Article

Optimizing Training Population Size and Content to Improve Prediction Accuracy of FHB-Related Traits in Wheat

Emmanuel Adeyemo ¹, Prabin Bajgain ¹ , Emily Conley ¹, Ahmad H. Sallam ² and James A. Anderson ^{1,*}

¹ Department of Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA; adeye017@umn.edu (E.A.); bajga002@umn.edu (P.B.); econley@umn.edu (E.C.)

² Department of Plant Pathology, University of Minnesota, 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA; sall0029@umn.edu

* Correspondence: ander319@umn.edu

Received: 10 March 2020; Accepted: 6 April 2020; Published: 9 April 2020

Abstract: Genomic selection combines phenotypic and molecular marker data from a training population to predict the genotypic values of untested lines. It can improve breeding efficiency as large pools of untested lines can be evaluated for selection. Training population (TP) composition is one of the most important factors affecting the accuracy of genomic prediction. The University of Minnesota wheat breeding program implements genomic selection at the F₅ stage for Fusarium head blight (FHB) resistance. This study used field data for FHB resistance in wheat (*Triticum aestivum* L.) to investigate the use of small-size TPs designed with and without stratified sampling for three FHB traits in three different F₅ populations (TP17, TP18, and TP19). We also compared the accuracies of these two TP design methods with the accuracy obtained from a large size TP. Lastly, we evaluated the impact on trait predictions when the parents of F₅ lines were included in the TP. We found that the small size TP selected randomly, without stratification, had the lowest predictive ability across the three F₅ populations and across the three traits. This trend was statistically significant ($p = 0.05$) for all three traits in TP17 and two traits in TP18. Designing a small-size TP by stratified sampling led to a higher accuracy than a large-size TP in most traits across TP18 and TP19; this is because stratified sampling allowed the selection of a small set of closely related lines. We also observed that the addition of parental lines to the TP and evaluating the TP in two replications led to an increase in predictive abilities in most cases.

Keywords: Fusarium head blight; genomic selection; wheat; training population optimization; stratified sampling

1. Introduction

Wheat (*Triticum aestivum* L.) is the most grown food crop in the world and considered the most important source of calories for humans [1]. Fusarium head blight (FHB), caused by *Fusarium graminearum* Schwabe, is a destructive fungal disease of wheat that threatens global wheat production and food security. It can cause a significant loss of grain yield [2,3] while also affecting grain quality due to the accumulation of mycotoxins, potentially making it unsafe for human and animal consumption [4]. No fungicide application, tillage system or crop rotation technique can completely eradicate FHB in wheat [5]. Fungicides can reduce FHB damage by as much as 40–70% [6,7] if application is within a few days following anthesis but the high cost of this additional input emphasizes the importance of developing wheat varieties resistant to FHB.

The success of a breeding program in improving FHB resistance depends on the availability of resistant germplasm, enough genetic variation, effective methods to evaluate resistance and select improved individuals [8]. While phenotyping for FHB resistance still requires considerable resources, advancement in sequencing technologies has significantly reduced the cost of genotyping [9]. Minimizing the cost of phenotyping for FHB resistance would ensure effective allocation of resources in a breeding program and genomic selection offers the potential to limit the need for phenotyping while also allowing breeders to broaden the pool for selection [10,11]. Genomic selection uses information from a training population that has been both genotyped and phenotyped to train a model that is applied to a breeding population of non-phenotyped individuals to predict their genotypic values that can be used in place of phenotypic values. Several studies have shown that genomic selection is a promising breeding strategy to improve FHB resistance [12–15].

The first step in the implementation of genomic selection in a breeding program is the establishment of a training population [16], which is the set of lines that are phenotyped and used to estimate marker effects. Since 2017, the hard-red spring wheat breeding program at the University of Minnesota has implemented genomic selection at the F₅ (pre-yield trial) stage to select for FHB resistance. The goal at this stage is to identify and eliminate highly susceptible lines before entering them in yield trials. Any line advanced to the preliminary yield trial stage or later will continue to undergo phenotypic assessment of FHB resistance in misted inoculated nurseries at two distinct Minnesota locations on an annual basis. The training population in the 2017 growing season was a subset of 500 F₅ lines selected by genomic and pedigree information (more details below). After 2017, our goal was to design an optimized training population (TP) that minimized the cost of phenotyping for FHB resistance while maximizing prediction accuracy for the untested lines. While simulation and empirical results have shown improved prediction accuracies when TP size is increased [17–21], other studies have achieved high accuracies with smaller TP sizes that were more closely related to the breeding population and lower accuracies as TP became more unrelated to the breeding population [22–25]. For example, Rincent et al. [22] found that an optimized set of 100 lines achieved the same prediction accuracy as a set of 200 lines selected at random.

Isidro et al. [26] introduced a TP optimization technique based on stratified sampling that allows the optimization of the TP based on genomic relationships, resulting in higher prediction accuracies in structured populations. This procedure maximizes relationships between the training and breeding population and allows for a TP to be designed around a breeding (target) population. The k-means clustering algorithm is a form of cluster analysis that partitions datapoints into groups to reflect underlying similarities. The assignment of an individual into a cluster is determined by its Euclidean distance relative to the mean of the cluster. Details on how the algorithm works are deferred to specialized texts [27]. However, the main aim is to reduce variability within clusters while increasing variability among clusters. This method is particularly useful when population structure is present as it tends to group lines into different subpopulations as stratified sampling has been shown to improve the efficiency of sampling compared to a simple random sample without stratification [28,29].

The objectives of this study are to determine if: (1) a small sized TP selected by stratified sampling will improve predictive abilities compared to a same size TP selected without stratification; (2) the addition of parents to the TP will improve the predictive abilities. In addition, we also evaluated the effects of multiple replications on predictive abilities.

2. Materials and Methods

2.1. Population Development

Each year the University of Minnesota's wheat breeding program evaluates between 1500–3000 F₅ lines, resulting from approximately 170 unique crosses. These two-way or three-way crosses are mostly elite × elite but it is not uncommon to make crosses with un-adapted parents for some special trait improvement. The F₅ lines are selected from F₄ head rows based on height, lodging, leaf and

stem rust resistance but have never been screened for FHB resistance. Usually, seven heads (~30 seeds per head) are harvested from F₄ headrows and are bulk threshed, except in 2017, in which the seven heads were threshed individually. Seeds from these heads were used to plant short rows (1.5 m) for phenotypic evaluation of disease resistance and agronomic traits in the F₅ plants. A single seed was used to plant an F₅ plant in the greenhouse, which was genotyped and seed from this plant was sent to our winter nursery for the F₆ generation. In 2017, a total of 1550 lines, selected from the 2016 F₄ head row nurseries, were advanced to the F₅ stage of the breeding program. Out of these, 500 lines were selected to serve as a genomic selection training population (TP17) and used to predict the remaining untested lines in the population (Table 1). These 500 individuals were selected based on their pedigree so that the wider genetic diversity of the breeding population was captured in the TP.

