Received: 8 April 2022 | Revised: 23 September 2022 | Accepted: 20 October 2022

DOI: 10.1111/pbr.13061

ORIGINAL ARTICLE

Plant Breeding WILEY

Evaluation of strategies to optimize training populations for genomic prediction in oat (*Avena sativa*)

Espen Sannes Sørensen[1,2] | Constantin Jansen[1] | Susanne Windju[1] | Jose Crossa[4] | Anna Kristina Sonesson[3] | Morten Lillemo[2] | Muath Alsheikh[1,2]

[1]Graminor AS Ridabu, Ridabu, Norway

[2]Department of Plant Science, Norwegian University of Life Sciences, Ås, Norway

[3]Nofima AS, Ås, Norway

[4]International Maize and Wheat Improvement Center, Texcoco, Mexico

Correspondence
Espen Sannes Sørensen, Department of Plant Science, Norwegian University of Life Sciences, Ås, Norway.
Email: espen.sorensen@nmbu.no

Funding information
Research Council of Norway and Graminor, Grant/Award Number: 286606

Abstract

Genomic selection is a promising breeding methodology that could increase selection accuracy and intensity and reduce generation interval. As the cost of genotyping decreases, it will be important to optimize training populations for costly phenotypic experiments for many complex traits. The aim of this research was to evaluate different optimization strategies, by using historical data from the Norwegian oat breeding programme at Graminor. In this paper, we focus on the optimization criteria: genetic diversity, phenotypic variance and genetic similarity between the training and testing populations. The four training population strategies—prediction core, diversity core, phenotypic selection and random selection—were applied to an oat candidate population of 1124 lines. An independent testing population was used to calculate the mean prediction abilities for the traits days to heading and plant height. Moreover, the strategies were tested in three independent wheat populations. The results showed that prediction core was the most promising strategy to select training populations with high genetic similarity to the testing set, high genetic diversity, and high phenotypic variance, which resulted in higher prediction ability across population sizes and traits.

KEYWORDS
genetic diversity, genetic similarity, optimization criteria, phenotypic variance, prediction ability, training population

1 | INTRODUCTION

The genetic gains per year of conventional breeding have been estimated to be 1% (Li et al., 2018). With the introduction of new molecular DNA-based technologies, breeders can increase selection accuracy and intensity and reduce the generation interval. This increased breeding efficiency (Heffner et al., 2010; Bhat et al., 2016; Xu et al., 2020) is key to increasing food production in the future. A promising marker-based breeding technique is genomic selection (GS; Meuwissen et al., 2001, Crossa et al., 2017, Wang et al., 2018), which uses whole-genome DNA markers and phenotypic information of a training population to predict the marker effects of a specific trait using statistical models. The marker effects are used to predict the breeding values of non-phenotyped individuals called testing population. GS has become a more available breeding methodology in recent years. As genotyping costs continue to decrease, cost of phenotyping will become the limiting factor of GS (Bhat et al., 2016).

Although GS in plant breeding was considered challenging (Desta & Ortiz, 2014), it has been successfully implemented in cereal crops, for example, wheat and barley (Ankamah-Yeboah et al., 2020;

Larkin et al., 2019), and has great potential for improved selection for yield and disease resistance in oats (Haikka, Knürr, et al., 2020; Haikka, Manninen, et al., 2020; Mellers et al., 2020). Genomics-resources and marker systems were for a long time limited in oats due to the complexity of the oat genome and reduced research investments compared with other major crops (Latta et al., 2019). However, the development of the 6K-SNP chip (Tinker et al., 2014) made GS more available for implementation in oat breeding and has already been implemented in Nordic breeding programmes (Ceplitis, 2014).

The composition of the training population highly affects the prediction ability, which is crucial for successful implementation of GS (Akdemir & Isidro-Sánchez, 2019; Berro et al., 2019; Crossa et al., 2016). The main criteria for training population optimization are (i) population size, (ii) genetic diversity, (iii) phenotypic diversity (Isidro et al., 2015), (iv) genetic relationship between the training and testing population (Crossa et al., 2014), and (v) degree of population structure (Werner et al., 2020). Some of these criteria could be more important than others for different traits, populations, and species (Crossa et al., 2010). Optimizing the training population is especially useful when phenotyping costs are high in traits with low heritability and in cases of high genotype by environment interaction. High heritability is also related to high prediction ability but is not something we try to optimize in this study.

By using the criteria mentioned above we have evaluated three different strategies for training population optimization. The first strategy preserves the genetic diversity and population structure from a larger population in smaller training populations (Crossa et al., 2016; Franco et al., 2005). The second strategy uses the genetic relationship between the training and testing population to identify individuals that have the lowest mean prediction error variance (PEV; Rincent et al., 2012, Isidro et al., 2015). The third strategy is based on selecting training populations with high phenotypic variation for a specific trait.

The goal of this study was to use historical data and breeding lines from the Norwegian oat breeding programme at Graminor, to develop an optimal training population for further research. The strategies mentioned above were applied to a large candidate population, with an independent breeding population as testing population. The main hypothesis is that an optimization strategy will give higher prediction abilities than a random selection. The optimization criteria genetic diversity, phenotypic diversity, and genetic similarity between training and testing population were analysed in all strategies. A wheat dataset from CIMMYT was used to validate the strategies in a completely independent breeding germplasm. The outcome of this study could contribute to the implementation of GS in commercial plant breeding programmes.

2 | MATERIALS AND METHODS

2.1 | Germplasm

Oat lines in this study were provided by Graminor plant breeding company and are listed in Table S1. Summary of the number of lines, SNP-markers, environments, and heritability is given in Table 1. Table 2 shows the number of lines tested in each location and year, and Table S8 shows the percentage of overlapping lines between the environments. All lines have been evaluated for the traits days to heading (DTH) and plant height (PH) by Graminor and the Norwegian University of Life Sciences from yield trials that were a randomized complete block design with plot size of 1.5 m × 5 m, and irrigated disease trials that were an alpha lattice design (Patterson & Williams, 1976) with plot size of 1.5 m × 1.25 m. Spatial variation was analysed by using nearest neighbour for yield trials (Cover & Hart, 1967) and alpha lattice for the disease trials. Plant height was collected by measuring the height of the plant from the ground to the top of the head 2–3 weeks after heading. Days to heading were recorded as the number of days from sowing until the date when 50% of the heads have emerged more than 50% from the flag leaf.

TABLE 1 Summary description of oat germplasm, genotypic data and phenotypic data used in this study

	Training population	Testing population
Population size	1124	257
Number of SNP markers	3022	3022
Locations	4	4
Years	5	1
Heritability of plant height	.71	.82
Heritability of days to heading	.62	.90

Note: The number of years and locations refers to the phenotypic trials of the training population candidates, which were used to calculate the heritability of the traits.

The training population candidates consisted of 65% F_9 and 16% F_{10}–F_{12} breeding lines from Graminor, and 19% are a collection of diverse material from Europe, North America and Australia. The testing population consisted of 257 Graminor F_9 breeding lines from 2019. The F_9 lines were tested for at least one year at three locations, the F_{10}–F_{13} lines were tested for at least two years at four locations, and the diverse materials were tested for at least two locations in 2016 and one in 2017.

