

## RESEARCH ARTICLE

## Gains through selection for grain yield in a winter wheat breeding program

Dennis N. Lozada<sup>1\*</sup>, Brian P. Ward<sup>2</sup>, Arron H. Carter<sup>1</sup>

**1** Crop and Soil Sciences Department, Washington State University, Pullman, WA, United States of America, **2** USDA-ARS Plant Science Research Unit, Raleigh, NC, United States of America

\* [den.lozada@wsu.edu](mailto:den.lozada@wsu.edu)

## Abstract

Increased genetic gain for complex traits in plant breeding programs can be achieved through different selection strategies. The objective of this study was to compare potential gains for grain yield in a winter wheat breeding program through estimating response to selection  $R$  values across several selection approaches including phenotypic (PS), marker-based (MS), genomic (GS), and a combination of PS and GS (PS+GS). Ten populations of Washington State University (WSU) winter wheat breeding lines including a diversity panel and F5 and double haploid lines evaluated from 2015 to 2019 growing seasons for grain yield in Lind and Pullman, WA, USA were used in the study. Selection was conducted by selecting the top 20% of lines based on observed yield (PS strategy), genomic estimated breeding values (GS), presence of yield “enhancing” alleles of the most significant single nucleotide polymorphism (SNP) markers identified from genome-wide association mapping (MS), and high observed yield and estimated breeding values (PS+GS). Overall, PS compared to other individual selection strategies (MS and GS) showed the highest mean response ( $R = 0.61$ ) within the same environment. When combined with GS, a 23% improvement in  $R$  for yield was observed, indicating that gains could be improved by complementing traditional PS with GS within the same environment. Validating selection strategies in different environments resulted in low to negative  $R$  values indicating the effects of genotype-by-environment interactions for grain yield. MS was not successful in terms of  $R$  relative to the other selection approaches; using this strategy resulted in a significant ( $P < 0.05$ ) decrease in response to selection compared with the other approaches. An integrated PS+GS approach could result in optimal genetic gain within the same environment, whereas a PS strategy might be a viable option for grain yield validated in different environments. Altogether, we demonstrated that gains through increased response to selection for yield could be achieved in the WSU winter wheat breeding program by implementing different selection strategies either exclusively or in combination.

## OPEN ACCESS

**Citation:** Lozada DN, Ward BP, Carter AH (2020) Gains through selection for grain yield in a winter wheat breeding program. PLoS ONE 15(4): e0221603. <https://doi.org/10.1371/journal.pone.0221603>

**Editor:** Aimin Zhang, Institute of Genetics and Developmental Biology Chinese Academy of Sciences, CHINA

**Received:** August 6, 2019

**Accepted:** March 26, 2020

**Published:** April 28, 2020

**Copyright:** © 2020 Lozada et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** AHC- Washington State University Hatch Project 1014919 (<http://fs.cahnrs.wsu.edu/home/research-and-grant-unit/grant-preparation/hatch-multi-state/>) and US Department of Agriculture National Institute of Food and Agriculture (<https://nifa.usda.gov/grants>) award number 2016-68004-24770. The funders had no role in study design,

data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

The challenge to develop higher yielding, climate resilient, disease- and pest-resistant, and more nutritious crops has never been more urgent considering anticipated continuing global population growth over the next 30 years [1]. As such, improving genetic gain or performance for important traits such as yield, disease resistance, and adaptation in staple crops such as wheat (*Triticum aestivum* L.) has been the goal of many breeding programs. Genetic gain is the predicted change in mean value of a trait within a population under selection [2] and is represented by the “breeder’s equation”,  $\Delta G = (\sigma_a)(i)(r)/L$ , where  $\Delta G$  is the change in genetic gain;  $\sigma_a$  is the additive genetic variation within the population,  $i$  is the selection intensity,  $r$  is the selection accuracy, and  $L$  is the number of cycles per year [3]. To increase genetic gain, an increase in the phenotypic variance, accuracy of selection, and selection intensity, or a decrease in generation time for cultivar development, is necessary [4]. Phenotypic, genomic, and marker-based selection approaches could be used to increase either of the factors mentioned to achieve improved gains.

In bread wheat, phenotypic selection for superior genotypes, characterized primarily by a “non-shattering” phenotype, began during its domestication [5]. This “unconscious” breeding resulted from the unintentional selection of lines that were more adapted and productive under early farming practices [6]. “Empirical” and “scientific” breeding followed the “unconscious”, which resulted in the development of wheat lines with improved characteristics in breeding programs [6]. Currently, plant breeders have access to advanced genome and phenotypic-based selection strategies to fast-track genetic improvement and increase gains for key traits in wheat [1].

Several studies have evaluated the gains which could be achieved by applying different selection strategies particularly for increasing resistance to specific diseases in wheat. Rutkoski et al. [7] compared gains for phenotypic and genomic selection for quantitative stem rust resistance and observed that genomic selection could perform as well as phenotypic selection for stem rust resistance improvement but can result in less genetic variance within a population over time. Significant gains using marker-assisted selection for Fusarium head blight (FHB) resistance were also observed in the University of Minnesota wheat breeding program due to the introgression of a major quantitative trait locus for FHB resistance (*Fhb1*). Using closely linked and diagnostic markers for *Fhb1* caused a 27% reduction in disease symptoms throughout the breeding program [8]. In another study, FHB severity in winter wheat was reduced by 6 and 5% using phenotypic and marker-aided selection, respectively [9]; whereas marker-assisted breeding for reduced disease severity and grain deoxynivalenol (DON) content resulted in higher gains in FHB resistance on an annual basis in spring wheat [10]. Both studies observed a large variation for FHB resistance in the marker-selected lines demonstrating the need to complement marker-based selection with phenotypic selection to further enhance gains.

Grain yield is a complex trait controlled mainly by many loci with small effects [11–13] making yield more difficult to select for than disease resistance. Improvement in grain yield, however, remains the prime emphasis of many wheat breeding programs [14], and therefore it is necessary to measure gains achieved through different breeding and selection strategies. As there are now several selection approaches available to plant breeders, we were interested in quantifying the possible gains for grain yield which could be attained when these methods are implemented either alone or in combination with others in a winter wheat breeding program. The objective of this study was to compare the projected gains for yield resulting from using different selection strategies in the Washington State University (WSU) winter wheat breeding program. Empirical datasets for grain yield collected from over 2,200 WSU winter wheat breeding lines grown from 2015 to 2019 were evaluated. The different selection strategies

assessed included phenotypic, marker-based, genomic, and a combination of phenotypic and genomic selection. Potential gains for yield represented as the response to selection  $R$  were calculated for these selection strategies. Validation for grain yield response  $R$  was also conducted using different environments for the different selection approaches.

## Materials and methods

### Winter wheat populations

A total of ten winter wheat breeding populations adapted to the US Pacific Northwest were used in the study. These populations included an association mapping panel (AMP), two F5 biparental populations, two double haploid (DH) biparental populations, and five populations of winter wheat lines evaluated for preliminary yield trials (Prel) as validation populations for yield. The AMP consisted of 456 lines evaluated in Lind (LND) and Pullman (PUL) WA, USA between 2015 and 2018. Significant soil crusting delayed the growth of the winter wheat lines in LND in 2016 and hence the AMP was not evaluated for this site-year. The F5 populations were comprised of 61 and 501 lines planted in 2017 in LND (LND17\_F5) and PUL (PUL17\_F5), WA respectively. The DH populations were evaluated in LND and PUL in 2018 and consisted of 447 (LND18\_DH) and 759 (PUL18\_DH) winter wheat breeding lines. The Prel validation populations comprised of lines selected and advanced from the 2017 and 2018 F5 and DH populations and were evaluated in LND and PUL, WA.

### Phenotypic data collection and analyses

Grain yield (in  $\text{t ha}^{-1}$ ) was assessed by harvesting whole plots using a Zurn<sup>®</sup> 150 combine (Waldenburg, Germany). Adjusted yields were calculated using an Augmented Complete Block Design (ACBD) with replicated checks and un-replicated test genotypes in each block using a custom R script [15]. Field plots were 2.5 m in length, with each entry covering  $\sim 3.7 \text{ m}^2$  area and  $\sim 260$  plants per  $\text{m}^2$  where each block contained  $\sim 37$  and 45 plots. The winter wheat line ‘Eltan’ [16] was used as a check in LND and ‘Madsen’ [17] was used as a check in PUL for the 2015–2018 growing seasons for the AMP. Checks for the LND17\_F5 included the lines ‘Bruehl’ [18], ‘Eltan’, ‘Otto’ [19], ‘Jasper’ [20], ‘Madsen’, and ‘Xerpha’ [21], whereas ‘Brundage’ [22], ‘Jasper’, ‘Madsen’, ‘Puma’ [23], ‘UI Bruneau’, and ‘Xerpha’ were used for the PUL17\_F5 population. ‘Jasper’, ‘Otto’, and ‘Xerpha’ were used as checks for LND18\_DH; whereas ‘Jasper’, ‘Madsen’, ‘Puma’, and ‘Xerpha’ were used as checks for the PUL18\_DH panel.

