G3, 2023, **13(4)**, jkad025
https://doi.org/10.1093/g3journal/jkad025
Advance Access Publication Date: 26 January 2023
Genomic Prediction

Results from rapid-cycle recurrent genomic selection in spring bread wheat

Susanne Dreisigacker [1], Paulino Pérez-Rodríguez [2], Leonardo Crespo-Herrera [1], Alison R. Bentley [1], José Crossa [1,2,*]

[1]International Maize and Wheat Improvement Center (CIMMYT), Km 45 Carretera México-Veracruz, Texcoco, Edo. de México, CP 56100, México
[2]Colegio de Postgraduados, Montecillos, Edo. de México, CP 56264, México

*Corresponding author: Email: j.crossa@cgiar.org
Dedicated to the memory of Jose Pablo Alva-Galindo and Paulino Miranda-Oliva.

Abstract

Genomic selection (GS) in wheat breeding programs is of great interest for predicting the genotypic values of individuals, where both additive and nonadditive effects determine the final breeding value of lines. While several simulation studies have shown the efficiency of rapid-cycling GS strategies for parental selection or population improvement, their practical implementations are still lacking in wheat and other crops. In this study, we demonstrate the potential of rapid-cycle recurrent GS (RCRGS) to increase genetic gain for grain yield (GY) in wheat. Our results showed a consistent realized genetic gain for GY after 3 cycles of recombination (C_1, C_2, and C_3) of bi-parental F_1s, when summarized across 2 years of phenotyping. For both evaluation years combined, genetic gain through RCRGS reached 12.3% from cycle C_0 to C_3 and realized gain was 0.28 ton ha^{-1} per cycle with a GY from C_0 (6.88 ton ha^{-1}) to C_3 (7.73 ton ha^{-1}). RCRGS was also associated with some changes in important agronomic traits that were measured (days to heading, days to maturity, and plant height) but not selected for. To account for these changes, we recommend implementing GS together with multi-trait prediction models.

Keywords: genomic-assisted breeding, molecular markers, pedigree information, rapid-cycle recurrent genomic selection, wheat, genomic prediction, GenPred, shared data resources

Introduction

The widespread adoption of genomic selection (GS) in plant and animal breeding has strongly been driven by new sequencing technologies that generate abundant and inexpensive molecular markers (Meuwissen *et al.* 2001; Bernardo and Yu 2007; Lorenzana and Bernardo 2009). GS significantly increases prediction accuracy over marker-assisted selection for low heritability traits (de los Campos *et al.* 2009, 2010, 2012; Crossa *et al.* 2010, 2011, 2013, 2014, 2017; González-Camacho *et al.* 2012, 2016; Heslot *et al.* 2012, 2014; Pérez-Rodríguez *et al.* 2012; Riedelsheimer *et al.* 2012; Windhausen *et al.* 2012; Zhao *et al.* 2012; Hickey *et al.* 2014; Dreisigacker *et al.* 2021). GS involves predicting breeding values that comprise the parental average (half the sum of the breeding values of both parents) and a deviation of Mendelian sampling. GS can therefore be applied in 2 different contexts: (1) predicting additive effects in early generations of a breeding program such that a rapid selection cycle with a short breeding interval (i.e. GS at the F_2 level of a bi-parental cross) is achieved (Crossa *et al.* 2014) and (2) predicting the genotypic values of individuals where both additive and nonadditive effects determine the final commercial value of the lines (i.e. lines established in a sparse multi-environment field evaluation). Gaynor *et al.* (2017) clearly suggested separating the use of GS for parental selection or population improvement for crosses based on breeding values from product development that consists of testing lines and deriving varieties based on total genetic values.

Gholami *et al.* (2021) emphasized that plant breeders traditionally focus on product development, rather than identifying parents for new crosses. In other words, plant breeders have been more inclined to use total genetic values comprising the complete genetic contribution to the phenotype than the additive genetic value necessary for line improvement and crossing of new parental lines.

The International Maize and Wheat Improvement Center (CIMMYT, https://www.cimmyt.org) has explored GS as a new applied breeding tool since 2009 (de los Campos *et al.* 2009; Crossa *et al.* 2010, 2019, 2021; Dreisigacker *et al.* 2021). Genomic estimated breeding values (GEBVs) are routinely implemented and used as a decision tool by breeders. Studies at CIMMYT have evaluated using GEBVs for germplasm that have not been included in trials, for applying GS in early selection to shorten cycle time, and for using sparse testing (Atanda *et al.* 2022). CIMMYT has also built the basis for a more informed screening of novel allelic diversity in germplasm collections by genotyping a substantial part of the available accessions from its gene banks (Sansaloni *et al.* 2020; Martini *et al.* 2021). Extensive studies utilizing the GEBVs of traits from wheat germplasm bank accessions (Crossa *et al.* 2016) were performed to explore its potential for harnessing genetic resources (Gholami *et al.* 2021; Martini *et al.* 2021). The practical application of GS has been studied and applied based on the

Received: November 11, 2022. **Accepted:** January 19, 2023

individual breeder's decision. However, most recently, there has been a clear focus on shortening the breeding cycle by advancing and selecting lines quickly up to the F_4 and F_5 generations, sparse testing these lines at several locations (some belonging to the target population of environments already defined), and recycling them based on total genetic values.

The CIMMYT Global Maize Program has been highly successful in achieving important genetic gains in bi-parental populations in drought environments (Beyene *et al.* 2015, 2019). Gains were achieved from significant decreases in the breeding cycle, and just as importantly, hybrids from lines developed using GS have proved to be productive, high yielding, and stable across several drought and optimal environments. The achievements reported by Beyene *et al.* (2015, 2019) concluded that genetic gain in maize hybrids developed with GS was remarkable, considering the commercial checks used in the studies were the best in the multi-environment trials. Beyene *et al.* (2015) concluded that "the average gain observed under drought in our study using GS was two- to four folds higher than what has been reported from conventional phenotypic selection". Moreover, Zhang *et al.* (2017) designed rapid-cycle recurrent GS (RCRGS) of multi-parental crosses with important significant gains per cycle in tropical maize in Mexico.

The CIMMYT Global Wheat Program started implementing GS as a routine breeding tool in 2013, and, since then, has made important contributions by developing and testing several new genome-enabled prediction models (Dreisigacker *et al.* 2021) including G×E interaction and multi-trait, multi-environment genome-based predictions. For decades, CIMMYT wheat breeders have been using a standard pedigree system for crosses, which makes it possible to accurately predict breeding values based on the additive relationship matrix (**A**) and its incorporation in the statistical analyses of multi-environment trials by modeling G×E interaction with the factor analytic model as shown in Crossa *et al.* (2006) and Burgueño *et al.* (2007). These authors concluded that epistatic interaction in wheat is important, and it is necessary to correctly assess additive, additive × additive, additive × environment, and additive × additive × environment interactions in wheat breeding.

Pérez-Rodríguez *et al.* (2012) assessed the predictive ability of linear and nonlinear models on the marker effects using high-density genotypic data in wheat. The linear models were Bayesian LASSO, Bayesian ridge regression, Bayes A, and Bayes B, whereas the nonlinear models were the reproduced kernel Hilbert space (RKHS) regression, Bayesian regularized neural networks (BRNN), and radial basis function neural networks (RBFNN). It was found that the 3 nonlinear models had consistently better prediction accuracy than the linear regression specification. Pérez-Rodríguez *et al.* (2012) concluded the importance of epistasis in wheat and coincided with the results of Crossa *et al.* (2006) and Burgueño *et al.* (2007) using the additive relationship matrix **A**. The results also agreed with Gianola *et al.* (2006), Long *et al.* (2010), and González-Camacho *et al.* (2012), which concluded that non-parametric treatment of markers may account for epistatic effects not captured by linear additive regression models and seemed to be useful for predicting quantitative traits with different complex underlying gene action under varying types of interaction in different environmental conditions.

