color represents a different parent).

Rice Population

Using both *indica* and *tropical japonica* subspecies as the TR population to predict all individuals resulted in a high predictive ability ($r = 0.8$, Fig. 5A). Group membership (i.e., *indica* or *tropical japonica*) was highly predictable but within-group predictions were low (Fig. 5B). Furthermore, when individuals from both subspecies were used to predict either *indica* ($r = 0.53$) or *tropical japonica* ($r = 0.37$), the predictive ability was low. Finally, when using only a subspecies to predict its own performance, the predictive ability of both *indica* ($r = 0.69$) and *tropical japonica* ($r = 0.54$, Fig. 5A) increased. There was a strong population structure among rice subspecies that was highly associated with GY (Fig. 5C).

Optimization of the TR to a Specific TE

The best OTR strategy was to use the average of the weighted additive relationship matrix with a stratified sampling accounting for genetic cluster ($\mathbf{W}_C$, Fig. 6). The **K** mean and **K** median were the worst optimization criteria (Fig.6). The CDmean and PEVmean stragtegies were no better than random selection but were better than **K** (Fig. 6).

DISCUSSION

Our study was able to first characterize the structure of small, highly predictive training sets to design an optimization strategy for identifying training sets for forward prediction of specific testing sets. We propose the use of a weighted relationship matrix in combination with stratified sampling to optimize the TR. This strategy is superior to random samples of the same size, the use of all available individuals, and other optimization strategies proposed in the literature. We discuss our findings in the context of population sizes, the diversity and structure of the TR, and relationship between the TR and the TE and propose conditions (i.e., traits and populations) where we believe our strategy would be superior.

Size of the TR

We found an increase in predictive ability with larger population sizes up to 600 individuals, which is similar to other studies (Lorenzana and Bernardo, 2009, Asoro et al., 2011). It has been widely established that predictive ability is higher when larger TRs are used (Muir, 2007; Meuwissen, 2009; Asoro et al., 2011; Lorenz, 2013; Isidro et al., 2015; Edwards et al., 2019); however, the relationship between TR size and other factors such as genetic relationship and population structure is less understood. We were able to identify smaller groups of individuals with high predictive ability (i.e., Group 1, AYT, short-cycle lines, and the $\mathbf{W}_C$ optimization strategy). These groups made better predictions than random samples of the same size and better predictions than all available individuals in the population. Therefore, there is a trade-off between population size and other factors that should be considered when optimizing the TR.

Diversity and Population Structure of the TR

There are a few processes that are relevant to how diversity affects the predictive ability of a TR. Diversity is a necessary condition for marker effect estimations (Norman et al., 2018) but how diversity is structured in the population is also relevant (Isidro et al., 2015). The genetic relationships among individuals are required for accurate genomic predictions (Habier et al, 2007; Asoro et al., 2011; Clark et al., 2012; Isidro et al., 2015; Edwards et al., 2019) because a relationship matrix is used to borrow information from relatives for prediction. Therefore, unrelated individuals can only be predicted by the mean performance of the population in a GBLUP context.

