

*Article*

# Accuracy of Single and Multi-Trait Genomic Prediction Models for Grain Yield in US Pacific Northwest Winter Wheat

Dennis N. Lozada \*, Arron H. Carter

Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA

\* Correspondence: Dennis N. Lozada, Email: den.lozada@wsu.edu.

## ABSTRACT

Incorporating secondary correlated traits collected from high-throughput phenotyping in genomic selection (GS) models for complex traits has been demonstrated to improve accuracy. The prediction ability of different single and multiple trait partial least square (PLS) regression models for grain yield were assessed for winter wheat lines evaluated in US Pacific Northwest environments. Different populations including a diversity panel, F5, and double haploid breeding lines were evaluated in Lind and Pullman, WA between 2015 and 2018 and were genotyped with genotyping by sequencing-derived SNP markers. Prediction ability was assessed under cross-validations and independent predictions. Multi-trait covariate models were advantageous in obtaining optimal predictions for yield, especially when there is less genetic relatedness between the training and test populations. Adding multiple traits in the model improved predictions for environments with low heritability. Cross-validations resulted in the highest prediction ability (0.16) whereas independent predictions using the diversity panel to predict F5 and double haploid winter wheat breeding lines obtained the lowest (0.002). Relatedness between populations, heritability of the secondary traits, and the type of PLS model used were among the principal factors affecting prediction ability. Our results showed the relevance of using multi-trait GS models to achieve increased predictions. Genetic architecture of the target trait and genetic relatedness between populations should be taken into consideration when choosing which type of models to implement in the breeding program. An increased prediction ability for the multi-trait models indicates the potential to attain improved genetic gains for yield in wheat breeding programs through these GS approaches.

## Open Access

Received: 12 June 2019

Accepted: 13 August 2019

Published: 15 August 2019

Copyright © 2019 by the author(s). Licensee Hapres, London, United Kingdom. This is an open access article distributed under the terms and conditions of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

**KEYWORDS:** genetic gain; genomic selection; grain yield; high-throughput phenotyping; partial least square regression; soft winter wheat

## INTRODUCTION

The wealth of genomic information available for important crops has enabled the use of marker data for molecular breeding and DNA-based

selection for plant improvement. In recent years, genomic approaches such as genome-wide association studies (GWAS) have been used to understand the genetic basis of important traits such as grain yield, disease resistance, and adaptation traits in wheat [1–4]. However, association mapping could not identify small effect loci and with such, the power of GWAS for the dissection of complex traits is limited [5,6]. In addition, the structure of breeding populations used in association mapping could cause the detection of spurious associations in genetic mapping [7]. On that regard, a complementary method, genomic selection (GS) has been explored in many crops and its potential to improve genetic gains through selection has been demonstrated [8–11].

In GS, a population with genotypic and phenotypic information is first used to train a prediction model (through a training population), and then used to predict the performance of lines in a validation set that have no phenotypic information [12]. The need to identify significant QTL-marker associations is removed, and instead GS considers numerous predictors simultaneously [13]. A measure of a model's ability to perform trait predictions is its prediction ability, which could be defined as the correlation between observed phenotype with predicted values [14]. Through GS, genomic estimated breeding values (GEBV) can also be calculated and these values can be used for performing selections and choosing which parents to cross. Depending on the trait, a high or low estimated breeding value would indicate that a line is predicted to perform better in succeeding field trials. GS can increase genetic gains by reducing the number of cycles and progenies that need to be phenotyped and by improving the intensity of selection [15,16].

In addition to whole genome-wide marker data available for plant breeders, emerging tools for high-throughput phenotyping (HTP) have been widely utilized to collect phenotypic information in the field [17,18]. Its rapid development in the past few years resulted in increased amounts of phenotypic information available that could lead to better understanding of the complexity of traits such as grain yield [19,20]. Accurate phenotyping remains crucial for the efficient building of prediction models for GS [15]. Phenotyping platforms are typically equipped with different sensors such as red, green, and blue (RGB) cameras, multiple spectral cameras, and LiDAR, among others [21]. Ground-based HTP platforms with cameras that collect reflectance data from plant canopies have been used to assess various traits in different crops such as wheat [22–24], cotton [25], and rice [26]. Vegetation indices such as Normalized Difference Vegetation Index (NDVI), Normalized Water Index (NWI), and Simple Ratio (SR) derived from the absorbance of plant tissues at specific wavelengths of light are associated with important traits such as grain yield, biomass, plant water status, and degree of senescence [27–29]. HTP traits provide additional information on the genetic predictor variables in GS models and thus can be used to improve accuracies for complex traits such as grain yield [15]. Previously, incorporating these traits in genomic

prediction models for grain yield resulted to improved predictions in wheat [30–32]. Combining information from HTP and genome-wide marker profile to identify lines with high genetic potential remain relevant for the improvement of grain yield especially in major wheat producing regions such as the Pacific Northwest area of the US.

Partial least square (PLS) regression takes advantage of correlated predictor variables  $X$  and response variables  $Y$  [33]. It relates both  $X$  and  $Y$  through a linear multivariate model and a dimension reduction approach which fits collinear genetic and environmental factors simultaneously by finding latent variables that explain the variance of the predictor variables as well as the covariance between predictor and response factors [34]. It is useful for predicting dependent variables from large number of predictors that might be highly correlated [35]. Moreover, PLS are computationally fast, statistically efficient, and can be used to address regression problems, predictions, classification, and survival analysis [36]. In Holstein bulls, PLS models were observed to be more accurate than pedigree based BLUP approaches in predicting traits with varying heritability [35]. Similarly, the robustness and superiority of these models over pedigree-based approaches for predicting milk yield, fat content, and somatic cell scores in dairy sheep has also been demonstrated [37]. PLS models have been used to predict grain yield [32]; yield, yield components and physiological traits [38]; and *Septoria tritici* blotch disease in wheat [39], among others. The feasibility of using these models for predicting yield in the presence of secondary traits in the model for Pacific Northwest winter wheat, nonetheless, has not been reported.

The objectives of this study were to compare single and multiple trait GS models for grain yield in US Pacific Northwest (PNW) winter wheat under different cross-validation and independent prediction scenarios, and to identify the prediction model resulting to optimal accuracies for yield. We evaluated the potential of using least square (PLS) regression models to predict grain yield for wheat in PNW growing conditions through using secondary traits collected from HTP field phenotyping.

## MATERIALS AND METHODS

### Winter Wheat Populations

The current study used five different populations of soft winter wheat adapted to the Pacific Northwest region of the US, including a diverse association mapping panel (DAP), F5, and double haploid (DH) populations described previously [40]. Briefly, DAP consisted of 456 lines evaluated in the Washington State University Dryland Research Station in Lind (LND) and in the Spillman Agronomy Farm near Pullman (PUL), WA during 2015–2018, whereas the F5 and DH populations were planted in LND and PUL in 2017 (F5\_LND17 and F5\_PUL17) and 2018 (DH\_LND18 and DH\_PUL18) growing seasons, respectively. F5\_LND17 consisted of 61

whereas F5\_PUL17 consisted of 501 lines. DH\_LND18 and DH\_PUL18 consisted of 449 and 761 winter wheat lines, respectively.

The winter wheat populations were planted in an augmented design [41] with repeated checks and genotypes (un-replicated) in each block. Plot length was 1.5 meter, each entry covered a 1 m<sup>2</sup> area, planted at ~100 plants per m<sup>2</sup>. Checks used were “Eltan” [42] and “Madsen” [43] in LND; and Madsen was used in PUL for 2015–2018 seasons for the DAP. Lines “Bruehl” [44], Eltan, “Otto” [45], “Jasper” [46], Madsen, and “Xerpha” [47] were used as checks for F5\_LND17; and “Brundage” [48], Jasper, Madsen, “Puma” [49], “UI Bruneau”, and “Xerpha” were used for the F5\_PUL17 population. Checks used for the DH\_LND18 lines included Jasper, Otto, and Xerpha; whereas Jasper, Madsen, Puma, and Xerpha were used as checks for the DH\_PUL18 panel. The DAP was not evaluated for SRI and grain yield in LND in 2016 as substantial soil crusting caused delayed emergence for the winter wheat lines.