2.2 | Phenotypic data

The phenotypic data used in the genomic prediction models come from a two-stage analysis. The first stage is the calculated adjusted mean values from field designs to account for the effect of replicate and block. The second stage is to use adjusted mean values in mixed linear models to account for the environmental effects of year, locations and experiment within the same environment. The following models were used in stage two:

$$\mathbf{y} = \mathbf{b_0} + \mathbf{b_1 x_g} + \mathbf{b_2 x_l} + \mathbf{b_3 x_y} + \mathbf{b_4 x_{ly}} + \mathbf{e} \quad (1)$$

TABLE 2 Number of lines tested in each year and location for the training population candidates

Year	Bjørke (60.80°N, 11.20°E)	Staur (60.73°N, 11.10°E)	Rød (59.34°N, 10.89°E)	Vollebekk (59.66°N, 10.75°E)
2014	304	304	304	34
2015	136	153	136	93
2016	174	344	174	289
2017	407	440	407	260
2018	337	356	337	357
2019	257	257	257	257

$$\mathbf{y} = \mathbf{b_0} + \mathbf{b_1 x_g} + \mathbf{b_2 x_l} + \mathbf{b_3 x_t(x_l)} + \mathbf{e} \quad (2)$$

$$\mathbf{y} = \mathbf{b_0} + \mathbf{b_1 x_g} + \mathbf{b_2 x_l} + \mathbf{e} \quad (3)$$

In the equations (Equations 1–3) **y** is the response phenotype, $\mathbf{b_0}$ the intercept, $\mathbf{b_1}$-$\mathbf{b_4}$ are coefficients, $\mathbf{x_g}$ the fixed effect of genotype, $\mathbf{x_l}$ the random effect of location, $\mathbf{x_y}$ the random effect of year, $\mathbf{x_t}$ the random effect of trial, $\mathbf{x_{ly}}$ the interaction between year and location, $\mathbf{x_t\,(x_l)}$ the effect of experiment nested in location, and **e** the error term.

Equation (1) was used on plant height in the training population, Equation (2) on days to heading in both training and testing population, and Equation (3) on plant height in the testing population. Plant height and days to heading were normalized with different models because data collected from the irrigated disease trials differed for days to heading compared with the other trials in the same year and location, so the factor of experiment was added to the model. The factor of experiment also contains the effect of year as the same experiments are only tested for one year. For plant height, it was sufficient to use year and location as factor. Material with phenotypic values two standard deviations from the mean were excluded as the distribution became skewed.

2.3 | Heritability

The broad sense heritability (h^2) was calculated as

$$h^2 = V_G/V_P \quad (4)$$

where V_G is the variance of genotype and V_P is the variance of phenotype. V_P is equal to the $V_G + V_e$, where V_e is the variance of error. V_G was estimated using the x_g term using the following mixed linear model:

$$\mathbf{y} = \mathbf{b_0} + \mathbf{b_1 x_g} + \mathbf{b_2 x_l} + \mathbf{b_3 x_t(x_l)} + \mathbf{e} \quad (5)$$

where **y** is the response phenotype, $\mathbf{b_0}$ is the intercept, $\mathbf{b_1}$–$\mathbf{b_3}$ are coefficients, $\mathbf{x_g}$ is the random effect of genotype, $\mathbf{x_l}$ is the fixed effect of location, $\mathbf{x_t\,(x_l)}$ is the fixed effect of trial nested in location, and **e** is the error using the Minitab software (Minitab, 2010). This calculation accounts for the fixed effect of environment, leaving only the effect of genotype and error in Equation (4).

2.4 | Genotyping

All lines were genotyped with a customized, unpublished 20 k SNP chip. The genetic data were analysed and filtered with a 10% missing values threshold and 5% MAF based on the training population candidates, resulting in 3022 polymorphic markers. The missing marker data were imputed with the 'impute' function and 'means' method with the package 'e1071' in the R statistical software (Meyer et al., 2021).

2.5 | Experimental design

Each optimization strategy was repeated 20 times for each population size of 100, 200, 300, 400 and 500. Average prediction ability was calculated as the average correlation between predicted and observed breeding values of the testing population. Bayesian ridge regression (BRR) was used to compute the marker effects, and the 'BGLR' function of the 'BGLR' package in the R software (Pérez & de los Campos, 2014) was used to calculate the genomic estimated breeding values of the testing population. The number of iterations were set to 30,000 and the burnin to 15,000.

2.6 | Training population optimization strategies

This study aimed to optimize known training population criteria. Each strategy was compared with a random selection. The correlation between the optimization criteria and the prediction abilities were calculated and tested for significance with ANOVA.

2.6.1 | Phenotypic selection

Phenotypic selection aims to maximize phenotypic variation in the training populations and is abbreviated to PheSe for the rest of the paper. Based on the MLM output data (Figure 1), equal proportions of lines with the most extreme highest and lowest adjusted breeding values were selected for the PheSe populations. This was done once for each population size and not replicated 20 times like the other strategies. A similar approach was proposed by Zhao et al. (2012) in a slightly different premise. They argue that a fraction of the training