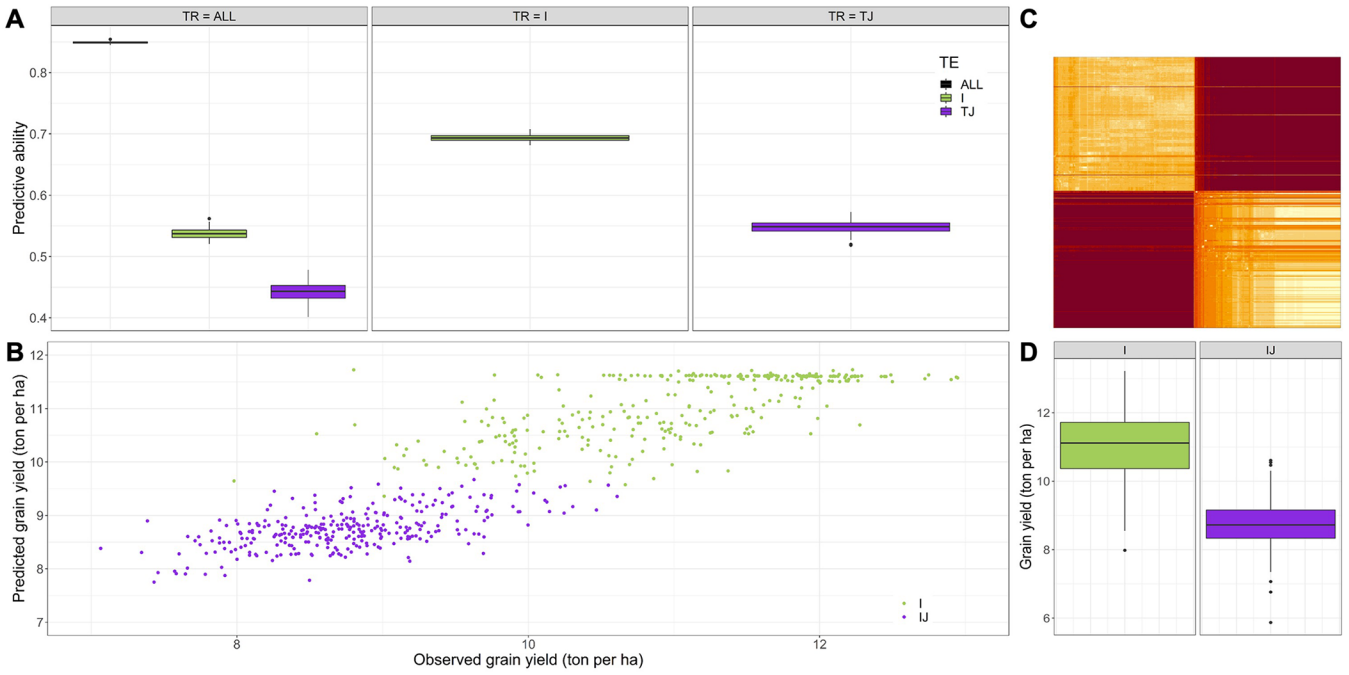

Fig. 5. (A) Predictive ability of different rice training population (TR) datasets to predict specific testing population (TE) datasets: ALL = all individuals; I = *indica* individuals; TJ = *tropical japonica* individuals. (B) One realization of a scatterplot of observed vs. predicted values for the dataset that used ALL individuals to predict ALL. (C) Heatmap of the realized additive relationship matrix of the rice population. (D) Box-plot of grain yield by rice subspecies.

In a broader interpretation, marker effects are genetic-background-dependent (Asoro et al., 2011; Toosi et al., 2010). Different levels of genetic relationship such as those created by family relationships can also create population structure (Würschum et al., 2017; Schmidt et al., 2016). The level of population structure determines the optimal prediction strategy (Isidro et al., 2015) and different strategies have been proposed to deal with population structure such as prediction within families (Würschum et al., 2017) or groups (Norman et al., 2018), prediction of group performance [i.e., compressed BLUP, (Wang et al., 2018)] or prediction of group performance and within-group deviation [structured GBLUP (de los Campos and Pérez, 2010)]. We evaluated these processes in two species with different levels of population structure and then proposed a new strategy for dealing with population structure that is more effective for forward prediction.

We found strong population structure in opposite phases in rice that was associated with the phenotype. This population structure overestimated the overall prediction accuracy (i.e., $r = 0.85$) by being accurate in predicting group membership and performance but being a poor predictor of within-group performance (i.e., $r < 0.5$). We show the effect of strong population structure on overall, group, and within-group predictions in a constructed example (Fig. 7). When population structure is associated with the phenotypic trait of interest, the model might be accurate at predicting group membership and group performance but extremely poor at predicting within-group performance. Group performance is relevant for predicting traits with low heritability that are hard to predict in general and where individuals within a group may act as pseudo-replications of family or groups, improving overall prediction accuracy (Wang et al., 2018). When the groups are families, average family performance can be easily predicted as the average performance of both parents and is therefore less relevant (Würschum et al., 2017). For example, in a broader case, identifying *indica* and *tropical japonica* individuals and predicting all individuals in each subspecies as the mean performance of their group had no practical relevance. Because population structure can create a challenge for predictions when markers are in opposite phases among subpopulations (Toosi et al., 2010; Asoro et al., 2011; Lopez-Cruz et al., 2015), within-family or within-group prediction might be more effective. Overestimation of the predictive accuracy was also observed by Schmidt et al. (2016), who combined spring and winter barley (*Hordeum vulgare* L.). Therefore, overall predictions with a strong population structure that is associated with the phenotype are challenging.

On the other hand, all the small groups that were highly predictive in our study (i.e., Group 1, AYT, and short-cycle lines) shared a low level of within-group population structure that was created by family relationships and was not associated with the phenotype. The presence of low levels of population structure that is not associated to the phenotype creates optimal prediction conditions (Fig. 7), where enough diversity is present to estimate marker effects and there is a high likelihood of each individual in the TE having a relative represented in the TR. Overall predictions in these cases can benefit from larger population sizes.