### Collection and Analysis of Phenotypic Data

Grain yield (in t·ha<sup>-1</sup>) was collected by harvesting whole plots using a Wintersteiger® Nursery combine (Ried im Innkreis, Austria). Spectral reflectance was collected using a CROPSCAN® multiple spectral radiometer (Rochester, MN, USA) attached to a pole and placed approximately 1 m above the canopy and in the middle of each plot. Filters measuring radiation at 16 different wavelengths of light (between 430 and 970 nm) were installed in the CROPSCAN. Data for spectral reflectance were taken approximately 10–15 days apart across three different growth stages- heading, early grain-fill and late grain-fill within a two-hour solar window (between 10:00 am and 2:00 pm) at clear and windless days. Spectral information for the plots were processed through the CROPSCAN MSR® software. Spectral reflectance indices, namely Normalized Difference Vegetative Index (NDVI), Normalized Water Index-1 (NWI-1), and Simple Ratio (SR) were calculated as shown in Table 1.

**Table 1.** Spectral reflectance indices measured for US Pacific Northwest soft winter wheat lines.

Index	Formula	Related traits	Reference(s)
Normalized Difference Vegetative Index (NDVI)	$(R800 - R680) / (R800 + R680)$	Biomass, vegetative greenness, degree of senescence	[50]
Normalized Water Index-1 (NWI-1)	$(R970 - R900) / (R970 + R900)$	Water status, root access to moisture	[27,51]
Simple Ratio (SR)	$R900/R680$	Biomass, vegetative greenness, degree of senescence	[52]

Adjusted means were calculated for an augmented design using the augmented complete block design (ACBD) in R program developed by Rodríguez *et al.* [53] for individual locations and combined across

environments. Best linear unbiased estimates (BLUEs) and predictors (BLUPs) were calculated for individual locations and for combined analyses across environments, respectively. The models used for calculations were (a)  $Y_{ij} = \mu + \text{Block}_i + \text{IDCheck} + \text{Gen} + \text{Check} + \varepsilon_{ij}$  and (b)  $Y_{ijkl} = \mu + \text{IDCheck} + \text{Gen} + \text{Check} + \text{Loc}_i + \text{Loc}_i \times \text{IDCheck} + \text{Loc}_i \times \text{Gen} + \text{Loc}_i \times \text{Check} + \text{Block}_k(\text{Loc}_i) + \varepsilon_{ijkl}$ , for individual environment, BLUE (a) and combined analyses across locations, BLUP (b), where  $Y$  is the trait of interest;  $\mu$  is the effect of the mean;  $\text{Block}_i$  is the effect of the  $i$ th block;  $\text{Gen}$  corresponds to the un-replicated genotypes;  $\text{Check}$  is the effect of the replicated checks on each block;  $\text{Loc}_i$  is the effect of the  $i$ th location;  $\text{IDCheck}$  is the identifier of the checks; and  $\varepsilon$  is the standard normal errors [40]. Effects were considered fixed and random when calculating BLUEs for individual locations and BLUPs for combined analyses across environments, respectively.

### Heritability and Genetic Correlations

Genetic correlation ( $r_G$ ) was calculated according to Falconer [54] using the following formula:  $r_G = (\text{Cov}_{xy}) / \sqrt{\text{Var}_x \times \text{Var}_y}$ ; where  $\text{Cov}_{xy}$  is the covariance between yield and SRI calculated using a multivariate approach in JMP v. 8.1 [55]; and  $\text{Var}_x$  and  $\text{Var}_y$  are the variances for grain yield and SRI, respectively, across all locations for each population (DAP, DH, and F5 lines). Broad-sense heritability for grain yield in individual environments was calculated as  $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$ , where  $\sigma_g^2$  and  $\sigma_e^2$  are the genotype and error variance components, respectively; whereas  $H^2$  across locations (combined analyses) was calculated as  $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{n} + \sigma_{e/nr}^2}$ , where  $\sigma_g^2$  is the genotype variance;  $\sigma_{ge}^2$  is the variance due to genotype by environment interaction;  $n$  is the number of environments;  $r$  is the number of replication per environment (*i.e.*, equal to 1 for an augmented design); and  $\sigma_e^2$  is the error variance using the augmented complete block design (ACBD) in R program [53].

### SNP Genotyping and Genomic Selection

Genotyping was conducted using genotyping-by-sequencing [11] using the restriction enzymes *MspI* and *PstI* through the NC State Genomics Sciences Laboratory in Raleigh, NC, USA. After filtering for minor allele frequency  $\text{MAF} > 0.05$  and quality control, 11,089 SNP markers, with 10,894 of these SNPs (98.2%) having known chromosome positions, were used for analyses. Imputation of missing data was done using the LDknni (linkage disequilibrium  $k$ -nearest neighbor joining imputation) function in TASSEL v. 5.0 [56]. Genetic relatedness between the training and validation populations was evaluated using Rogers genetic distance [57] through the “Population Measures” function in JMP v.8.1.

Genomic predictions were conducted using partial least square (PLS) regression with spectral reflectance data incorporated in the model

following the methods of Crain *et al.* [32]. Prediction scenarios included cross-validations (CV) using a diversity association mapping panel (DAP); and two independent prediction (IP) schemes. In the CV, 80% of the lines was used to train a model to predict the remaining 20% under a five-fold validation. In the IP1 scenario, the DAP was used to predict grain yield of Washington State University F5 and DH winter wheat breeding lines; whereas in the second independent validation (IP2), predictions were conducted within the breeding lines. The BLUE and BLUP datasets for grain yield used for GS analyses are shown in Table 2. Mean spectral reflectance values across the three developmental stages were used as fixed effects in the PLS prediction models for yield.

**Table 2.** Genotypic variance and broad-sense heritability for grain yield across different populations and datasets of US Pacific Northwest soft winter wheat.

Population	No. of lines	Dataset	Mean yield (t·ha <sup>-1</sup> )	Genotypic variance	Heritability <sup>b</sup>
DAP <sup>a</sup>	458	LND15	2.50	0.18	0.70
		LND17	3.12	0.27	0.59
		LND18	4.86	0.29	0.43
		LND_Com	3.50	0.08	0.39
		PUL15	6.97	0.27	0.32
		PUL16	5.84	0.38	0.55
		PUL17	7.77	0.09	0.23
		PUL18	10.15	0.38	0.59
		PUL_Com	7.65	0.15	0.50
F5	61	LND17	3.34	0.08	0.15
	501	PUL17	7.66	0.12	0.13
DH	449	LND18	4.43	1.01	0.56
	761	PUL18	8.57	1.41	0.53

<sup>a</sup> Diversity association mapping panel, <sup>b</sup> Broad-sense heritability; calculated as  $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$  (for individual locations); and  $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{n} + \sigma_{e/er}^2}$  (for combined analyses).

The first PLS regression model used was a univariate model with the spectral reflectance measurements included as covariate predictor traits (Cov). The second model was a multivariate (MV) approach predicting yield together with the spectral reflectance traits. The basic form of the Cov model was: Grain yield =  $\mu + X\beta + Zu + \varepsilon$ ; where  $X$  is the matrix of individual observations (*i.e.*, individual wheat lines);  $\beta$  is the fixed effects of spectral traits;  $Z$  is an ( $n \times m$ ) matrix assigning markers to genotypes;  $u$  is a ( $1 \times n$ ) array of random effects for the markers; and  $\varepsilon$  is the error [32]. Cov model, therefore, is a univariate model that includes predictor traits (in the form of spectral measurements) as covariates.