Table 1. Number of lines, crosses and range of lines per cross in 2017, 2018 and 2019 F₅ populations.

	2017	2018	2019
Number of Crosses	229	184	169
Number of Lines	1550	2590	2715
Range of Lines/Cross (mean)	1–30 (7)	1–64 (14)	1–81 (16)
Training Population Size	500	545	544
TP † Selection Method			
Pedigree	500	-	-
Stratified Sampling	-	200	200
Pedigree + Random	-	300	300
Parents in TP	0	45	44

† TP = Training Population.

In the 2018 growing season, the F₅ population consisted of 2590 lines selected from the 2017 F₄ head row nursery (Table 1). However, instead of selecting 500 lines to serve as TP, we selected an optimized set of 200 lines based on stratified sampling with the k-means clustering method (details below). The selection process used a k-means clustering analysis performed on all 2590 F₅ lines so that the lines were divided into three clusters with each cluster containing lines that are more closely related to each other. To determine the value of *k* (i.e., the number of clusters), we tried different values for *k* until we were satisfied with the cluster separation and we chose that value as the optimum number of *k*, which in this case was three. After grouping into different clusters, a simple random sample was performed within each cluster.

Since we wanted a total of 200 F₅ lines in the TP18 selected by stratified sampling, few lines were sampled within each cluster to capture the entire genetic space of the cluster. The number of lines selected from each cluster was proportional to the size of the cluster in the entire F₅ population. We sampled 95 lines from cluster-one, 46 lines from cluster-two and 59 lines from cluster-three to make a total of 200 lines selected by genomic relationship. The lines sampled from each cluster served as the TP for the remaining untested lines in that cluster. To serve as both buffer and basis of comparison, a different set of 300 lines were also selected from the remaining F₅ population. These lines were first selected so that each pedigree is sampled at least once and then more lines from each pedigree were added randomly in proportion to the family size. This method allows for the sampling of lines across the entire genetic distribution of the F₅ population. In a structured population, random sampling of lines across subpopulations to make up the TP will reduce the similarity across lines in the TP. Lastly, 45 parental lines also were evaluated to make a total of 545 phenotyped lines.

The training population for the 2019 growing season (TP19) was selected similar to TP18 and consisted of 544 lines (Table 1). In summary, TP17 was sampled to maximize genetic diversity without stratification while TP18 and TP19 were first divided into clusters to maximize relationships within clusters and sampling was done within each cluster. Also, TP18 and TP19 had parental lines evaluated in the field while parental lines were not evaluated for TP17.

The 500 TP17 lines were grown as a single replication in misted, inoculated FHB nurseries in St. Paul and Crookston, Minnesota in 2017. For TP18 and TP19, the 200 lines selected by stratified

sampling, and the parents were tested in two replications while the 300 lines sampled randomly without stratification were tested in one replication in the same FHB nurseries. Unfortunately, the irrigation system in the Crookston 2018 nursery failed, resulting in biased data, thus, the data from this location was omitted and only the data from St. Paul was used for further analysis in 2018. In all FHB nurseries five checks (Alsen (MR, [30], MN00269 (S), Roblin (S), Rollag (MR, [31]), and Wheaton (S, [32])) were planted for every one hundred plots. Procedures for misting, inoculation and data collection for the FHB traits disease index (DIS), visually scabby kernels (VSK) and micro-test weight (MicroTwt) were as described by Fuentes-Granados et al. [33]. For each plot, the number of infected spikelets per head was counted on each of 20 randomly selected main heads. FHB severity was calculated as the average percentage of infected spikelets per head for each plot. FHB incidence was calculated as the proportion of heads that had any visible infection out of the 20 heads used for FHB severity assessment. DIS was calculated as follows:

$$\text{FHB disease index (DIS)} = \frac{\text{FHB severity} \times \text{FHB incidence}}{100} \quad (1)$$

To rate VSK, approximately 200–300 kernels were assessed visually to determine the percentage of kernels that showed signs of infection [34]. For MicroTwt, cleaned seed samples were poured into a 15.7 mL copper vessel (20 mm in diameter and 50 mm in height). A ruler was used to level the sample at the top of the vessel and then the sample was weighed.

2.2. Training Population Comparison

We assessed trait predictive abilities (correlation between the observed phenotypic values and the predicted genotypic values) in an optimized TP selected by stratified sampling compared with the same size TP selected without stratification. We also compared the two methods with training a prediction model on the entire TP (entire TP is a total of 500 F₅ lines—200 selected by stratified sampling and 300 selected by pedigree information). For the set of 200 lines selected by stratified sampling in TP18 and TP19, cross-validation was performed within each cluster and the average predictive ability across clusters was used to assess this method of TP selection. For the 300 lines sampled by pedigree information and without stratification, we further sampled 200 lines randomly to remove confounding effects based on TP size. The process was repeated 100 times and a cross-validation performed each time. The average predictive ability across 100 cross-validations was used to assess this method of TP selection. Lastly, we performed a cross-validation within the entire F₅ TP (i.e., 500 F₅ lines) and compared the predictive ability to the smaller sized TPs (Figure 1).

Figure 1. A graphical framework of the training population composition for model comparison.

To determine the effect of adding parents to the TP on the predictive abilities, we included parental lines with the F₅ lines in each cluster. For example, there are 95 F₅ lines in cluster one of TP18, and the addition of 45 parents increased the TP size to 140 lines and the addition of 45 parents to cluster two increased the TP size to 91 lines and so on. In the same vein, cross-validation was performed in each cluster and the average predictive abilities across the clusters were used to assess the effect of parental addition on the F₅ lines in the TPs selected by stratified sampling. Furthermore, after randomly selecting 200 F₅ lines from the 300 selected by pedigree information and without stratification, we added the parental lines to make a total of 245 lines in TP18 and 244 lines in TP19. The process was repeated 100 times with cross-validation performed each time. The average predictive abilities across the 100 cross-validations was used to assess how well this method performed. Finally, we added parental lines to the entire F₅ TP so that there were 545 lines in TP18 and 544 in TP19. We compared the predictive ability of this larger sized TP with the smaller ones.

Since we have implemented a stratified sampling method for TP18 and TP19, we decided to also divide TP17 into clusters using the k-means clustering algorithm and then sample a few lines from each cluster to make a total of 200 F₅ lines selected by stratified sampling. The performance of this selection method was assessed as the average predictive ability across clusters. From the remaining 300 lines, 200 were sampled randomly without stratification and a cross-validation was repeated 100 times. We assessed this selection method as the average predictive ability across the 100 cross-validations. Again, we compared the performance of the small-sized TP by using the entire set of 500 F₅ lines as the TP.