Furthermore, structured GBLUP models outperformed overall predictions in most of our situations by including a group effect as well as within-group

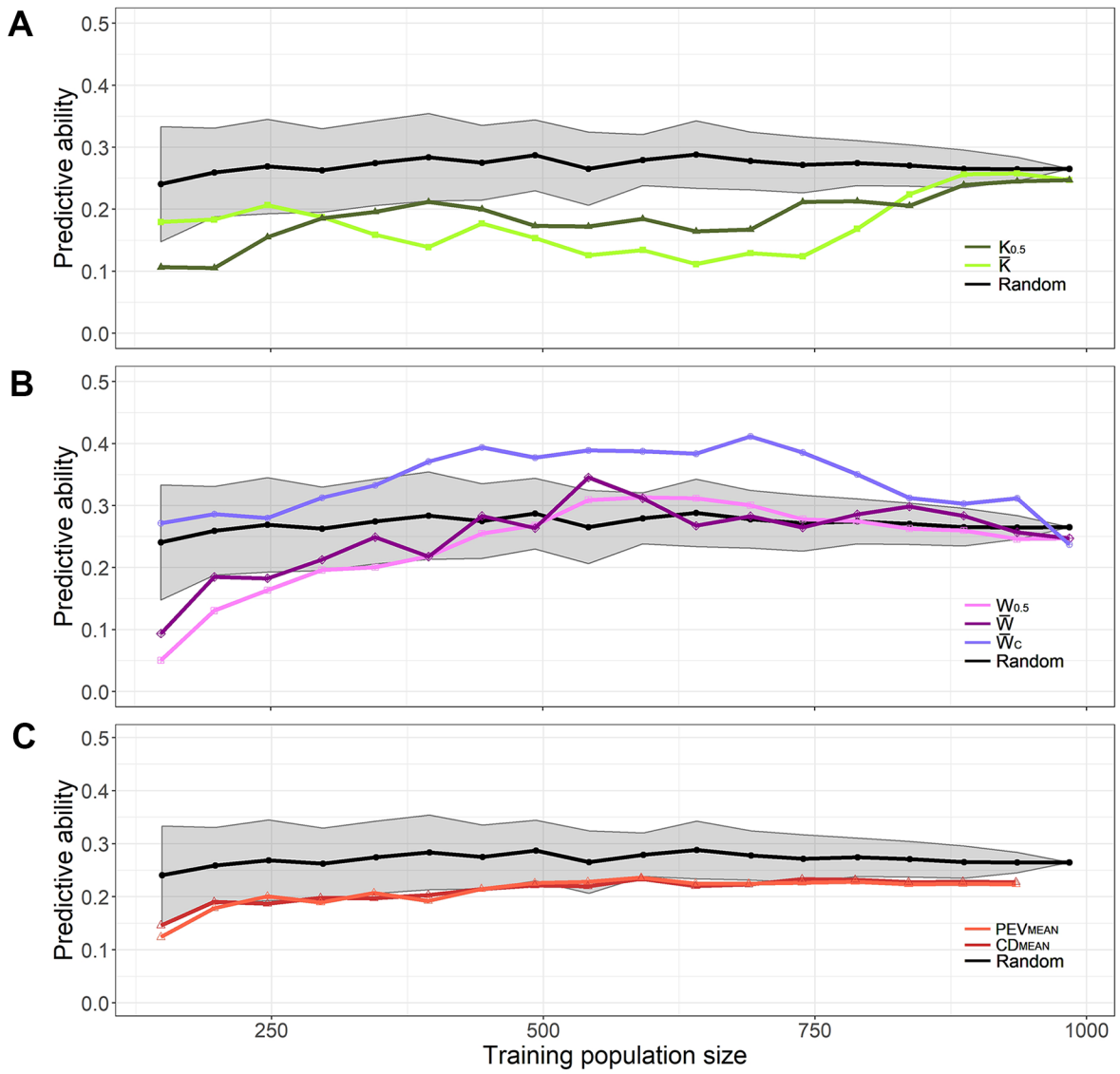

Fig. 6. Optimization of the training population (TR) through three overall strategies. (A) Strategy based on the average ($\bar{K}$) or median ($K_{0.5}$) realized genetic relationship of TR individuals to the testing population (TE). (B) Strategy based on the average ($\bar{W}$) or median ($W_{0.5}$) weighted genetic relationship of TR individuals to the TE or the average W with a stratified sampling considering genetic cluster (W_C, and C) on the basis of the the average coefficient of determination (CDmean) and average prediction error variance (PEVmean) optimization criteria. All optimized TRs are compared with random samples of the same size (the mean performance of 100 random samples is in black and the range is shaded).

deviations. Marker-independent groups are more effective than groups based on marker-only information, probably because these groups provide redundant information that is already accessible from the relationship matrix. Other studies have shown that considering population structure improves prediction ability (Isidro et al., 2015; Rincent et al., 2017; Norman et al., 2018); therefore, population structure should play a key role in the strategies used to build OTRs for genomic selection (Asoro et al., 2011; Crossa et al., 2014; Isidro et al., 2015, Lorenz and Nice, 2017; Rincent et al., 2017).

