



# Evaluation of the Potential for Genomic Selection to Improve Spring Wheat Resistance to Fusarium Head Blight in the Pacific Northwest

Haixiao Dong<sup>1,2</sup>, Rui Wang<sup>3</sup>, Yaping Yuan<sup>1</sup>, James Anderson<sup>4</sup>, Michael Pumphrey<sup>2</sup>, Zhiwu Zhang<sup>2\*</sup> and Jianli Chen<sup>3\*</sup>

<sup>1</sup> College of Plant Sciences, Jilin University, Changchun, China, <sup>2</sup> Department of Crop and Soil Sciences, Washington State University, Pullman, WA, United States, <sup>3</sup> Department of Plant Sciences, University of Idaho, Aberdeen, ID, United States, <sup>4</sup> Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, United States

## OPEN ACCESS

### Edited by:

Rodomirol Ortiz,  
Swedish University of Agricultural  
Sciences, Sweden

### Reviewed by:

Fernando Martinez,  
Universidad de Sevilla, Spain  
Agostino Fricano,  
Consiglio per la Ricerca in Agricoltura  
e l'Analisi dell'Economia  
Agraria (CREA), Italy

### \*Correspondence:

Zhiwu Zhang  
zhiwu.zhang@wsu.edu  
Jianli Chen  
jchen@uidaho.edu

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 02 April 2018

**Accepted:** 08 June 2018

**Published:** 03 July 2018

### Citation:

Dong H, Wang R, Yuan Y,  
Anderson J, Pumphrey M, Zhang Z  
and Chen J (2018) Evaluation of the  
Potential for Genomic Selection  
to Improve Spring Wheat Resistance  
to Fusarium Head Blight in the Pacific  
Northwest. *Front. Plant Sci.* 9:911.  
doi: 10.3389/fpls.2018.00911

Fusarium Head Blight (FHB) has emerged in spring wheat production in Pacific Northwest during the last decade due to factors including climate changes, crop rotations, and tillage practices. A breeding population with 170 spring wheat lines was established and screened over a 2-year period in multiple locations for FHB incidence (INC), severity (SEV), and deposition of the mycotoxin, deoxynivalenol (DON). A genome-wide association study suggested that the detectable number of genetic loci and effects are limited for marker-assisted selection. In conjunction with the success of breeding on FHB resistance in other programs, genomic selection (GS) was suggested as a better option. To evaluate the prediction accuracy of GS in the current breeding population, we conducted a variety of validations by varying proportions of testing populations and cohorts based on both FHB resistance and market class, including soft white spring (SWS), hard white spring (HWS), and hard red spring (HRS). We found that INC had higher heritability, higher correlation across years and locations, and higher prediction accuracy than SEV and DON. Prediction accuracy varied among the scenarios that restricted the testing population to a certain cohort. For a small set of newly developed or introduced lines (<17), prediction accuracy will be about 60% if the lines have similar genetic relationships as those among the current 170-line training population. However, we expect a lower prediction accuracy if new lines are selected for a specific characteristic, such as FHB resistance or market class. With the exception of DON in the SWS lines, the current training population is capable of making reasonably accurate predictions for FHB-resistant lines in most of the major market classes. For SWS, adding more lines or further phenotyping is required to improve prediction accuracy. These results demonstrate the potential and challenges of GS, especially for developing FHB-resistant varieties in the SWS market class.

**Keywords:** wheat, fusarium head blight, genomic selection, disease resistance, cross validation, molecular breeding, cultivars, Pacific Northwest

## INTRODUCTION

Fusarium head blight (FHB) is a devastating disease that affects corn and small grain crops such as wheat in humid conditions (Schroeder and Christensen, 1963; Wang et al., 1982; Snijders, 1990). FHB causes shriveled kernels, significant yield losses, and deposition of the mycotoxin, deoxynivalenol in the infected seeds, which renders the grain unsuitable for human consumption and animal feed (McMullen et al., 1997). Historically, during the last three decades, FHB has been an endemic disease in the north-central and eastern regions of the United States. However, in recent years, the disease has emerged in the Pacific Northwest (PNW), coinciding with increased corn production, reduced tillage farming practices, and changing climate in the region (Marshall et al., 2012). Due to higher temperatures and greater humidity during flowering time in the PNW, spring wheat is more susceptible to FHB than winter wheat. Most of the wheat cultivars currently grown in the PNW are susceptible to FHB and, when infected, often produce high levels of deoxynivalenol (Marshall, 2014). Therefore, development of FHB-resistant wheat cultivars for the region is critically needed to reduce the impending negative consequences, including reduced crop yield and quality and increased costs for fungicides.