Early GS studies utilizing CIMMYT wheat datasets have already shown that molecular markers increased genome-wide prediction abilities over the pedigree-derived models (de los Campos *et al.* 2009; Crossa *et al.* 2010). Furthermore, when molecular markers and pedigree information are jointly considered, the prediction abilities are slightly, but consistently, superior to the marker or pedigree-derived models on their own. The CIMMYT Global Wheat Program has not yet applied GS at early breeding stages for population improvement. Nevertheless, as early as 2009, an extensive proof-of-concept experiment was established with the objective of incorporating genomic predictions for increased grain yield in an early breeding generation (Bonnett *et al.* 2022) to compare the realized response to selection based on 3 prediction models. Experiment 2 in the study of Bonnett *et al.* (2022) compared the predictive ability of the different GEBV calculation methods in F_2 using a set of single plant-derived $F_{2:4}$ lines from randomly selected F_2 plants. Results showed a significant positive correlation between the observed yield of $F_{2:4}$ lines and the predicted yield GEBVs of F_2 single plants based on the nonlinear RKHS method. For the first time in wheat, results showed the potential for the application of GS in early generations of wheat breeding and the importance of using the appropriate statistical model for GEBVs calculation.

Based on the initial results of Bonnett *et al.* (2022), a second RCRGS experiment was established. The main objective of this study was to perform 3 genomic-assisted recurrent selection cycles in the greenhouse based on a training population of F_4 lines and to estimate realized genetic gains for grain yield in each cycle and across cycles.

Materials and methods

Developing the training population (C_0)

The training population (C_0) consisted of 1,609 F_4 lines derived from 14 F_2 families, which were based on 16 parents from the CIMMYT spring bread wheat breeding program. Eleven F_2 families comprised 94 to 95 F_4 lines and 3 F_2 families included 186 to 190 lines. F_2 individuals were genotyped with the Infinium iSelect 90K SNP genotyping array (Wang *et al.* 2014) and genotype calling was performed using GenomeStudio Software v2011.1 (https://www.illumina.com). The F_2 individuals were phenotyped as F_4 lines at the Campo Experimental Norman E. Borlaug (CENEB) in Ciudad Obregón, northern Mexico. The F_4 trial was sown as an augmented block design with 2 replications, including the line "BORLAUG100 F2014" as a repeated check. The phenotypic data included grain yield (GY, ton ha^{-1}), days to heading (DTH, days), and plant height (PH, cm). Agronomic traits (DTH and PH) were only measured in one replication. Best linear unbiased estimators (BLUEs) for GY were assessed for all genotypes. A numerical relationship matrix (**A**) derived from the pedigree was also available for all individuals in the training set. This relationship matrix was computed by the "coefficient of parentage (COP)" using the BROWSE software (McLaren *et al.* 2000, 2005).

Cycle 0 phenotypic selection and formation of cycle 1

In cycle 0 (C_0), the 10 highest yielding lines of 6 F_4 families each were selected as parents to form cycle 1 (C_1). The 6 F_4 families were selected based on several criteria: their rank in GY (BLUEs) in the training population, GY heritability, and the estimated genomic prediction ability within and between F_4 families. The agronomic traits (DTH and PH) were not considered when making selections. The 60 selected F_4 lines were planted in the greenhouse on 3 different dates (3 pots with 6–7 plants per F_4 line at each date). C_1 was formed by intermating the F_4s (Fig. 1). Six crosses were performed within each selected F_4 family (36 crosses) and 10 crosses between 11 pairs of F_4 families (110 crosses), which were chosen based on the average genomic prediction ability between them. Each F_4 family was used in intercrosses at least 3 times. The F_1 seed of each cross was harvested and threshed to form C_1.

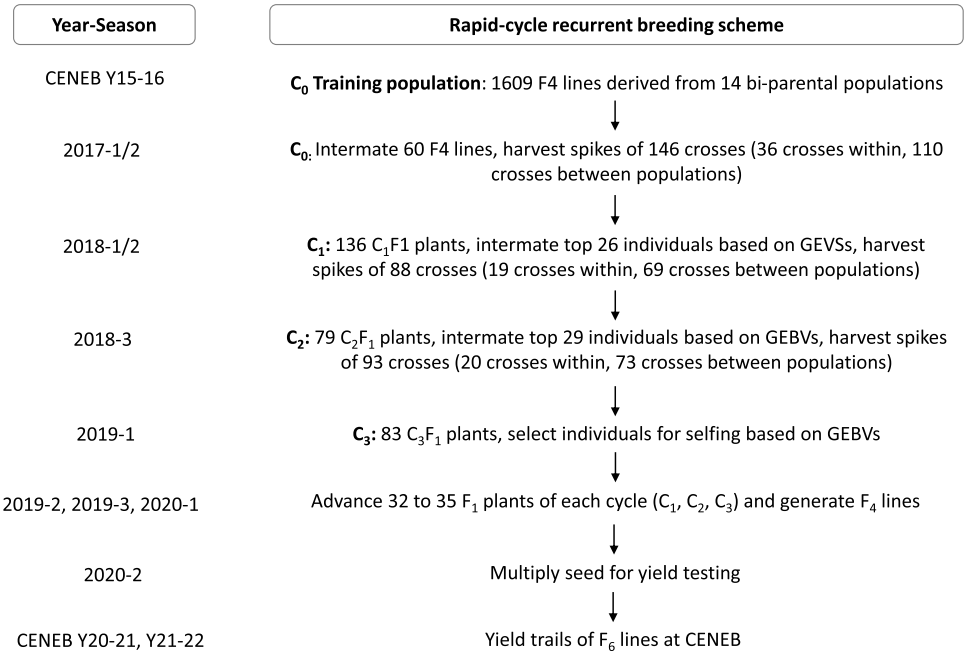

Fig. 1. Rapid-cycle recurrent breeding scheme starting with the formation of the training population to performing final yield trials.

Recombination using GS in cycle 1, cycle 2, and cycle 3

In C_1, 136 F_1s were planted at 2 different dates in the greenhouse (1 pot with 3 plants per F_1 at each date). DNA was extracted from bulked tissue and shipped to TraitGenetics GmbH, Germany, for genotyping with the Illumina 20K microarray (https://www.traitgenetics.com). GEBVs were calculated for all 136 F_1s. The top 26 C_1F_1s were selected and intermated to form the cycle 2 (C_2) population. Like C_1, crosses were performed within and between families (19 and 69 crosses, respectively), C_2F_1 seed of each cross was harvested and threshed. In C_2, the recombination protocol was repeated. The top 29 C_2F_1s were intermated to form cycle 3 (C_3). The number of F_1s planted per cycle, the number of parents selected for next cycle recombination, and the number of crosses performed are shown in Fig. 1. After C_2 recombination, C_3F_1 were genotyped and GEBVs calculated, but they were not recombined.

Line development and phenotypic evaluation of the selection cycles

After recurrent selection, 32 to 35 F_1s from each cycle (C_1 to C_3) were selfed to derive F_6 lines. Not all F_1s within a selection cycle had sufficient seed for selfing. Therefore, a distributed set of F_1s was chosen. The GEBVs of the selfed F_1s ranked between 1 and 58 in each cycle, and about 50% of F_1s were used in recombination, while the residual F_1s were not crossed (Supplementary Table 1). F_1s were advanced to the F_4 breeding generation in the greenhouse via "selected bulks." Residual F_1 seed was planted in trays in the greenhouse, and several visually "good" spikes were bagged to derive F_2 seed. This advancement procedure was repeated up to the F_4 generation. F_5 seed was harvested and planted for seed multiplication at CIMMYT in El Batan, Mexico, in 0.5 m plots; plant off-types were eliminated.