The Relationship between TR and TE

One of the main factors driving the trade-offs among population size, diversity, and population structure is the relationship between the TR and the TE (Habier et al., 2007, 2013; Crossa et al., 2010, 2014; Lorenz et al., 2012; Pszczola et al., 2012). Training sets that are more related to the testing sets have higher predictive ability (Crossa et al., 2014; Lorenz and Smith, 2015; Riedelsheimer et al., 2013). Furthermore, Isidro et al. (2015) concluded that an optimal design for the TR should minimize the relationships among the genotypes in the TR (i.e., the high diversity within the TR) while maximizing the relationship between the TE and the TR. Several methods have been proposed to optimize the TR in this context (Rincent et al., 2012, 2017; Isidro et al., 2015) with trait- and population-structure-dependent results. We used both a within-TR cross-validation and an independent forward TE approach to evaluate these effects and found that genetic relationship between TR and TE is one of the most relevant properties of the TR for increasing

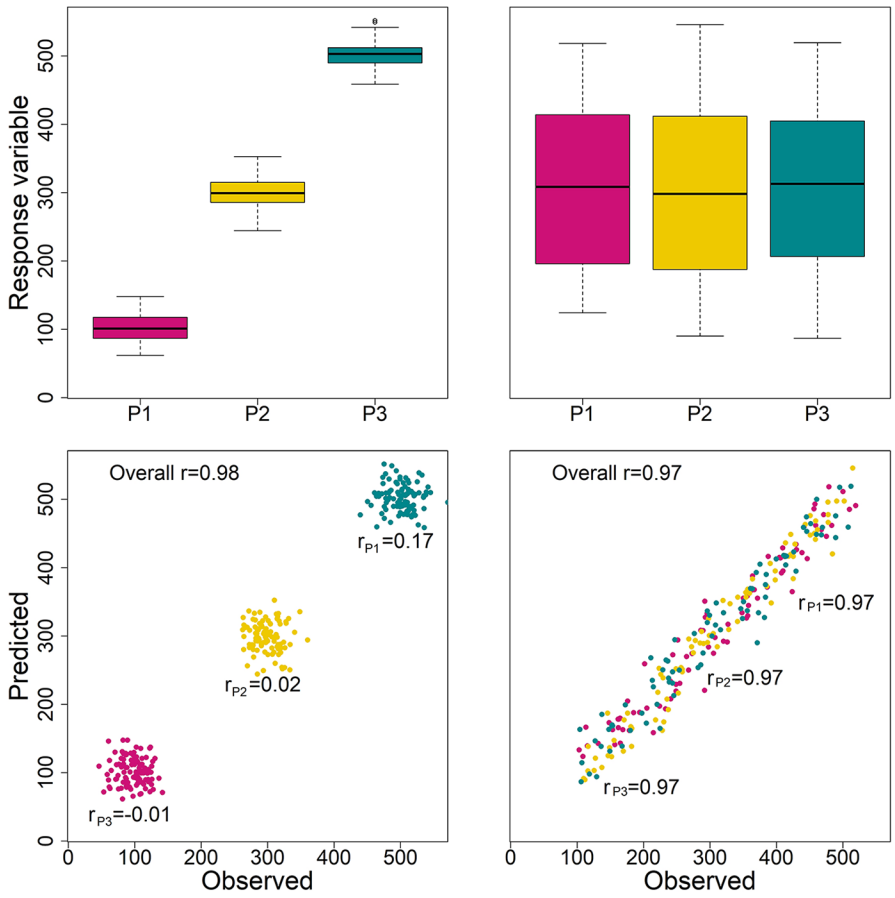

Fig. 7. Example constructed to show the consequences of overestimation when the phenotypic trait is associated with population structure, showing the case where population structure is associated with the phenotype (left panels) and a case where population structure is not associated with the phenotype (right panels). Boxplots for phenotypic values for a hypothetical trait are shown in the upper panels and the correlations between predicted and observed values are shown in the lower panels. Overall predictions might be high when population structure is associated with the phenotypic traits as the result of an artifact of mean performance for each group. This model will excel at predicting group membership and group performance (i.e., which population every individual belongs to and what that group's overall performance is) but might be a very poor predictor of within-group performance (i.e., which individual within that group or family will be superior). On the other hand, when population structure is not associated with the phenotype, overall predictions will reflect within-group predictions, with the advantage that larger populations are used.

predictive ability. We proposed an optimization strategy that outperformed previous methods. The best optimization strategy was the weighted relationship matrix with stratified sampling. The CDmean and PEVmean methods outperformed the relationship matrix (**K**) method, as in Rincent et al. (2012). However, the weighted relationship matrix with stratified sampling was superior to CDmean and PEVmean. Both CDmean and PEVmean have shown limited success when population structure is present (Isidro et al., 2015). We believe that proper modeling of the population structure is one of the main reasons for the success of our strategy. Estimations of the predictive ability of the CDmean strategy can be biased in the optimization phase if the search space is not large enough in terms of the number of starting points and iterations. We believe that our CDmean estimations are appropriate, the number of iterations used were based on achieving a plateau in the predictive ability for each population size, and the number of starting points was increased until consistent estimations of the predictive ability were found for each population size (Supplemental Fig. S1).