Adjusted values for yield were calculated employing two statistical models following Lozada and Carter [24]. Briefly, the models used were:

$$Y_{ij} = \mu + B_i + G + C + I + \varepsilon_{ij} \quad (1)$$

$$Y_{ijkl} = \mu + G + C + I + E_i + I \times E_i + G \times E_i + C \times E_i + B_k(E_i) + \varepsilon_{ijkl} \quad (2)$$

where  $Y$  is the trait of interest;  $\mu$  is the effect of the mean;  $B_i$  is the effect of the  $i^{\text{th}}$  block;  $G$  corresponds to the un-replicated genotypes;  $C$  is the effect of the replicated checks on each block;  $E_i$  is the effect of the  $i^{\text{th}}$  environment;  $I$  is the effect of the identifier of the checks; this was used to differentiate the effects of one check over the other checks, as well as the number of checks present on each block;  $I \times E_i$ ,  $G \times E_i$ , and  $C \times E_i$  are the effects of check identifier by environment, genotype by environment, and check by environment interactions, respectively;  $B_k(E_i)$  is the effect of block nested within each environment; and  $\varepsilon$  is the standard normal error [15]. Best linear unbiased estimates (BLUEs) were calculated for individual environments (Eq (1)), whereas best linear unbiased predictors (BLUPs) were computed for the combined analyses

across locations (Eq (2)). Factors were considered fixed when calculating BLUEs whereas effects were regarded as random for calculating BLUPs.

### Genome-wide association study and genomic predictions

SNP genotyping was conducted using genotyping-by-sequencing (GBS) using the protocol of Poland et al. [25,26] through the NC State University Genomics Sciences Laboratory in Raleigh, NC, USA. The restriction enzymes *MspI* and *PstI* were used for GBS. Short-read sequences were aligned to the Chinese Spring IWGSC RefSeq v1.0 [27] using the Burrow-Wheeler Aligner (BWA) 0.7.17 [28] followed by SNP calling using TASSEL-GBS v. 5.2.43 [29,30]. SNP markers were filtered for minor allele frequency (MAF) of  $> 0.05$  and 10% missing data and were designated with their chromosome location and base pair position, e.g. “S1A\_497083519”. Imputation of missing data was done using the linkage disequilibrium  $k$ -nearest joining imputation (LD-kNNi) function [31] in TASSEL v.5.2.25. After filtering and quality control, 16,233 markers (genotype data 1, GD1; S1 File), where 15,853 (97.7%) of which aligned to contigs which were mapped to chromosomes, remained and were used for genome-wide association study (GWAS) using a fixed and random effects circulating probability unification (FarmCPU; [32]) kinship ( $K$ ) only with no PC included in the model in R [33]. The optimal number of PCs to be included in the GWAS model was determined using the “model selection” function in GAPIT [34], which uses a Bayesian information criterion to select which is the optimal model to use. Based on analyses, a PC = 0 was the optimal number across the datasets (S1 Table), and therefore a  $K$  only model was used for GWAS. Minimal deviations from the quantile-quantile (QQ) plots were also observed, indicating that the GWAS model used already sufficiently accounted for familiar relatedness and population structure, and hence PC was excluded in the model (S1 Fig). SNP loci were declared to be significant under a Benjamini-Hochberg false discovery rate (FDR) [35] threshold of 0.05 to control false positive results. The percent phenotypic variation explained ( $R^2$ ) by each significant SNP locus was calculated using a stepwise regression model in JMP<sup>®</sup> Genomics v.8.1 [36], where the  $R^2$  value when a marker was removed from the regression model was subtracted from the total  $R^2$  when all the significant SNP markers were fitted in the model.

Genomic predictions and genomic estimated breeding value (GEBV) calculations were implemented in the iPAT (Intelligent Prediction and Association Tool) package [37], where a ridge regression best linear unbiased prediction (RRBLUP) selection model [38] was trained using the AMP to predict the yield performance of WSU F5 and DH winter wheat breeding lines for independent validations. This prediction model shrinks marker effects towards zero with a common variance [38]. RRBLUP uses the ‘mixed.solve’ function in the form:  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ ,  $\mathbf{u} \sim N(0, \mathbf{K}\sigma^2_u)$ , where  $\mathbf{X}$  is a full-rank design matrix for the fixed effects,  $\boldsymbol{\beta}$ ;  $\mathbf{Z}$  is the design matrix for the random effects  $\mathbf{u}$ ,  $\mathbf{K}$  is a semidefinite covariance matrix, obtained from markers using the ‘A.mat’ (additive relationship matrix function); residuals are normal with a mean of zero and constant variance; and  $\mathbf{u}$  and  $\boldsymbol{\varepsilon}$  independent [38].

A total of 11,089 high-quality GBS-derived SNP markers common to both the AMP and the validation sets (genotype data 2, GD2; S2 File) were used for genomic predictions. GD2 was a subset of GD1 which was used to perform association analyses using the AMP. Phenotypic data for yield in the validation populations (F5 and DH breeding lines) were masked by representing them as “NA” during each analysis. Two GS scenarios were implemented, namely, a standard GS (GS1) considering all markers as contributing equally to the polygenic background effect, and a GWAS-assisted GS scheme (GS2) using the AMP as training population. Predictive ability for the independent validations were calculated as the Pearson correlation between GEBV and adjusted yield for the F5 and DH wheat breeding lines. For the

GWAS-assisted GS (GS2), the top five most significant SNP markers based on an FDR of 0.05 were fitted in an RRBLUP genomic prediction model as fixed effects in iPAT. A total of seven BLUE and two BLUP yield datasets were used for GWAS and genomic predictions. Relatedness between the diversity training panel and winter wheat test lines were assessed using Rogers genetic distances calculated in JMP Genomics v.8.0.

### Correlation between GEBV for one year and observed yield in the succeeding year

The relationships between calculated breeding values from one year and corresponding adjusted yield values in the succeeding year were evaluated by calculating GEBV of the lines in the AMP and comparing them to their adjusted yield in the next growing season (e.g. GEBV for PUL2015 was compared to adjusted yield in PUL2016), which in essence, is also test of genotype-by-environment interactions. GEBVs were calculated by performing a five-fold cross-validation for the AMP, where 80% of the lines were used to predict the remaining 20% using an RRBLUP model in iPAT for the GS1 scenario. The Pearson correlation coefficients between GEBV and adjusted yield were calculated.

### Selection strategies and response to selection

Different selection approaches for grain yield, namely phenotypic (PS), marker-based (MS), genomic (GS), and phenotypic + genomic (PS+GS) selection were compared in this study. For PS, the top 20% of the F5 and DH lines based on adjusted values for yield were selected. In MS, lines having five yield “enhancing” loci identified from association mapping using the AMP were selected. These loci represented the five most significant SNP markers based on a Benjamini-Hochberg FDR of 0.05 across datasets. In the GS approach, the top 20% of the breeding lines having the highest GEBV were identified through independent predictions by training the AMP to predict yield of the F5 and DH breeding lines (GS1). In another GS scenario, five of the most significant markers identified from association mapping using the AMP were included in the selection model as fixed effects to predict yield for the breeding lines using an RRBLUP model (GS2). Finally, for the PS+GS approach, lines having the top 20% highest adjusted grain yield and the highest GEBV were selected for both GS1 (PS+GS1) and GS2 (PS+GS2). The average of the adjusted yield of the corresponding lines selected for each of the selection strategy was reported. Comparisons between mean yield achieved by applying the different selection approaches were also compared to the mean of the check lines.