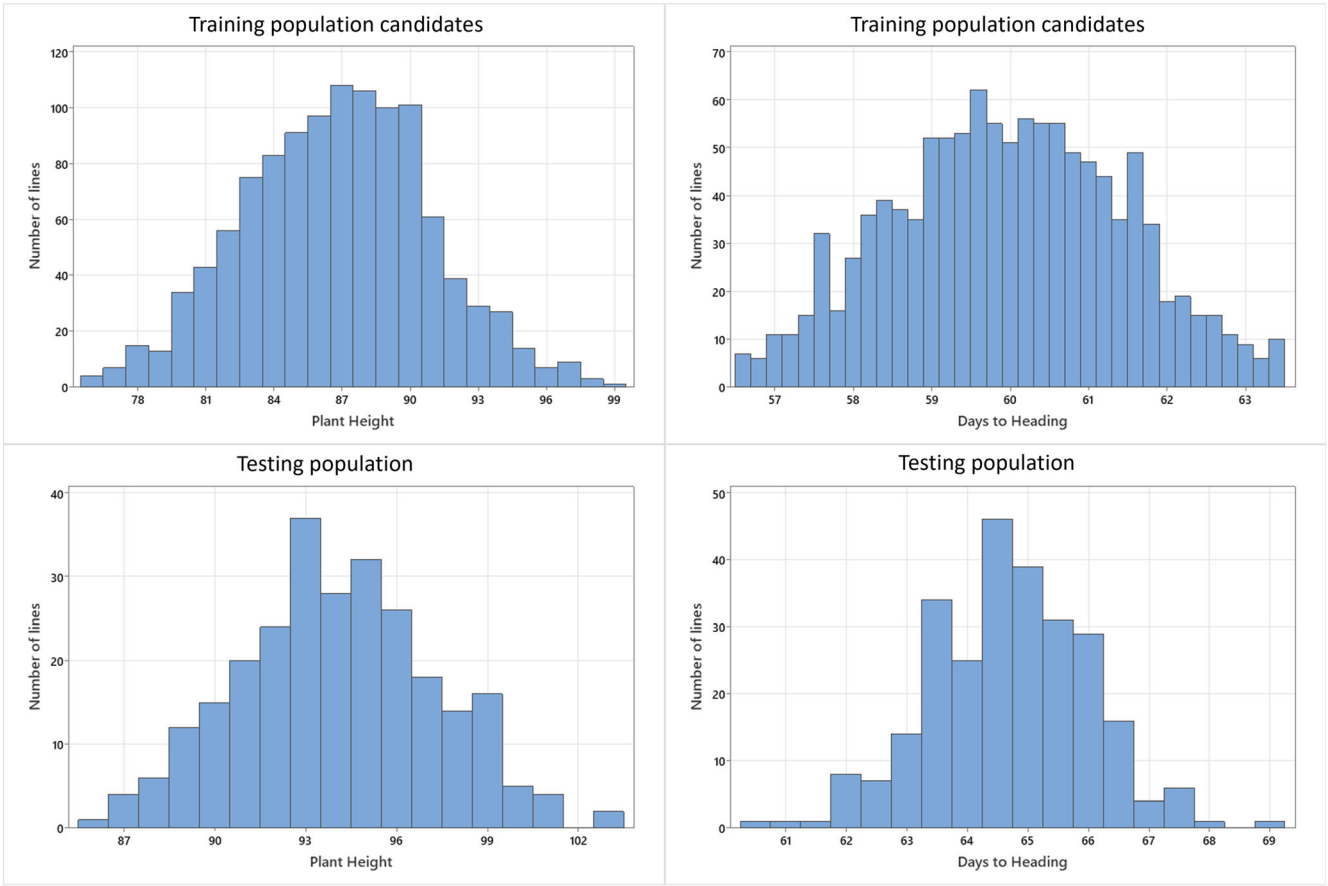

FIGURE 1 Phenotypic distribution of the training population candidates and testing population for days to heading and plant height after applying the models in Equations (1)–(3) [Color figure can be viewed at wileyonlinelibrary.com]

population should consist of inferior material to increase prediction accuracy (Zhao et al., 2012).

2.6.2 | Prediction core

The prediction core strategy aims to optimize the genetic relationship between the training and testing population by minimizing the PEV of the testing population, while also maintaining high diversity in the training population, which is done by calculating an optimization criterion called coefficient of determination (CD; Laloë, 1993). In the rest of the paper prediction core is abbreviated to PreCo. The strategy was published in 2012 by Rincent et al. (2012) and made into an R package by Akdemir (2018).

PreCo populations were selected by performing principal component analyses (PCA) on the genetic markers of the training and testing populations (Akdemir et al., 2015). The first 100 principal components (PCs) were used as input for a selection algorithm using the function 'GenAlgForSubsetSelection' in the R package 'STPGA', which starts off with a random sample, calculates the CD values and replaces one genotype at the time until it finds one that increases or gives the same the CD value. This process is repeated until no further increase in CD values is achieved (Akdemir, 2018). CDMEAN2 was used as selection criterion. The arguments of the function were set to npop = 300, nelite = 20, niterations = 5000, and minitbefstop = 1000. 'Npop' refers to the number of crosses in the testing population, and 'nelite' refers to the number of parents used. We chose higher parameters than required in order to give the algorithm more power and better solutions. The 'niterations' argument is the maximum number of iterations the selections algorithm use to find the optimal solution, whereas the 'minitbefstop' argument is the number of equal solution required for the algorithm to stop before the maximum is reached.

2.6.3 | Diversity core

The diversity core strategy aim to preserve the genetic diversity and population structure from the total candidate population in smaller populations (Crossa et al., 2016; Franco et al., 2005). Hereafter, diversity core is abbreviated to DivCo.

DivCo populations were selected by performing a structure analysis with the software STRUCTURE (Hubisz et al., 2009), and the structure harvester (Earl & vonHoldt, 2012) to determine the optimal number of clusters. A dendrogram was created with the 'hclust' function which performs a hierarchical clustering of a distance matrix based on the genetic markers. The Ward.D2 method was used in the

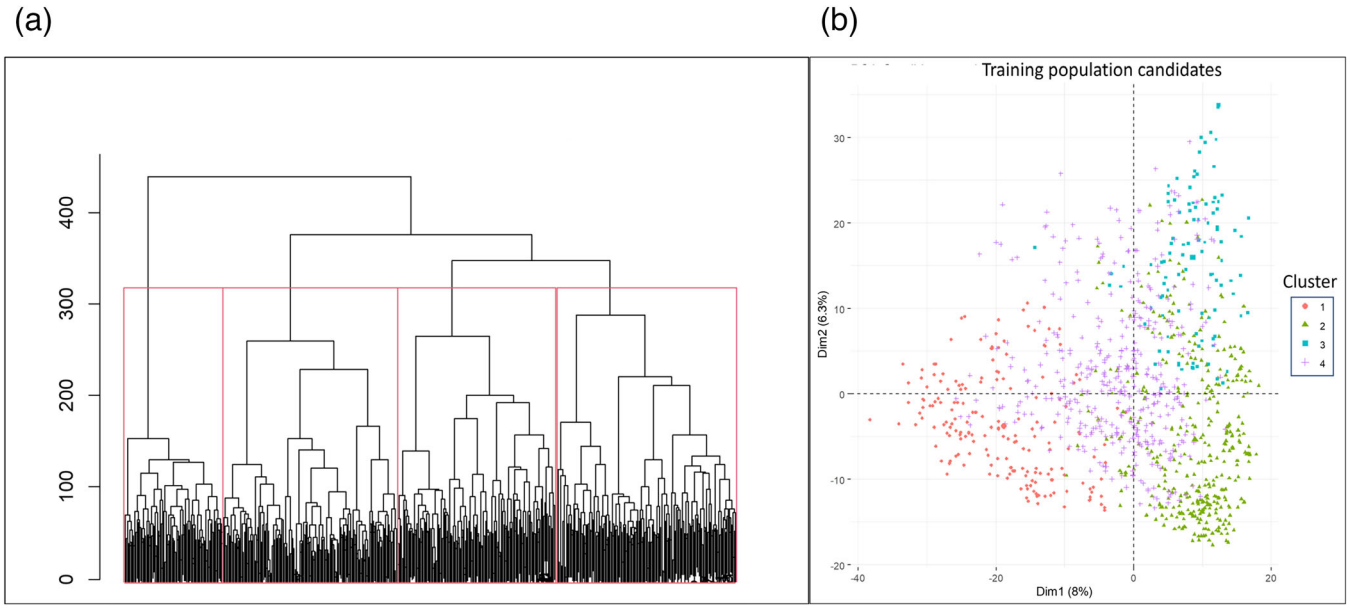

FIGURE 2 (a) Dendrogram of the training population candidates separated into four clusters. The height of the dendrogram is given as the total sum of squares between individuals and each cluster (b) principal component analysis of the training population candidates separated into four clusters [Color figure can be viewed at wileyonlinelibrary.com]

clustering to ensure that the within-group distance is low, and the between-group distance is high (Ward, 1963). The dendrogram was separated into the optimal number of clusters from the structure analysis with the 'rect.hclust' function in R, which isolates the clusters with the highest genetic distance to each other (Figure 2a). Figure 2b shows the four clusters in the PCA.

The mean distance (MD) of each cluster was summed up, and the number of genotypes selected from each cluster were proportional to sum of MD from all clusters. A stratified random sampling was done 1000 times in each cluster, and the subsamples with the highest average mean distance were selected for the DivCo populations.

2.6.4 | Random selection

Random selection was included as a control to represent random unoptimized training populations and is abbreviated to RanSe for the rest of the paper. Populations were selected by using the 'sample_n' function from the 'dplyr' package in the R software (Wickham et al., 2021), which randomly selects a given number of random rows from a dataframe.

2.7 | Statistical analysis

ANOVA was used to identify significant effects of optimization strategy on prediction ability, and the equation is stated as:

$$\mathbf{y} = \mathbf{b_0} + \mathbf{b_1 x_1} + \mathbf{b_2 x_2} + \mathbf{b_3 x_{12}} + \mathbf{e} \quad (6)$$

where **y** is the response prediction ability, $\mathbf{b_0}$ the intercept, $\mathbf{b_1}$–$\mathbf{b_3}$ are coefficients, $\mathbf{x_1}$ the fixed effect of optimization strategy, $\mathbf{x_2}$ the fixed effect of population size, $\mathbf{x_{12}}$ the interaction between population size and optimization strategy, and **e** the error term.