The region of the genome where individuals are more similar is more relevant than being similar across the genome. We found that a weighted genetic relationship matrix (**W**) outperformed the genetic relationship matrix (**K**). This shows that the kinship matrix *per se* is not a good indicator of groups of individuals with different marker effect responses for relevant phenotypic traits. Zhang et al. (2010) proposed and explored the idea of a weighted relationship matrix but in a different context. They found that the use of a trait-specific marker information matrix (taBLUP) for predictions improved the predictive ability over the GBLUP model. We used the same general idea of the taBLUP but with a few modifications. Genome-wide estimations of marker effects were used instead of grouping markers through previous identification of the relevant markers by genome-wide association studies. Furthermore, instead of using the weighted matrix for the prediction model, we used it in the sampling strategy of the TR to decide which individuals are the most related to the TE (i.e., **W** strategy). This OTR has higher predictive ability than random samples of the same size and higher

predictive ability than including all available individuals from the TR. This strategy is probably better suited for complex quantitative traits because it uses a genome-wide approach for marker effect estimation (Zhang et al., 2010). Other strategies for weighting the relationship matrix, such as taBLUP and sBLUP will be more effective for oligogenic traits (Wang et al., 2018). Furthermore, because bin-optimized quantitative trait nucleotides from genome-wide association studies are optimized to account for linkage disequilibrium structures in sBLUP, sBLUP will perform better than taBLUP (Wang et al., 2018). The **W** TR optimization strategy that we proposed could be used with the sBLUP weighted matrix instead of the genomic-weighted matrix to predict oligogenic traits in future studies.

The best optimization strategy in our study included a weighted matrix to evaluate relationship among individuals, along with stratified sampling ($\mathbf{W}_C$). We used the general ideas of the compressed BLUP (Wang et al., 2018) and structured GBLUP (de los Campos and Pérez, 2010) approaches but modified them to optimize the TRs. Instead of using the clusters to predict group performance, as in the compressed BLUP (Wang et al., 2018), or group and within-group performance, as in the structured GBLUP (de los Campos and Pérez, 2010), we used a clustering strategy to obtain a stratified sample of the TR that was more similar to Isidro et al (2015). The stratified sampling approach is superior to sampling the most related individuals overall because it produces a better representation of all the linkage disequilibrium structures in the TR that are relevant for the TE. However, when the TR and TE are conceptually the same as in *x*-fold validations or replaced phenotyping strategies, the $\mathbf{W}_C$ method performs similar to the PEVmean and CDmean methods. We showed the effect of different sampling strategies for forward and *x*-fold validation strategies (Supplemental Fig. S2). When the TE is conceptually different from the TR or has a lower level of relationship with the TR (i.e., similar to when forward predictions are used in GS to predict future populations), the $\mathbf{W}_C$ is the best strategy (Supplemental Fig. S2A). However, when an *x*-fold validation is used, such as in replaced phenotyping strategies, the $\mathbf{W}_C$ strategy has no advantage over the other methods (Supplemental Fig. 2B). Stratified sampling with the weighted relationship matrix ($\mathbf{W}_C$) is therefore especially relevant when forward predictions are attempted with future TEs, as in our case. In our study, we optimized the TR for a forward prediction three generations apart (i.e., a future PYT). This strategy was able to capture the relevant genomic regions and better estimate the marker effects increasing predictive ability. Therefore, smaller, OTR can be used to improve predictive ability. This indicates that neither the general relationship nor the population size *per se* are the best indicators of predictive ability. Optimized populations perform better than random samples of the population of the same size but are more relevant and they have higher predictive ability than the use of all individuals available from the TR.

CONCLUSIONS

In summary, we proposed a new strategy to optimize TRs to predict specific TEs in a forward approach. Our strategy of using a weighted relationship matrix in combination with a stratified sampling approach was the best approach for optimizing the TR. This strategy performed better than a random sample of the population of equal size, use of all the individuals available in the training population, or use of the relationship matrix, the CDmean, or PEVmean to choose the individuals to make up the TR. This strategy will be superior for complex quantitative traits and when small levels of population structure are present, such as with the familial relationship structures common in plant breeding populations. For oligogenic traits, a similar sampling strategy could be implemented but with the sBLUP weighted matrix instead of genome-wide marker effects. Our strategy will perform similar to the CDmean or PEVmean when population structure is absent or if the TE has conceptually the same structure as the TRs.