Gains achieved through each selection approach were represented as the response to selection,  $R$ , calculated as  $R = H^2S$  [39], where  $H^2$  is the broad-sense heritability calculated as  $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$ , where  $\sigma_g^2$  and  $\sigma_e^2$  are the variances due to genotype and error, respectively; and  $S$  is the selection differential, calculated as  $S = \mu_{\text{Selected}} - \mu_{\text{Unselected}}$ , where  $\mu_{\text{Selected}}$  is the mean yield for the lines with a selection strategy implemented and  $\mu_{\text{Unselected}}$  is the mean yield of the lines without selection applied [40]. Both values for broad and narrow-sense heritability can be used to predict  $R$  [41]; however to capture the maximum variation due to genetic effects arising from using different populations of wheat breeding lines from the breeding program,  $H^2$  values were used to calculate response  $R$ . Significance tests using  $t$ -test were conducted to compare the mean response  $R$  of the different selection strategies for yield.

### Validation of selection response for grain yield in different environments

Different selection strategies, namely, PS, GS, and PS+GS were validated in other environments by comparing the mean yield achieved by the top 20% of lines selected from the test

populations (LND17\_F5, LND18\_DH, PUL17\_F5, and PUL18\_DH) through the observed phenotypic values (PS), genomic breeding values (GS), and a combination of observed and genomic values (PS+GS) with their performance in the next growing seasons. To assess the performance of lines selected in each population in different environment, five different validation populations of winter wheat breeding lines evaluated in preliminary yield trials (Prel) from the WSU winter wheat breeding program were used, including LND18\_F5\_Prel, LND19\_DH\_Prel, PUL18\_F5\_Prel, PUL18\_F5\_Prel 2, and PUL19\_DH\_Prel.

## Results

### Distribution of markers and significant marker-trait associations

SNP markers used were distributed in all chromosomes of wheat, with genome B having the largest number of markers (7,201; 44.4%), followed by genome A (6,244; 38.5%), and D (2,408; 14.8%) (S2 Table). Chromosomes 2B, 5B, and 7A had the greatest number of markers (1,306, 1,223, and 1,179 SNP markers, respectively), whereas 4D (92), 6D (284), and 1D (304) had the least among the chromosomes. There were 380 SNP markers (2.3%) that aligned to contigs which were not mapped to any chromosome. A total of 24 significant marker-trait associations (MTAs) distributed across 14 chromosomes were identified for yield in the AMP under a kinship model and an FDR of 0.05 (Table 1). FDR adjusted *P*-values for the significant markers ranged between 6.43E-06 (S1A\_535858090) and 0.048 (S3B\_482345832), whereas allele effects ranged between -0.39 and 0.26. The significant MTAs had an average minor allele frequency of 0.32. No SNP locus was identified to be significant across all locations; nevertheless, five loci were identified as having the most significant *p*-value and being significant across most of the locations (Table 1). These five SNP markers, located on chromosomes 1A, 3B, 6B, and 7A, were subsequently used to validate the MS approach.

### Predictive ability and genomic estimated breeding values for grain yield

Prediction ability for the GS1 scenario under independent validations were low, ranging from -0.21 (PUL16 predicting LND17\_F5) to 0.21 (PUL15 predicting LND17\_F5) across the wheat breeding lines (Fig 1). Overall, higher accuracies were observed for predicting the F5 lines compared with the DH populations (accuracy of 0.03 vs. 0.0002). No significant differences were observed for accuracies when models were trained using the LND and PUL datasets (0.01 vs. 0.02). Predicting LND17\_F5 and LND18\_DH wheat breeding lines using LND datasets resulted in a mean prediction ability of -0.01 whereas using PUL17\_F5 and PUL18\_DH as validation populations resulted in a mean predictive ability of 0.01 (S3 Table). Across environment predictions using the LND yield datasets to predict PUL17\_F5 and PUL18\_DH populations resulted in a mean of 0.04, whereas using PUL datasets to predict LND17\_F5 and LND\_18 DH resulted in a mean of 0.02. BLUP datasets showed an advantage over BLUE datasets for predictions (0.02 vs. 0.01) across different validation populations. Mean grain yield GEBV for all the breeding lines across each dataset ranged between 2.22 (LND15 as training dataset) and 9.99 (PUL18 as training dataset) for GS1 (S4 Table).

Predicting grain yield using the five SNP markers identified from GWAS as fixed effects in the model (GS2) did not result in significant differences in mean accuracy overall, although it resulted in an increase in predictive ability (0.05 vs. 0.02). Significant differences ( $P < 0.05$ ) for mean prediction ability, nonetheless, were observed for PUL17\_F5 and PUL18\_DH. Prediction ability for GS2 ranged between -0.09 (PUL18 predicting PUL17\_F5) and 0.27 (PUL15 predicting PUL18\_DH). Highest mean prediction ability across datasets was observed for PUL18\_DH (0.19), followed by LND17\_F5 (0.05), LND18\_DH (0.02), and PUL17\_F5 (-0.04). Predicting yield using BLUP datasets did not give advantage over to using BLUEs for

Table 1. SNP markers associated with grain yield identified in a diverse training panel of US Pacific Northwest winter wheat lines (N = 456 lines).

SNP	Dataset	P-value	FDR adj. P-value, $q^a$	Minor allele frequency	Percent variation explained, $R^2$	Reference
S1A_497083519	PUL15	1.91E-06	0.01	0.38	0.02	
S1A_535858090 <sup>b</sup>	PUL18	<b>4.13E-10</b>	<b>6.43E-06</b>	<b>0.34</b>	<b>0.03</b>	
S1B_8150831	PUL18	3.90E-06	0.01	0.14	0.05	[43]
S2A_752287563	LND17	7.92E-06	0.02	0.36	0.01	[43]
S2B_239862383	LND17	1.38E-06	0.01	0.37	1.0E-04	[43]
S2B_775486161	PUL18	5.71E-06	0.01	0.41	0.02	[43,44]
S2D_639821303	LND17	1.52E-05	0.03	0.18	0.02	
S2D_642029978	LND17	5.66E-06	0.02	0.08	7.0E-04	
S3A_22831895	LND18	4.68E-06	0.04	0.42	0.02	[42,43]
S3A_567971108	PUL15	3.04E-06	0.01	0.19	0.01	[42,43]
S3B_482345832	PUL18	2.51E-05	0.05	0.47	2.5E-03	[43]
S3B_561570016	PUL18	<b>4.49E-07</b>	<b>3.0E-03</b>	<b>0.26</b>	<b>0.02</b>	[43]
S3B_818284683	PUL15	<b>6.39E-07</b>	<b>5.0E-03</b>	<b>0.47</b>	<b>0.02</b>	[43]
S3D_325690	LND17	2.25E-06	0.01	0.21	3.3E-03	
S5B_29125444	LND17	4.18E-07	0.01	0.08	5.0E-04	[43]
S5B_47592949	PUL18	2.83E-06	0.01	0.31	0.01	[43]
S5B_679577399	LND18	8.40E-06	0.04	0.32	0.02	[43]
S6A_601959488	LND17	1.42E-05	0.04	0.21	1.2E-03	[43,44]
S6B_118986455	LND18	<b>1.05E-08</b>	<b>1.0E-04</b>	<b>0.15</b>	<b>0.03</b>	[44]
S6B_33331876	PUL18	4.43E-06	0.01	0.50	8.0E-04	[44]
S7A_545581556	PUL15	1.44E-05	0.03	0.46	0.06	[42-44]
S7A_61774265	PUL15	<b>6.86E-08</b>	<b>1.0E-03</b>	<b>0.36</b>	<b>0.02</b>	[42-44]
S7B_711208053	PUL15	1.29E-06	0.01	0.45	0.02	
S7D_635365239	PUL15	6.48E-06	0.02	0.46	0.01	[44]

<sup>a</sup> FDR- False discovery rate

<sup>b</sup> Significant SNP markers in bold text were included in the prediction model as fixed effects for a GWAS-assisted genomic selection scenario (GS2)

<https://doi.org/10.1371/journal.pone.0221603.t001>

predictions. In contrast to GS1, within environment predictions resulted in a 50% gain in mean prediction ability compared to predicting across environments. Similar with the GS1

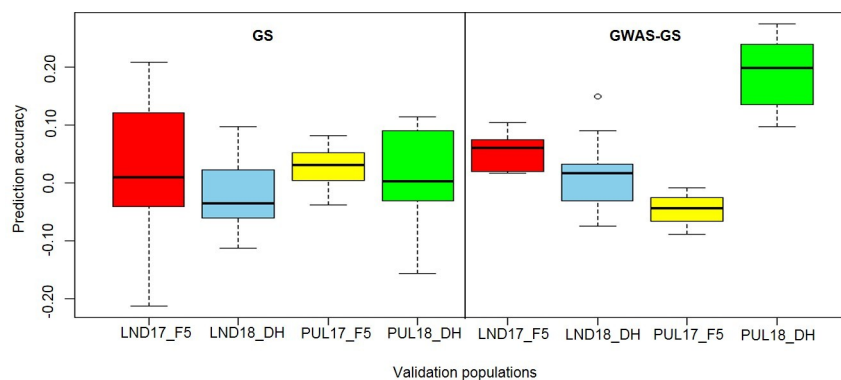


Fig 1. Box plots for prediction ability across a standard genomic selection approach using RRBLUP (GS1) and a GWAS-assisted GS scheme (GS2) for grain yield in a winter wheat breeding program using the AMP as training population.