A total of 118 lines were field tested together with the cv. "BORLAUG100 F2014" as standard check for 2 crop cycles (2020–2021 and 2021–2022) at CENEB. The lines were grown in 2 trials of 60 entries (59 lines + check) in an incomplete block design with 2 replications. The evaluations were conducted under optimal conditions, i.e. 500 mm of irrigation water, mechanized and chemical control of weeds, diseases, and pests. The 240 plots were of dimension 2.8 × 1.6 m and sown at a seed rate of 120 kg ha^{-1}. The same agronomic traits as in the training population were measured, GY, DTH, and PH, in addition days to maturity (DTM). Lines included 101 F_6 lines derived from the recurrent GS cycles and 17 of the 60 C_0 parents and the check. Means of GY BLUEs of each recurrent selection cycle were compared, and differences were determined using the least significant difference (LSD at 5% significance). The heritability of the trials was estimated from the variance components using the equation:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma^2}{re}}$$

with r = number of replications, e = number of environments (years), σ^2 = error variance, σ_g^2 = genotypic variance, and σ_{ge}^2 = G × Y variance.

SNP genotyping and filters

As described above, the initial training population was genotyped with the Infinium iSelect 90 K SNP genotyping array for wheat and the recurrent selection F_1 plants with the lower-density Illumina Infinium 20 K wheat SNP array. A total of 7,815 markers overlapped between platforms. We imputed missing marker genotypes at random according to allele frequencies and subsequently removed monomorphic markers and markers with a minor allele frequency smaller than 0.05. After this quality control, 7,139 markers were available for further analysis. The 101 derived F_6 lines were also genotyped with the Illumina Infinitum array (TraitGenetics) to compute the genetic diversity maintained in each selection cycle.

Genomic prediction models

We considered 4 different prediction models to fit the training population for selecting the best parents to be crossed and initiate

the RCRGS as well as for selecting the best F_1s in each recurrent cycle: (1) the genomic best linear unbiased prediction model (GBLUP), (2) GBLUP including the pedigree information (P + GBLUP), (3) Reproducing Kernel Hilbert Spaces with Kernel Averaging (RKHS-KA) method, and (4) Reproducing Kernel Hilbert Spaces with Kernel Averaging and Pedigree (P + RKHS-KA). While all models were computed, selections during the RCRGS scheme were based only on Model 4.

For the prediction (GEBV) and selection of the best parental candidates for the next cycle, we focused on (1) assessing additive effects by including the pedigree (P) information (numerator relationship matrix) and thus emphasizing between family variance, and (2) including genotypic values of individuals where both additive and nonadditive are included on the genomic information (RKHS-KA). Using pedigree and marker information together has been successful in decreasing the interval cycle at the early stages of population improvement (Crossa *et al.* 2017; Bonnet et al. 2022).

GBLUP model

The GBLUP model has become widely used in genomic prediction (e.g. Endelman 2011). The model can be written as:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{u} + \mathbf{e}, \quad (1)$$

where $\mathbf{y}$ is a vector with the response variable of dimension $n \times 1$ (phenotypes), μ is an intercept, $\mathbf{1}$ is a vector of ones, $\mathbf{u} \sim N(\mathbf{0}, \sigma^2_u\mathbf{K})$ corresponds to the random effect of wheat lines, σ^2_u is the variance parameter associated to the wheat lines, and $\mathbf{K} = \mathbf{MM}'/p$ is a genomic relationship matrix derived from markers (e.g. Lopez-Cruz *et al.* 2015) with $\mathbf{M}$ the matrix of markers centered and standardized by columns and p the number of markers, $\mathbf{e} \sim N(\mathbf{0}, \sigma^2_e\mathbf{I})$ the vector of random error terms, with σ^2_e the variance parameter associated to the error, and $\mathbf{I}$ denotes the identity matrix.

The pedigree + molecular marker model (P + GBLUP)

The pedigree + molecular marker model takes into account the pedigree information of the wheat lines represented by the numerical relationship matrix ($\mathbf{A}$) and the genomic relationship matrix derived from markers described before (Eq. 1). The full genetic model is:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{a} + \mathbf{u} + \mathbf{e}, \quad (2)$$

where $\mathbf{a} \sim N(\mathbf{0}, \sigma^2_a\mathbf{A})$ is the vector of additive random effects for wheat lines whose variance–covariance matrix is obtained from the numerator relationship matrix ($\mathbf{A}$) derived from the coefficient of co-ancestry between the wheat lines and σ^2_a is a variance parameter associated with the additive relationship matrix derived from pedigree information and the rest of the terms has been described before.

Reproducing kernel Hilbert spaces with kernel averaging

The Gaussian kernel commonly used in genomic prediction is $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_{i'}) = \exp(-hd^2_{ii'})$ (e.g. Pérez-Rodríguez *et al.* 2012), where $d_{ii'}$ is the distance based on markers between individuals i, i' ($i = 1, \ldots, n$), the bandwidth parameter controls how fast the covariance function drops as a function of the distance. The estimation of the bandwidth parameter is computationally demanding and to overcome this problem, de los Campos *et al.* (2010) proposed to fit a model that includes several kernels, each one with its own bandwidth parameter. The RKHS-KA with 3 kernels is given by:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 + \mathbf{e}, \quad (3)$$

where $\mathbf{u}_1 \sim N(\mathbf{0}, \sigma^2_{u1}\mathbf{K}_1)$, $\mathbf{u}_2 \sim N(\mathbf{0}, \sigma^2_{u2}\mathbf{K}_2)$, $\mathbf{u}_3 \sim N(\mathbf{0}, \sigma^2_{u3}\mathbf{K}_3)$ and $\mathbf{e}$ distributed independently, with 3 different Gaussian kernels computed with bandwidth parameters $h_1 = \frac{1}{5m}$, $h_2 = \frac{1}{m}$, $h_3 = \frac{5}{m}$, with m that corresponds to the median squared Euclidean distances between lines without including the diagonal entries (Pérez-Rodríguez and de los Campos 2014), σ^2_{u1}, σ^2_{u2}, σ^2_{u3} corresponds to variance parameters associated to $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ respectively. The rest of the terms has been already described.

P + RKHS-KA model

This model is an extension of model (3) where we include a random effect to take the additive relationship matrix derived from pedigree into account, the model is given by:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{a} + \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 + \mathbf{e}, \quad (4)$$

where all terms have been described previously and $\mathbf{a} \sim N(\mathbf{0}, \sigma^2_a\mathbf{A})$ distributed independently from $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, and $\mathbf{e}$.

Models (1)–(4) were fitted using the "BGLR" statistical package (Pérez-Rodríguez and de los Campos 2014) using the R Software (R Core Team 2018).

Assessing the genetic diversity in each selection cycles

Based on the genomic data, we computed Nei's standard genetic distance D (Nei 1972) between the 60 parents in C_0 and the phenotypically evaluated F_6 lines from the different selection cycles C_1, C_2, and C_3. Principal component analysis (PCA) was performed to assess the genetic relationship between lines. The matrix of genetic distances and PCA were generated with the packages "adegenet" and "ggplot2" using the R Software (R Core Team 2018).