A Tukey pairwise comparison test was used for each pair of strategies to identify if they were significantly different from each other.

Bootstrapping was also used to calculate the significant differences between the strategies within each defined population size. Bootstrapping was used because the PheSe strategy was not replicated. The sample closest to the mean prediction ability in each strategy and population size was compared with each other. The bootstrapping was conducted by removing a random set of lines from the testing population and calculating the prediction ability of the remaining lines in the testing population for both training populations. The procedure was done using the R package 'GRousellet/bootcorci' (Rousselet et al., 2019) and the function 'twocorci.ov' by removing a random set of lines from the testing population, and calculating the prediction ability for the remaining lines. The bootstrapping was done with a significance level of $\alpha = .05$ and 2000 iterations.

2.8 | Optimization criteria

The genetic diversity was calculated as the mean expected heterozygosity by using the R package 'diveRsity' and the function 'Divbasic' in the R software which calculates the frequencies of the alleles of each marker using the formula (2)*p*q, where p and q is the frequencies of the different alleles. Then then mean 2pq is calculated for all markers (Keenan et al., 2013).

The genetic similarity between the training and testing populations was calculated as the proportion of shared alleles per pair of populations based on the allele frequencies, summed and averaged across all loci. The calculations were done by using the R package 'PopGenReport' and the function 'pairwise.propShared' in the R software (Adamack & Gruber, 2014).

The Phenotypic variance was calculated by using the 'var' function of the 'base' R software.

2.9 | Wheat (*Triticum aestivum*) validation sets

Four different datasets were provided by CIMMYT to validate the results of this study and is described in Table 3. The largest were chosen as the training population candidates, whereas the rest were used as testing populations.

TABLE 3 Summary description of wheat germplasm, genotypic data and phenotypic data used in this study (Montesinos-López et al., 2019)

Training population candidates	980 lines
Testing population 1	766 lines
Testing population 2	775 lines
Testing population 3	964 lines
Number of SNP markers	9285
Locations	6 per year for each population
Years	4, 1 year per population
Traits	Plant height, days to heading and grain yield

3 | RESULTS

3.1 | Prediction ability

For plant height (Figure 3a), RanSe and DivCo showed similar prediction abilities of .26 using population sizes 300–500, whereas the prediction abilities of PheSe and PreCo were higher at .33 and .35, respectively. PreCo performed significantly better than DivCo in size 300 and 400, whereas PheSe performed significantly better than DivCo at size 400 and RanSe at 400 and 500 (Table 4). PreCo performed approximately .025 points better than PheSe in size 200–500, but this difference was not significant. The prediction ability of all

TABLE 4 Significant *p*-values values from bootstrapping tests for plant height for the optimization strategies prediction core (PreCo), diversity core (DivCo), phenotypic selection (PheSe) and random selection (RanSe)

Optimization strategies	Population size	$p < .05$
DivCo vs. PreCo	300	.036
DivCo vs. PreCo	400	.018
DivCo vs. PheSe	400	.019
RanSe vs. PheSe	400	.038
RanSe vs. PheSe	500	.046

Note: Days to heading had no significant differences in bootstrapping.

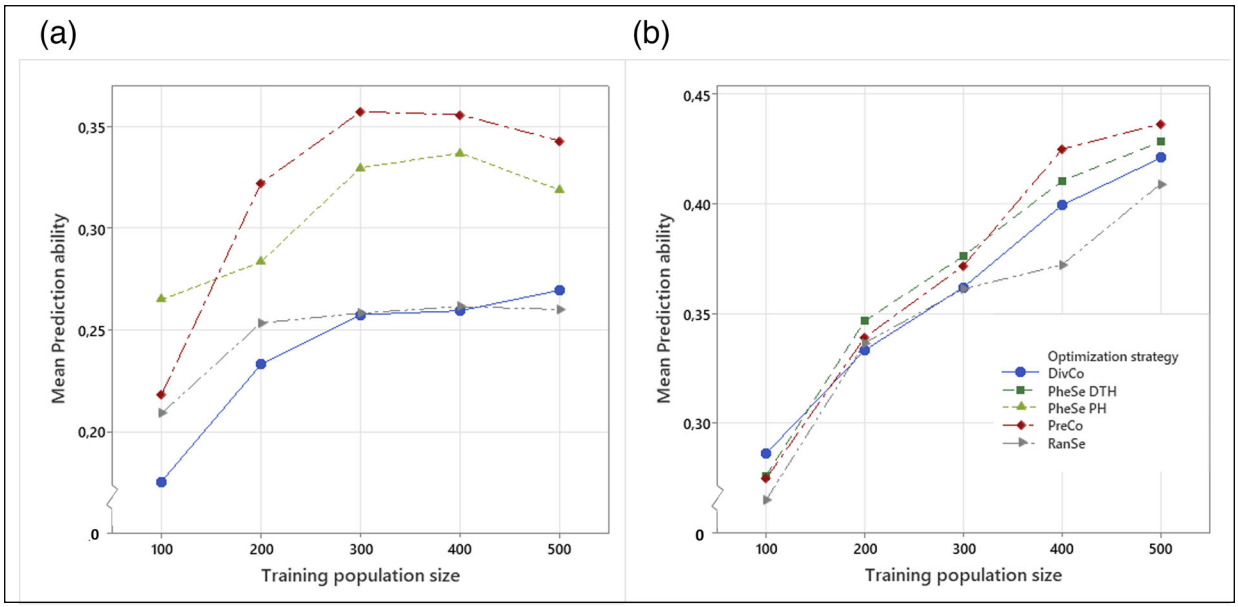

FIGURE 3 Average prediction abilities for (a) plant height (PH) and (b) days to heading (DTH in oats for the different optimization strategies prediction core (PreCo), diversity core (DivCo), phenotypic selection (PheSe) and random selection (RanSe) across different training population sizes [Color figure can be viewed at wileyonlinelibrary.com]

lines plateaued at size 300, with a slight decrease in larger populations for PreCo and PheCo as the prediction ability of the total candidate population was .33 (data not shown).

For days to heading (Figure 3b), all strategies performed similarly for population sizes 100–300 but differed more for population sizes 400 and 500, where the PreCo showed the highest prediction ability followed by PheSe, DivCo and RanSe. However, none of these differences were significant in a bootstrapping test. Prediction ability increased linearly with population size with a maximum of .44 for PreCo at size 500. Using all lines in the candidate population resulted in a prediction ability of .49 (data not shown).

3.2 | ANOVA

The ANOVA results (Table 5) showed that population size contributed to about 62% of the variation in prediction ability for days to heading and 20% for plant height. Optimization strategy contributed to less than 2% of the variation for days to heading, and 21% for plant height. Both factors were significant for both traits, whereas the interaction term was not significant. The PreCo populations yielded significantly higher prediction ability than RanSe for days to heading, and significantly better than DivCo and RanSe for plant height (Table 6). No other significant differences were detected using the Tukey test.

3.3 | Genetic diversity, similarity and phenotypic variance

The optimizations criteria (Figures 4 and 5) showed that the RanSe populations had the lowest phenotypic and genetic diversity, and intermediate genetic similarities. DivCo populations had high genetic diversity, low genetic similarity, and intermediate phenotypic diversity. PreCo populations had the highest genetic diversity, high genetic similarity, and high phenotypic diversity. PheSe populations had very high phenotypic diversity for both traits, intermediate genetic diversity for days to heading and low for plant height, and the highest genetic similarity for plant height and the lowest for days to heading.