The traditional genetic marker-based breeding method is called marker-assisted selection, which utilizes known, previously identified quantitative trait loci (QTL)-associated markers. In the last two decades, more than 200 FHB-resistant QTL have been identified across the entire wheat genome (Buerstmayr et al., 2009; Liu et al., 2009). However, none of these QTL confer promising resistance to FHB. For example, although one of the major QTL, *Fhb1* identified on chromosome 3BS (Anderson et al., 2001; Liu and Anderson, 2003) could reduce disease occurrence by 20–25% (Pumphrey et al., 2007), this QTL was originally identified from the Chinese line, Sumai 3, and the favorable allele has a very low frequency in North American germplasm (Sneller et al., 2010; Bernardo et al., 2012). Introducing this allele may not only take time, but also risk the diversity among North American germplasm. Additional QTL were also confirmed on chromosomes 2D (Shen et al., 2003a; Yang et al., 2005; Cuthbert et al., 2007), 3A (Shen et al., 2003a; Steiner et al., 2004; Yang et al., 2005), 5AS (Buerstmayr et al., 2003; McKendry et al., 2004; Chen et al., 2006), and 6B (Shen et al., 2003b; Yang et al., 2005; Cuthbert et al., 2007). Similar to *Fhb1*, most of these QTL were originally derived from Chinese wheat germplasm.

Efforts to identify FHB resistance in native North American winter wheat germplasm have been undertaken (Sneller et al., 2010; Arruda et al., 2016). Both Sneller et al. (2010) and Arruda et al. (2016) showed that FHB resistance is conferred by many small effect QTL in soft red winter wheat lines grown in the eastern United States. A research has been conducted on other native spring wheat germplasm without Sumai 3 backgrounds, which included 170 lines developed by wheat breeders at PNW and the International Maize and Wheat Improvement Center (CIMMYT) (Wang et al., 2017). This study also suggested that FHB resistance in spring wheat is controlled by a relatively large number of QTL with small effects. Thus, there is a critical need to

assemble the total genetic effects of individual lines—accounting not only for QTL with large or moderate effects, but also QTL with small effects, which collectively control a large proportion of the total genetic variance. This method is commonly known as genomic selection (GS).

GS was introduced to plant breeding in 1994 to evaluate yield potential in maize inbred lines as the Best Linear Unbiased Prediction (BLUP), using restriction fragment length polymorphism (RFLP) markers covering the maize genome (Bernardo, 1994). The method is now known as genomic BLUP, or gBLUP. GS has been used to predict genetic merit in animals (Hayes et al., 2009; Guo et al., 2011) and plants (Heffner et al., 2009; Jannink et al., 2010). GS has also been applied in winter wheat programs for FHB resistance in the mid-western and eastern United States and Canada. Rutkoski et al. (2012) compared the accuracy of different GS models for FHB-related traits using 170 winter wheat lines from 18 different breeding programs and more than 2,000 diversity array technology markers and single-sequence repeat markers. Arruda et al. (2015) evaluated different factors affecting the accuracy of genomic prediction using 273 winter wheat breeding lines and 5,054 genotyping-by-sequencing markers. Both studies demonstrated that GS is a very promising breeding strategy for FHB resistance in winter wheat.

Our study aimed to fill the knowledge gaps on FHB in spring wheat by evaluating prediction accuracy of using the existing 170 lines under three conditions: (1) no restriction on testing cohort of market class and segmentation of FHB resistance; (2) restriction of testing cohort to market classes; and (3) restriction of testing cohort to segmentation of FHB resistance. This knowledge was intended to provide guidance for developing a GS pipeline for breeding FHB-resistant wheat in the PNW. Additionally, we expect the new knowledge and the approach used to benefit similar breeding programs in other traits, geographic areas, and breeding programs for other crops.

## MATERIALS AND METHODS

### Plant Materials and Disease Evaluation

The spring wheat panel used in this study and the FHB disease evaluation were described in Wang et al. (2017). In short, a total of 170 spring wheat cultivars and elite lines developed from the breeding programs in the PNW and CIMMYT were used in this study. The 170 lines (PNW panel, hereafter) include 26 lines from Washington State University (WSU), 33 from University of Idaho (UI), 34 from University of California, Davis (UCD), 25 from Montana State University (MSU), 49 from CIMMYT, two from Limagrain Cereal Seeds (LCS), and one from SSK (Agriculture and Agri-Food Canada, Saskatchewan). Over 50% of lines have been used as parental lines in one or more breeding programs. Therefore, this PNW panel is representative of the genetic diversity in currently used germplasm. Based on the end-use products, the panel contains three main market classes of spring wheat: soft white spring (SWS), hard white spring (HWS), and hard red spring (HRS) wheat.