To assess the effect of replications, we used the lines selected by stratified sampling combined with parental lines in TP18 and TP19, since both sets of lines were tested in two replications. For comparison, we used data from the first replication (Rep1) as the second replication will not have existed if the study was not replicated. Thus, we compared the predictive abilities when phenotypic data from Rep1 and the mean across both replications were used to train the model. To assess significant differences in accuracies, we applied a paired *t*-test after Fisher's Z transformation.

2.3. Phenotypic Analysis

Since the 2018 cohort had data from just a single location, the 200 lines selected by genomic relationship and the 45 parental lines were averaged across two replications. The observed data for this single location was modeled as:

$$y_{ik} = \mu + g_i + \varepsilon_{ik} \quad (2)$$

where y_{ik} is the k th observation of the i th genotype, μ is the intercept, g_i is the effect for the i th genotype, and ε_{ik} is the plot error effect for y_{ik} . Assuming g_i , and ε_{ik} are random and independent with a mean of zero and variances of σ^2_g , and σ^2_ε , the phenotypic variance can be calculated as:

$$\sigma^2_P = \sigma^2_g + \sigma^2_\varepsilon/n_r \quad (3)$$

where σ^2_g is the genetic variance, σ^2_ε is the plot error variance, and n_r is the number of replications.

For 2017 and 2019 cohort with multiple locations, the observed data was modeled as:

$$y_{ijk} = \mu + g_i + l_j + (gl)_{ij} + \varepsilon_{ijk} \quad (4)$$

where y_{ijk} is the k th observation of the i th genotype at the j th location, μ is the intercept, g_i is the main effect for the i th genotype, l_j is the main effect for the j th location, $(gl)_{ij}$ is the i th genotype-by-location interaction effect, and ε_{ijk} is the plot error effect for y_{ijk} . Phenotypic values for each genotype was then estimated as:

$$\bar{Y}_{i.} = \text{BLUE}(\mu + g_i) \quad (5)$$

Assuming g_i , l_j , $(gl)_{ij}$, and ε_{ijk} are random, independent, each has a mean of zero and variances of σ^2_g , σ^2_l , σ^2_{gl} , and σ^2_ε respectively, phenotypic variance was calculated as:

$$\sigma^2_P = \sigma^2_g + \sigma^2_{gl}/n_l + \sigma^2_{\epsilon}/n_l n_r \quad (6)$$

where σ^2_g is the genetic variance, σ^2_{gl} is the variance for genotype $\times$ location interaction, σ^2_{ϵ} is the plot error variance, n_l is the number of locations, and n_r is the number of replications. For the 2019 population, the 200 lines and 46 parents were first averaged across replications and then the above linear model was used to calculate the best linear unbiased estimates (BLUEs) of each line. The emmeans package version 1.4.2 [35] in R was used to estimate marginal means (BLUEs) for each line. To estimate phenotypic variances, the number of replications was set to one in both 2018 and 2019 cohorts to remain conservative.

Broad-sense heritability for TP18 with one location was calculated as:

$$H = \sigma^2_g / (\sigma^2_g + \sigma^2_{\epsilon}/n_r) \quad (7)$$

While the broad-sense heritability of TP17 and TP19 with two locations was calculated as:

$$H = \sigma^2_g / (\sigma^2_g + \sigma^2_{gl}/n_l + \sigma^2_{\epsilon}/n_l n_r) \quad (8)$$

2.4. Genotyping

The populations were genotyped using the genotyping by sequencing method [36]. Reads were aligned to the *Triticum aestivum* IWGSC RefSeq v1.0 [37] using bwa [38], and samtools and bcftools [39] for single nucleotide polymorphism (SNP) calling. This resulted in 7102 markers for the 2017 set, 4934 markers for the 2018 set, and 3046 markers for the 2019 set. SNP markers with minor allele frequency of <5% and >20% missing data were discarded. Missing markers were imputed with the linkage disequilibrium k-nearest neighbor (LD-kNNi) genotype imputation algorithm [40] implemented in TASSEL [41] with default parameters.

2.5. Stratified Sampling with k-means Clustering Algorithm

Euclidean distances estimated from SNP markers in each population were used to generate partitions that represent the underlying groups present in the population, based on the specified number of clusters. After the clustering process was performed on the entire F₅ population (i.e., about 1500–3000 lines), a principal component analysis was carried out using the genotypic data. The first two principal components for each line with its corresponding cluster assignment were used to visualize the cluster partitioning of the data (Figure 2a–c). The cluster analysis is based on Euclidean distances derived from genotypic data so lines that fall within a cluster were expected to be more genetically similar to each other versus lines in a different cluster. Since we are trying to exploit the power of genetic relationships and its effects on prediction accuracy, individuals selected from a cluster will serve as the training population for the remaining untested individuals in that particular cluster. In addition, the number of individuals that represent a cluster as the training set depends on the number of lines in that cluster, i.e., clusters with more individuals will have more lines selected as the training set and vice-versa. This training population selection method tends to define a training set based on its relationship to the breeding set (untested individuals) and not the other way around.

2.6. Linkage Disequilibrium (LD) Analysis

We evaluated the LD between pairs of markers for the entire TP and for each cluster. The extent of LD in each TP was indicated as r^2 which is the squared allele frequency correlation between pairs of markers with known genomic positions. Generally, the larger the r^2 value, the larger the degree of association among loci in the genome. This estimation was performed in TASSEL v5.2.60 [41].

2.7. Genomic Selection Model

Three genomic selection models, Bayesian LASSO, Reproducing Kernel Hilbert Space (RKHS), and Ridge-Regression Best-Linear Unbiased Prediction (RR-BLUP), were used for model training. However, RR-BLUP performed similarly or slightly better than the two other models and had the fastest computational time; therefore, only results from RR-BLUP are presented in this study.

An RR-BLUP model was used for all genomic predictions with models trained as follows:

$$y = 1\mu + Zu + \varepsilon \quad (9)$$

$$u \sim \text{MVN}(0, I\sigma_u^2) \quad (10)$$

$$\varepsilon \sim \text{MVN}(0, I\sigma_\varepsilon^2) \quad (11)$$

where y is a vector of BLUEs of the lines in the training population; 1 is an $n \times 1$ vector with elements equal to 1; μ is the grand mean; Z is an $n \times p$ design matrix with elements of 1 or -1 (where -1 represents the minor allele); u is a $p \times 1$ vectors of marker effects; ε is an $n \times 1$ vector of residuals; σ_u^2 is the variance of marker effects; I is an identity matrix; and σ_ε^2 is the error variance. σ_u^2 and σ_ε^2 were estimated using restricted maximum likelihood (REML). The RR-BLUP model was implemented with the rrBLUP package version 4.6 [42] in R. Predictive abilities were calculated using a five-fold cross validation method with 500 iterations within the rrBLUP package. All analyses were performed using R version 3.5.2.