Results

Variation for heritability and prediction ability of GY differs between families in the training population

The average GY of the evaluated F_4 families in the training population ranged from 5.91 to 7.23 ton ha^{-1} (Table 1). Low-to-high GY heritability was observed in the F_4 families, with an h^2 of 0.01 for family 8 (WAXBI/3/ATTILA*2/PBW65*2//MURGA) to an h^2 of 0.73 for family 4 (NELOKI//KACHU/KIRITATI). To select the parents for rapid cycle recombination, we implemented random cross-validation within and between the F_4 families. Prediction abilities using the P + RKHS-KA model within the families were significantly higher than between the families but varied widely. The highest prediction ability within families was 0.496 for family 12 (MUTUS*2/JUCHI/6/COPIO) and, between families, it was 0.310 for family 3 (NELOKI//KFA/2*KACHU). We further assessed the COP between families. Mean COPs had an overall lower range compared with prediction abilities, the most distinct family being family 2 (NELOKI/WAXBI) with a value of 0.483 (Table 1). Out of the 14 F_4 families, we selected 6 families (families 2, 3, 4, 11, 12, and 13) with the overall highest GY heritability and good prediction ability within and between families. The 10 highest-yielding F_4 lines of each family were selected as parents in C_0 (Supplementary Table 2). The average GY of the 10 F_4 lines per family varied,

Table 1. Summary statistics of 14 F_4 families including 1,609 F_4 lines in the training population.

No.	Cross/family name	No. F_4 lines	Mean—highest yielding 10 F_4 lines			Mean—all F4 lines				Prediction within families	Prediction between families	COP between families
			GY[a]	PH	DTH	GY	σ_g^2	σ_e^2	h^b			
1	PAURAQ/3/ATTILA*2/PBW65*2//W485/HD29	94	7.45	101.1	72.9	6.94	0.085	0.132	0.39	0.005	−0.028	0.353
2	**NELOKI/WAXBI[b]**	**190**	**7.65**	**98.0**	**78.0**	**7.03**	**0.072**	**0.092**	**0.44**	**0.353**	**0.037**	**0.483**
3	**NELOKI//KFA/2*KACHU**	**190**	**7.21**	**93.4**	**74.4**	**6.43**	**0.138**	**0.107**	**0.56**	**0.267**	**0.301**	**0.410**
4	**NELOKI//KACHU/KIRITATI**	**95**	**7.04**	**94.4**	**76.3**	**5.91**	**0.354**	**0.134**	**0.73**	**0.338**	**0.067**	**0.413**
5	COPIO/6/MUTUS*2/AKURI	95	7.39	97.5	73.9	6.74	0.022	0.315	0.07	0.184	−0.113	0.181
6	PARUS/FRANCOLIN#1//KFA/2*KACHU	95	7.23	101.6	75.1	6.60	0.096	0.077	0.55	0.021	0.230	0.326
7	ATTILA*2/PBW65//MUU#1/3/FRANCOLIN#1/4/KACHU/KINDE	95	7.11	97.5	76.2	6.49	0.081	0.103	0.44	−0.001	0.085	0.360
8	WAXBI/3/ATTILA*2/PBW65*2//MURGA	95	7.36	95.5	77.6	6.81	0.003	0.211	0.01	0.125	0.088	0.383
9	WAXBI/4/ATTILA*2/PBW65//MUU#1/3/FRANCOLIN#1	94	7.75	95.3	73.8	7.23	0.014	0.321	0.04	0.031	0.033	0.396
10	WAXBI//KFA/2*KACHU	95	7.54	97.3	74.8	7.00	0.084	0.177	0.32	0.093	0.218	0.345
11	**SUP152/BAJ#1//KFA/2*KACHU**	**95**	**7.66**	**99.8**	**72.1**	**6.97**	**0.121**	**0.130**	**0.48**	**0.303**	**0.250**	**0.340**
12	**MUTUS*2/JUCHI//COPIO**	**95**	**7.26**	**95.5**	**78.1**	**6.64**	**0.137**	**0.112**	**0.55**	**0.496**	**0.059**	**0.168**
13	**KACHU/KINDE//SUP152**	**95**	**7.03**	**97.7**	**72.7**	**6.33**	**0.094**	**0.117**	**0.44**	**0.179**	**0.091**	**0.434**
14	KACHU/KINDE//NELOKI	186	7.63	103.0	76.4	6.75	0.063	0.160	0.28	0.010	0.098	0.401

[a] GY: Grain yield (ton ha^{-1}), PH: Plant height (cm), DTH: Days to heading (days), σ_g^2: genotypic variance, σ_e^2: residual variance.
[b] From the families marked in bold, the 10 highest yielding lines were selected as parents to from cycle C_1.

with 2 families each showing higher, medium, and lower GY when compared to all families.

GS accuracies varied between set of lines

F_1 individuals in each recombination cycle were predicted using the entire training population, which was not updated throughout the study. GEBVs calculated using the P + RKHS-KA model ranged from 6.42 to 7.50 ton ha^{-1} among F_1s across cycles. The mean of the GEBVs constantly increased with each cycle from 6.71 ton ha^{-1} in C_0 to 7.20 ton ha^{-1} in C_3, with an average increase of 0.16. GEBV means between cycles were, however, not always significantly different (Supplementary Table 3). This steady increase of GEBVs was not apparent for the F_1 individuals that were selected as parents and the individuals that were selected to be advanced to F_6 for yield evaluation (Supplementary Tables 1–3). In both sets of selected lines, the mean GEBVs were slightly lower in C_2 compared to C_1 and increased again in C_3. In recombination cycle C_2, the smallest number of F_1s was generated, and the selected sets of parents and individuals advanced included only 30–40% of the total number of F_1s, which might explain this result.

In addition to the P + RKHS-KA model, which was the only model applied for the selection of parents in each recombination cycle, GEBVs using the RKHS-KA model without the numerical relationship matrix **A** and the standard GBLUP model with and without **A** were calculated to corroborate the correlation between GEBVs of different models in an RCRGS setting (Supplementary Table 3). The GEBVs of the 3 additional models showed very similar trends across the recombination cycles regarding the P + RKHS-KA model, while the significance between cycles varied. Interestingly, the GEBVs of the models without **A** showed lower predicted GY values and lower and nonsignificant means in cycle C_3 compared to cycle C_1 for the selected parents and the individuals that were advanced. The P + GBLUP model predicted the highest yields. The correlations between the GEBVs of the models were positive and ranged on average from 0.32 to 0.94. The correlations were highest in C_0 (0.91) and declined in the subsequent cycles. Correlations also declined in the selected subsets of F_1s in comparison to the entire F_1 population.

Rapid cycling recombination GS for grain yield increases realized genetic gains

Four groups of entries derived from C_0, C_1, C_2, and C_3 and the repeated check were used for field evaluation at CENEB across 2 crop cycles. The mean GY for each cycle and the average gain per cycle are shown in Fig. 2 and also presented in Table 2. Overall, GY in the trial was slightly lower in Year 1 (2020–2021), reaching 7.42 ton ha^{-1}, but not significantly different from Year 2 (2021–2022) with an average of 7.51 ton ha^{-1} (Table 2). In Year 1 and over the 2 years combined, the entries of the base selection cycle C_0, (using 17 out of the 60 initial parents) had the lowest GY, with an average of 6.88 ton ha^{-1} across years. The same 17 parents revealed an average GY of 7.31 ton ha^{-1} in the original training population, which was the same as the 60 initial parents used in C_0 (7.31 ton ha^{-1}) and higher than the average GY (6.71 ton ha^{-1}) across all entries in the training population (1,609 entries). In Year 2, C_0 entries showed a higher GY average than C_1 and C_2 entries (Table 2).

In both years, the performance of the 35 C_3 entries surpassed the GY of all the other cycles. The average GY among C_3 entries was 7.73 ton ha^{-1} in both years and across years. The higher GY in C_3 was not significant in Year 1 but in Year 2 and for both years combined.

The average gain per cycle for each year and combined across years ranged from −0.39 ton ha^{-1} to 1.47 ton ha^{-1}. Across both years, the realized genetic gains were 0.28 ton ha^{-1}, with the highest gains in C_1 followed by C_3 and C_2.