There was a significant positive correlation between prediction ability and genetic similarity for both traits, with r values of .48 for plant height and .4 for days to heading. There was also a significant positive correlation between prediction ability and genetic and phenotypic diversity for plant height, but with a low r of .25 for both criteria. Population size had a high significant positive correlation with prediction ability with r values of .77 for days to heading and .36 for plant height.

3.4 | Wheat validation results

Population size had a significant large effect on the variation in prediction ability for all three testing populations in all traits (Table 7). Selection strategy showed a significant contribution in two out of the three

TABLE 6 Results from the pairwise comparisons of the prediction abilities of the four optimization strategies prediction core (PreCo), diversity core (DivCo), phenotypic selection (PheSe) and random selection (RanSe) for the traits days to heading and plant height

Strategies compared		Difference in means	*p*-values
Days to heading			
PheSe	DivCo	0.007	.980
PreCo	DivCo	0.009	.384
RanSe	DivCo	−0.012	.177
PreCo	PheSe	0.002	1.000
RanSe	PheSe	−0.019	.740
RanSe	PreCo	−0.020**	.002
Plant height			
PheSe	DivCo	0.068	.062
PreCo	DivCo	0.080**	<.001
RanSe	DivCo	0.010	.673
PreCo	PheSe	0.013	.968
RanSe	PheSe	−0.058	.140
RanSe	PreCo	−0.071**	<.001

*$p < .05$.**$p < .01$.

TABLE 5 ANOVA for oats with prediction ability as response variable and population size (size), optimization strategy (strategy) and the interaction term Size*Strategy as factors for days to heading and plant height

Source	*df*	Contribution	Adj SS	Adj MS	*F*-value	*p*-value
Days to heading						
Size	4	61.51%	0.187	0.047	28.95	<.001
Strategy	3	1.65%	0.022	0.007	4.45	.004
Size*Strategy	12	1.56%	0.020	0.002	1.05	.404
Error	285	35.27%	0.461	0.002		
Plant height						
Size	4	20.33%	0.074	0.019	5.25	<.001
Strategy	3	21.43%	0.392	0.131	37	<.001
Size*strategy	12	3.23%	0.059	0.005	1.4	.167
Error	285	55.01%	1.007	0.004		

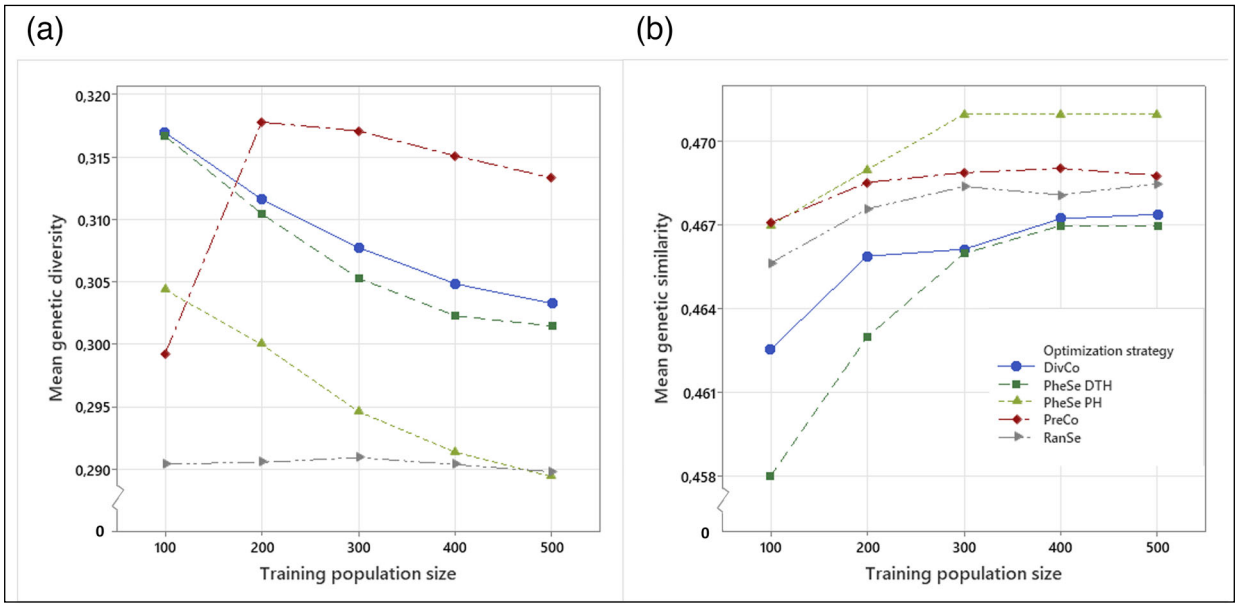

FIGURE 4 (a) Mean genetic diversity and (b) mean genetic similarity for the optimization strategies diversity core (DivCo), prediction core (PreCo), phenotypic selection (PheSe) for plant height (PH) and days to heading (DTH) and random selection (RanSe) in population size 100–500 [Color figure can be viewed at wileyonlinelibrary.com]

populations for days to heading, and one out of the three populations for plant height and grain yield. In all cases of with non-significant contribution of selection strategy for prediction ability of days to heading and plant height, the Size × Strategy interaction term was significant. For grain yield prediction, the error term was always non-significant.

DivCo were the best strategy in terms of prediction ability in three of four significant pairwise comparisons in testing population 1 (Table S5), PreCo were best in three of four comparisons in population 2 (Table S6), and PheSe were best in three of five comparisons in population 3 (Table S7). Full ANOVA tables are available in Tables S2–S4 along with plots of the mean prediction ability for the optimization strategies (Figure S1).

4 | DISCUSSION

Our study compared three training population optimization strategies (prediction core, diversity core and phenotypic selection) to random selection. The training population criteria optimized were genetic diversity, phenotypic variance and genetic similarity. The four strategies were validated for their prediction ability, that is, their ability to predict the phenotypes of a given testing population, and analysed for their genetic diversity, genetic similarity between the training and testing population, and phenotypic diversity.

The broad sense heritability was high for the traits days to heading (.62) and plant height (.71) which is expected for these traits. Studies have shown that smaller training populations are needed for traits with high heritability (Kaler et al., 2022; Zhang et al., 2017), and others have shown that high prediction ability can be achieved for plant height and days to heading with small training populations (Baertschi et al., 2021; Haikka, Knürr, et al., 2020). A study done on unbalanced agronomic traits showed that the standard broad sense heritability calculation overestimates the heritability (Schmidt et al., 2019), which is also likely true for the dataset of this research. This can explain the relatively low prediction abilities in this study. But the overestimation does not likely affect the ranking of the strategies as an adjustment of the heritability as suggested by Schmidt et al. (2019) would shrink the heritability for both traits equally. There are also large G × E effects on unbalanced data, which could also have contributed to the low prediction abilities. The observation that the maximum prediction ability was reached at population size 300 for plant height, whereas it was not yet reached at size 500 for days to heading is likely an effect of the difference in heritability between the two traits.