2.8. Accuracy of Genomic Selection for Line Advancement

Prior to implementing genomic selection, approximately 25% of the most susceptible F_5 lines were discarded each year based on phenotypic evaluation. We assessed the effectiveness of genomic prediction in discarding susceptible lines and selecting moderately resistant lines. The likelihood of making a correct decision was calculated as the proportion of observed susceptible lines that the model predicted to be susceptible (true positives) and the proportion of lines observed to be moderately resistant and also predicted to be moderately resistant (true positives) expressed as a percentage of the entire cohort as shown below:

$$\text{Correct decision index (CDI)} = \frac{\text{Number of true positives} + \text{Number of true negatives}}{\text{Total number of lines}} \times 100 \quad (12)$$

The higher the correct decision index, the more likely it is to advance moderately resistant lines and discard susceptible lines and vice-versa.

3. Results

3.1. Principal Component and Cluster Analyses for F_5 Populations

All 2590 lines in the 2018 F_5 population were divided into three clusters (Figure 2b). Cluster one had the highest number of lines with 1122, while cluster two had the lowest number with 679 F_5 lines. Principal component (PC) analysis showed that the first two components accounted for 7.2% and 5.8% of genetic variance in the TP18 dataset. After stratification, 200 (8.5%) lines were selected to capture the genetic diversity of each cluster. Ninety-five lines were selected across cluster one, 46 across cluster two and 59 across cluster three. In 2019, the 2715 F_5 lines also were divided into three clusters (Figure 2c). The cluster sizes ranged from 314 F_5 lines in cluster three to 1406 F_5 lines in cluster one with 90, 65, and 45 lines selected from cluster one, two and three, respectively. PC analysis revealed that the first component explained 9.3% of variation while the second component explained 7.4% of variation in the TP19 dataset. To evaluate the TP sampling with and without stratification on the 2017 set, the k-means clustering algorithm divided the 500 F_5 lines into four clusters (Figure 2a). PC analysis showed that about 6.7% and 5.5% of variance were explained by the first and second components respectively, the

lowest among the three populations. The size of the clusters ranged from 73 lines in cluster four to 201 lines in cluster one. From each cluster, 40% of the lines were selected to achieve a total population size of 200 and ranged from 30 lines in cluster four to 80 lines in cluster one.

(A)

(B)

Figure 2. Cont.

Figure 2. Principal components and cluster analyses of F_5 populations with (A) 7102 SNP markers in 2017; (B) 4935 SNP markers in 2018; (C) 3046 SNP markers in 2019. Solid dots represent different F_5 lines and colors indicate cluster membership.

3.2. Trait Correlations, Phenotypic Variances, Heritabilities and $G \times E$ Interaction

Heritability estimates were significantly different from zero for all sets and traits. The mean heritabilities across years were 0.39 for DIS, 0.40 for VSK, and 0.53 for MicroTwt (Table 2). Genotype by environment interaction was significant for all traits in TP17 and TP19 except MicroTwt weight in TP19. Across the 2018 and 2019 cohorts, 18 parental lines were common in both populations and their differential performance in the three environments (St. Paul 2018, St. Paul 2019, and Crookston 2019) are shown in Figure 3. Trait distribution for DIS, VSK and MicroTwt in the three TPs are presented in Figure 4.

FHB traits were significantly correlated in each population (Table 3, $p < 0.05$). As expected, MicroTwt had a negative relationship with DIS and VSK with a stronger association to VSK than DIS. Correlations between DIS and VSK were highest in TP17 at $r = 0.57$ and lowest in TP18 at 0.40 while the correlation between DIS and MicroTwt ranged from -0.47 in TP17 to -0.27 in TP18. Lastly, the correlation between VSK and MicroTwt weight ranged from -0.75 in TP18 to -0.61 in TP17.

Table 2. Estimates of phenotypic variances, broad-sense heritability and standard error for the 2017, 2018, and 2019 training populations.

Set/Trait	Phenotypic Variance (SE)			H (SE)		
	DIS †	VSK	MicroTwt	DIS	VSK	MicroTwt
TP17 ‡	53.87 (4.25)	52.66 (4.19)	0.30 (0.02)	0.39 (0.06)	0.38 (0.06)	0.56 (0.04)
TP18	272.02 (14.15)	359.15 (18.85)	0.61 (0.06)	0.34 (0.06)	0.39 (0.06)	0.45 (0.03)
TP19	208.59 (12.6)	129.76 (9.08)	0.26 (0.02)	0.45 (0.04)	0.42 (0.05)	0.59 (0.03)

† DIS = FHB-disease index; VSK = visual scabby kernels; MicroTwt = micro-test weight. ‡ TP17 = 2017 training population; TP18 = 2018 training population; TP19 = 2019 training population.

A)

B)

Figure 3. Cont.

Figure 3. G × E interactions observed in (A) FHB disease index (DIS); (B) visual scabby kernels (VSK); (C) micro-test weight (MicroTwt) for 18 parental lines shared across 2018 and 2019 cohorts. StP18 shows the performance of each line in St. Paul 2018, StP19 is the performance in St. Paul 2019, and Crk19 is the performance in Crookston 2019.

Table 3. Trait correlations for 2017, 2018, and 2019 F₅ populations.

Trait	VSK †			MicroTwt		
	TP17 ‡	TP18	TP19	TP17	TP18	TP19
FHB	0.57 *	0.40 *	0.50 *	−0.47 *	−0.27 *	−0.40 *
VSK	-	-	-	−0.61 *	−0.75 *	−0.65 *

† DIS = FHB-disease index; VSK = visual scabby kernels; MicroTwt = micro-test weight. ‡ TP17 = 2017 training population; TP18 = 2018 training population; TP19 = 2019 training population. * significant at 0.05 probability levels.

(A)

(B)

Figure 4. Cont.

(C)

Figure 4. Trait distribution of FHB disease index (DIS), visual scabby kernels (VSK) and micro-test weight (MicroTwt) for: (A) 2017 F₅ population; (B) 2018 F₅ population; (C) 2019 F₅ population. Red bars indicate the position of the moderately resistant check line Alsen and yellow bars indicate the position of the susceptible check line Wheaton.

3.3. Linkage Disequilibrium

As expected, genome-wide pairwise LD (r^2) was highest in the 2017 set that had the largest marker density (7102) with genome-wide r^2 values ranging from 0.23 in the entire TP17 to 0.28 in TP17 selected by stratified sampling. Lowest LD estimates were observed in TP19 as it had the least number of markers (3046) with genome-wide r^2 values ranging from 0.06 in the entire TP19 to 0.14 in TP19 lines selected by stratified sampling (Table 4). Across all the three populations, r^2 values were the lowest in the entire TP and the highest in sets selected by stratified sampling (Table 4).

Table 4. Linkage disequilibrium measured as r^2 and number of markers in each sub-genome and genome-wide in the three F₅ populations.