Unselected flowering and height traits decreased across recombination cycles

In the training population, genetic correlations of PH, DTH, and DTM with GY were in general low (−0.07, 0.11, and 0.02, respectively) and high levels of indirect selection were not expected. The effects of GS on the unselected agronomic traits PH, DTH, and DTM are presented in Supplementary Table 4. All 3 traits were only evaluated in one replication in each of the yield trials. Some plots showed segregation for one of the traits, likely since lines were derived from selected bulks. On average, lines flowered

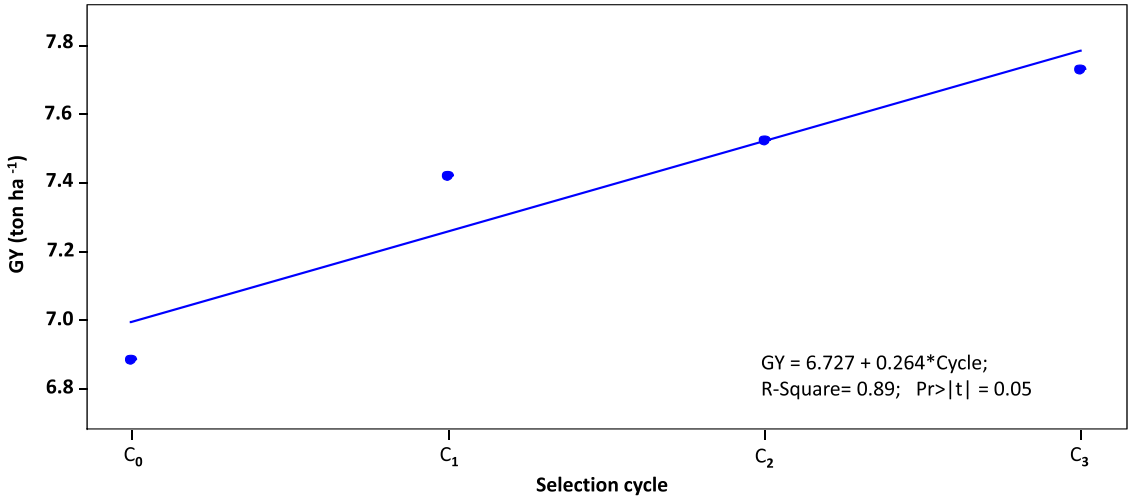

Fig. 2. Mean GY (ton ha^{-1}) for each selection cycle: C_0, C_1, C_2, and C_3.

significantly earlier (81.9 and 79.3 days) and matured earlier (130.1 and 127.7 days) as well as having a lower height (109.5 and 102.3 cm) in Year 2 when compared to Year 1, respectively. Across recombination cycles, GS on average shortened the growing cycle of the selected lines. In Year 2, no significant differences were observed between recombination cycles. Across both years combined, DTH and DTM decreased by 2.8 and 1.5 days with respect to cycle C_0, including the subset of the initial parents. During the first recombination cycle, GS produced significantly taller plants (on average 3.7 cm from C_0 to C_1), while in the subsequent cycles, PH slightly increased, but with no significant change. For all 3 traits, C_0 showed nonsignificant values compared to the check.

Genetic diversity was maintained in each of the selection cycles

The genetic diversity in each of the selection cycles computed by Nei's standard genetic distance is displayed in Table 3. The overall mean genetic distances within and between cycles were very similar. Mean genetic distances between the initial parents in C_0 were higher than between the F_6 lines in each recombination cycle, but mean distances between the F_6 lines did not significantly decline from cycle C_1 to C_3. The largest genetic distance was observed between the group of C_0 parents and the F_6 lines in C_3. Principal component analysis at a 2-dimensional scale depicted 3 groups for the initial 60 parents (Fig. 3). The 2 smaller groups (groups 1 and 3) each comprised the 10 selected parents of one F_4 family (families 2 and 12). The larger group (group 2) included the parents of the 4 additional F_4 families (families 3, 4, 11, and 13) characterized by their common parent "KACHU." Lines in each of the selection cycles follow approximately the same, but wider patterns driven by intercrossing.

Table 2. Mean yield comparison and least significant difference (LSD) of recurrent cycles evaluated at CENEB during 2020–2021 (year-1) and 2021–2022 (year-2) growing seasons.

Cycle	No. of lines	Mean GY (ton ha^{-1})	Tukey grouping	
Year 1 (Y2020–2021)		[LSD (0.05) = 0.212]		
C_0	17	6.11	A	
C_1	32	7.58	B	
C_2	34	7.60	B	
C_3	35	7.73	B	
Year 2 (Y2021–2022)		[LSD (0.05) = 0.221]		
C_0	17	7.65	A	B
C_1	32	7.26	C	
C_2	34	7.45	C	B
C_3	35	7.73	A	
Combined		[LSD (0.05) = 0.153]		
C_0	17	6.88	A	
C_1	32	7.42	B	
C_2	34	7.52	B	
C_3	35	7.73	C	

Discussion

Accelerating the genetic progress of major cultivated crops such as wheat, maize, and rice is necessary to increase production in response to the global food crisis (Bentley *et al.* 2022). In autogamous crops, bulk and pedigree methods of breeding, which are based on inbred line selection, are commonly used in genetic improvement programs. These methods, however, produce limited novel combinations of genes in a breeding population. Recurrent selection promotes recombination and produces novel combinations of genes in a breeding population, but it requires accurate single-plant evaluation. GS, which can predict the breeding value of individuals based on their marker genotype, provides the potential to give a higher reliability of single-plant evaluations and to, therefore, be effective in recurrent selection. In this study, we implemented RCRGS in bi-parental spring wheat populations of CIMMYT spring bread wheat to evaluate its feasibility and estimate potential realized genetic gain. RCRGS was applied for GY in

Table 3. Mean and standard deviation of Nei's standard genetic distance within and between lines of each selection cycle.

Cycle	No. of lines	Mean distance within cycle	Mean distance between cycles		
			C0	C1	C2
C_0	60	0.352 (±0.13)			
C_1	29	0.327 (±0.07)	0.344 (±0.09)		
C_2	32	0.320 (±0.07)	0.345 (±0.08)	0.330 (±0.07)	
C_3	31	0.333 (±0.07)	0.354 (±0.10)	0.332 (±0.08)	0.328 (±0.07)

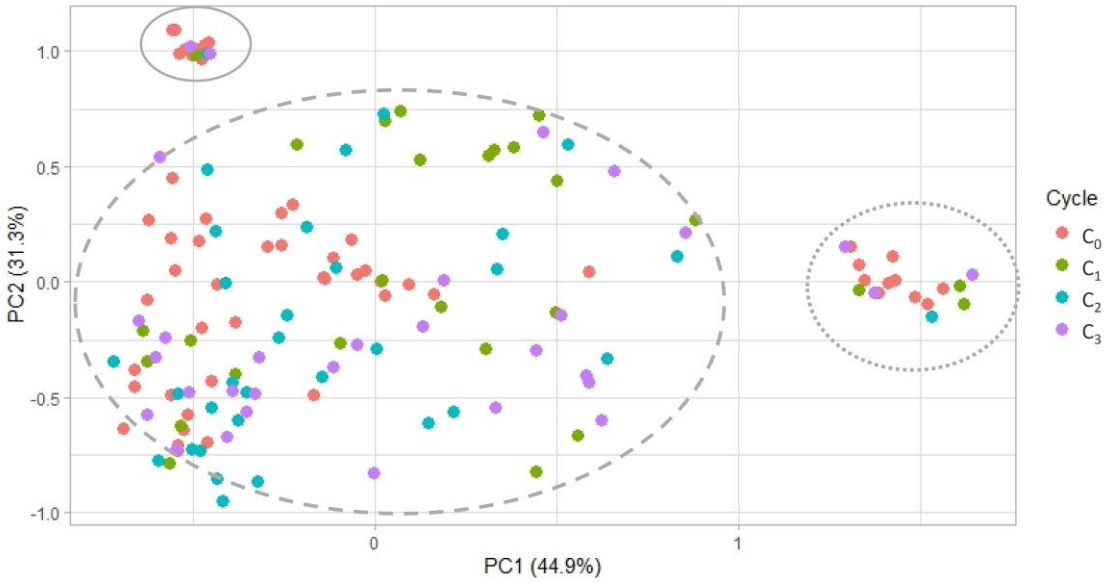

Fig. 3. Principal component analysis based on Nei's standard distance including all parents in C_0 and F_6 lines derived from selection cycles C_1, C_2, and C_3. Circles display distinct groups, group 1 (solid line), group 2 (dashed line), and group 3 (round dotted line).

recombined F_1s that originated from 14 CIMMYT crosses, based on 16 elite breeding lines.