The main factors effecting the prediction abilities were population size and genetic similarity, which has been highlighted as important training population criteria in several studies (Liu et al., 2018; Lorenz & Nice, 2017; Zhang et al., 2017). Genetic and phenotypic diversity were however less important since increasing these criteria alone would decrease the genetic similarity as the testing population has low genetic diversity. Other studies also found that genetic diversity is more important when population structure is present (Berro et al., 2019; Isidro et al., 2015). One study on diversity core and prediction core found that they gave similar prediction abilities (Crossa et al., 2016), which is different from what this study concludes. But in their study, the authors used diverse landrace populations with a much higher degree of genetic diversity and population structure. These populations are better suited for the diversity core strategy than the more narrow breeding population used in this study.

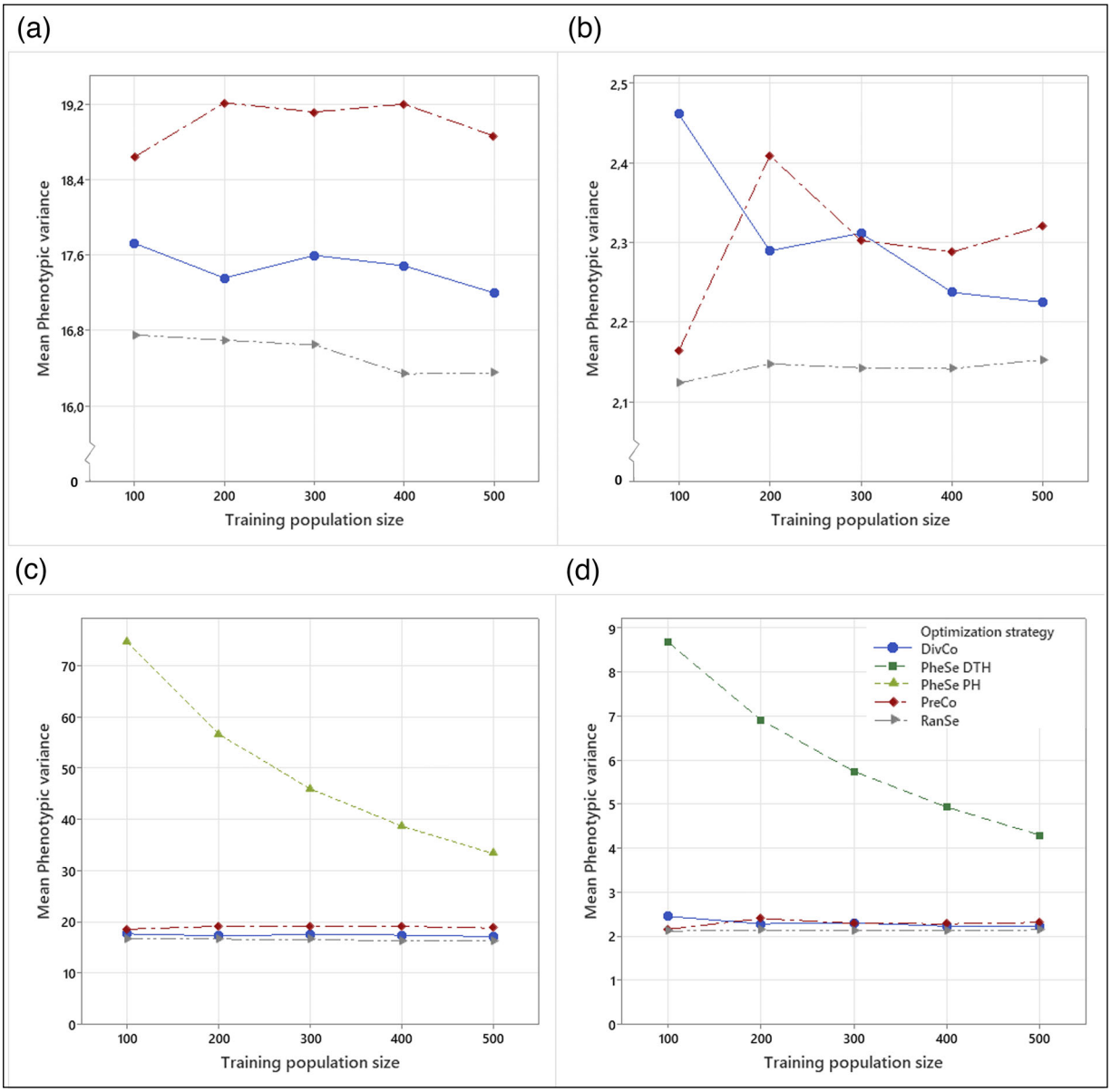

FIGURE 5 Mean phenotypic diversity of the optimization strategies diversity core (DivCo), prediction core (PreCo), phenotypic selection (PheSe) and random selection (RanSe) in population size 100–500 for the traits (a) plant height (PH) without PheSe, (b) days to heading (DTH) without PheSe, (c) plant height with PheSe and (d) days to heading with PheSe [Color figure can be viewed at wileyonlinelibrary.com]

Another study (Akdemir et al., 2015) showed that the prediction core consistently gave better prediction abilities than random selection across population sizes and different traits, which is similar to the results as presented in this paper.

4.1 | Diversity core

The diversity core strategy worked as intended. It produced training populations with similar genetic diversity as the total candidate population and selected relatively equal number of lines from all four clusters. DivCo populations performed very similar to RanSe in both traits and all population sizes (Figure 3). DivCo populations showed lower genetic similarity than RanSe (Figure 4b), which along with the fact that it showed intermediate genetic and phenotypic diversity (Figures 4a and 5a,b) explains the low prediction abilities. The DivCo strategy is not optimal for our data because of the lack of population structure and genetic diversity in the testing population. Further research on this strategy in populations with more diversity and population structure would be useful to properly evaluate it. However, when exotic material is introduced into a breeding programme DivCo could be more useful. As the DivCo strategy does not depend on a specific testing population it is reasonable to think that it would give more stable prediction abilities.

TABLE 7 *p*-values and contribution percentage from ANOVA with prediction ability as response variable population size (size), optimization strategy (strategy) and the interaction term Size*Strategy as factors

	Days to heading		Plant height		Grain yield	
Source	Contribution	*p*-value	Contribution	*p*-value	Contribution	*p*-value
Validation set 1						
Size	12.86%	<.001	36.65%	<.001	29.77%	<.001
Strategy	1.57%	.034	5.61%	<.001	.44%	.395
Size*Strategy	20.60%	<.001	13.70%	<.001	2.85%	.151
Error	64.98%		44.04%		66.95%	
Validation set 2						
Size	19.55%	<.001	45.35%	<.001	25.18%	<.001
Strategy	.23%	.648	.14%	.666	20.05%	<.001
Size*Strategy	4.23%	.049	4.75%	.001	.74%	.864
Error	75.99%		49.76%		54.03%	
Validation set 3						
Size	28.50%	<.001	50.9 5%	<.001	9.99%	<.001
Strategy	8.20%	<.001	.22%	.502	1.17%	.143
Size*Strategy	1.34%	.628	2.70%	.037	3.76%	.132
Error	61.96%		46.12%		85.09%	

Note: The traits analysed were days to heading, plant height and grain yield. The wheat data used for training and testing populations are described in Section 2.7.

4.2 | Prediction core

The PreCo strategy worked as intended and produced training populations highly related to the testing population (Figure 4b). The PreCo populations also showed high genetic diversity (Figure 4a) and relatively high phenotypic variance (Figure 5a,b). This is likely because CD values in addition to minimizing PEV also maintains high genetic distance between individuals in the training population. The combination of high values for the three optimization criteria, and especially the genetic similarity likely explains why PreCo gave the highest prediction abilities. For the sake of this study, it would have been useful to also include populations only selected based on the PEV values. This would likely have decreased the genetic and phenotypic diversity. The PreCo strategy with the CD criteria works well when you know the genotypes you want to predict. However, further research is needed into the PreCo strategy to see whether these prediction abilities are stable across different testing populations. If the testing population is a good representation of the genetic diversity of the breeding programme, then the training population should work for the next breeding cycles as well.