TP Design Method (No of Lines)	Genome	r^2 (Number of Markers)		
		TP17 ‡	TP18	TP19
Stratified Sampling (200)	A	0.25 (2860)	0.20 (2025)	0.14 (1181)
	B	0.26 (3174)	0.22 (2199)	0.17 (1202)
	D	0.34 (1068)	0.22 (710)	0.11 (663)
	Genome-wide	0.28 (7102)	0.21 (4934)	0.14 (3046)
Pedigree (300)	A	0.21	0.17	0.09
	B	0.19	0.19	0.11
	D	0.28	0.20	0.06
	Genome-wide	0.23	0.19	0.09
Entire TP (500)	A	0.22	0.17	0.09
	B	0.20	0.19	0.12
	D	0.28	0.19	0.07
	Genome-wide	0.23	0.18	0.09

‡ TP17 = 2017 training population; TP18 = 2018 training population; TP19 = 2019 training population.

3.4. Training Population Comparison

For TP17, predictive ability was the lowest for VSK and MicroTwt at 0.34 and highest for DIS at 0.45 when all 500 lines were used (Table 5). We further divided the 500 F₅ lines into clusters to test the ability of TPs selected with and without stratified sampling. The average predictive abilities of TP17 designed by stratified sampling was lowest for MicroTwt at 0.29 and highest for VSK at 0.32. This range was lower than what we observed when the entire TP17 of 500 F₅ lines was used to train the prediction model. However, we observed even lower predictive abilities (0.005 for DIS, 0.03 for VSK, and 0.02 for MicroTwt) when the same size TP17 was designed by random sampling without stratification with prediction abilities that were not significantly different from zero (Table 5).

Average predictive abilities for TP18 designed by stratified sampling ranged from 0.09 for DIS to 0.49 for MicroTwt while predictive abilities for TP18 selected randomly without stratification were not significantly different from zero for all traits. When the entire TP18 was used, predictive ability was the lowest for DIS at 0.30 while MicroTwt was the highest at 0.34. The predictive ability for DIS in the entire TP18 (0.30) was significantly higher ($p < 0.05$) than that of small size TP18 (0.09) selected by stratified sampling. We observed poorer predictive abilities for TP19 compared to TP17 and TP18. Average predictive ability in TP19 designed by stratified sampling was lowest for MicroTwt at 0.01 and highest for DIS at 0.10 while predictive abilities for TP19 selected randomly without stratification were not significantly different from zero for the three traits. Finally, when the entire TP19 was used, predictive abilities were also not significantly different from zero for all traits (Table 5).

3.5. Adding Parents to the 2018 and 2019 Training Populations

Eighteen parental lines were common to both the 2018 and 2019 F₅ populations. When parents of the 2018 F₅ lines were added to TP18 designed by the stratified sampling method, the average predictive abilities ranged from 0.25 for DIS to 0.53 for MicroTwt, a significant improvement ($p < 0.05$) in predictive ability of up to 16 percentage points for DIS, the trait that had the lowest predictive ability without parents in the TP. Although we also observed an increase in predictive ability with the addition of parents to the TP for the other two traits, the increments were rather small and non-significant (i.e., one and four percentage points for VSK and MicroTwt, respectively). While the addition of F₅ parents to the lines selected randomly without stratification resulted in improved predictive abilities for DIS, VSK and MicroTwt by 13, 11 and 3 percentage points respectively, those increments were only significant for DIS ($p < 0.05$). When the parental lines were added to the entire TP18, small, non-significant improvements in predictive ability were observed (Table 5).

The addition of 2019 F₅ parents to the TP19 designed by stratified sampling led to a significant ($p < 0.05$) increase in predictive ability for DIS with a 34 percentage point gain. Although non-significant, we also observed an eight percentage point increase in predictive ability for VSK and a four percentage point increase for MicroTwt. For the TP19 designed without stratification, we observed an 11 percentage point improvement in predictive ability for DIS, 2 for VSK and 5 for MicroTwt. However, none of those increases were statistically significant. When parental lines were added to the entire TP19, the only significant increase ($p < 0.05$) in predictive ability was observed for DIS (15 percentage points), while VSK and MicroTwt had a non-significant 4 percentage point increase and 6 percentage point decrease, respectively (Table 5).

Table 5. Predictive abilities of F₅ lines, with and without parents, selected with different TP design (Stratified Sampling, Non-Stratified Sampling, and Entire Training Population) for FHB disease index (DIS), visual scabby kernels (VSK) and micro-test weight (MicroTwt).

Trait	Predictive Ability (Confidence Interval at 95%)					
	Stratified Sampling *		Non-Stratified Sampling		Entire Training Population	
	F5 only§	F5 + P	F5 only	F5 + P	F5 only	F5 + P
TP17 ‡	DIS †	0.30 (0.22–0.38)	-	0.005 (−0.08–0.09)	-	0.45 (0.38–0.52)
	VSK	0.32 (0.24–0.40)	-	0.03 (−0.06–0.12)	-	0.34 (0.26–0.42)
	MicroTwt	0.29 (0.21–0.37)	-	0.02 (−0.06–0.1)	-	0.34 (0.26–0.42)
TP18	DIS	0.09 (0.002–0.18)	0.25 (0.17–0.33)	−0.01 (−0.1–0.08)	0.12 (0.03–0.21)	0.30 (0.22–0.38)
	VSK	0.44 (0.36–0.51)	0.45 (0.38–0.52)	0.03 (−0.06–0.12)	0.12 (0.03–0.21)	0.34 (0.26–0.42)
	MicroTwt	0.49 (0.42–0.55)	0.53 (0.46–0.59)	0.02 (−0.06–0.1)	0.05 (−0.04–0.14)	0.30 (0.22–0.38)
TP19	DIS	0.10 (0.01–0.19)	0.44 (0.35–0.52)	0.04 (−0.05–0.13)	0.15 (0.06–0.24)	0.005 (−0.08–0.09)
	VSK	0.06 (−0.03–0.15)	0.14 (0.05–0.23)	0.04 (−0.05–0.13)	0.06 (−0.03–0.15)	0.02 (−0.06–0.1)
	MicroTwt	0.01 (−0.081–0.1)	0.05 (−0.04–0.14)	0.003 (−0.09–0.09)	0.05 (−0.04–0.14)	−0.004 (−0.09–0.08)

* TP selected by stratified sampling (done retrospectively for TP17); † DIS = FHB-disease index; VSK = visual scabby kernels; MicroTwt = micro-test weight. ‡ TP17 = 2017 training population; TP18 = 2018 training population; TP19 = 2018 training population. § F₅ only—TP comprised only F₅ lines; F₅ + P = TP comprised both F₅ lines and F₅ parents.

3.6. Single Replication vs. Multiple Replications

The effect of replication in our FHB disease nurseries was assessed with the combined set of F₅ parents and lines selected by stratified sampling in TP18 and TP19. In the single location (St. Paul) for TP18, predictive abilities in Rep1 was highest for MicroTwt with 0.41 and lowest for DIS with 0.11. When the predictive abilities of phenotypic data averaged across replications was compared to Rep1, we observed significant ($p < 0.05$) increases up to 14 percentage points in predictive ability for DIS, 28 percentage points for VSK and 12 percentage points for MicroTwt (Table 6).