Our results showed a consistent genetic gain for GY when summarized across 2 years of phenotyping. Genetic gain varied in percentage per cycle and in individual years and was not significant from C_1 to C_2. The highest gain was revealed from C_0 to C_1 and the lowest from C_1 to C_2. For the 2 years of phenotyping combined, genetic gain reached 12.3% from C_0 to C_3 and realized gain was 0.28 ton ha^{-1} per cycle. This genetic gain was slightly higher than expected from the GEBVs reported in each selection cycle, with an estimated realized genetic gain of 0.26 ton ha^{-1}. These differences in genetic gain are anticipated and might be explained by additional G×E interactions during field evaluation, as well as other factors for example the choice of the prediction model and the genotyping platform. GEBVs indicated a slight decline for GY from C_1 to C_2 for the F_1 individuals that were selected as parents and advanced to F_6. However, this decline was not apparent for the observed GY in field evaluations in Year 1 and across the 2 years combined, while in Year 2, GY was lower in C_1 and C_2.

To further compute the realized genetic gain per year (ton ha^{-1} year^{-1}), it is necessary to account for the number of cycles per year (2–3 cycles per year in this study) and for the time from the initial cross to the last cycle (theoretically 3.5 years from F_1 development to the harvest of the C_3F_6 lines in this study, but extended to 5 years as 2 cycles were repeated due to logistical constraints). Therefore, given that GY from C_0 (6.88 ton ha^{-1}) to C_3 (7.73 ton ha^{-1}) increased by 12.3%, the average genetic gain of 0.28 t/ha per cycle is equivalent to 0.187 ton ha^{-1} year^{-1} [i.e. (3 · 0.28)/5] under our conditions and equivalent to 0.242 ton ha^{-1} year^{-1} under optimal theoretical conditions. Crespo-Herrera *et al.* (2017) analyzed genetic gain in CIMMYT Elite Spring Wheat Yield Trial in a period of 8 years from 2006–2007 to 2014–2015 and across 426 international locations classified in 3 target populations of environments. The highest genetic gain reached 0.102 ton ha^{-1} year^{-1} in optimally irrigated environments relative to a widely grown cultivar "ATTILA" and 0.044 ton ha^{-1} year^{-1} relative to several local checks. Mondal *et al.* (2020) reported the grain yield progress in CIMMYT spring bread wheat over 50 years determined in field trials during 5 crop seasons performed at CENEB under simulated field conditions. The highest genetic gains per year accounted for 0.035 and 0.031 ton ha^{-1} year^{-1} under irrigated and rainfed (limited drip irrigation) conditions, respectively. Therefore, the short-term genetic gain from RCRGS observed in the populations used in this study (0.187 ton ha^{-1} year^{-1}) is higher (up to six times) than observed in previous CIMMYT studies under phenotypic selection, which, however, were longer-term studies between 8 and 50 years and were achieved in national trials in the case of Crespo-Herrera *et al.* (2017). The results we obtained reinforce the results by Bonnett *et al.* (2022) and highlight the potential of GS-assisted recombination at early breeding generations for achieving high genetic gain for GY.

Our study presents the second empirical report of RCRGS in wheat and the first for a complex trait such as GY. In a previous study, Veenstra *et al.* (2020) reported the improvement in nutritional quality of wheat via recurrent GS. The authors determined the realized genetic gain from GS for wheat grain fructan content by applying truncated selection (TS) and optimized contribution selection (OCS). GS led to a 25 ± 12% and 24 ± 6.4% increase in wheat grain fructans using TS and OCS, respectively. OCS showed a simultaneously greater retention of genetic variance and lower inbreeding levels.

High rates of inbreeding per breeding cycle with GS have been observed in simulations and empirical studies (Jannink *et al.* 2010; Rutkoski *et al.* 2015; Lin *et al.* 2017; Gorjanc *et al.* 2018). Rutkoski *et al.* (2015) found significant increases in inbreeding after 1 and 2 cycles of GS when compared with C_0, significantly greater than the expected value under random genetic drift for all populations. Several methods to control the rate of inbreeding have, therefore, been proposed, including OCS (Meuwissen 1997) or optimal cross-selection (Gorjanc *et al.* 2018), in the population improvement context to improve selection and crossing plans. In this study, we only used TS and did not specifically consider maintaining genetic diversity in our crossing plans. Genetic diversity declined comparing cycle C_0 with the subsequent recombination cycles but was well maintained from cycle C_1 to C_3. Thus, our results only partially agree with the findings reported in earlier studies with a reduced genetic variation. Zhang *et al.* (2017) reported similar results deploying rapid-cycle GS in multi-parental tropical maize populations, with a slightly narrowed genetic

diversity only during the last GS cycles (C_3 and C_4). In this study, we balanced crosses within and between bi-parental F1s for each recombination cycle, generating 3 times more crosses between bi-parental F1s than within F1s, which we propose to be a reason that genetic diversity remained at a similar level throughout recombination cycles.

RCRGS was associated with a change in agronomic traits that were measured (DTH, DTM, and PH) in our study. Lines in cycle C_3 had a shorter crop cycle, and lines in cycles C_1 to C_3 were taller. We, therefore, suggest that selection should be applied through a selection index to optimize selection of multiple traits. Each additional trait added to a selection index usually takes away some of the selective pressure that can be applied to other traits. Furthermore, tradeoffs exist between progress in one trait versus others. Nonetheless, several recently published studies show an increase in the prediction accuracy of genomic multi-trait selection over genomic single-trait selection (Montesinos-López *et al.* 2016, 2019). It will be important to further investigate multi-trait selection to optimize the predictive power of RCRGS.

We calculated GEBVs based on the nonlinear Gaussian kernel function, including the relationship matrix **A** (P + RKHS-KA) for the selection of new parents in each cycle. We favored this model because it demonstrated a significant positive correlation between observed yields of $F_{2:4}$ lines and predicted GEBVs of F_2 single plants in the study of Bonnett *et al.* (2022). The predictions of F_2s in Bonnett *et al.* (2022) derived from crosses between inbreds that were part of the training population showed very little to no correlation between models. In contrast, the correlation of GEBVs derived from different models in this study was positive and still moderate after 3 cycles of recombination. Models including the pedigree information predicted higher yields and genetic gain calculated from the GEBVs in each selection cycle. These were closer to the observed realized genetic gain than models only using the marker information, which underlines our previous results that when molecular markers and pedigree information are considered jointly, prediction abilities are slightly but consistently superior to the marker or pedigree-derived models alone (de los Campos *et al.* 2009; Crossa *et al.* 2010).