The MV model used was in the form:

$$\begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} X & 0 \\ \vdots & \vdots \\ 0 & Xn \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} Z & 0 \\ \vdots & \vdots \\ 0 & Zn \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (1)$$

where  $n$  is the number of traits,  $t_1$  is a vector of trait values for grain yield;  $X$  is the matrix for fixed effects which simplifies to a vector of 1 for each trait representing the mean as only the markers were used in this model;  $Z$  is the random marker matrix for each trait; and  $\varepsilon_{1...n}$  is the error term [32]. This model follows a scenario where grain yield is masked in the prediction model for the validation set, and therefore represents a breeding program where only spectral traits have been collected and the plants were not yet harvested, and hence lines are chosen primarily based on predicted values for grain yield [32]. Prediction ability was represented as the Pearson correlation coefficient between the predicted and actual grain yield values obtained by fitting HTP traits in the prediction model. Single trait prediction models included only a single spectral reflectance index, whereas multi-trait models included at least two spectral measurements for GS. All analyses were implemented under the partial least square (“pls”) package [58] in R. Additive and non-additive genetic variances for grain yield for each of the dataset were estimated using the “sommer” package through the “mmer” function [59] in R, using phenotypic and SNP marker data. The proportion of additive to the total genetic variance (*i.e.*, additive + non-additive) was calculated by dividing the additive variance by the total genetic variance.

## RESULTS

### Heritability, Phenotypic, and Genetic Correlations

Broad-sense heritability for grain yield was low to moderate, ranging between 0.13 (F5\_PUL17) and 0.70 (DAP\_LND15). Yield had higher heritability in the DH lines (0.53 and 0.56 for DH\_PUL18 and DH\_LND18, respectively) compared with the F5 lines (0.13 (F5\_PUL17) and 0.15 (F5\_LND17)). In general, for the DAP, significant phenotypic correlations with grain yield were observed for the spectral traits (Table 3). Non-significant phenotypic correlations, nonetheless, were observed between SR and yield for DAP\_LND18 and all the PUL datasets. Likewise, non-significant phenotypic correlations were observed for NDVI in PUL16 and PUL\_Com, and PUL16, PUL18, and PUL\_Com datasets for NWI-1. Significant phenotypic correlation between spectral traits and grain yield were observed for all the F5 and DH winter wheat breeding lines. Overall, higher  $H^2$  values were observed for the HTP traits. Heritability for the spectral traits ranged between 0.52 (NWI-1, DAP\_LND18 and SR, F5\_PUL17) and 0.95 (SR, DAP\_LND17) dataset (Table 4). Genetic correlations using all locations for the different wheat populations ranged between -0.72 (NWI-1; DAP and F5 lines) and 0.78 (NDVI and SR, DAP) (Table 5). Genetic correlations between environments were low to

moderate, ranging between −0.14 and 0.69 (DAP); −0.18 and 0.21 (DAP and the F5 and DH wheat breeding lines); and −0.13 and 0.30 (within the breeding lines) (Supplementary Tables S1–S3, respectively).

**Table 3.** Phenotypic correlation of spectral reflectance traits with grain yield across different US Pacific Northwest soft winter wheat populations.

Population	Dataset	NDVI <sup>a</sup>	NWI-1 <sup>b</sup>	SR <sup>c</sup>
DAP <sup>d</sup>	LND15	0.27 ***	−0.23 ***	0.19 ***
	LND17	0.58 ***	−0.59 ***	0.58 ***
	LND18	0.13 *	−0.11 *	0.09
	LND_Com	0.44 ***	−0.42 ***	0.38 ***
	PUL15	0.11 *	−0.11 *	0.08
	PUL16	−0.06	0.02	−0.07
	PUL17	−0.09 *	0.10 *	−0.08
	PUL18	−0.11 *	0.06	−0.06
	PUL_Com	0.04	−0.07	0.04
F5	LND17	0.77 ***	−0.77 ***	0.70 ***
	PUL17	0.43 ***	−0.46 ***	0.42 ***
DH	LND18	0.70 ***	−0.70 ***	0.64 ***
	PUL18	0.74 ***	−0.71 ***	0.65 ***

<sup>a</sup> NDVI, Normalized Difference Vegetation Index; <sup>b</sup> NWI-1, Normalized Water Index-1; <sup>c</sup> SR, Simple Ratio; <sup>d</sup> Diversity association mapping panel; \*\*\*, significant at  $P < 0.0001$ ; \*\*, significant at  $P < 0.001$ ; \*, significant at  $P < 0.05$ .

**Table 4.** Heritability of spectral reflectance traits across different US Pacific Northwest soft winter wheat populations.

Population	Dataset	NDVI <sup>a</sup>	NWI-1 <sup>b</sup>	SR <sup>c</sup>
DAP <sup>d</sup>	LND17	0.91	0.53	0.95
	LND18	0.77	0.52	0.81
	PUL17	0.78	0.68	0.68
	PUL18	0.77	0.66	0.86
F5	LND17	0.72	0.72	0.63
	PUL17	0.67	0.63	0.52
DH	LND18	0.77	0.77	0.84
	PUL18	0.90	0.90	0.88

<sup>a</sup> NDVI, Normalized Difference Vegetation Index; <sup>b</sup> NWI-1, Normalized Water Index-1; <sup>c</sup> SR, Simple Ratio; <sup>d</sup> Diversity association mapping panel.

### Principal Component Analysis and Genetic Distances within the Breeding Lines

Analysis of principal components using genome-wide marker data revealed a 5.2% variation caused by the first PC (PC1) and 4.5% by PC2 within the F5 breeding lines, with no distinct clustering observed (Figure 1A). PC1 and PC2 caused 16.2 and 4.2% of variation, respectively within the DH lines (Figure 1B) with more differentiation observed compared with the F5, where the DH\_LND lines formed a separate cluster