4.3 | Phenotypic selection

The PheSe strategy selected training populations with very high phenotypic variance (Figure 5c,d). The days to heading populations showed similar genetic diversity as DivCo, but also the lowest genetic similarity (Figure 4). The plant height populations showed low genetic diversity but the highest genetic similarity. Both PheSe populations gave relatively high prediction abilities, indicating that phenotypic variance is an important criterion to optimize, despite their low genetic diversity in the plant height populations, and low genetic similarity in the days to heading populations. This can be either due to overfitting of the marker effects or increased diversity for the relevant markers. Our study suggests that PheSe is a good strategy for selecting training populations when no genotype data is available, and that the inclusion of material with low breeding value is important to increase prediction accuracy. In our study we maximized this by selecting 50% lines with low breeding values, whereas Zhao et al. concludes that 30% is enough to ensure high accuracy without underfitting of marker effects (Zhao et al., 2012).

4.4 | Wheat validation

The wheat validation sets were inconclusive in determining which strategy works best, as they rank differently in the different testing populations for different traits. We can see that DivCo worked best in population 1, PreCo in population 2 and PheSe in population 3. We did not do any further analysis into the optimization criteria of the wheat datasets. Further research can show why the optimization strategies worked differently for the different testing populations. A likely reason could be that the phenotypic data for the three populations were collected from different years, which increases the G × E effect. It is shown that the correlation between environments can vary a lot for the same trait (Cooper & DeLacy, 1994), which could explain the low prediction abilities in the validation sets.

4.5 | Conclusion

Of the three strategies analysed, prediction core had the highest average prediction ability in most population sizes for both traits and produced training populations with high genetic diversity, high genetic similarity to the testing population and high phenotypic variance compared with random selection. Genetic similarity along with population size were the most important criteria to optimize in the training populations. More research is needed to evaluate how well the prediction core strategy works over several breeding cycles, but our research points to prediction core as the best strategy to optimize training populations in cereals.

ACKNOWLEDGEMENTS

This research was funded by the Research Council of Norway and Graminor (NFR # 286606) (the research funds for agriculture and the food industry - FFL/JA). The authors wish to thank Graminor breeders for providing the material for this research and the technical staff at Graminor and NMBU for helping in field experiments and data collection. We would also like to thank Prof. Åsmund Bjørnstad and Dr. Trond Buraas for sharing their experience and knowledge about oats. Further, we acknowledge the International Maize and Wheat Improvement Center for providing wheat material and support on genomic predictions.

CONFLICTS OF INTEREST

All authors have no conflicts of interest to declare that are relevant to the content of this article.

AUTHOR CONTRIBUTIONS

ESS collected and analysed historical data from Graminor; MA provided funding and support in data analysis; SW and JC provided support in genomic predictions; ML and AKS provided support in genetic analysis; CJ provided support in phenotypic analysis; ESS wrote the manuscript; and MA, SW, JC, ML, AKS and CJ participated in revising the scientific work and writing.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Espen Sannes Sørensen https://orcid.org/0000-0003-4646-8250
Morten Lillemo https://orcid.org/0000-0002-8594-8794

REFERENCES

Adamack, A. T., & Gruber, B. (2014). PopGenReport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, *5*(4), 384–387. https://doi.org/10.1111/2041-210X.12158

Akdemir, D. (2018). STPGA: Selection of Training Populations by Genetic Algorithm. R package version 5.2.1. https://CRAN.R-project.org/package=STPGA

Akdemir, D., & Isidro-Sánchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. *Scientific Reports*, *9*, 1446. https://doi.org/10.1038/s41598-018-38081-6

Akdemir, D., Sanchez, J. I., & Jannink, J. L. (2015). Optimization of genomic selection training populations with a genetic algorithm. *Genetics Selelection Evolution*, *47*, 38. https://doi.org/10.1186/s12711-015-0116-6

Ankamah-Yeboah, T., Janss, L. L., Jensen, J. D., Hjortshøj, R. L., & Rasmussen, S. K. (2020). Genomic selection using pedigree and marker-by-environment interaction for barley seed quality traits from two commercial breeding programs. *Frontiers in Plant Science*, *11*, 539. https://doi.org/10.3389/fpls.2020.00539

Baertschi, C., Cao, T. V., Bartholomé, J., Ospina, Y., Quintero, C., Frouin, J., Bouvet, J.-M., & Grenier, C. (2021). Impact of early genomic prediction for recurrent selection in an upland rice synthetic population. *G3 Genes|Genomes|Genetics*, *11*(12), jkab320. https://doi.org/10.1093/g3journal/jkab320

Berro, I., Lado, B., Nalin, R. S., Quincke, M., & Gutiérrez, L. (2019). Training population optimization for genomic selection. *The Plant Genome*, *12*(3), 190028. https://doi.org/10.3835/plantgenome2019.04.0028

Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., Tyagi, A., Mushtaq, M., Jain, N., Singh, P. K., Singh, G. P., & Prabhu, K. V. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Frontiers in Genetics*, *7*, 221. https://doi.org/10.3389/fgene.2016.00221

Ceplitis, A. (2014). Lantmännen spring oat breeding [PowerPoint slides]. American Oat Workers Conferance, Ottawa, 2014. https://oatnews.org/oatnews_pdfs/AOWC2014/breeders/Ceplitis_breedingAOWC2014.pdf

Cooper, M., & DeLacy, I. H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics*, *88*(5), 561–572. https://doi.org/10.1007/BF01240919

Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Crossa, J., Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., Makumbi, D., Singh, R. P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., & Braun, H. J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, *186*(2), 713–724. https://doi.org/10.1534/genetics.110.118521

Crossa, J., Jarquín, D., Franco, J., Pérez-Rodríguez, P., Burgueño, J., Saint-Pierre, C., Vikram, P., Sansaloni, C., Petroli, C., Akdemir, D., Sneller, C., Reynolds, M., Tattaris, M., Payne, T., Guzman, C., Peña, R. J., Wenzl, P., & Singh, S. (2016). Genomic prediction of Gene Bank wheat landraces. *G3: Genes, Genomes, Genetics*, *6*(7), 1819–1834. https://doi.org/10.1534/g3.116.029637

Crossa, J., Pérez, P., Hickey, J., Burgueño, J., Ornella, L., Ceròn-Rojas, J., Zhang, X., Dreisigacker, S., Babu, R., Li, Y., Bonnett, D., & Mathews, K. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity*, *112*(1), 48–60. https://doi.org/10.1038/hdy.2013.16

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*, *22*(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Desta, Z. A., & Ortiz, R. (2014). Genomic selection: Genome-wide prediction in plant improvement. *Trends in Plant Science*, *19*(9), 592–601. https://doi.org/10.1016/j.tplants.2014.05.006

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–336. https://doi.org/10.1007/s12686-011-9548-7

Franco, J., Crossa, J., Taba, S., & Shands, H. A. (2005). Sampling strategy for conserving genetic diversity when forming core subsets. *Crop Science*, *45*(3), 1035–1044. https://doi.org/10.2135/cropsci2004.0292