Supplemental Information

Supplemental File S1. The optimization algorithm code used for the CDmean and PEVmean optimization. This code was adapted from a code provided by R. Rincent (pers. comm., 29 May 2017) and was implemented in R (R Development Core Team, 2018).

Supplemental Fig. S1. (A) Coefficients of determination in each iteration and by training population size. (B) Predicted ability for CDmean, PEVmean, and random selection for each training population size.

Supplemental Fig. S2. Optimization of the training population using two overall strategies for two partitions of the TR: (i) predicting the least related individuals and (ii) predicting individuals from the same conceptual population as in an *x*-fold validation or replaced phenotyping. (A) Principal component of the TR, highlighting the partition strategies. (B) Predictions based on the average ($\overline{K}$) or median ($\mathbf{K}_{0.5}$) realized genetic relationship of TR individuals to the TE. (C) Predictions based on the average ($\overline{\boldsymbol{W}}$ or median ($\mathbf{W}_{0.5}$) weighted genetic relationship of TR individuals to the TE, or the average **W** with a stratified sampling considering genetic cluster ($\mathbf{W}_C$). All OTRs are compared with random samples of the same size (the mean performance of 100 random samples is in black and the range is shaded).

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

Author Contributions

IB, BL, RN and LG conducted statistical analyses. MQ designed the phenotyping experiments. BL performed genotyping analysis. IB and LG wrote the paper. LG designed the study and hypothesis. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This research was partially funded by a Graduate College Fellowship (CAP_2014) awarded to IB from University de la Republica, Uruguay; a Young Scientist Investigator Grant from the Comision Sectorial de Investigacion Cientifica (CSIC_2015_M2_12), Uruguay, awarded to IB and BL; and an internship from the Comision Sectorial de Investigacion Cientifica awarded to IB. We thank Regan Hoefler for proofreading an earlier manuscript. We thank the two anonymous reviewers for providing suggestions that improved the content of the manuscript.

REFERENCES

Araus, J.L., and J.E. Cairns. 2014. Field high-throughput phenotyping: The new crop breeding frontier. Trends Plant Sci. 19:52–61. doi:10.1016/j.tplants.2013.09.008

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. Plant Genome 4:132–144. doi:10.3835/plantgenome2011.02.0007

Bates, D., and D. Sarkar. 2010. Linear mixed-effects models using S4 classes. R Package lme4. Univ. of Auckland. http://ftp.auckland.ac.nz/software/CRAN/doc/packages/lme4.pdf (accessed 19 Oct. 2019).

Bernardo, R., and J. Yu. 2007. Prospects for genome wide selection for quantitative traits in maize. Crop Sci. 47:1082–1090. doi:10.2135/cropsci2006.11.0690

Blanco, P., Pérez de Vida, F., Piriz, M. 1993. Inia-Tacuarí Nueva variedad de arroz precoz de alto rendimiento. Bol. de Divulgación 31. Instituto Nacional de Investigación Agropecuaria, Montevideo, Uruguay.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of 505 breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci. 52:707–719. doi:10.2135/cropsci2011.06.0299

Clark, S.A., J.M. Hickey, H.D. Daetwyle, and J.H.J. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet. Sel. Evol. 44:4. doi:10.1186/1297-9686-44-4

Combs, E., and R. Bernardo. 2013. Accuracy of genome wide selection for different traits with constant population size, heritability, and number of markers. Plant Genome 6:1–7. doi:10.3835/plantgenome2012.11.0030

Cooper, M., C.D. Messina, D. Podlich, L.R. Totir, A. Baumgarten, N.J. Hausmann, D. Wright, et al. 2014. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci. 65:311–336. doi:10.1071/CP14007

Crossa, J., G. de los Campos, P. Pérez, D. Gianola, and J. Burgueño. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–124. doi:10.1534/genetics.110.118521

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, and L. Ornella. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112:48–60. doi:10.1038/hdy.2013.16

De Leon, N., J.-L. Jannink, J.W. Edwards, and S.M. Kaeppler. 2016. Introduction to a special issue on genotype by environment interaction. Crop Sci. 56:2081–2089. doi:10.2135/cropsci2016.07.0002in

de los Campos, G., and P. Pérez. 2010. BGLR: Bayesian generalized linear regression R package. R Foundation for Statistical Computing. http://cran.r-project.org/web/packages/BGLR/BGLR.pdf (accessed 19 Oct. 2019)

de los Campos, G., J.M. Hickey, R. Pong-Wong, H.D. Daetwyler, and M. Calus. 2012. Whole genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345. doi:10.1534/genetics.112.143313