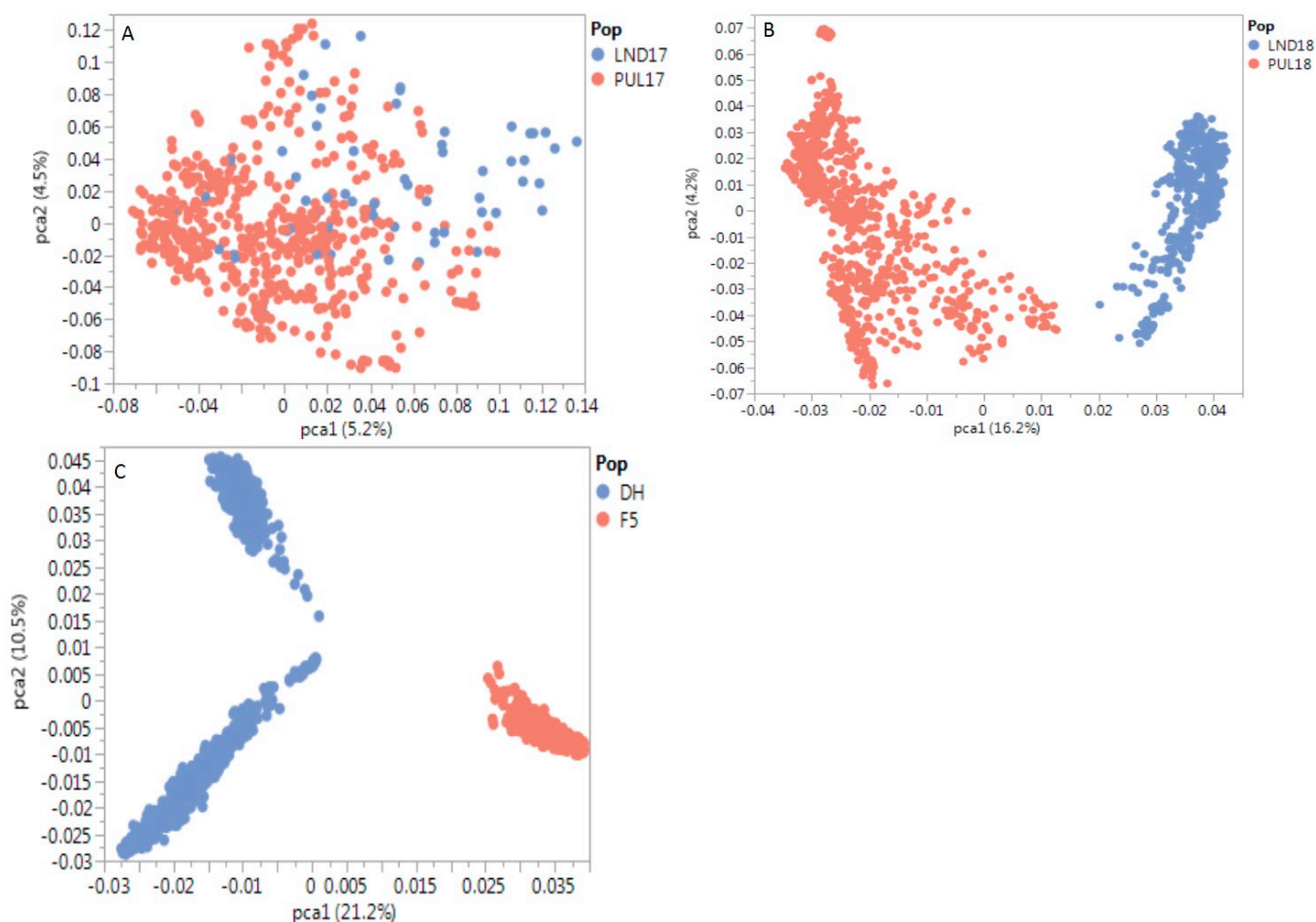


with that of the DH\_PUL lines. When combined on a single PCA biplot, the F5 lines formed a separate group from that of the DH lines, where the first PC caused 21.2% and the second PC caused 10.5% of variation. The DH entries formed separate clusters, indicating a higher genetic differentiation among them (Figure 1C). Genetic (Rogers) distances among the breeding lines were 0.05 (F5), 0.13 (DH) and 0.40 (F5 and DH).

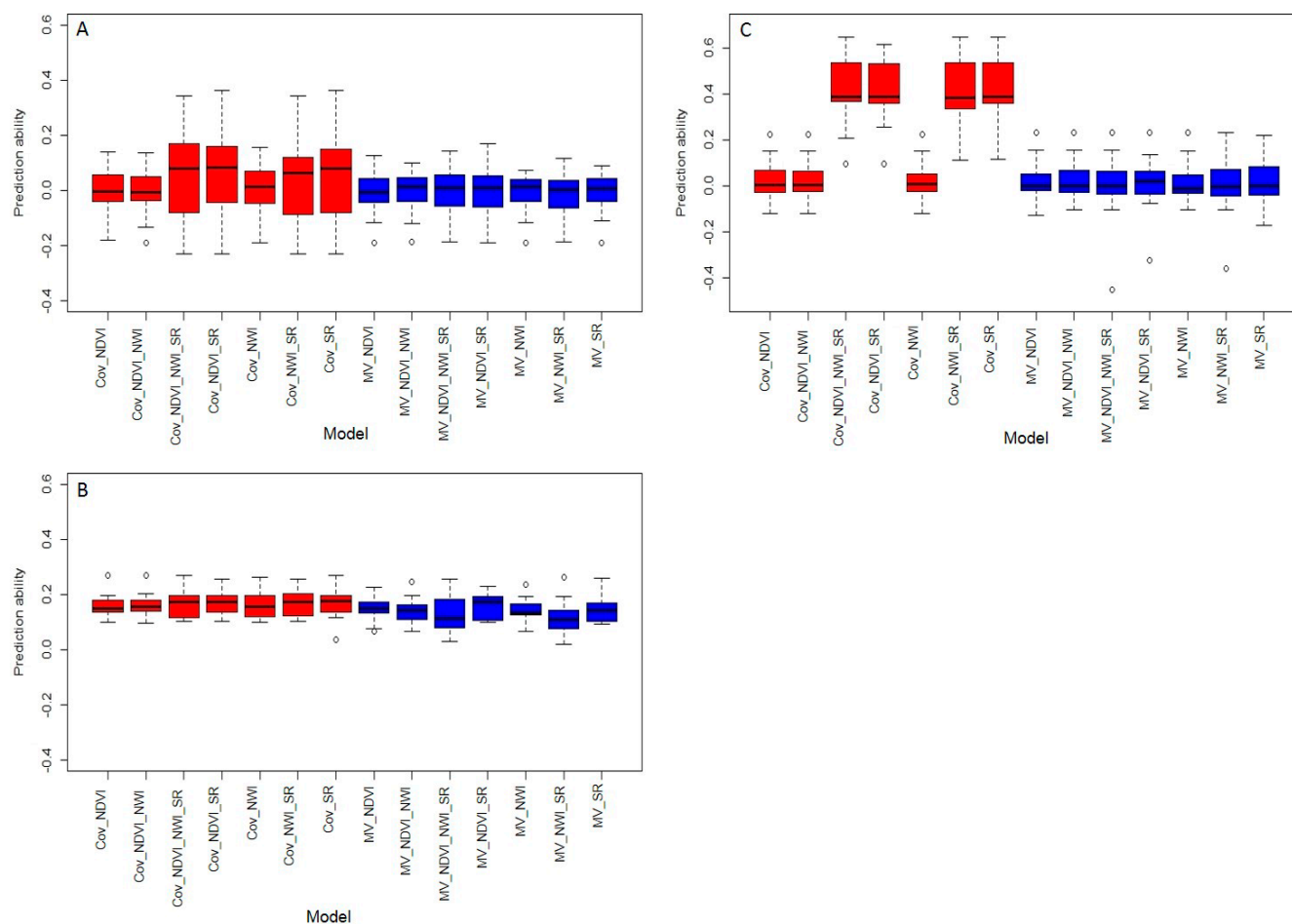
**Table 5.** Variances and genetic correlation of the spectral reflectance traits with grain yield across locations for the winter wheat diversity panel and breeding lines.

Population	$\sigma^2_{GY}$	$\sigma^2$			Genetic correlation		
		NDVI <sup>a</sup>	NWI-1 <sup>b</sup>	SR <sup>c</sup>	NDVI	NWI-1	SR
DAP <sup>d</sup>	8.01	0.033	0.00074	68.529	0.78	-0.72	0.78
F5 lines	3.05	0.009	0.0003	11.647	0.70	-0.72	0.68
DH lines	7.02	0.007	0.0002	7.017	0.54	-0.58	0.55

<sup>a</sup> NDVI, Normalized Difference Vegetation Index; <sup>b</sup> NWI-1, Normalized Water Index-1; <sup>c</sup> SR, Simple Ratio; <sup>d</sup> Diversity association mapping panel; All genetic correlations were significant at  $P < 0.0001$ .



**Figure 1.** PCA biplots for winter wheat breeding lines used for predicting grain yield using partial least square regression models. (A) F5 breeding lines; (B) DH lines; (C) F5 + DH breeding lines.

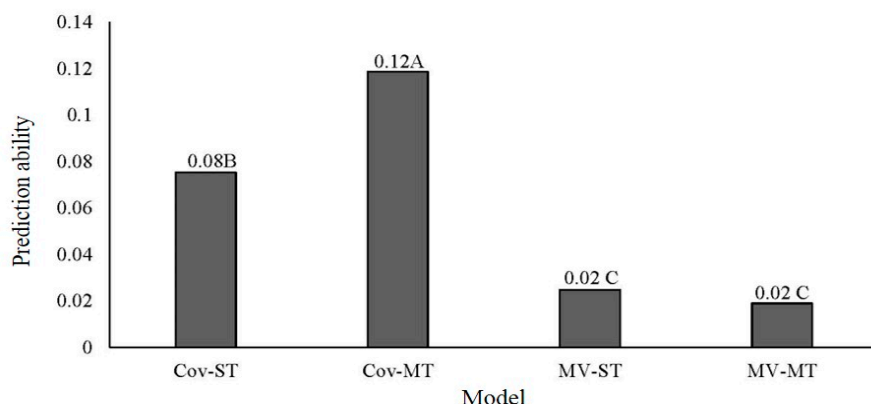


**Figure 2.** Boxplots of prediction ability for grain yield under different genomic selection scenarios using covariate (Cov) and multivariate (MV) partial least square regression models incorporating spectral reflectance traits. **(A)** cross validations using a diversity association mapping panel (DAP); **(B)** independent predictions using DAP to predict F5 and DH winter wheat breeding lines; and **(C)** independent predictions within the breeding lines. Actual values for prediction ability are reported in Supplementary Table S5. NDVI, Normalized Difference Vegetative Index; NWI-1, Normalized Water Index-1; SR, Simple Ratio.