For TP19, predictive abilities when the phenotypic data from Rep 1 was used ranged from 0.06 for DIS to 0.11 for MicroTwt in St. Paul and from 0.08 for MicroTwt to 0.19 for DIS in Crookston. In St. Paul, when the predictive abilities of Rep1 were compared with the average across both replications, we found that the use of phenotypic data averaged across the two replications led to a significant 36 percentage point increase ($p < 0.05$) for DIS and non-significant 8 percentage point reduction for VSK and 8 percentage point increase for MicroTwt. In Crookston, when the predictive abilities of phenotypic data averaged across replications were compared to the Rep1, we observed a significant 16 percentage points increase ($p < 0.05$) in predictive abilities for DIS and no significant difference for both VSK and MicroTwt (Table 6).

3.7. Accuracy of Genomic Selection for Line Advancement

As expected, the ability to make correct selection decisions increases with increasing predictive ability. TP17 was highest for correct decision index with values ranging from 81% for VSK to 86% for DIS. The percentage of lines that were rightly discarded were between 52% for VSK and 72% for DIS. In TP18, the correct decision index was lowest for DIS at 73% and highest for MicroTwt at 80% with about 47% (DIS) to 59% (MicroTwt) lines correctly discarded. TP19 that had the lowest predictive ability also had the lowest correct decision index with values varying between 56% for VSK and 68% for DIS. The number of lines correctly discarded ranged from 30% for VSK to 36% for DIS and MicroTwt (Figure 5).

Table 6. Predictive abilities of F₅ lines selected by stratified sampling with parents for FHB disease index (DIS), visual scabby kernels (VSK) and micro-test weight (MicroTwt) when phenotyped in one or two replications.

Year	TP Data	Predictive Ability (Confidence Interval at 95%)					
		St. Paul			Crookston		
		DIS †	VSK	MicroTwt	DIS†	VSK	MicroTwt
TP18 ‡	Rep 1 *	0.11 (0.02–0.2)	0.17 (0.08 –0.25)	0.41 (0.33–0.48)	-	-	-
	Rep mean	0.25 (0.17–0.33)	0.45 (0.38–0.52)	0.53 (0.46–0.59)	-	-	-
TP19	Rep 1	0.06 (–0.03–0.15)	0.11 (0.02–0.2)	0.07 (–0.02–0.16)	0.19 (0.10–0.27)	0.18 (0.09–0.26)	0.08 (–0.01–0.17)
	Rep mean	0.42 (0.35–0.50)	0.03 (–0.06–0.12)	0.15 (0.06–0.24)	0.35 (0.27–0.43)	0.18 (0.09–0.26)	0.06 (–0.03–0.15)

† DIS = FHB-disease index; VSK = visual scabby kernels; MicroTwt = micro-test weight. ‡ TP18 = 2018 training population; TP19 = 2019 training population. * Rep 1 = TP had phenotypic data from the first rep; Rep mean = TP had phenotypic data that was averaged across the two reps.

Figure 5. Scatterplots of observed vs predicted for FHB disease index (DIS), visual scabby kernels (VSK) and micro-test weight (MicroTwt). (A) 2017 cohort, (B) 2018 cohort, and (C) 2019 cohort. The red dashed line represents the cutoff limit to discard genotypes based on the observed field data so that every genotype to the right side of the red dashed line would be discarded for traits in which lower values are desirable (i.e., DIS and VSK) while the genotypes to the left of the red dashed line are discarded for MicroTwt. The green dashed line is the cutoff limit based on genomic predicted values, every genotype above the green dashed line would be discarded for DIS and VSK while genotypes below the green dashed line would be discarded for MicroTwt.

4. Discussion

4.1. Small Population Size

Our results clearly showed that higher predictive abilities can be obtained when prediction models were trained on a small sized TP selected by stratified sampling compared to a TP of similar size selected without stratification. Except for the three traits in TP17 and DIS in TP18, the small sized population selected by stratified sampling also had higher predictive abilities than when the entire TP was used. Similar to our findings, several studies [23–25,43] have shown that a small TP with closely related lines had a greater predictive ability than a same-sized or larger TP with less-related lines.

With complete linkage between marker and QTL, increasing the size of the TP should increase the predictive abilities [44,45]; hence, using the entire TP should lead to a higher accuracy than a small optimized set. However, a large sized TP can have a lower predictive ability if the marker and QTL are not in complete linkage disequilibrium (LD). A smaller set of closely related lines should have increased precision in estimating additive genetic relationships that will increase predictive abilities [45]. A less-related TP (in our case, the small TP selected randomly without stratification and the entire TP) should have poorer marker and QTL associations and will require large TP size and marker density to have high accuracies [46]. This may explain why our lowest prediction accuracies were observed for the small TP selected randomly without stratification compared to when the entire TP was used. From theoretical expectations and field studies [17–21], we expected the predictive ability to be greatest when the entire TP was used; however, in this study, predictive abilities were only highest in TP17 when the entire TP was used. Comparatively, predictive ability for the entire TP was the highest in TP17 that also had the largest number of markers, and the lowest in TP19 that had less than half the marker density of TP17. The large marker density and relatively high r^2 values in TP17 suggests that genome-wide marker coverage adequately captured genomic relationships within the lines, leading to higher accuracies. Increasing marker density in TP19 would increase the probability that markers will be in LD with QTL, which might improve accuracies [46,47]. Furthermore, the low predictive ability observed in TP19 could also be due to the atypical environmental conditions in 2019 (cold season and late harvesting due to rains). Rainfalls late in the season caused seed bleaching that might have confounded the VSK ratings for TP19.

The greater predictive ability of a small optimized TP over the entire TP in 2018 and 2019 indicate the presence of population or family structure in the breeding germplasm. Since each cluster had closely related lines, the large DNA segments shared among these lines coupled with a higher accuracy to separate true signal from noise, due to the absence of less-related lines, allowed for a more precise estimate of realized genomic relationships that increases the power to accurately predict genotypic values, leading to higher predictive abilities.

4.2. Addition of Parents to TP

Across different populations, traits, and TP design methods, we found that the addition of parental lines to the TP led to improved accuracies for all but one trait in the TPs investigated. Several studies [48–50] also reported similar findings across different crop species. Interestingly, we observed the largest improvements in predictive abilities for DIS across the different TP design methods tested. While we are not sure why the addition of parents to DIS led to such large improvement in accuracies, a possible contributing factor for this might be that the level of inbreeding confounds our ability to accurately phenotype field-based traits like DIS. In our breeding program, parents are advanced breeding lines that are almost completely homogeneous (F₇-derived) and homozygous. With such a high level of homogeneity and homozygosity, the parental lines are less likely to segregate within a row, thus ensuring more accurate disease scoring in the field. However, the F₅ lines phenotypically evaluated are F₃-derived, in theory; only ~94% homozygous and highly heterogenous. It is not uncommon to observe F₅ lines segregating within a row for heading date or height in the field nurseries, the two traits that affect the amount and effectiveness of inoculum that reaches the spike and subsequently,

the level of infection and DIS scoring. Hence, the addition of high quality DIS data of the parental lines likely augments the data for the F₅ lines that are not well characterized; therefore, improving the prediction accuracy.