Edwards, S.M., J.B. Buntjer, R. Jackson, A.R. Bentley, J. Lage, E. Byrne, et al. 2019. The effect of training population design on genomic prediction accuracy in wheat. Theor. Appl. Genet. 132(7):1943–1952. doi:10.1007/s00122-019-03327-y

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, and E.S. Buckler. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. doi:10.1371/journal.pone.0019379

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255. doi:10.3835/plantgenome2011.08.0024

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346. doi:10.1371/journal.pone.0090346

Gorjanc, G., M. Battagin, J.F. Dumasy, R. Antolin, R.C. Gaynor, and J.M. Hickey. 2017. Prospects for cost-effective genomic selection via accurate within-family imputation. Crop Sci. 57(1):216–228. doi:10.2135/cropsci2016.06.0526

Habier, D., R.L. Fernando, and J.C. Dekkers. 2009. Genomic selection using low-density marker panels. Genetics 182(1):343–353. doi:10.1534/genetics.108.100289

Habier, D., R.L. Fernando, and J.C.M. Dekkers. 2007. The Impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397. doi:10.1534/genetics.107.081190

Habier, D., R.L. Fernando, and D.J. Garrick. 2013. Genomic BLUP decoded: A look into the black box of genomic prediction. Genetics 194:597–607. doi:10.1534/genetics.113.152207

Hayes, B.J., P.J. Bowman, A.J. Chamberlian, and M.E. Goddard. 2009. Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433–443. doi:10.3168/jds.2008-1646

Heffner, E.L., M. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1–12. doi:10.2135/cropsci2008.08.0512

Heffner, E.L., J.L. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65–75. doi:10.3835/plantgenome2010.12.0029

Heslot, N., H. Yang, M.E. Sorrells, and J.L. Jannink. 2012. Genomic selection in plant breeding: A comparison of models. Crop Sci. 52:146–160. doi:10.2135/cropsci2011.06.0297

Hickey, J., M.S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, et al. 2015. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci. 54:1476–1488. doi:10.2135/cropsci2013.03.0195

Instituto Nacional de Semillas. 2017. Registro nacional de cultivares y registro de propiedad de cultivares. Instituto Nacional de Semillas. http://www.inase.uy/EvaluacionRegistro/ (accessed 19 Oct. 2019).

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

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: From theory to practice. Brief. Funct. Genomics 9:166–177. doi:10.1093/bfgp/elq001

Kaufman, L., and P.J. Rousseeuw. 1990. Cluster analysis methods. R package 'cluster' version 2.0.1. R Foundation for Statistical Computing. http://cran.r-project.org/web/packages/cluster/cluster.pdf (accessed 19 Oct. 2019).

Lado, B., S. Battenfield, C. Guzmán, M. Quincke, R.P. Singh, S. Dreisigacker, et al. 2017. Strategies for selecting crosses using genomic prediction in two wheat breeding programs. Plant Genome 10. doi:10.3835/plantgenome2016.12.0128

Lado, B., P. González Barrios, M. Quincke, P. Silva, and L. Gutiérrez. 2016. Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Sci. 56:2165–2176. doi:10.2135/cropsci2015.04.0207

Lado, B., D. Vázquez, M. Quincke, P. Silva, I. Aguilar, and L. Gutiérrez. 2018. Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. Theor. Appl. Genet. 131(12):2719–2731. doi:10.1007/s00122-018-3186-3

Laloë, D. 1993. Precision and information in linear models of genetic evaluation. Genet. Sel. Evol. 25:557–576. doi:10.1186/1297-9686-25-6-557

Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359. doi:10.1038/nmeth.1923

Legarra, A., C. Robert-Granié, E. Manfredi, and J.-M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611–618. doi:10.1534/genetics.108.088575

Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J. Jannink, et al. 2015. Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 (Bethesda) 5:569–582. doi:10.1534/g3.114.016097

Lorenz, A.J., S. Chao, F. Asoro, E. Heffner, T. Hayashi, H. Iwata, et al. 2011. Genomic selection in plant breeding: Knowledge and prospects. In: D.L. Sparks, editor, Advances in agronomy. Academic Press, San Diego. p. 77–123.

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 Sci. 52:1609–1621. doi.org:doi:10.2135/cropsci2011.09.0503

Lorenz, A.J. 2013. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. G3 (Bethesda) 3:481–491. doi:10.1534/g3.112.004911

Lorenz, A.J., and K.P. Smith. 2015. Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci. 55:2657–2667. doi:10.2135/cropsci2014.12.0827

Lorenz, A.J., and L. Nice. 2017. Training population design and resource allocation for genomic selection in plant breeding. In: R.K. Varshney, M. Roorkiwal, and M.E. Sorrells, editors, Genomic selection for crop Improvementi Springer International, Cham, Switzerland. p. 7–22.

Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151–161. doi:10.1007/s00122-009-1166-3

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

Meuwissen, T.H. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:35. doi:10.1186/1297-9686-41-35

Molina, F., P. Blanco, and F. Pérez de Vida. 2011. Nuevo cultivar de arroz INIA L5502 PARAO: Características y comportamiento. INIA (Instituto Nacional de Investigación Agropecuaria). Arroz. 68:26–32.

Monteverde, E., J.E. Rosas, P. Blanco, F. Pérez de Vida, V. Bonnecarrère, G. Quero, L. Gutiérrez, and S. McCouch. 2018. Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of rice. Crop Sci. 58:1519–1530. doi:10.2135/cropsci2017.09.0564

Muir, W.M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342–355. doi:10.1111/j.1439-0388.2007.00700.x

Norman, A., J. Taylor, J. Edwards, and H. Kuchel. 2018. Optimizing genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. G3 (Bethesda) 8:2889–2899. doi:10.1534/g3.118.200311

Pinheiro, J., D. Bates, S.D. Roy, and D. Sarkar. 2007. Linear and nonlinear mixed effects models. R Package nlme. Univ. of Bayreuth. http://ftp.uni-bayreuth.de/math/statlib/R/CRAN/doc/packages/nlme.pdf (accessed 19 Oct. 2019).

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102. doi:10.3835/plantgenome2012.05.0005

Pszczola, M., T. Strabel, H.A. Mulder, and M.P.L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. Dairy Science 95:389–400. doi:10.3168/jds.2011-4338

Quero, G., L. Gutiérrez, E. Monteverde, P. Blanco, F. Pérez de Vida, J. Rosas, et al. 2018. Genome-wide association study using historical breeding populations discovers genomic regions involved in high-quality rice. Plant Genome 11:170076. doi:10.3835/plantgenome2017.08.0076

R Development Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/ (accessed 19 Oct. 2019).

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

Rincent, R., D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, et al. 2012. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). Genetics 192:715–728. doi:10.1534/genetics.112.141473

Rincent, R., A. Charcosset, and L. Moreau. 2017. Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured population. Theor. Appl. Genet. 130:2231–2247. doi:10.1007/s00122-017-2956-7

Saghai-Maroof, M.A., K.M. Soliman, R.A. 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 81:8014–8018. doi:10.1073/pnas.81.24.8014

Schmidt, M., S. Kollers, A. Maasberg-Prelle, J. Grober, B. Schinkel, A. Tomerius, et al. 2016. Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. Theor. Appl. Genet. 129:203–213. doi:10.1007/s00122-015-2639-1

Smith, K.P., W. Thomas, L. Gutierrez, and H. Bull. 2018. Genomics-based barley breeding. In: N. Stein and G.J. Muehlbauer, editors, The barley genome. Compendium of plant genomes. Springer International Publishing AG, New York.

Solberg, T.R., A.K. Sonesson, J.A. Woolliams, and T.H. Meuwissen. 2008. Genomic selection using different marker types and densities. Anim. Sci. 86:2447–2454. doi:10.2527/jas.2007-0010

Spindel, J., M. Wright, C. Chen, J. Cobb, J. Gage, S. Harrington, et al. 2013. Bridging the genotyping gap: Using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor. Appl. Genet. 126:2699–2716. doi:10.1007/s00122-013-2166-x

Toosi, A., R.L. Fernando, and J.C.M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. Anim. Sci. 88:32–46. doi:10.2527/jas.2009-1975

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980

Wang, J., Z. Zhou, Z. Zhang, H. Li, D. Liu, Q.P.J. Zhang, et al. 2018. Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. Heredity 121:648–662. doi:10.1038/s41437-018-0075-0

Wang, Q., F. Tian, Y. Pan, E.S. Buckler, and Z. Zhang. 2014. A SUPER powerful method for genome wide association study. PLoS One 9:e107684. doi:10.1371/journal.pone.0107684

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

Würschum, T., H.P. Maurer, S. Weissmann, V. Hahn, and W.L. Leiser. 2017. Accuracy of within- and among-family genomic prediction in triticale. Plant Breed. 136:230–236. doi:10.1111/pbr.12465

Xu, S. 2003. Estimating polygenic effects using markers of the entire genome. Genetics 163:789–801.

Yan, W., J.N. Rutger, R.J. Bryant, H.E. Bockelman, R.G. Fjellstrom, and M.H. Chen. 2007. Development and evaluation of a core subset of the USDA rice germplasm collection. Crop Sci. 47:869–876. doi:10.2135/cropsci2006.07.0444

Zhang, Z., J. Liu, X. Ding, P. Bijma, D.-J. de Koning, and Q. Zhang. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. PLoS One 5:1–8. doi:10.1371/journal.pone.0012648