### Prediction Ability for Grain Yield

Boxplots showing prediction ability for the different GS schemes are shown in Figure 2A,C. Overall, prediction ability for yield ranged between -0.45 and 0.66. Prediction ability under a CV scenario using the DAP ranged between 0.02 and 0.27 (mean of 0.16) for single and multiple trait models. Using DAP to predict F5 and DH breeding lines (IP1) resulted to an average prediction ability of 0.06 and 0.07, respectively. Predictions within the breeding lines under independent validations (IP2) resulted to a mean prediction ability of 0.13, where predicting within the same population resulted in a 36% increase in average prediction ability compared with using the F5 to predict the DH lines (and vice versa). Prediction ability using covariate models was significantly higher ( $P < 0.05$ ) than the predictions for multivariate models under single and multiple traits and across all prediction scenarios (Figure 3). Similarly, significant ( $P < 0.05$ ) differences were observed for the mean prediction ability using multiple

traits compared with using only a single trait in the model (0.15 vs 0.09) across all GS scenarios. Predicting within similar environments resulted in a 13% advantage compared to across environment predictions. Overall, a positive correlation ( $r = 0.35$ ) between the proportion of additive variance to the total genetic variance and mean prediction ability across datasets was observed (Supplementary Table S4).



**Figure 3.** Mean prediction ability across all genomic selection scenarios for grain yield using partial least square covariate (Cov) and multivariate (MV) regression models in the presence of single (ST) and multiple (MT) traits collected from high-throughput phenotyping. Prediction scenarios included cross-validations and independent predictions. Means followed by the same letter are not significantly different, LSD < 0.05.

## DISCUSSION

The current study reports the prediction ability of single and multiple trait GS models for grain yield incorporating spectral reflectance measurements derived from high-throughput field phenotyping. Different prediction scenarios for empirical datasets in a winter wheat breeding program were evaluated for the prediction ability of grain yield. Prediction ability of covariate and multivariate PLS regression models for yield were compared across different datasets and prediction scenarios in winter wheat grown in the PNW region of the US.

### Prediction Ability for Single and Multiple Trait Genomic Selection Models

Previous studies have shown the relevance of using secondary traits to predict a focal trait, which is often difficult to phenotype or measure (e.g., grain yield). In US Holsteins, GS accuracy improved when switching from single- to multi-trait models across different traits [60]. Similarly, in African cassava, multi-trait and multi-environment mixed models were recommended for selection, as using these models resulted in a 40% improvement in prediction ability of target agronomic traits [61]. The advantage of multi-trait models over the single trait relies mainly on the use of genetically correlated traits for predictions [62]. Thus, genetic correlation between the primary and the predictor traits is crucial to achieve optimal GS accuracies in the breeding program. In the present

work, the genetic correlations between the measured SRI across locations and grain yield were high, indicating their potential to be used as secondary traits to achieve increased prediction ability for yield in US PNW winter wheat.

Multi-trait prediction models showed a 40% advantage over the single-trait model across GS scenarios in the current study. Increasing the number of traits in the model was relevant in increasing prediction ability, where using all three traits resulted to the highest mean and was significantly different ( $P < 0.05$ ) with the predictions using a single-trait model. In some scenarios, however, having three indices in the model did not necessarily result to higher predictions compared to the two-trait models. It was previously observed that including more traits did not show advantage in predicting yield and protein content in rye, as additional traits could introduce issues in co-linearity [63]. Furthermore, it was noted that if the aim of a breeding program is to improve the prediction ability, particularly for a scarcely phenotyped trait, using a two-trait instead of a three-trait model could offer a greater advantage [63]. In this context, it would thus be beneficial to identify only a few of the highly heritable, strongly correlated predictor traits and use these routinely for indirect selection and prediction of the target or primary trait. In our case, using NDVI and SR in combination under a covariate model as fixed effect predictor traits showed the highest average prediction ability across all scenarios (0.15), and therefore could be used to improve predictions for grain yield in US PNW winter wheat.

Other studies observed that using multiple traits in the prediction model did not result in improved predictions [64,65] demonstrating that in some cases, multi-trait models are not advantageous compared to single trait approaches. Overall, although we observed generally low values for the prediction ability of grain yield in the current study, integrating GS strategies in the breeding program should still be considered. In the long run, the gains achieved using these approaches through increased selection intensity and faster breeding cycles should give additional advantage over the traditional marker or phenotypic selection [15,16]. Furthermore, prediction ability could be improved by optimizing different factors that affect GS accuracy such as the genetic architecture of the trait, heritability, number of markers, genetic and phenotypic correlations among the traits, and the percentage of missing data, either alone or in combination [66–68]. Ultimately, the success of genomic predictions in breeding programs does not all depend on the calculated prediction ability [69] but on how breeders will use this information in performing guided decisions on which lines to advance or used as parents. There were no significant differences between predictions when single and multiple traits were used in CV and IP1 scenarios. Nevertheless, when predictions were conducted within the breeding lines (IP2), where genetic relatedness is more apparent than in IP1, using multiple traits resulted in a significant ( $P < 0.05$ ) increase in prediction ability for grain yield, where a 67%

advantage over using only a single trait in the GS model was observed. When there is less relatedness between the training and test populations, the use of multi-trait models should be preferred over single trait approaches [70].

### **Prediction Ability across Different GS Scenarios**

Cross-validations (CV) represent a common prediction scenario where a single population is divided into a training and a testing set, whereas in independent predictions, a model is trained in one population and is later used to predict another comprising of untested genotypes in untested environments [71]. Overall, using CV resulted to the highest average predictions (0.16) followed by IP2 (0.13) and IP1 (0.01), similar with previous results comparing accuracy for CV and independent validations in DH wheat [72]. The increased prediction ability in the CV scenario could be the result of using the same environmental conditions in which both the training and validation populations are evaluated [73]. Nonetheless, independent validations present a more realistic approach to plant breeding, as this GS strategy aims to predict the performance of lines that are yet to be evaluated in different environments or trials. In CV, we observed significant differences in prediction ability between using combined (BLUP) and individual (BLUE) datasets for predictions (0.19 vs 0.12), showing the advantage of combining environments across different years to capture variability for grain yield. Moreover, predicting within the same environment (e.g., LND predicting LND) for the IP1 scenario was significantly higher compared to across environment predictions ( $P < 0.0001$ ), demonstrating the relevance of predicting within the same environment as effects of QTLs could vary across different environments and lead to poor prediction accuracies [74]. The variability of prediction ability values observed across different scenarios and datasets could be primarily due to the differences across the environments used for predictions, as indicated by low to moderate genetic correlations among the environments (Supplementary Tables S1–S3). The proportion of additive variance to the total genetic variance across different datasets was also related to improved prediction ability. In general, it was observed that the higher the proportion of additive variance, the higher the mean prediction ability across datasets. A positive correlation ( $r = 0.35$ ) between prediction ability and additive variance to total genetic variance ratios was observed across the datasets used for analyses.

### **Prediction Ability for Covariate and Multivariate Models**

Covariate models incorporating reflectance indices as fixed effect predictor traits resulted in higher accuracies compared to multivariate across all prediction scenarios (0.10 vs 0.02). Using covariate multi-trait models further resulted in a 50% gain in prediction ability compared to using covariate single-trait model, whereas no significant differences were observed within multivariate models regardless of the number of



secondary traits. In contrast with our results, the superiority of multivariate models for predicting yield has been previously demonstrated in wheat [32,75]. Our results thus rendered some interesting observations regarding the use of these types of models for predicting grain yield in the presence of secondary traits across different CV and IP scenarios for PNW winter wheat.

First, multivariate models might not be advantageous for predicting lines with little or no genetic relatedness even when there is a high genetic correlation between the predictor and target traits, as in our case. Similar with the IP1 scenario, when a breeding panel was previously used to predict wheat DH lines, reduced predictions were observed in wheat, which was attributed to the presence of opposite linkage phases between SNPs and QTL for yield for genetically distant populations [76]. We observed the relevance of relatedness between the training and validation panels to achieve maximum prediction ability, particularly under the IP2 scenario, consistent with previous studies which highlighted the importance of genetic relationships between the populations used for predictions [77–80]. Among the breeding lines, an increased genetic (Rogers) distance between the training and test panels was related to a decrease in accuracy. Altogether, highest mean prediction ability was observed for the F5 lines (0.16), followed by the DH (0.14), and the across population predictions (0.11), with Rogers distance of 0.05, 0.13, and 0.40, respectively. PCA revealed less genetic differentiation among the F5 lines, whereas the DH lines formed two distinct clusters, based on location. Predicting within the populations (*i.e.*, F5 predicting F5; DH predicting DH lines) showed a 36% advantage compared to across population predictions, indicating the need to use genetically related populations to achieve optimal prediction ability.