<https://doi.org/10.1371/journal.pone.0221603.g001>

scenario, the highest mean GEBV for yield was observed for PUL18 (7.68) whereas the lowest was observed for LND15 (1.74) (S5 Table).

Correlations between GEBV and adjusted yield for the winter wheat breeding lines were low to high, ranging between 0.08 (LND18) and 0.71 (PUL combined across years, PUL-Com). Likewise, significant associations ( $P < 0.0001$ ) between GEBV and yield were observed across growing seasons for the diverse population of US Pacific Northwest winter wheat lines (AMP) (Fig 2). Correlation coefficients ranged from 0.003 (PUL15GEBV\_PUL16GY) to 0.22 (PUL17GEBV\_PUL18GY).

### Response to selection across different selection strategies

Selection of the top 20% of winter wheat lines was based on adjusted yield and GEBV for the F5 and DH wheat breeding populations in the same environment. This resulted in selecting 91 (AMP), 12 (LND17\_F5), 90 (LND18\_DH), 100 (PUL17\_F5), and 150 lines (PUL18\_DH) (Table 2). The highest average value for response to selection,  $R$ , was 0.63 for PS+GS1. Using paired  $t$ -test, the mean  $R$  of the PS+GS (PS+GS1 and PS+GS2) approaches were observed to be significantly ( $P < 0.0001$ ) different with that of the means for GS1, GS2, and MS. The mean  $R$  for PS+GS, however, was not significantly different than that of PS, although there was a 5% increase in  $R$  for PS+GS1. Negative mean values for selection response were observed for both GS1 (-0.003) and MS (-0.35) (Tables 2 and 3). No line was selected under the LND17\_F5 population using an MS approach, whereas there were four, 86, and 11 lines containing five favorable alleles for the most significant SNP markers identified from GWAS for LND18\_DH, PUL17\_F5, and PUL18\_DH, respectively. Validating grain yield in different environments by comparing the yield of the selected 20% in each population evaluated in another environment (i.e. the following growing season) based on PS resulted in low to negative selection responses (Table 2), where  $R$  values ranged between -0.28 and 0.26. Using LND18\_F5\_Prel for validating LND17\_F5 yield through a PS approach resulted in significant ( $P < 0.0001$ ) differences in mean  $R$  compared to other populations under a paired  $t$ -test. Mean  $R$  for both GS and PS+GS for yield validated in different environments were also low, ranging between -0.29 and 0.08 (S3 Table).

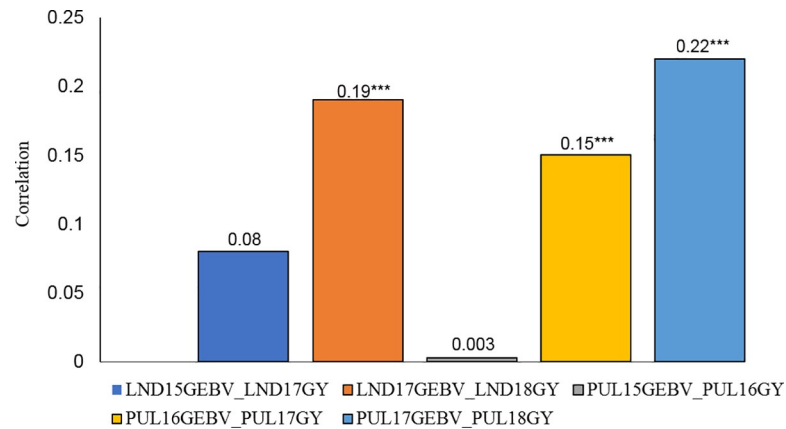
Using both PS+GS1 and PS+GS2 strategies, with mean  $R$  of 0.63 and 0.53 respectively, were more advantageous in terms of response than MS, GS1, and GS2 within the same environment (Table 4). Using GWAS-derived SNP markers as fixed effects in the prediction model in the GS2 scenario resulted in higher mean  $R$  (0.10) compared to GS1 (-0.003). The number of lines selected on both PS and GS ranged from 0 to 44 for both PS+GS1 and PS+GS2 approaches. There were no breeding lines selected for both PS and GS scenarios when PUL16 was used to predict LND17\_F5. There were 16 values for  $R$  (44%) for the PS+GS1 that were greater than the  $R$  value using the PS alone. On the other hand, only 13  $R$  values (36%) for the PS+GS2 were greater than the  $R$  for PS (Table 4, underscored and boldfaced values).

Significant differences ( $P < 0.05$ ) were observed between the mean  $R$  values for PS, GS, and MS when the mean of the checks was compared to the mean yield for the population under selection (S6–S8 Tables). Mean  $R$  values for PS and PS+GS1 both resulted in a 56% gain in response when compared to the mean of the checks. A total of 16 selection response values (44%) for the PS+GS1 showed higher  $R$  compared to the PS, whereas no  $R$  value for the PS+GS2 was observed to be greater than that for PS alone (S8 Table).

### Discussion

This study reports significant marker-trait associations for grain yield, and the potential gains, represented as the response to selection  $R$ , which could be achieved through employing





**Fig 2. Correlation between genomic estimated breeding values (GEBV) and adjusted yield for consecutive growing seasons for a diverse association mapping population (AMP) of US Pacific Northwest winter wheat evaluated in Lind (LND) and Pullman (PUL), WA from 2015–2018. \*\*\* - Significant correlation at  $P < 0.0001$ .**

<https://doi.org/10.1371/journal.pone.0221603.g002>

different selection strategies for grain yield in a winter wheat breeding program. Among the selection strategies evaluated were phenotypic (PS), marker-based (MS), genomic (GS), and the combination of PS and GS (PS+GS) under independent predictions within the same environment and in different environments. The potential of integrating selection approaches to achieve increased gains for grain yield in winter wheat breeding programs was observed when evaluating within the same environment, whereas a PS approach might be advantageous for achieving optimal gains for yield when selecting across different environments.

### Significant marker-trait associations for grain yield

A GWAS approach identified 24 significant SNP markers associated with grain yield distributed across 14 chromosomes. Many of these coincided with previously identified MTAs in wheat [42–44], indicating the role of these genomic regions in controlling yield. A locus on chromosome 3A (S3A\_22831895) was within ~1.7Mbp of SNP 3A\_21102523, which was associated with yield in a population of spring wheat lines evaluated under high-temperature conditions [43]. The SNP markers on chromosomes 1A, 2D, 3D, and 7B identified in this study could be potentially novel loci controlling grain yield, as they have not previously been reported for wheat. All these associations were environment specific, and no SNP markers were associated across all environments or with the yield BLUP values. In augmented designs,

**Table 2. Response to selection  $R$  based on phenotypic selection (PS) for grain yield validated in different environments for US Pacific Northwest winter wheat.**

Population	Validation pop.	Pop. mean (without selection)	Mean (with selection)	Selection differential <sup>a</sup>	$H^2$ <sup>b</sup>	Response to Selection <sup>c</sup>
LND17_F5	LND18_F5_Prel	3.58	5.33	1.75	0.15	0.26
LND18_DH	LND19_DH_Prel	4.57	4.06	-0.50	0.56	-0.28
PUL17_F5	PUL18_F5_Prel	8.66	8.80	0.14	0.13	0.02
PUL17_F5	PUL18_F5_Prel 2	8.66	8.72	0.06	0.13	0.01
PUL18_DH	PUL19_DH_Prel	9.62	9.27	-0.35	0.53	-0.19

<sup>a</sup> Calculated as the difference between the mean yield of lines with selection and mean yield without selection,  $S = \mu_{\text{Sel}} - \mu_{\text{Unselected}}$

<sup>b</sup> Broad-sense heritability

<sup>c</sup> Calculated as  $R = H^2S$

<https://doi.org/10.1371/journal.pone.0221603.t002>

**Table 3. Response to selection, *R* for GEBV-based selection (GS1 and GS2) strategies within the same environment for grain yield in US Pacific Northwest winter wheat. *R* values calculated based on the mean of population without selection applied.**