Haikka, H., Knürr, T., Manninen, O., Pietilä, L., Isolahti, M., Teperi, E., Mäntysaari, E. A., & Strandén, I. (2020). Genomic prediction of grain yield in commercial Finnish oat (Avena sativa) and barley (Hordeum vulgare) breeding programmes. *Plant Breeding*, *139*(3), 550–561. https://doi.org/10.1111/pbr.12807

Haikka, H., Manninen, O., Hautsalo, J., Pietilä, L., Jalli, M., & Veteläinen, M. (2020). Genome-wide association study and genomic prediction for *fusarium graminearum* resistance traits in Nordic oat (*Avena sativa* L.). *Agronomy*, *10*(2), 174. https://doi.org/10.3390/agronomy10020174

Heffner, E. L., Lorenz, A. J., Jannink, J. L., & Sorrells, M. E. (2010). Plant breeding with genomic selection: Gain per unit time and cost. *Crop Science*, *50*(5), 1681–1690. https://doi.org/10.2135/cropsci2009.11.0662

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*(5), 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Isidro, J., Jannink, J. L., Akdemir, D., Poland, J., Heslot, N., & Sorrells, M. E. (2015). Training set optimization under population structure in genomic selection. *Theoretical and Applied Genetics*, *128*(1), 145–158. https://doi.org/10.1007/s00122-014-2418-4

Kaler, A. S., Purcell, L. C., Beissinger, T., & Gillman, J. D. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biology*, *22*, 87 (2022). https://doi.org/10.1186/s12870-022-03479-y

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782–788. https://doi.org/10.1111/2041-210X.12067

Laloë, D. (1993). Precision and information in linear models of genetic evaluation. *Genetics Selection Evololution*, *25*, 557. https://doi.org/10.1186/1297-9686-25-6-557

Larkin, D. L., Lozada, D. N., & Mason, R. E. (2019). Genomic selection—Considerations for successful implementation in wheat breeding programs. *Agronomy*, *9*(9), 479. https://doi.org/10.3390/agronomy9090479

Latta, R. G., Bekele, W. A., Wight, C. P., & Tinker, N. A. (2019). Comparative linkage mapping of diploid, tetraploid, and hexaploid Avena species suggests extensive chromosome rearrangement in ancestral diploids. *Scientific Reports*, *9*(1), 12298. https://doi.org/10.1038/s41598-019-48639-7

Li, H., Rarsheed, A., Hickey, L. T., & He, Z. (2018). Rapid breeding cycles: Reaping the low hanging fruit of increasing genetic gain. *Trends in Plant Science*, *23*, 184–186. https://doi.org/10.1016/j.tplants.2018.01.007

Liu, X., Wang, H., Wang, H., Guo, Z., Xu, X., Liu, J., Wang, S., Li, W. X., Zou, C., Prasanna, B. M., Olsen, M. S., Huang, C., & Xu, Y. (2018). Factors affecting genomic selection revealed by empirical evidence in maize. *The Crop Journal*, *6*(4), 341–352. https://doi.org/10.1016/j.cj.2018.03.005

Lorenz, A., & Nice, L. (2017). Training Population Design and Resource Allocation for Genomic Selection in Plant Breeding. In R. Varshney, M. Roorkiwal, & M. Sorrells (Eds.). *Genomic selection for crop improvement*. Springer. https://doi.org/10.1007/978-3-319-63170-7_2

Mellers, G., Mackay, I., Cowan, S., Griffiths, I., Martinez-Martin, P., Poland, J. A., Bekele, W., Tinker, N. A., Bentley, A. R., & Howarth, C. J. (2020). Implementing within-cross genomic prediction to reduce oat breeding costs. *The Plant Genome*, *13*(1), e20004. https://doi.org/10.1002/tpg2.20004

Meuwissen, T. H., Hayes, N. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. (2021). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7–9, https://CRAN.R-project.org/package=e1071

Minitab 21 Statistical Software. (2010). *Computer software*. Minitab, Inc. www.minitab.com

Montesinos-López, O. A., Martín-Vallejo, J., Crossa, J., Gianola, D., Hernández-Suárez, C. M., Montesinos-López, A., Juliana, P., & Singh, R. (2019). New deep learning genomic-based prediction model for multiple traits with binary, ordinal, and continuous phenotypes. *G3: Genes, Genomes, Genetics*, *9*(5), 1545–1556. https://doi.org/10.1534/g3.119.300585

Patterson, H. D., & Williams, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika*, *63*, 83–92. https://doi.org/10.1093/biomet/63.1.83

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

Rincent, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., Rodríguez, V. M., Moreno-Gonzalez, J., Melchinger, A., Bauer, E., Schoen, C. C., Meyer, N., Giauffret, C., Bauland, C., Jamin, P., Laborde, J., Monod, H., Flament, P., Charcosset, A., & Moreau, L. (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*(2), 715–728. https://doi.org/10.1534/genetics.112.141473

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2019). A practical introduction to the bootstrap: A versatile method to make inferences by using data-driven simulations. https://doi.org/10.31234/osf.io/h8ft7

Schmidt, P., Hartung, J., Rath, J., & Piepho, H. (2019). Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. *Crop Science*, *59*(2), 525–536. https://doi.org/10.2135/cropsci2018.06.0376

Tinker, N. A., Chao, S., Lazo, G. R., Oliver, R. E., Huang, Y. F., Poland, J. A., Jellen, E. N., Maughan, P. J., Kilian, A., & Jackson, E. W. (2014). A SNP genotyping Array for Hexaploid oat. *The Plant Genome*, *7*(3), plantgenome2014.03.0010. https://doi.org/10.3835/plantgenome2014.03.0010

Wang, X., Xu, Y., Hu, Z., & Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal*, *6*(4), 330–340. https://doi.org/10.1016/j.cj.2018.03.001

Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Werner, C. R., Gaynor, R. C., Gorjanc, G., Hickey, J. M., Kox, T., Abbadi, A., Leckband, G., Snowdon, R. J., & Stahl, A. (2020). How population structure impacts genomic selection accuracy in Cross-validation: Implications for practical breeding. *Frontiers in Plant Science*, *11*, 592977. https://doi.org/10.3389/fpls.2020.592977

Wickham, H., François, R., Henry, H. & Müller, K. (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr

Xu, Y., Liu, X., Fu, J., Wang, H., Wang, J., Huang, C., Prasanna, B. M., Olsen, M. S., Wang, G., & Zhang, A. (2020). Enhancing genetic gain through genomic selection: From livestock to plants. *Plant Communications*, *1*(1), 100005. https://doi.org/10.1016/j.xplc.2019.100005

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., Cui, Z., Ruan, Y., Burgueño, J., Vicente, F. S., Olsen, M., Prasanna, B. M., Crossa, J., Yu, H., & Zhang, X. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Frontiers in Plant Science*, *8*, 1916. https://doi.org/10.3389/fpls.2017.01916

Zhao, Y., Gowda, M., Longin, F. H., Würschum, T., Ranc, N., & Reif, J. C. (2012). Impact of selective genotyping in the training population on accuracy and bias of genomic selection. *Theoretical and Applied Genetics*, *125*(4), 707–713. https://doi.org/10.1007/s00122-012-1862-2

SUPPORTING INFORMATION

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

How to cite this article: Sørensen, E. S., Jansen, C., Windju, S., Crossa, J., Sonesson, A. K., Lillemo, M., & Alsheikh, M. (2023). Evaluation of strategies to optimize training populations for genomic prediction in oat (*Avena sativa*). *Plant Breeding*, *142*(1), 41–53. https://doi.org/10.1111/pbr.13061