Second, incorporating multiple traits is essential to achieve optimal predictions when the target trait has low heritability. A 43% and 64% increase in prediction ability was observed for multi-trait models for the low heritability datasets LND17\_F5 ( $H^2 = 0.15$ ) and PUL17\_F5 ( $H^2 = 0.13$ ). Likewise, Guo *et al.* [81] previously observed that multi-trait approaches performed better than single trait models when evaluating traits with low heritability and lots of missing data using simulations. Furthermore, multi-trait analysis enabled more accurate predictions of breeding values for low heritable traits (heritability of 0.10) that are correlated with highly heritable traits (heritability of 0.80) by utilizing correlation structure between traits under a Bayesian regression model [82]. In barley, PLS did not perform well on traits with medium or high heritability and a smaller training set [83]. Overall, although the observed heritability values for grain yield in the current study were within the range of published values in wheat [4,84,85], we still observed a weak correlation ( $r = 0.06$ ) between heritability and mean prediction ability across datasets for the cross-validations. There was also no significant difference in average accuracies in using the DH datasets (where grain yield had higher heritability values)



for predictions in the IP2 scenario. Taken together, these observations suggest that the heritability of the target trait was secondary only to genetic relatedness and the type of prediction model affecting GS accuracy for yield in PNW winter wheat. It was interesting to note, nevertheless, that the heritability of the secondary spectral traits was related to improved predictions in the covariate PLS models, where using SR ( $H^2$  between 0.52 and 0.95) for GS significantly improved prediction ability for yield (Figure 2C) under the IP2 scenario. This suggests that the prediction ability of a lowly heritable target trait could be improved by using one or multiple secondary predictor traits with higher heritability when using genetically related populations under covariate models. Lastly, size of the training population was relevant to attain improved accuracies, particularly in the IP2 scenario where predictions were done within the breeding lines (with differing size of the TP), consistent with previous studies demonstrating the importance of TP size in achieving optimal accuracy [67,80,86,87]. Using larger population sizes as TP (PUL17 and PUL18) resulted to an average of 14% increase in the prediction ability for yield in the IP2 scenario.

The multivariate approach implemented in the current study resembles a breeding program where plots are yet to be harvested, similar with earlier methods used in wheat [32,75]. This scenario then allows breeders to select for yield based on the predicted values from GS models that incorporate secondary spectral reflectance traits. Previously, it has been shown that these secondary spectral traits have a huge potential for the indirect selection of yield in US Pacific Northwest winter wheat, since they have high genetic and phenotypic correlations [24,40]. Additionally, it has been shown that the inclusion of these traits in prediction models could improve the accuracy for yield [30,32,75]. Breeders could therefore use these HTP traits, which could be collected in either single plants or in plots, as proxy measurements for yield potential, as a basis of selecting higher yielding lines throughout the growing season as they could be predictive of the final yield observed in the field.

## CONCLUSIONS

The prediction ability of single and multiple trait prediction models for yield in PNW winter wheat was evaluated. Our results showed the feasibility of using least square regression models incorporating secondary traits to predict yield in soft winter wheat. Covariate models were superior to multivariate in predicting grain yield in the presence of secondary traits from high-throughput field phenotyping. Multi-trait models also showed an advantage over single trait prediction models for grain yield. The influence of relatedness and size of the training population on genomic selection accuracy was observed, whereas there was no apparent relationship between heritability and prediction ability. When using related populations, using of highly heritable HTP traits such as simple ratio, either alone or in combination with other spectral traits

significantly improved prediction ability for yield under the covariate models. Predicting within the same environments resulted to higher accuracies. Overall, the results presented herein demonstrated the power of combining different genomic and phenotypic approaches to accelerate plant breeding and the potential to improve genetic gains for complex traits in wheat breeding programs. Other prediction models such as multivariate Bayesian, genomic BLUP, and machine learning approaches will also be explored for these GS scenarios for grain yield in the WSU winter wheat breeding program.

### **SUPPLEMENTARY MATERIALS**

The following supplementary materials are available online at <https://doi.org/10.20900/cbgg20190012>:

- Supplementary Table S1: Genetic correlation for grain yield between different datasets for a winter wheat diversity association mapping panel,
- Supplementary Table S2: Genetic correlation for grain yield between different datasets for a winter wheat diversity association mapping panel and F5 and DH wheat breeding lines,
- Supplementary Table S3: Genetic correlation for grain yield between different datasets for the F5 and DH winter wheat breeding lines,
- Supplementary Table S4: Additive and total genetic variances observed for grain yield across different US Pacific Northwest soft winter wheat populations,
- Supplementary Table S5: Prediction ability of partial least square (PLS) regression models for grain yield in US Pacific Northwest winter wheat across different prediction scenarios.

### **DATA AVAILABILITY STATEMENT**

The dataset of this study is available from the authors upon reasonable request.

### **AUTHOR CONTRIBUTIONS**

DNL performed genomic prediction analyses and wrote a draft of the manuscript; AHC edited the manuscript and obtained funding for the project.

### **CONFLICTS OF INTEREST**

The authors declare no conflict of interest.

### **FUNDING SOURCES**

The study was funded the National Institute of Food and Agriculture (NIFA) of the U.S. Department of Agriculture (Award number 2016-68004-24770) and Hatch project 1014919.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Jared Crain (Kansas State University, Manhattan, KS, USA) for assistance in the genomic selection analyses; and Dr. Jayfred Godoy, Gary Shelton, Kyall Hagemeyer, and Jason Wigen for collection of phenotypic data.

## REFERENCES

1. 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 Apr;19(1):168. doi: 10.1186/s12870-019-1781-3
2. Emebiri L, Singh S, Tan M-K, Singh PK, Fuentes-Dávila G, Ogonnaya F. Unravelling the Complex Genetics of Karnal Bunt (*Tilletia indica*) Resistance in Common Wheat (*Triticum aestivum*) by Genetic Linkage and Genome-Wide Association Analyses. *G3 (Bethesda).* 2019;9(5):1437-47. doi: 10.1534/g3.119.400103
3. Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor Appl Genet.* 2015 Feb;128(2):353-63. doi: 10.1007/s00122-014-2435-3
4. Lozada DN, Mason RE, Babar MA, Carver BF, Guedira GB, Merrill K, et al. Association mapping reveals loci associated with multiple traits that affect grain yield and adaptation in soft winter wheat. *Euphytica.* 2017 Sep;213(9):222. doi: 10.1007/s10681-017-2005-2
5. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods.* 2013 Jul 22;9:29. doi: 10.1186/1746-4811-9-29
6. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 2019;20:467-84. doi: 10.1038/s41576-019-0127-1
7. Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A. Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol Breed.* 2011 Jan;27(1):37-58. doi: 10.1007/s11032-010-9411-7
8. He S, Schulthess AW, Mirdita V, Zhao Y, Korzun V, Bothe R, et al. Genomic selection in a commercial winter wheat population. *Theor Appl Genet.* 2016 Mar;129(3):641-51. doi: 10.1007/s00122-015-2655-1
9. Mirdita V, He S, Zhao Y, Korzun V, Bothe R, Ebmeyer E, et al. Potential and limits of whole genome prediction of resistance to Fusarium head blight and Septoria tritici blotch in a vast Central European elite winter wheat population. *Theor Appl Genet.* 2015 Dec;128(12):2471-81. doi: 10.1007/s00122-015-2602-1
10. 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 Genome.* 2012;5:51-61. doi: 10.3835/plantgenome2012.02.0001