Rapid generation advance or speed breeding can achieve up to 6 generations by year for spring wheat (Watson *et al.* 2018) with adequate infrastructure and trained staff in place. In this study, we achieved 3 to 4 crop cycles per year by taking several practical considerations in the RCRGS scheme into account. A very fast crop cycle provides only a short time from planting to flowering. In an RCRGS breeding scheme, breeding teams need to acquire DNA from seedling tissue, receive genotypic data, and run the statistical models to make parental predictions prior to the plants reaching the flowering stage, demonstrating a logistical challenge that requires careful planning and good communication within the team and with external genotyping providers, which are regularly used in public breeding programs. In addition, for repeated crossing in recombination cycles, male and female parents need to be sown at 2 to 3 different dates, to match the flowering of the selected parents. This extends the length of the greenhouse cycle, and some crosses might fail, making a full standardization of the scheme (with a constant number of crosses and offspring) difficult. Also, greenhouse-grown plants in pots are usually smaller and produce less seed. For the 3 recombination cycles in this study, the seed of the F_1 individuals was limited in several cases. Being potential new parents, F_1s were sown at 3 dates for subsequent crossing, and insufficient seed remained for selfing. It could, therefore, be recommended to apply GS at the F_2 or F_3 breeding generation to bypass the limited amount of seed for selfing (Gorjanc *et al.* 2018). We performed 3 recurrent GS cycles and evaluated derived lines at the end of the experiment. In a 2-part strategy as suggested by Gaynor *et al.* (2017), selected plants should be advanced directly for product development. This implies that recurrent cycles in the greenhouse must be aligned to the crop cycles of product development in the field. Overall, these and other logistical constraints remain a barrier to the practical application and implementation of RCRGS for many breeding programs.

Conclusions

Different GS strategies are likely to be relevant in individual breeding programs and each program must determine which strategy is the best choice. This will be specific to the biological specificities of a crop, the breeding organization itself, and its economic context. Wheat breeding programs tend to use GS to control for G × E interaction, predicting the total genetic values of individuals, where both additive and nonadditive effects determine the final commercial value of the lines. Practical implementation of rapid-cycling GS strategies in wheat is still lacking and our study shows the potential of RCRGS to increase genetic gains for GY. Further work is needed to evaluate and optimize these GS strategies in wheat and other crop species in order to support the acceleration of current breeding progress.

Data availability

The genotypic and phenotypic data of the training population and final F_6 lines underlying this article are available in the CIMMYT data repository Dataverse (https://hdl.handle.net/11529/10548816).

Supplemental material available at G3 online.

Acknowledgments

The authors are grateful to their CIMMYT colleagues for intensive discussions on the subject, CIMMYT field, and lab technicians who helped to record phenotypic and genotypic data. We especially wish to express our sincere gratitude to Alma Yuyutzy Lopez-Garcia and Jose Pablo Alva-Galindo, who took care of the recombination cycles in the greenhouse.

Funding

Support for this study was provided by Monsanto's Beachell-Borlaug International Scholars Program and the CGIAR Research Program on Wheat. We are also thankful for the financial support provided by the Bill & Melinda Gates Foundation and the Foreign, Commonwealth & Development Office [INV-003439, BMGF/FCDO, Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods], USAID [USAID Amend. No. 9 MTO 069033 USAID-CIMMYT Wheat/AGGMW, Supplementary Project], the Foundation for Research Levy on Agricultural Products (FFL) and the Agricultural Agreement Research Fund (JA) through the Research Council of Norway for grants 301835 and 320090 and the One CGIAR Accelerated Breeding Initiative.

Conflicts of interest

None declared.

Literature cited

Atanda SA, Govindan V, Singh R, Robbins KR, Crossa J, Bentley AR. Sparse testing using genomic predication improves selection for breeding targets in elite spring wheat. Theor Appl Genet. 2022; 135(6):1939–1950. doi:10.1007/s00122-022-04085-0.

Bentley AR, Donovan J, Sonder K, Baudron F, Lewis JM, Voss R, Rutsaert P, Poole N, Kamoun S, Saunders DG, *et al.* Near- to long-term measures to stabilize global wheat supplies, and food security. Nat Food. 2022;3(7):483–486. doi:10.1038/s43016-022-00559-y.

Bernardo R, Yu J. Prospects for genome-wide selection for quantitative traits in maize. Crop Sci. 2007;47(3):1082–1090. doi:10.2135/cropsci2006.11.0690.

Beyene Y, Gowda M, Olsen M, Robbins KR, Pérez-Rodríguez P, Alvarado G, Dreher K, Gao SY, Mugo S, Prasanna BM, *et al.* Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections. Front Plant Sci. 2019;10:1502. doi:10.3389/fpls.2019.01502.

Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, *et al.* Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci. 2015; 55(1):154–163. doi:10.2135/cropsci2014.07.0460.

Bonnett D, Li Y, Crossa J, Dreisigacker S, Basnet B, Pérez-Rodríguez P, Alvarado G, Jannink JL, Poland J, Sorrells M. Response to early generation genomic selection for yield in wheat. Front Plant Sci. 2022;12:718611. doi:10.3389/fpls.2021.718611.

Burgueño J, Crossa J, Cornelius PL, Trethowan R, McLaren G, Krishnamachari A. Modeling additive × environment and additive × additive × environment using genetic covariances of relatives of wheat genotypes. Crop Sci. 2007;47(1):311–320. doi:10.2135/cropsci2006.09.0564.

Crespo-Herrera LA, Crossa J, Huerta-Espino J, Autrique E, Mondal S, Velu G, Vargas M, Braun HJ, Singh RP. Genetic yield gains in CIMMYT'S International elite spring wheat yield trials by modeling the genotype × environment interaction. Crop Sci. 2017;57(2): 789–801. doi:10.2135/cropsci2016.06.0553.

Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, de los Campos G, Burgueño J, Windhausen VS, Buckler E, *et al.* Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 (Bethesda). 2013;3(11):1903–1926. doi:10.1534/g3.113.008227.

Crossa J, Burgueño J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A. Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci. 2006;46(4): 1722–1733. doi:10.2135/cropsci2005.11-0427.

Crossa J, de los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, *et al.* Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics. 2010;186(2):713–724. doi: 10.1534/genetics.110.118521.

Crossa J, Fritsche-Neto R, Montesinos-Lopez OA, Costa-Neto G, Dreisigacker S, Montesinos-Lopez A, Bentley AR. The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. Front Plant Sci. 2021;12:651480. doi:10.3389/fpls.2021.651480.

Crossa J, Jarquín D, Franco J, Pérez-Rodríguez P, Burgueño J, Saint-Pierre C, Vikram P, Sansaloni C, Petroli C, Akdemir D, *et al.* Genomic prediction of gene bank wheat landraces. G3 (Bethesda). 2016;6(7):1819–1834. doi:10.1534/g3.116.029637.

Crossa J, Martini JWR, Gianola D, Pérez-Rodríguez P, Jarquín D, Juliana P, Montesinos-López O, Cuevas J. Deep kernel and deep learning for genome-based prediction of single traits in multi-environment breeding trials. Front Genet. 2019;10:1168. doi:10.3389/fgene.2019.01168.

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, Jarquín D, de Los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, *et al.* Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 2017;22(11): 961–975. doi:10.1016/j.tplants.2017.08.011.

Crossa J, Pérez P, de los Campos G, Mahuku G, Dreisigacker S, Magorokosho C. Genomic selection and prediction in plant breeding. J Crop Improv. 2011;25(3):239–261. doi:10.1080/15427528.2011.558767.

Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, *et al.* Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity (Edinb). 2014;112(1):48–60. doi:10.1038/hdy.2013.16.

de los Campos G, Gianola D, Rosa GJM, Weigel KA, Crossa J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet Res. 2010;92(4):295–308. doi:10.1017/S0016672310000285.

de los Campos G, Klimentidis YC, Vazquez AI, Allison DB. Prediction of expected years of life using whole-genome markers. PLoS One. 2012;7(7):e40964. doi:10.1371/journal.pone.0040964.

de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics. 2009;182(1):375–385. doi:10.1534/genetics.109.101501.