Test population	<i>H</i> <sup>2</sup>	Training population (AMP) <sup>a</sup>								
		LND15	LND17	LND18	LND_Com	PUL15	PUL16	PUL17	PUL18	PUL_Com
<b>GS1<sup>b</sup></b>										
LND17_F5	0.15	3.0E-03	-0.02	0.01	0.03	0.05	-0.07	-0.02	0.02	0.02
LND18_DH	0.56	0.08	-0.10	-0.03	-0.12	-0.02	0.0	0.13	-0.12	-0.12
PUL17_F5	0.13	-0.02	-9.0E-03	0.02	-0.01	0.02	0.02	0.01	-0.01	0.03
PUL18_DH	0.53	-0.13	0.11	0.11	0.02	0.27	0.04	-0.10	-0.19	-0.01
<b>GS2<sup>c</sup></b>										
LND17_F5	0.15	0.003	-0.01	-0.03	-0.05	0.06	0.02	4.0E-04	0.09	0.08
LND18_DH	0.56	0.08	-0.10	-0.03	-0.13	-0.02	-2.5E-03	0.13	-0.12	-0.12
PUL17_F5	0.13	-0.02	-9.0E-03	0.02	-0.01	0.02	0.02	0.01	-4.0E-03	0.19
PUL18_DH	0.53	-0.13	0.11	0.06	0.02	0.27	1.10	1.09	1.09	-0.02

<sup>a</sup> AMP-Association mapping panel

<sup>b</sup> GS1- standard genomic selection

<sup>c</sup> GS2- GWAS-assisted genomic selection

<https://doi.org/10.1371/journal.pone.0221603.t003>

BLUP trait values derived from combining trials from multiple locations may be limited by the estimation of variance components [45]. Only minor effect loci were detected across datasets, although five SNP markers were still identified as being the most significantly associated with grain yield and therefore used in subsequent MS approaches. Altogether, our GWAS results further support the complex genetic architecture for grain yield in wheat [11–13].

### Response to selection across different approaches

Response to selection *R* is one of the measures of genetic gain [46]. In the current study, phenotypic selection (PS) showed an advantage over MS and GS approaches in terms of *R*. Selecting a portion of lines (i.e. top 20%) based only on the adjusted yield for the F5 and DH wheat breeding lines showed a potential 24% gain on yield relative to the mean of the unselected population. It was nevertheless observed that combining PS with different GS approaches (PS

**Table 4. Response to selection, *R*, for phenotypic + genomic (PS+GS1 and PS+GS2) selection strategies and number of lines selected in combining both approaches for selection (in parentheses) of yield within the same environment in US Pacific Northwest winter wheat. *R* values calculated based on the mean of population without selection applied.**

Test population	<i>H</i> <sup>2</sup>	Training population (AMP) <sup>a</sup>								
		LND15	LND17	LND18	LND_Com	PUL15	PUL16	PUL17	PUL18	PUL_Com
<b>PS+GS1</b>										
LND17_F5	0.15	0.11 (1)	0.14 (1)	0.15 (2)	<b>0.17</b> (3)	<b>0.17</b> (2)	0	<b>0.18</b> (2)	0.16 (3)	0.15 (2)
LND18_DH	0.56	<b>1.06</b> <sup>b</sup> (24)	0.96 (19)	<b>1.05</b> (21)	0.97 (18)	<b>1.06</b> (18)	0.96 (13)	<b>1.00</b> (26)	<b>1.01</b> (16)	<b>1.01</b> (16)
PUL17_F5	0.13	0.17 (15)	0.17 (20)	0.18 (27)	0.19 (16)	<b>0.20</b> (29)	0.18 (26)	0.18 (29)	0.19 (19)	0.19 (31)
PUL18_DH	0.53	<b>1.14</b> (32)	1.09 (38)	<b>1.12</b> (30)	<b>1.12</b> (35)	<b>1.11</b> (44)	<b>1.10</b> (29)	1.07 (27)	1.05 (23)	<b>1.11</b> (28)
<b>PS+GS2</b>										
LND17_F5	0.15	0.11 (1)	0.14 (1)	-0.02 (2)	0 (3)	0.08 (2)	- (0)	<b>0.18</b> (2)	0.10 (3)	0.14 (2)
LND18_DH	0.56	<b>1.06</b> (24)	0.96 (19)	<b>1.05</b> (21)	0.97 (18)	<b>1.06</b> (18)	0.96 (13)	<b>1.00</b> (26)	<b>1.01</b> (16)	<b>1.01</b> (16)
PUL17_F5	0.13	0.17 (15)	0.17 (20)	0.18 (27)	0.19 (16)	<b>0.20</b> (29)	0.18 (26)	0.18 (29)	0.19 (19)	0.19 (31)
PUL18_DH	0.53	<b>1.14</b> (32)	1.09 (38)	<b>1.17</b> (39)	<b>1.12</b> (35)	<b>1.10</b> (44)	<b>1.10</b> (29)	-0.20 (30)	-0.37 (38)	1.09 (28)

<sup>a</sup> AMP-Association mapping panel

<sup>b</sup> Values in boldface and underlined indicate that the response is greater than that of response for PS within the same environment

<https://doi.org/10.1371/journal.pone.0221603.t004>

+GS1 and PS+GS2) under independent predictions for some of the datasets resulted in improved  $R$  relative to that of the PS (Table 4). This indicates the possibility of achieving increased gains when selecting for lines having high observed yield and high estimated breeding values (GEBV) within the same environment. Our results were consistent with previous observations for increased  $R$  values when GS was combined with a PS approach in selecting for winter wheat lines with high grain yield and tolerance to snow mold, where a 10% and 7% gain in response were observed for the traits, respectively, compared to using PS alone [47,48]. Therefore, when performing selections within the same environment for traits with lower heritability, breeders could consider both information from PS and GS (through GEBV) to select lines. Selecting entries having high observed yield and high breeding values could give an opportunity to choose lines that are likely to do well across environments and years in comparison to lines selected based on phenotype alone in a single year [49]. One caveat for using the PS+GS approach for selection in low heritability traits, however, is that in some instances there would be no lines that have both high GEBV and high observed yield, as in the case of using the PUL16 dataset for predictions. This issue could be circumvented by evaluating more lines and increasing the selection intensity in the breeding program, which could improve the chances of selecting lines having both high phenotypic value and high GEBV.

Selection responses achieved by integrating GWAS-derived markers as fixed effects in the prediction model (GS2) were not significantly different than that of a standard GS approach (GS1) considering only a polygenic background effect, although 17% improvement in the mean  $R$  was observed. This demonstrated the potential to increase gains by incorporating fixed effect markers in the model, consistent with previous studies [50,51]. It should be noted that the markers used as fixed effects in the selection model were identified to be significant only in the training population (AMP) to disregard the effect of “inside trading,” which was previously observed to cause overestimated accuracies for FHB resistance in wheat [40]. These inflated accuracies under “inside trading” are attributed to the bias caused by using significant markers that are identified using data obtained from the validation set of lines for which predictions are being generated [40]. Using simulations, Bernardo [52] previously showed that incorporating markers with  $R^2$  greater than 10% in the model should give an advantage in increasing the accuracy. In the present study, significant loci with  $R^2$  greater than 10% were not identified. Nevertheless, we still observed a positive effect of including significant markers on the predictive ability for grain yield. In addition to using GWAS-derived markers for prediction, the inclusion of genetically correlated, highly heritable traits from high-throughput field phenotyping in the prediction model have been observed to improve selection accuracy for grain yield in wheat [53–56].

Negative responses were observed for marker selection (MS) for wheat breeding lines using independent SNP markers identified from association mapping using the AMP, indicating the inefficiency of using this approach exclusively for the selection of grain yield. Some reasons for this could be the low heritability of the trait, allelic variation, variable marker effects among populations, and linkage between the marker and trait of interest. Additionally, there were no LND17\_F5 lines having favorable allele combinations for the most significant yield-related SNP loci, which further demonstrates the difficulty of performing selections based on an MS approach (Table 2; S9 Table). Nevertheless, its potential for selecting lines with high genetic potential particularly when combined with other approaches should not be discounted. Previously, some QTL validation studies for grain yield in wheat showed the potential of using allele specific assays such as KASP<sup>®</sup> [57] to select for lines with high yield potential. Lozada et al. [58] developed marker assays for yield and component traits and used a diverse panel of spring wheat lines from the International Maize and Wheat Improvement Center (CIMMYT) in Mexico to validate the effects of yield-related loci previously identified in southern US winter

wheat. They eventually showed the potential of developing molecular marker assays that could select for spring wheat lines with improved yield potential. In the context of genomic predictions for FHB related traits in wheat, the use of independent SNP markers (i.e. markers identified using a different mapping population) was previously observed to have neutral or reducing effects on selection accuracy [40].