11. Poland J, 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(3):103-13. doi: 10.3835/plantgenome2012.06.0006
12. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, et al. Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci*. 2017 Nov;22(1):961-75. doi: 10.1016/j.tplants.2017.08.011
13. Desta ZA, Ortiz R. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci*. 2014;19(9):592-601. doi: 10.1016/j.tplants.2014.05.006
14. Charmet G, Storlie E, Oury FX, Laurent V, Beghin D, Chevarin L, et al. Genome-wide prediction of three important traits in bread wheat. *Mol Breed*. 2014 Dec;34(4):1843-52. doi: 10.1007/s11032-014-0143-y
15. Morota G, Jarquin D, Campbell MT, Iwata H. Statistical methods for the quantitative genetic analysis of high-throughput phenotyping data. arXiv:1904.12341v1 [Preprint]. 2019 Apr. Available from: <https://arxiv.org/abs/1904.12341>. Accessed 2019 May 9.
16. Würschum T. Modern Field Phenotyping Opens New Avenues for Selection. In: Miedaner T, Korzun V, editors. *Applications of Genetic and Genomic Research in Cereals*. Cambridge (UK): Woodhead Publishing; 2018. p. 164-71.
17. Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS, Cairns JE. Translating High-Throughput Phenotyping into Genetic Gain. *Trends Plant Sci*. 2018;23(5):451-66. doi: 10.1016/j.tplants.2018.02.001
18. Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Luis Araus J. High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge. *J Integr Plant Biol*. 2012;54(5):312-20. doi: 10.1111/j.1744-7909.2012.01116.x
19. van Eeuwijk FA, Bustos-Korts D, Millet EJ, Boer MP, Kruijer W, Thompson A, et al. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci*. 2019;282:23-39. doi: 10.1016/j.plantsci.2018.06.018
20. Crain JL, Wei Y, Barker J, Thompson SM, Alderman PD, Reynolds M, et al. Development and deployment of a portable field phenotyping platform. *Crop Sci*. 2016;56(3):965-75. doi: 10.2135/cropsci2015.05.0290
21. Walter J, Edwards J, Cai J, McDonald G, Miklavcic SJ, Kuchel H. High-Throughput Field Imaging and Basic Image Analysis in a Wheat Breeding Programme. *Front Plant Sci*. 2019;10:449. doi: 10.3389/fpls.2019.00449
22. Pradhan S, Sehgal VK, Bandyopadhyay KK, Sahoo RN, Panigrahi P, Parihar CM, et al. Comparison of Vegetation Indices from Two Ground Based Sensors. *J Indian Soc Remote Sens*. 2018 Feb;46(2):321-6. doi: 10.1007/s12524-017-0671-0
23. Barker J, Zhang N, Sharon J, Steeves R, Wang X, Wei Y, et al. Development of a field-based high-throughput mobile phenotyping platform. *Comput Electron Agric*. 2016;122:74-85. doi: 10.1016/j.compag.2016.01.017

24. Gizaw SA, Garland-Campbell K, Carter AH. Use of spectral reflectance for indirect selection of yield potential and stability in Pacific Northwest winter wheat. *Field Crops Res.* 2016;196:199-206. doi: 10.1016/j.fcr.2016.06.022
25. Andrade-Sanchez P, Gore MA, Heun JT, Thorp KR, Carmo-Silva AE, French AN, et al. Development and evaluation of a field-based high-throughput phenotyping platform. *Funct Plant Biol.* 2014;41(1):68-79. doi: 10.1071/FP13126
26. Jo SH, Ko JH. Determining canopy growth conditions of paddy rice via ground-based remote sensing. *Korea J Remote Sens.* 2015;31:11-20. doi: 10.7780/kjrs.2015.31.1.2
27. Prasad B, Carver BF, Stone ML, Babar MA, Raun WR, Klatt AR. Genetic Analysis of Indirect Selection for Winter Wheat Grain Yield Using Spectral Reflectance Indices. *Crop Sci.* 2007;47:1416-25. doi: 10.2135/cropsci2006.08.0546
28. Gutierrez M, Reynolds MP, Raun WR, Stone ML, Klatt AR. Spectral Water Indices for Assessing Yield in Elite Bread Wheat Genotypes under Well-Irrigated, Water-Stressed, and High-Temperature Conditions. *Crop Sci.* 2010;50:197-214. doi: 10.2135/cropsci2009.07.0381
29. Babar MA, Reynolds MP, van Ginkel M, Klatt AR, Raun WR, Stone ML. Spectral Reflectance to Estimate Genetic Variation for In-Season Biomass, Leaf Chlorophyll, and Canopy Temperature in Wheat. *Crop Sci.* 2006;46:1046-57. doi: 10.2135/cropsci2005.0211
30. Sun J, Rutkoski JE, Poland JA, Crossa J, Jannink J-L, Sorrells ME. Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *Plant Genome.* 2017 Jul;10:2. doi: 10.3835/plantgenome2016.11.0111
31. Juliana P, Montesinos-López OA, Crossa J, Mondal S, González Pérez L, Poland J, et al. Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theor Appl Genet.* 2019 Jan;132(1):177-94. doi: 10.1007/s00122-018-3206-3
32. 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 Mar;11:1. doi: 10.3835/plantgenome2017.05.0043
33. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst.* 2001;58(2):109-30. doi: 10.1016/S0169-7439(01)00155-1
34. Monteverde E, Gutierrez L, Blanco P, de Vida FP, Rosas JE, Bonnacarrere V, et al. Integrating Molecular Markers and Environmental Covariates To Interpret Genotype by Environment Interaction in Rice (*Oryza sativa* L.) Grown in Subtropical Areas. *G3 (Bethesda).* 2019 May 7;9(5):1519-31. doi: 10.1534/g3.119.400064
35. Colombani C, Croiseau P, Fritz S, Guillaume F, Legarra A, Ducrocq V, et al. A comparison of partial least squares (PLS) and sparse PLS regressions in genomic selection in French dairy cattle. *J Dairy Sci.* 2012 Apr;95(4):2120-31. doi: 10.3168/jds.2011-4647

36. Boulesteix A-L, Strimmer K. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform.* 2006 May 26;8(1):32-44. doi: 10.1093/bib/bbl016
37. Duchemin SI, Colombani C, Legarra A, Baloché G, Larroque H, Astruc J-M, et al. Genomic selection in the French Lacaune dairy sheep breed. *J Dairy Sci.* 2012;95(5):2723-33. doi: 10.3168/jds.2011-4980
38. Garriga M, Romero-Bravo S, Estrada F, Escobar A, Matus IA, del Pozo A, et al. Assessing Wheat Traits by Spectral Reflectance: Do We Really Need to Focus on Predicted Trait-Values or Directly Identify the Elite Genotypes Group? *Front Plant Sci.* 2017 Mar 9;8:280. doi: 10.3389/fpls.2017.00280
39. Yu K, Anderegg J, Mikaberidze A, Karisto P, Mascher F, McDonald BA, et al. Hyperspectral canopy sensing of wheat *Septoria tritici* blotch disease. *Front Plant Sci.* 2018;9:1195. doi: 10.3389/fpls.2018.01195
40. Lozada DN, Godoy JGV, Carter AH. Genomic prediction and indirect selection for grain yield using spectral reflectance indices from high-throughput phenotyping. *Euphytica.* 2019 unpublished.
41. Federer WT, Raghavarao D. On augmented designs. *Biometrics.* 1975;29-35. doi: 10.2307/2529707
42. Peterson CJ, Allan RE, Rubenthaler GL, Line RF. Registration of 'Eltan' Wheat. *Crop Sci.* 1991;31:1704. doi: 10.2135/cropsci1991.0011183X003100060075x
43. Allan RE, Peterson CJ, Rubenthaler GL, Line RF, Roberts DE. Registration of 'Madsen' wheat. *Crop Sci.* 1989;29(6):1575-6. doi: 10.2135/cropsci1989.0011183X002900060068x
44. Jones SS, Murray TD, Lyon SR, Morris CF, Line RF. Registration of 'Bruehl' wheat. *Crop Sci.* 2001;41(6):2006-8. doi: 10.2135/cropsci2001.2006
45. Carter AH, Jones SS, Lyon SR, Balow KA, Shelton GB, Higginbotham RW, et al. Registration of 'Otto' wheat. *J Plant Regist.* 2013;7(2):195-200. doi: 10.3198/jpr2012.07.0013crc
46. 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(3):263-8. doi: 10.3198/jpr2016.09.0051crc
47. Jones SS, Lyon SR, Balow KA, Gollnick MA, Murphy KM, Kuehner JS, et al. Registration of 'Xerpha' wheat. *J Plant Regist.* 2010;4(2):137-40. doi: 10.3198/jpr2009.06.0306crc
48. Zemetra RS, Souza EJ, Lauver M, Windes J, Guy SO, Brown B, et al. Registration of 'Brundage' wheat. *Crop Sci.* 1998;38(5):1404. doi: 10.2135/cropsci1998.0011183X003800050056x
49. 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(3):273-8. doi: 10.3198/jpr2013.12.0074crc
50. Rouse JW Jr, Haas RH, Schell JA, Deering DW. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Texas (US): Texas A&M University Remote Sensing Center; 1972.
51. Babar MA, Reynolds MP, Van Ginkel M, Klatt AR, Raun WR, Stone ML. Spectral reflectance indices as a potential indirect selection criteria for wheat yield under irrigation. *Crop Sci.* 2006;46(2):578-88. doi: 10.2135/cropsci2005.0059