After correcting for bias due to heterosis, Liang et al. [49] observed a moderate increase in prediction accuracy when inbred parents were combined with their hybrid population compared to a scenario where data from inbred parents were excluded in pearl millet. Minamikawa et al. [50] studying Japanese pear, also showed that training a genomic prediction model on both parental and breeding populations led to a greater accuracy than when each population was used separately. Across eight growth and wood traits of eucalyptus, prediction accuracy increased in seven out of eight traits when all parental lines were included as part of the TP compared to a model trained on random lines [48]. The improvement in predictive ability when parental and F₅ populations are combined as the TP over using the F₅ lines alone indicates that the parental lines added more information for accurate estimation of genotypic effects.

5. Conclusions

In this study, we found that a small size TP of 200 F₅ lines selected by stratified sampling had higher predictive abilities compared to the same-size TP selected randomly without stratification or using a larger TP of 500 F₅ lines. We also found that combining both parental lines and the F₅ lines as the TP for model training led to higher accuracies than when the F₅ lines were used as the TP alone. Our findings indicate that a training population consisting of a few hundred lines and their parents, tested in two replications, generate satisfactory predictions for the unphenotyped lines. This is especially important in traits such as FHB that are difficult and expensive to phenotype. In fact, genomic selection was implemented at the F₅ stage of our breeding program to help reduce the amount of resources expended at this stage. Smaller breeding programs, particularly those with a limited capacity to operate FHB disease nurseries desiring to implement genomic selection can also benefit by developing TPs that are closely related to the target populations.

Author Contributions: Methodology, P.B., A.H.S., J.A.A.; software, E.A.; validation, E.A.; formal analysis, E.A.; data curation, E.A., P.B., E.C.; writing—original draft preparation, E.A.; writing—review and editing, J.A.A., P.B., E.C., A.H.S., E.A.; supervision, J.A.A.; funding acquisition, J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported by the U.S. Department of Agriculture, under Agreement Nos. 59-0206-8-202. This is a cooperative project with the U.S. Wheat & Barley Scab Initiative. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

Acknowledgments: E.A. was supported by a Project Aggrad Fellowship. We thank Susan Reynolds for helping with field experiments and Ruth Dill-Macky for providing pathology support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Food and Agriculture Organization of the United Nations. Available online: <http://www.fao.org/resources/infographics/infographics-details/en/c/240943/> (accessed on 27 January 2020).
2. McMullen, M. Impacts of FHB on the North American agriculture community—The power of one disease to catapult change. In *Fusarium Head Blight of Wheat and Barley*; Leonard, K.J., Bushnell, W.R., Eds.; American Phytopathological Society: St. Paul, MN, USA, 2003; pp. 484–503.
3. Nganje, W.; Kaitibie, S.; Wilson, W.; Leistriz, F.; Bangsund, D. Economic impacts of Fusarium head blight in wheat and barley: 1993–2001. *Agribus. Appl. Econ.* **2004**. Available online: <http://ageconsearch.umn.edu/bitsream/23627/1/aer538.pdf> (accessed on 25 January 2020).
4. Desjardins, A.E. *Fusarium Mycotoxins, Chemistry, Genetics, and Biology*; American Phytopathological Society: St. Paul, MN, USA, 2006.

5. Paul, P.A.; McMullen, M.P.; Hershman, D.E.; Madden, L.V. Metaanalysis of the effects of triazole-based fungicides on wheat yield and test weight as influenced by Fusarium head blight intensity. *Phytopathology* **2010**, *100*, 160–171. [CrossRef] [PubMed]
6. Paul, P.; Lipps, P.; Hershman, D.; McMullen, M.; Draper, M.; Madden, L. Efficacy of triazole-based fungicides for Fusarium head blight and deoxynivalenol control in wheat: A multivariate meta-analysis. *Phytopathology* **2008**, *98*, 999–1011. [CrossRef] [PubMed]
7. Willyerd, K.; Li, C.; Madden, L. Efficacy and stability of integrating fungicide and cultivar resistance to manage Fusarium head blight and deoxynivalenol in wheat. *Plant Dis.* **2012**, *96*, 957–967. [CrossRef] [PubMed]
8. Steiner, B.; Buerstmayr, M.; Michel, S.; Schweiger, W.; Lemmens, M.; Buerstmayr, H. Breeding strategies and advances in line selection for Fusarium head blight resistance in wheat. *Trop. Plant Pathol.* **2017**, *42*, 165–174. [CrossRef]
9. Wetterstrand, K.A. *DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program*; National Human Genome Research Institute: Bethesda, MD, USA, 2016. Available online: <http://www.genome.gov/sequencingcosts> (accessed on 25 January 2020).
10. Heffner, E.; Lorenz, A.; Jannink, J.-L.; Sorrells, M. Plant breeding with Genomic selection: Gain per unit time and cost. *Crop. Sci.* **2010**, *50*, 1681–1690. [CrossRef]
11. Crossa, J.; Pérez, P.; Hickey, J.; Buvargueno, 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* **2014**, *112*, 48–60. [CrossRef]
12. Rutkoski, J.; Benson, J.; Jia, Y.; Brown-Guedira, G.; Jannink, J.-L.; Sorrells, M. Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Gen.* **2012**, *5*, 51–61. [CrossRef]
13. Jiang, Y.; Zhao, Y.; Rodemann, B.; Plieske, J.; Kollers, S.; Korzun, V.; Ebmeyer, E.; Argillier, O.; Hinze, M.; Ling, J.; et al. Potential and limits to unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.). *Heredity* **2015**, *114*, 318–326. [CrossRef]
14. Jiang, Y.; Schulthess, A.; Rodemann, B.; Ling, J.; Plieske, J.; Kollers, S.; Ebmeyer, E.; Korzun, V.; Argillier, O.; Stiewe, G.; et al. Validating the prediction accuracies of marker-assisted and genomic selection of Fusarium head blight resistance in wheat using an independent sample. *Theor. Appl. Genet.* **2017**, *130*, 471–482. [CrossRef]
15. Arruda, M.; Brown, P.; Lipka, A.; Krill, A.; Thurber, C.; Kolb, F. Genomic Selection for Predicting Fusarium Head Blight Resistance in a Wheat Breeding Program. *Plant Genome* **2015**, *8*, 1–12. [CrossRef]
16. Jannink, J.L.; Lorenz, A.J.; Iwata, H. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genom.* **2010**, *9*, 166–177. [CrossRef] [PubMed]
17. Daetwyler, H.; Villanueva, B.; Woolliams, J. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* **2008**, *3*, e3395. [CrossRef] [PubMed]
18. Daetwyler, H.; Pong-Wong, R.; Villanueva, B.; Woolliams, J. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* **2010**, *185*, 1021–1031. [CrossRef] [PubMed]
19. Combs, E.; Bernardo, R. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome* **2013**, *6*, 1–7. [CrossRef]
20. Norman, A.; Taylor, J.; Edwards, J.; Kuchel, H. Optimising Genomic Selection in Wheat: Effect of Marker Density, Population Size and Population Structure on Prediction Accuracy. *G3 Gens Genomes Genet.* **2018**, *8*, 2889–2899. [CrossRef] [PubMed]
21. Lozada, D.; Mason, R.; Sarinelli, J.; Brown-Guedira, G. Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genet.* **2019**, *20*, 82. [CrossRef]
22. Rincint, R.; Laloë, D.; Nicolas, S.; Altmann, T.; Brunel, D.; Revilla, P.; Rodríguez, V.M.; Moreno-Gonzalez, J.; Melchinger, A.; Bauer, E.; et al. 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* **2012**, *192*, 715–728. [CrossRef]
23. Lorenz, A.; Smith, K. Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. *Crop. Sci.* **2015**, *55*, 2657–2667. [CrossRef]