Improvement for grain yield using MS approaches remains a challenge due to its genetic complexity, heritability, and the effects of genotype-by-environment interactions compared to disease resistance traits which are controlled by relatively few QTL with major effects [59]. In contrast, marker-assisted validation, marker-aided backcrossing, and marker-assisted gene pyramiding have been successfully implemented for different traits such as leaf rust resistance, powdery mildew resistance, and pre-harvest sprouting tolerance, among other traits [60]. In the present study, using MS alone did not result in improved gains, though the utility of MS in combination with GS and PS is an area for potential future study.

Validating grain yield of winter lines selected either through the PS, GS, or an integrated PS +GS approach in different environments altogether did not result in an increase for  $R$  across the datasets. This indicates the relevance of using similar environments when evaluating selections [24,53,61] as differences of QTL effects in environments used for predictions could affect prediction accuracy [61]. With the diversity of environments and years for the multilocation trials in the Pacific Northwest, differences in the effects of QTL for yield were expected and genotype-by-environment interactions were present strongly affecting  $R$ . While some positive responses were seen using one location to predict the other, in general the low heritability of grain yield and the high environmental variation between environments and across years makes this selection method difficult. In other production regions with less environmental variation, prediction across locations may be more feasible.

### Prediction accuracy and GEBV-based selection for grain yield

Predictive ability for grain yield under independent validations was also low. The genetic relatedness between the training and test populations was the main driver of predictive ability in the current study. Average Rogers' genetic coefficient between the training and test populations was 0.31, indicating genetic differences among the populations used (S10 Table) which could have resulted in low prediction accuracy. These results were consistent with a related study [24], where low prediction ability was also observed for grain yield in winter wheat using the same set of winter wheat lines evaluated under partial least square regression prediction models. This demonstrates the complexity of using training panels for prediction within breeding programs. Although the AMP consisted of 456 lines from regional Pacific Northwest breeding programs (which are routinely used as parents in cross-hybridizations), and with over half of the lines coming from the WSU program, it was still limited in usefulness for genomic prediction due to low genetic relatedness with the F5 and DH validation panels that were being selected.

Using GEBV alone for selection was not successful relative to the PS and PS+GS approaches in terms of response to selection for within and across environment selections for grain yield. Negative  $R$  values were observed for almost 50% of the datasets for both GS1 and GS2. Relying exclusively on GEBV for performing selections should therefore be taken with care, as some lines predicted to have high GEBV could have low yield. Correlations between GEBV and observed yield between a year and the next growing season under cross-validations using the AMP were in general low, indicating that high GEBV sometimes do not translate to high observed phenotypic values. This is especially true when evaluating across years due to the possible effects of genotype-by-environment interactions, especially for low heritability traits. In

the context of selecting new parental lines based on GEBV alone, it was recently observed that selecting for high FHB resistance in winter wheat was not reliable, as only 19% of lines (9 out of 47) predicted by GEBV belong to the best 10% for FHB resistance [62]. In another study, negative GEBV for yield were observed for synthetic hexaploid spring bread wheat lines evaluated across heat-stressed and irrigated environments [63]. Selection for drought tolerance in maize using GEBV, in contrast, has resulted in rapid genetic gain and positive selection responses through using molecular markers associated with high yield under drought stress [64].

While selecting lines based on GEBV alone should be considered with caution, the implementation of genomic selection in breeding programs should help increase the rate of genetic gain through a faster breeding cycle, higher selection intensity, and efficiency of genomic prediction approaches in integrating novel genetic material in wide-crosses and pre-breeding programs [65]. For GEBV to be more relevant in breeding programs, strategies that could help increase the selection accuracy, such as using genetically related populations, utilizing optimal training population composition and sizes, and employing ideal number of markers for predictions [66–69] should be implemented. Altogether, our results demonstrated that GEBV could still be used as a selection criterion for grain yield in winter wheat breeding, particularly for within environment selection for grain yield.

### Response to selection: Implications for breeding

Using different selection approaches in combination could result in increased response  $R$ , and hence genetic gain for complex traits. In the context of the WSU winter wheat breeding program, evaluating the selection approaches proposed here on breeding materials from earlier stages before the actual yield trials could also be performed to reduce costs associated with phenotyping and maximize gain achieved from genomic selection. Nevertheless, limitations were found when trying to select lines from one environment and predict their grain yield in another environment. Selection within the same environment was more successful. In earlier generations of selection, a PS+GS approach could result in better genetic gain for selections within the same environment. A PS strategy might still be more advantageous for selecting grain yield across different environments after the initial selection within a given environment.

### Conclusions

Gains in terms of response to selection  $R$  was compared for different selection strategies in a winter wheat breeding program. Phenotypic selection (PS) showed favorable responses to selection compared to genomic (GS) and marker-based selection (MS) approaches. Combining PS with GS showed a great potential in achieving higher  $R$  values compared to using either method alone for lines evaluated in the same environment. Validating selection responses in other environments, in contrast, resulted in negative  $R$ . We observed that GS, when combined with traditional PS for yield, could facilitate an increased response to selection within the same environment and ultimately could improve genetic gain in the WSU winter wheat breeding program. Breeders should therefore make important decisions based on the combination of one or more selection strategies to achieve optimal gains in plant breeding programs. For example, for complex traits evaluated in diverse environments, PS might be advantageous, whereas when evaluating within the same environment, an integrated PS+GS approach, where selections are based on both phenotypic and GEBV information, could result in optimal gains for grain yield. Careful consideration on which selection strategies to implement, depending

on the traits being evaluated, cost, target environments, and available resources should altogether facilitate improved genetic gain for complex traits in winter wheat breeding programs.

## Supporting information

**S1 Fig. Quantile-quantile plots for grain yield.** There were minimal deviations from the diagonal line (in red) indicating that the GWAS model ( $K$  only) was already able to capture population structure and genetic relatedness, and hence, principal components were excluded in the model. The SNP markers that deviate from the diagonal on the upper right-hand section of the graph are the loci significantly associated with the trait.

(TIF)

**S1 Table. Bayesian information criterion (BIC) values for different number of principal components (PC = 0 to PC = 3) across different yield datasets in the US Pacific Northwest association mapping panel (AMP).**

(XLSX)

**S2 Table. Number of SNP markers and marker proportion for each chromosome of wheat used for genomewide association mapping.**

(XLSX)

**S3 Table. Response to selection based on genomic (GS) and phenotypic and genomic selection (PS+GS) strategies for grain yield validated in different environments for US Pacific Northwest winter wheat.**

(XLSX)

**S4 Table. Genomic estimated breeding values (GEBV) for the F5 and DH winter wheat breeding lines under a standard genomic selection (GS1) scenario.**

(XLSX)

**S5 Table. Genomic estimated breeding values (GEBV) for the F5 and DH winter wheat breeding lines with GWAS-derived markers included as fixed effects in the prediction model (GS2).**

(XLSX)

**S6 Table. Response to selection,  $R$  based on phenotypic selection (PS) and marker-based selection (MS) for grain yield in US Pacific Northwest winter wheat.  $R$  calculated relative to the mean of the check lines.**

(XLSX)

**S7 Table. Response to selection,  $R$  for GEBV-based selection (GS1 and GS2) strategies for grain yield in US Pacific Northwest winter wheat.  $R$  calculated relative to the mean of the check lines.**

(XLSX)

**S8 Table. Response to selection,  $R$  for phenotypic + genomic (PS+GS1 and PS+GS2) selection strategies and the number of lines selected in combining both approaches for selection (in parentheses) of yield in US Pacific Northwest winter wheat.  $R$  calculated relative to the mean of the check lines.**

(XLSX)

**S9 Table. Winter wheat breeding lines selected under a marker-based selection (MS) strategy using the five most significant SNPs identified using a GWAS approach for a diverse**

winter wheat mapping panel.

(XLSX)

**S10 Table. Roger's genetic coefficient between the association mapping training panel (AMP) and the winter wheat breeding lines across each chromosome.**

(XLSX)

**S1 File. Genotype data (16,233 SNP markers) for the winter wheat association mapping panel (AMP) used for genomewide association study.**

(XLSX)

**S2 File. Genotype data (11,089 SNP markers) for the winter wheat breeding lines used for genomic predictions for grain yield.** This panel is a subset of the 16,233 markers used for the AMP ([S1 File](#)).