52. Stenberg P, Rautiainen M, Manninen T, Voipio P, Smolander H. Reduced simple ratio better than NDVI for estimating LAI in Finnish pine and spruce stands. *Silva Fennica*. 2004;38(1):3-14. doi: 10.14214/sf.431
53. Rodríguez F, Alvarado G, Pacheco Á, Burgueño J. ACBD-R. Augmented Complete Block Design with R for Windows. Version 4.0. Mexico (Mexico): CIMMYT Research Data & Software Repository Network; 2018. Available from: <http://hdl.handle.net/11529/10855>. Accessed 2019 Jan 23.
54. Falconer DS. Introduction to Quantitative Genetics. 3rd ed. New York (US): Longman Scientific and Technical; 1989.
55. SAS Institute. SAS System Options: Reference. 2nd ed. Cary (US): SAS Institute; 2011.
56. 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 Oct 1;23(19):2633-5. doi: 10.1093/bioinformatics/btm308
57. Roger JS. Measure of genetic similarity and genetic distance. *Studies in Genetics VII*. Univ Texas Publ. 1972;7213:145-53.
58. Wehrens R, Mevik B-H. The pls package: principal component and partial least squares regression in R. *J Statiscal Softw*. 2007;18(2). doi: 10.18637/jss.v018.i02
59. Covarrubias-Pazaran G. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. *PLoS One*. 2016;11:e0156744
60. Tsuruta S, Misztal I, Aguilar I, Lawlor T. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *J Dairy Sci*. 2011;94:4198-204. doi: 10.3168/jds.2011-4256
61. Okeke UG, Akdemir D, Rabbi I, Kulakow P, Jannink J-L. Accuracies of univariate and multivariate genomic prediction models in African cassava. *Genet Sel Evol*. 2017 Dec 4;49(1):88. doi: 10.1186/s12711-017-0361-y
62. Jia Y, Jannink J-L. Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics*. 2012;192(4):1513-22. doi: 10.1534/genetics.112.144246
63. Schulthess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y. Multiple-trait-and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor Appl Genet*. 2016 Feb;129(2):273-87. doi: 10.1007/s00122-015-2626-6
64. Fernandes SB, Dias KOG, Ferreira DF, Brown PJ. Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theor Appl Genet*. 2018 Mar;131(3):747-55. doi: 10.1007/s00122-017-3033-y
65. Bao Y, Kurle JE, Anderson G, Young ND. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. *Mol Breed*. 2015;35(6):128. doi: 10.1007/s11032-015-0324-3
66. 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 Jan 12;12(1):e0169606. doi: 10.1371/journal.pone.0169606
67. Werner CR, Voss-Fels KP, Miller CN, Qian W, Hua W, Guan C-Y, et al. Effective Genomic Selection in a Narrow-Genepool Crop with Low-Density Markers: Asian Rapeseed as an Example. Plant Genome. 2018;11. doi: 10.3835/plantgenome2017.09.0084
  68. 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 May;182(1):355-64. doi: 10.1534/genetics.108.098277
  69. 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;8:2735-47.
  70. Schulthess AW, Zhao Y, Longin CFH, Reif JC. Advantages and limitations of multiple-trait genomic prediction for Fusarium head blight severity in hybrid wheat (*Triticum aestivum* L.). Theor Appl Genet. 2018 Mar;131(3):685-701. doi: 10.1007/s00122-017-3029-7
  71. Jiang Y, Schulthess AW, Rodemann B, Ling J, Plieske J, Kollers S, 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 Mar;130(3):471-82. doi: 10.1007/s00122-016-2827-7
  72. Thavamanikumar S, Dolferus R, Thumma BR. Comparison of Genomic Selection Models to Predict Flowering Time and Spike Grain Number in Two Hexaploid Wheat Doubled Haploid Populations. G3 (Bethesda). 2015 Oct 1;5(10):1991. doi: 10.1534/g3.115.019745
  73. Đorđević V, Čeran M, Miladinović J, Balešević-Tubić S, Petrović K, Miladinov Z, et al. Exploring the performance of genomic prediction models for soybean yield using different validation approaches. Mol Breed. 2019 May;39(5):74. doi: 10.1007/s11032-019-0983-6
  74. 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;58:2274-88. doi: 10.2135/cropsci2017.10.0638
  75. 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 Jul 6;6(9):2799-808. doi: 10.1534/g3.116.032888
  76. Haile JK, N'Diaye A, Clarke F, Clarke J, Knox R, Rutkoski J, et al. Genomic selection for grain yield and quality traits in durum wheat. Mol Breed. 2018 May;38(6):75. doi: 10.1007/s11032-018-0818-x
  77. Lorenz AJ, Smith KP. Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci. 2015;55(6):2657-67. doi: 10.2135/cropsci2014.12.0827
  78. Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L. Accuracy and training population design for genomic selection on quantitative traits in elite North

- American oats. *Plant Genome*. 2011;4(2):132-44. doi: 10.3835/plantgenome.2011.02.0007
79. Edwards SM, Buntjer JB, Jackson R, Bentley AR, Lage J, Byrne E, et al. The effects of training population design on genomic prediction accuracy in wheat. *Theor Appl Genet*. 2019;132:1943-52. doi: 10.1007/s00122-019-03327-y
80. Norman A, Taylor J, Edwards J, Kuchel H. Optimising Genomic Selection in Wheat: Effect of Marker Density, Population Size and Population Structure on Prediction ability. *G3 (Bethesda)*. 2018;8:2889-99. doi: 10.1534/g3.118.200311
81. Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet*. 2014;15(1):30. doi: 10.1186/1471-2156-15-30
82. Hayashi T, Iwata H. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics*. 2013;14:34. doi: 10.1186/1471-2105-14-34
83. Iwata H, Jannink J-L. Accuracy of Genomic Selection Prediction in Barley Breeding Programs: A Simulation Study Based On the Real Single Nucleotide Polymorphism Data of Barley Breeding Lines. *Crop Sci*. 2011;51:1915-27 doi: 10.2135/cropsci2010.12.0732
84. Michel S, Ametz C, Gungor H, Epure D, Grausgruber H, Löschenberger F, et al. Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theor Appl Genet*. 2016 Jun;129(6):1179-89. doi: 10.1007/s00122-016-2694-2
85. 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;132:1705-20. doi: 10.1007/s00122-019-03309-0
86. Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y. Genomic selection in sugar beet breeding populations. *BMC Genet*. 2013 Sep;14(1):85. doi: 10.1186/1471-2156-14-85
87. Zhang A, Wang H, Beyene Y, Semagn K, Liu Y, Cao S, et al. Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. *Front Plant Sci*. 2017;8:1916. doi: 10.3389/fpls.2017.01916

How to cite this article:

Lozada DN, Carter AH. Accuracy of Single and Multi-Trait Genomic Prediction Models for Grain Yield in US Pacific Northwest Winter Wheat. *Crop Breed Genet Genom*. 2019;1:e190012. <https://doi.org/10.20900/cbgg20190012>