24. Rutkoski, J.; Singh, R.; Huerta-Espino, J.; Bhavani, S.; Poland, J.; Jannink, J.L.; Sorrells, M. Efficient Use of Historical Data for Genomic Selection: A Case Study of Stem Rust Resistance in Wheat. *Plant Genome* **2015**, *8*, 1–10. [CrossRef]
25. Edwards, S.M.; Buntjer, J.; Jackson, R.; Bentley, A.; Lage, J.; Byrne, E.; Burt, C.; Jack, P.; Berry, S.; Flatman, E.; et al. The effects of training population design on genomic prediction accuracy in wheat. *Theor. Appl. Genet.* **2019**, *132*, 1943–1952. [CrossRef]
26. Isidro, J.; Jannink, J.-L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* **2015**, *128*, 145–158. [CrossRef] [PubMed]
27. Macqueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *5th Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1967; pp. 281–297.
28. Lohr, S.L. *Sampling: Design and Analysis*; Duxbury Press: Belmont, CA, USA, 1999.
29. Scheaffer, R.; Mendenhall, W.; Ott, R.; Gerow, K. *Elementary Survey Sampling*, 7th ed.; Duxbury Press: Belmont, CA, USA, 2012.
30. Froberg, R.; Stack, R.; Mergoum, M. Registration of ‘Alsen’ wheat. *Crop. Sci.* **2006**, *46*, 2311–2312. [CrossRef]
31. Anderson, J.; Wiersma, J.; Linkert, G.; Reynolds, S.; Kolmer, J.; Jin, Y.; Dill-Macky, R.; Hareland, G. Registration of ‘Rollag’ spring wheat. *J. Plant Reg.* **2015**, *9*, 201–207. [CrossRef]
32. Busch, R.; McVey, D.; Rauch, T.; Baumer, J.; Elsayed, F. Registration of Wheaton wheat. *Crop. Sci.* **1984**, *24*, 622. [CrossRef]
33. Fuentes, R.; Mickelson, H.; Busch, R.; Dill-Macky, R.; Evans, C.; Thompson, W.; Wiersma, J.; Xie, W.; Dong, Y.; Anderson, J. Resource allocation and cultivar stability in breeding for Fusarium head blight resistance in spring wheat. *Crop. Sci.* **2005**, *45*, 1965–1972. [CrossRef]
34. Jones, R.; Mirocha, C. Quality parameters in small grains from Minnesota affected by fusarium head blight. *Plant Dis.* **1999**, *83*, 506–511. [CrossRef] [PubMed]
35. Lenth, R. Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.2. 2019. Available online: <https://CRAN.R-project.org/package=emmeans> (accessed on 5 November 2019).
36. Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **2012**, *7*, e32253. [CrossRef]
37. IWGSC; Appels, R.; Eversole, K.; Stein, N.; Feuillet, C.; Keller, B.; Rogers, J.; Pozniak, C.J.; Choulet, F.; Distelfeld, A.; et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [CrossRef]
38. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
39. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [CrossRef]
40. Money, D.; Gardner, K.; Migicovsky, Z.; Schwaninger, H.; Zhong, G.; Myles, S. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3* **2015**, *5*, 2383–2390. [CrossRef] [PubMed]
41. Bradbury, P.; Zhang, Z.; Kroon, D.; Casstevens, T.; Ramdoss, Y.; Buckler, E. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef] [PubMed]
42. Endelman, J.B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* **2011**, *4*, 250–255. [CrossRef]
43. Clark, S.; Hickey, J.; Daetwyler, H.; van der Werf, J. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol. GSE* **2012**, *44*, 4. [CrossRef]
44. Goddard, M. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* **2009**, *136*, 245–257. [CrossRef]
45. De Los Campos, G.; Vazquez, A.; Fernando, R.; Klimentidis, Y.; Sorensen, D. Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet.* **2013**, *9*, e1003608. [CrossRef]
46. Hickey, J.; Dreisigacker, S.; Crossa, J.; Hearne, S.; Babu, R.; Prasanna, B.M.; Grondona, M.; Zambelli, A.; Windhausen, V.S.; Mathews, K.; et al. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop. Sci.* **2015**, *54*, 1476–1488. [CrossRef]

47. Heffner, E.L.; Sorrells, M.; Jannink, J.-L. Genomic Selection for Crop Improvement. *Crop. Sci.* **2009**, *49*, 1–12. [[CrossRef](#)]
48. Tan, B.; Grattapaglia, D.; Martins, G.; Ferreira, K.Z.; Sundberg, B.; Ingvarsson, P.K. Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F1 hybrids. *BMC Plant Biol.* **2017**, *17*, 110. [[CrossRef](#)]
49. Liang, Z.; Gupta, S.; Yeh, C.; Zhang, Y.; Ngu, D.; Kumar, R.; Patil, H.T.; Mungra, K.D.; Yadav, D.V.; Rathore, A.; et al. Phenotypic Data from Inbred Parents Can Improve Genomic Prediction in Pearl Millet Hybrids. *G3 Genes Genomes Genet.* **2018**, *8*, 2513–2522. [[CrossRef](#)]
50. Minamikawa, M.; Takada, N.; Terakami, S.; Saito, T.; Onogi, A.; Kajiyama-Kanegae, H.; Hayashi, T.; Yamamoto, T.; Iwata, H. Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (*Pyrus pyrifolia* Nakai). *Sci. Rep.* **2018**, *8*, 11994. [[CrossRef](#)] [[PubMed](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).