(XLSX)

**S3 File. Adjusted yield (t/ha) for each site-year combination for the winter wheat association mapping panel (AMP).**

(XLSX)

**S4 File. Adjusted yield (t/ha) for the F5 and DH winter wheat breeding lines.**

(XLSX)

## Acknowledgments

The authors would like to thank Gary Shelton and Kyall Hagemeyer for assistance in the collection of yield data.

## Author Contributions

**Conceptualization:** Dennis N. Lozada.

**Data curation:** Dennis N. Lozada, Brian P. Ward.

**Formal analysis:** Dennis N. Lozada.

**Funding acquisition:** Arron H. Carter.

**Resources:** Arron H. Carter.

**Writing – original draft:** Dennis N. Lozada.

**Writing – review & editing:** Dennis N. Lozada, Brian P. Ward, Arron H. Carter.

## References

1. Hickey LT, N. Hafeez A, Robinson H, Jackson SA, Leal-Bertioli SCM, Tester M, et al. Breeding crops to feed 10 billion. *Nat Biotechnol.* 2019; <https://doi.org/10.1038/s41587-019-0152-9> PMID: [31209375](https://pubmed.ncbi.nlm.nih.gov/31209375/)
2. Moose SP, Mumm RH. Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.* 2008; 147: 969–977 <https://doi.org/10.1104/pp.108.118232> PMID: [18612074](https://pubmed.ncbi.nlm.nih.gov/18612074/)
3. Cobb JN, Juma RU, Biswas PS, Arbelaez JD, Rutkoski J, Atlin G, et al. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theor Appl Genet.* 2019; 132: 627–645. <https://doi.org/10.1007/s00122-019-03317-0> PMID: [30824972](https://pubmed.ncbi.nlm.nih.gov/30824972/)
4. Li H, Rasheed A, Hickey LT, He Z. Fast-Forwarding Genetic Gain. *Trends Plant Sci.* 2018; 23: 184–186. <https://doi.org/10.1016/j.tplants.2018.01.007> PMID: [29426713](https://pubmed.ncbi.nlm.nih.gov/29426713/)
5. Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai Y-S, Gill BS, et al. Molecular Characterization of the Major Wheat Domestication Gene Q. *Genetics.* 2006; 172: 547–555. <https://doi.org/10.1534/genetics.105.044727> PMID: [16172507](https://pubmed.ncbi.nlm.nih.gov/16172507/)

6. Venske E, dos Santos RS, Busanello C, Gustafson P, de Oliveira A. Bread wheat: a role model for plant domestication and breeding. *Hereditas*. 2019; 156: 16. <https://doi.org/10.1186/s41065-019-0093-9> PMID: [31160891](https://pubmed.ncbi.nlm.nih.gov/31160891/)
7. Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J, Jannink J-L, et al. Genetic Gain from Phenotypic and Genomic Selection for Quantitative Resistance to Stem Rust of Wheat. *Plant Genome*. 2015; 8. <https://doi.org/10.3835/plantgenome2014.10.0074>
8. Anderson JA, Chao S, Liu S. Molecular Breeding Using a Major QTL for Fusarium Head Blight Resistance in Wheat. *Crop Sci*. 2007; 47: S-112–S-119. <https://doi.org/10.2135/cropsci2007.04.0006IPBS>
9. Miedaner T, Wilde F, Korzun V, Ebmeyer E, Schmolke M, Hartl L, et al. Marker selection for Fusarium head blight resistance based on quantitative trait loci (QTL) from two European sources compared to phenotypic selection in winter wheat. *Euphytica*. 2009; 166: 219–227. <https://doi.org/10.1007/s10681-008-9832-0>
10. Wilde F, Korzun V, Ebmeyer E, Geiger HH, Miedaner T. Comparison of phenotypic and marker-based selection for Fusarium head blight resistance and DON content in spring wheat. *Mol Breed*. 2007; 19: 357–370. <https://doi.org/10.1007/s11032-006-9067-5>
11. Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, et al. A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring xSQ1 and its use to compare QTLs for grain yield across a range of environments. *Theor Appl Genet*. 2005; 110: 865–880. <https://doi.org/10.1007/s00122-004-1902-7> PMID: [15719212](https://pubmed.ncbi.nlm.nih.gov/15719212/)
12. Li F, Wen W, Liu J, Zhang Y, Cao S, He Z, et al. Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol*. 2019; 19: 168. <https://doi.org/10.1186/s12870-019-1781-3> PMID: [31035920](https://pubmed.ncbi.nlm.nih.gov/31035920/)
13. Garcia M, Eckermann P, Haefele S, Satija S, Sznajder B, Timmins A, et al. Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. *PLoS One*. 2019; 14: e0211730. <https://doi.org/10.1371/journal.pone.0211730> PMID: [30716107](https://pubmed.ncbi.nlm.nih.gov/30716107/)
14. Green AJ, Berger G, Griffey CA, Pitman R, Thomason W, Balota M, et al. Genetic yield improvement in soft red winter wheat in the Eastern United States from 1919 to 2009. *Crop Sci*. 2012; 52: 2097–2108
15. Rodríguez F, Alvarado G, Pacheco Á, Burgueño J. ACBD-R. Augmented Complete Block Design with R for Windows. Version 4.0. CIMMYT Research Data & Software Repository Network; 2018. doi: [11529/10855](https://doi.org/10.11529/10855)
16. Peterson CJ, Allan RE, Rubenthaler GL, Line RF. Registration of ‘Eltan’ Wheat. *Crop Sci*. 1991; 31: 1704. <https://doi.org/10.2135/cropsci1991.0011183X003100060075x>
17. Allan RE, Peterson CJ, Rubenthaler GL, Line RF, Roberts DE. Registration of ‘Madsen’ wheat. *Crop Sci*. 1989; 29: 1575–1576
18. Jones SS, Murray TD, Lyon SR, Morris CF, Line RF. Registration of ‘Bruehl’ wheat. (Registrations of Cultivars). *Crop Sci*. 2001; 41: 2006–2008
19. Carter AH, Jones SS, Lyon SR, Balow KA, Shelton GB, Higginbotham RW, et al. Registration of ‘Otto’ wheat. *J Plant Regist*. 2013; 7: 195–200
20. Carter AH, Jones SS, Balow KA, Shelton GB, Burke AB, Lyon S, et al. Registration of ‘Jasper’ soft white winter wheat. *J Plant Regist*. 2017; 11: 263–268
21. Jones SS, Lyon SR, Balow KA, Gollnick MA, Murphy KM, Kuehner JS, et al. Registration of ‘Xerpha’ wheat. *J Plant Regist*. 2010; 4: 137–140
22. Zemetra RS, Souza EJ, Lauver M, Windes J, Guy SO, Brown B, et al. Registration of ‘Brundage’ wheat. *Crop Sci*. 1998; 38: 67
23. Carter AH, Jones SS, Cai X, Lyon SR, Balow KA, Shelton GB, et al. Registration of ‘Puma’ soft white winter wheat. *J Plant Regist*. 2014; 8: 273–278
24. Lozada DN, Carter AH. Accuracy of single and multi-trait genomic prediction models for grain yield in US Pacific Northwest winter wheat. 2019. *Crop Breeding, Genetics, and Genomics* 2019; 1:e190012. <https://doi.org/10.20900/cbagg20190012>
25. Poland JA, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, et al. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*. 2012; 5. <https://doi.org/10.3835/plantgenome2012.06.0006>
26. Poland JA, Brown PJ, Sorrells ME, 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: 1–8. <https://doi.org/10.1371/journal.pone.0032253> PMID: [22389690](https://pubmed.ncbi.nlm.nih.gov/22389690/)
27. Appels R, Eversole K, Stein N, Feuillet C, Keller B, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. 2018; 361: eaar7191. <https://doi.org/10.1126/science.aar7191> PMID: [30115783](https://pubmed.ncbi.nlm.nih.gov/30115783/)



28. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
29. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23: 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
30. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS One*. 2014; 9: e90346. <https://doi.org/10.1371/journal.pone.0090346> PMID: 24587335
31. Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong G-Y, Myles S. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3 (Bethesda)*. 2015; 5: 2383–2390. <https://doi.org/10.1534/g3.115.021667> PMID: 26377960
32. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genet*. 2016; 12: e1005767 <https://doi.org/10.1371/journal.pgen.1005767> PMID: 26828793
33. R Development Core Team. R: A Language and Environment for Statistical Computing. 2018. Vienna, Austria. <https://www.r-project.org/>
34. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012; 28: 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444> PMID: 22796960
35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995; 289–300
36. Institute SAS. SAS System Options: Reference, 2nd ed. Cary, NC: SAS Institute; 2015
37. Chen CJ, Zhang Z. iPat: intelligent prediction and association tool for genomic research. *Bioinformatics*. 2018; 34: 1925–1927. <https://doi.org/10.1093/bioinformatics/bty015> PMID: 29342241
38. Endelman JB. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*. 2011; 4: 250–255
39. Falconer DS. *Introduction to Quantitative Genetics*. 3rd ed. New York: Longman Scientific and Technical; 1989
40. Arruda MP, Lipka AE, Brown PJ, Krill AM, Thurber C, Brown-Guedira G, et al. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol Breed*. 2016; 36: 84. <https://doi.org/10.1007/s11032-016-0508-5>
41. Piepho H-P, Möhring J. Computing Heritability and Selection Response From Unbalanced Plant Breeding Trials. *Genetics*. 2007; 177: 1881–1888. <https://doi.org/10.1534/genetics.107.074229> PMID: 18039886
42. Bhatta M, Morgounov A, Belamkar V, Baenziger PS. Genome-Wide Association Study Reveals Novel Genomic Regions for Grain Yield and Yield-Related Traits in Drought-Stressed Synthetic Hexaploid Wheat. *Int J Mol Sci*. 2018; 19: 3011. <https://doi.org/10.3390/ijms19103011> PMID: 30279375
43. Jamil M, Ali A, Gul A, Ghafoor A, Napar AA, Ibrahim AMH, et al. Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biol*. 2019; 19: 149. <https://doi.org/10.1186/s12870-019-1754-6> PMID: 31003597
44. Ward BP, Brown-Guedira G, Kolb FL, Van Sanford DA, Tyagi P, Sneller CH, et al. Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. *PLoS One*. 2019; 14: e0208217. <https://doi.org/10.1371/journal.pone.0208217> PMID: 30794545
45. Santos AH, Bearzoti E, Ferreira DF, Silva Filho JL da. Simulation of mixed models in augmented block design. *Sci Agric*. 2002; 59: 483–489
46. Rutkoski JE. Chapter Four—A practical guide to genetic gain. In: Sparks DL, editor. Academic Press; 2019. pp. 217–249. <https://doi.org/10.1016/bs.agron.2019.05.001>
47. Lozada DN, Mason RE, Sarinelli JM, Guedira GB. Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genet*. 2019; 20:82 <https://doi.org/10.1186/s12863-019-0785-1> PMID: 31675927
48. Lozada DN, Godoy JV, Murray TD, Ward BP, Carter AH. Genetic Dissection of Snow Mold Tolerance in US Pacific Northwest Winter Wheat Through Genome-Wide Association Study and Genomic Selection. *Front Plant Sci*. 2019; 10: 1337. <https://doi.org/10.3389/fpls.2019.01337> PMID: 31736994
49. Belamkar V, Guttieri MJ, Hussain W, Jarquín D, El-basyoni I, Poland J, et al. Genomic Selection in Preliminary Yield Trials in a Winter Wheat Breeding Program. *G3 (Bethesda)*. 2018; 8: 2735–2747
50. Michel S, Kummer C, Gallee M, Hellinger J, Ametz C, Akgöl B, et al. Improving the baking quality of bread wheat by genomic selection in early generations. *Theor Appl Genet*. 2018; 131: 477–493. <https://doi.org/10.1007/s00122-017-2998-x> PMID: 29063161

51. Galiano-Carneiro AL, Boeven PHG, Maurer HP, Würschum T, Miedaner T. Genome-wide association study for an efficient selection of Fusarium head blight resistance in winter triticale. *Euphytica*. 2018; 215: 4. <https://doi.org/10.1007/s10681-018-2327-8>
52. Bernardo R. Genomewide Selection when Major Genes Are Known. *Crop Sci*. 2014; 54: 68–75. <https://doi.org/10.2135/cropsci2013.05.0315>
53. Rutkoski J, Poland J, Mondal S, Autrique E, Pérez LG, Crossa J, et al. Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. *G3 (Bethesda)*. 2016; 6: 2799–2808. <https://doi.org/10.1534/g3.116.032888> PMID: 27402362
54. Crain J, Mondal S, Rutkoski J, Singh RP, Poland J. Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding. *Plant Genome*. 2018; 11. <https://doi.org/10.3835/plantgenome2017.05.0043> PMID: 29505641
55. Sun J, Poland JA, Mondal S, Crossa J, Juliana P, Singh RP, et al. High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. *Theor Appl Genet*. 2019; <https://doi.org/10.1007/s00122-019-03309-0> PMID: 30778634
56. Lozada DN, Godoy JV, Ward BP, Carter AH. Genomic prediction and indirect selection for grain yield using spectral reflectance indices from high-throughput phenotyping. *Int J Mol Sci*. 2019; 21(1), 165. <https://doi.org/10.3390/ijms21010165> PMID: 31881728
57. Semagn K, Babu R, Hearne S, Olsen M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol Breed*. 2014; 33: 1–14. <https://doi.org/10.1007/s11032-013-9917-x>
58. Lozada DN, Mason RE, Sukumaran S, Dreisigacker S. Validation of grain yield QTLs from soft winter wheat using a CIMMYT spring wheat panel. *Crop Sci*. 2018; 58:1964–1971. <https://doi.org/10.2135/cropsci2018.04.0232>
59. Hospital F. Challenges for effective marker-assisted selection in plants. *Genetica*. 2009; 136: 303–310. <https://doi.org/10.1007/s10709-008-9307-1> PMID: 18695989
60. Gupta PK, Langridge P, Mir RR. Marker-assisted wheat breeding: present status and future possibilities. *Mol Breed*. 2010; 26: 145–161. <https://doi.org/10.1007/s11032-009-9359-7>
61. Huang M, Ward B, Griffey C, Van Sanford D, McKendry A, Brown-Guedira G, et al. The Accuracy of Genomic Prediction between Environments and Populations for Soft Wheat Traits. *Crop Sci*. 2018; <https://doi.org/10.2135/cropsci2017.10.0638>
62. Herter CP, Ebmeyer E, Kollers S, Korzun V, Miedaner T. An experimental approach for estimating the genomic selection advantage for Fusarium head blight and Septoria tritici blotch in winter wheat. *Theor Appl Genet*. 2019; <https://doi.org/10.1007/s00122-019-03364-7> PMID: 31144000
63. Jafarzadeh J, Bonnett D, Jannink J-L, Akdemir D, Dreisigacker S, Sorrells ME. Breeding Value of Primary Synthetic Wheat Genotypes for Grain Yield. *PLoS One*. 2016; 11: e0162860. <https://doi.org/10.1371/journal.pone.0162860> PMID: 27656893
64. Vivek BS, Krishna GK, Vengadessan V, Babu R, Zaidi PH, Kha LQ, et al. Use of Genomic Estimated Breeding Values Results in Rapid Genetic Gains for Drought Tolerance in Maize. *Plant Genome*. 2017; 10. <https://doi.org/10.3835/plantgenome2016.07.0070> PMID: 28464061
65. Hickey JM, Chiurugwi T, Mackay I, Powell W, et al. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat Genet*. 2017; 49: 1297. <https://doi.org/10.1038/ng.3920> PMID: 28854179
66. Lorenz AJ. Resource Allocation for Maximizing Prediction Accuracy and Genetic Gain of Genomic Selection in Plant Breeding: A Simulation Experiment. *G3 (Bethesda)*. 2013; 3: 481–491. <https://doi.org/10.1534/g3.112.004911> PMID: 23450123
67. Zhong S, Dekkers JCM, Fernando RL, Jannink J-L. Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. *Genetics*. 2009; 182: 355–364. <https://doi.org/10.1534/genetics.108.098277> PMID: 19299342
68. Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, et al. Genomic Selection and Association Mapping in Rice (*Oryza sativa*): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. *PLoS Genet*. 2015; 11: 1–25. <https://doi.org/10.1371/journal.pgen.1004982> PMID: 25689273
69. Cericola F, Jahoor A, Orabi J, Andersen JR, Janss LL, Jensen J. Optimizing Training Population Size and Genotyping Strategy for Genomic Prediction Using Association Study Results and Pedigree Information. A Case of Study in Advanced Wheat Breeding Lines. *PLoS One*. 2017; 12: e0169606. <https://doi.org/10.1371/journal.pone.0169606> PMID: 28081208