Dreisigacker S, Crossa J, Pérez-rodríguez P, Montesinos-López O, Rosyara U, Juliana P, Mondal S, Crespo-Herrera L, Govindan V, Singh RP, *et al.* Implementation of genomic selection in the CIMMYT global wheat program, findings from the past 10 years. Crop Breed Genet Genomics. 2021;3(2):e210005. doi:doi.org/10.20900/cbgg20210005.

Endelman JB. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome. 2011;4(3):3. doi:10.3835/plantgenome2011.08.0024.

Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R, Mackay IJ, Hickey JM. A two-part strategy for using genomic selection to develop inbred lines. Crop Sci. 2017;57(5):2372–2386. doi:10.2135/cropsci2016.09.0742.

Gholami M, Wimmer V, Sansaloni C, Petroli C, Hearne SJ, Covarrubias-Pazaran G, Rensing S, Heise J, Pérez-Rodríguez P, Dreisigacker S, *et al.* A comparison of the adoption of genomic selection across different breeding institutions. Front Plant Sci. 2021;12:728567. doi:10.3389/fpls.2021.728567.

Gianola D, Fernando RL, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics. 2006; 173(3):1761–1776. doi:10.1534/genetics.105.049510.

González-Camacho JM, Crossa J, Pérez-Rodríguez P, Ornella L, Gianola D. Genome-enabled prediction using probabilistic neural network classifiers. BMC Genomics. 2016;17(1):1–16. doi:10.1186/s12864-016-2553-1.

González-Camacho JM, de los Campos G, Pérez P, Gianola D, Cairns JE, Mahuku G, Babu R, Crossa J. Genome-enabled prediction of genetic values using radial basis function neural networks. Theor Appl Genet. 2012;125(4):759–771. doi:10.1007/s00122-012-1868-9.

Gorjanc G, Gaynor RC, Hickey JM. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theor Appl Genet. 2018;131(9):1953–1966. doi:10.1007/s00122-018-3125-3.

Heslot N, Akdemir D, Sorrells ME, Jannink JL. Integrating environmental covariates and crop modeling into the genomic selection

framework to predict genotype by environment interactions. Theor Appl Genet. 2014;127(2):463–480. doi:10.1007/s00122-013-2231-5.

Heslot N, Yang H, Sorrells ME, Jannink J. Genomic selection in plant breeding: a comparison of models. Crop Sci. 2012; 52(1):146–160. doi:10.2135/cropsci2011.06.0297.

Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, Grondona M, Zambelli A, Windhausen VS, Mathews K, *et al.* Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci. 2014;54(4):1476–1488. doi:10.2135/cropsci2013.03.0195.

Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics. 2010;9(2):166–177. doi:10.1093/bfgp/elq001.

Lin Z, Wang J, Cogan NO, Pembleton L, Badenhorst P, Forster JW, Spangenberg GC, Hayes BJ, Daetwyler HD. Optimizing resource allocation in a genomic breeding program for perennial ryegrass to balance genetic gain, cost, and inbreeding. Crop J. 2017;57(1): 243–252. doi:10.2135/cropsci2016.07.0577.

Long N, Gianola D, Rosa GJM, Weigel KA, Kranis A, González-Recio O. Radial basis function regression methods for predicting quantitative traits using SNP markers. Genet Res (Camb). 2010;92(3): 209–225. doi:10.1017/S0016672310000157.

Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink JL, Singh RP, Autrique E, de los Campos G. Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 (Bethesda). 2015;5(4): 569–582. doi:10.1534/g3.114.016097.

Lorenzana RE, Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet. 2009;120(1):151–161. doi:10.1007/s00122-009-1166-3.

Martini JWR, Molnar TL, Crossa J, Hearne SJ, Pixley KV. Opportunities and challenges of predictive approaches for harnessing the potential of genetic resources. Front Plant Sci. 2021;12:674036. doi: 10.3389/fpls.2021.674036.

McLaren CG, Bruskiewich RM, Portugal AM, Cosico AB. The international rice information system. A platform for meta-analysis of rice crop data. Plant Physiol. 2005;139(2):637–642. doi:10.1104/pp.105.063438.

McLaren CG, Ramos L, Lopez C, Eusebio W. Applications of the genealogy management system. In: McLaren C, White J, Fox P, editors. International Crop Information System. Technical Development Manual, version VI. Mexico DF: CIMMYT; 2000. p. 8–13.

Meuwissen TH. Maximizing the response of selection with a predefined rate of inbreeding. J Anim Sci. 1997;75(4):934–940. doi:10.2527/1997.754934x.

Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome wide dense marker map. Genetics. 2001;157(4): 1819–1829. doi:10.1093/genetics/157.4.1819.

Mondal S, Dutta S, Crespo-Herrera L, Huerta-Espino J, Braun HJ, Singh RP. Fifty years of semi-dwarf spring wheat breeding at CIMMYT: grain yield progress in optimum, drought and heat stress environments. Field Crop Res. 2020;250:107757. doi:10.1016/j.fcr.2020.107757.

Montesinos-López OA, Montesinos-López A, Crossa J, Toledo FH, Pérez-Hernández O, Eskridge KM, Rutkoski J. A genomic Bayesian multi-trait and multi-environment model. G3 (Bethesda). 2016;6(9):2725–2774. doi:10.1534/g3.116.032359.

Montesinos-López OA, Montesinos-López A, Hernández MV, Ortiz-Monasterio I, Pérez-Rodríguez P, Burgueño J, Crossa J. Multivariate Bayesian analysis of on-farm trials with multiple-trait and multiple-environment data. Agron J. 2019;111(6): 2658–2669. doi:10.2134/agronj2018.06.0362.

Nei M. Genetic distance between populations. Am Naturalist. 1972; 106(949):283–292. doi:10.1086/282771.

Pérez-Rodríguez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. Genetics. 2014; 198(2):483–495. doi:10.1534/genetics.114.164442.

Pérez-Rodríguez P, Gianola D, González-Camacho JM, Crossa J, Manès Y, Dreisigacker S. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3 (Bethesda). 2012;2(12):1595–1605. doi:10.1534/g3.112.003665.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. https://cran.r-project.org/.

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci U S A. 2012;109(23):8872–8877. doi:10.1073/pnas.1120813109.

Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J, Jannink JL, Sorrells ME. Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. Plant Genome. 2015;8(2):eplantgenome2014.10.0074. doi:10.3835/plantgenome2014.10.0074.

Sansaloni C, Franco J, Santos B, Percival-Alwyn L, Singh S, Petroli C, Campos J, Dreher K, Payne T, Marshall D, *et al.* Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. Nat Commun. 2020;11(1):4572. doi:10.1038/s41467-020-18404-w.

Veenstra LD, Poland J, Jannink JL, Sorrells ME. Recurrent genomic selection for wheat grain fructans. Crop Sci. 2020;60(3):1499–1512. doi:10.1002/csc2.20130.

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J. 2014;12(6): 787–796. doi:10.1111/pbi.12183.

Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Asyraf Md Hatta M, Hinchliffe A, Steed A, Reynolds D, *et al.* Speed breeding is a powerful tool to accelerate crop research and breeding. Nat Plants. 2018;4(1):23–29. doi:10.1038/s41477-017-0083-8.

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink J, Sorrells ME, Raman B, Cairns JE, Tarekegne A, Semagn K, *et al.* Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 (Bethesda). 2012;2-(11):1427–1436. doi:10.1534/g3.112.003699.

Zhang X, Pérez-Rodríguez P, Burgueño J, Olsen M, Buckler E, Atlin G, Prasanna BM, Vargas M, San Vicente F, Crossa J. Rapid cycling genomic selection in a multiparental tropical maize population. G3 (Bethesda). 2017;7(7):2315–2326. doi: 10.1534/g3.117.043141.

Zhao Y, Gowda M, Liu W, Ranc N, Reif JC. Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet. 2012;124(4):769–776. doi:10.1007/s00122-011-1745-y.

Editor: A. Lipka