The phenotypes of the 170 lines were collected from three field nurseries [Saint Paul, MN (StP), Crookston, MN (CrK), and Aberdeen, ID (AB)] and one greenhouse (GH) over 2 years (2015 and 2016). Fields were divided into zones according to field orientation. Checks were placed in each zone for validation. The checks included the FHB resistant cultivars Alsen, BacUp, W14, and AC Barrie (Chen et al., 2006; Zhang et al., 2016) and the FHB susceptible cultivars Roblin, Wheaton, and UI Platinum (Chen and Marshall, 2012; Zhang et al., 2016). In MN fields, 190 lines, including the 170 lines that were genotyped and used in this study, were randomly assigned to two zones by using an unreplicated augmented complete block design. In each zone, 95 lines and five checks (Alsen, BacUp, W14, Roblin and Wheaton) were randomly assigned. In Aberdeen fields, the 190 lines were randomly assigned to five zones. In each zone, 38 lines and two checks (AC Barrie and UI Platinum) were randomly assigned with two replications. In GH, two checks (AC Barrie and UI Platinum) were used, and each line had four replications. The inoculum for StP and CrK nurseries was prepared by balanced mixing 30–39 *F. graminearum* isolates, as used in Zhang et al. (2016).

The inoculum used in AB nursery and GH experiment was isolated as single spore isolate from natural infected seeds in Idaho. This isolate is virulent to most wheat cultivars in Idaho. The plots in the StP and AB nurseries were individually sprayed two to three times with the prepared inoculum using a CO<sub>2</sub>-pressure backpack sprayer. The plants in the GH experiment were point inoculated with approximately 5 µl of macroconidia suspension at the concentration of 8–10 × 10<sup>4</sup> ml<sup>-1</sup> using a pipette dropper. For both methods (spray and point inoculation), to facilitate even distribution and adherence of fungal spores on the plant, a surfactant (Tween 20: polyoxyethylene-20-sorbitan monolaurate) was added (100 µl per liter) to the inoculum suspension before use. Corn spawn method was applied to the CrK nursery. The colonized corn grains were spread throughout the nursery twice at the wheat jointing stage and at 1 week later. For all nurseries and GH experiments, the plants were misted immediately after inoculation to promote infection.

Due to variable disease pressure, disease assessment was performed at 21 days after inoculation in the StP and CrK experiment locations, and at 28 days after inoculation in the AB and GH experiment locations. Incidence (INC), severity (SEV), and deoxynivalenol concentration (DON) were assessed as FHB disease reactions. INC was recorded as the percentage of infected spikes in a headrow plot. SEV was visually diagnosed as the percentage of infected spikelets in each spike. DON (ppm) was measured using grain samples from 30 randomly selected heads for each line in the StP and CrK field nurseries and from the composite of all plants in the AB nursery (Supplementary Table S1). The procedure to quantify DON is described in Fuentes et al. (2005) and Mirocha et al. (1998) and briefly summarized in Wang et al. (2017). In short, the sample was prepared and passed through a column and then the dried filtrate was derivatized for gas chromatography-mass spectrometry (GC/MS) analysis. The characteristic ions of DON with fragment ion (m/z value) of 235.10 was detected as target while the fragment ions of 259.10 and 422.10 as reference ions.

## Phenotypic Data Analysis

In this study, we treated each year-location combination as an environment. We evaluated FHB SEV in five environments and assessed FHB INC and DON in four environments. In each environment, the raw phenotypic values within replicates were adjusted by using a spatial analysis. The analyses were performed by using R package “lme4” (Bates et al., 2015) with zone fitted as random effect. For the phenotypic data collected from GH, no spatial adjustment was applied. The arithmetic mean of all replications in each trial was calculated and used as the single environment phenotypic data value. We applied R package “lme4” to estimate the BLUP values for the three phenotypes (INC, SEV, and DON) across different environments (Supplementary Table S1). The lines and environments were considered random effects in the mixed model. Raw phenotypes, means and BLUPs were displayed by the pairs.panel function in R package “psych” (Revelle, 2017) to illustrate the distributions with trends indicated by the correlations ellipse.

## Genotypic Markers

Of the total 170 lines in the PNW panel, 143 belong to the Spring Wheat Association Mapping (SWAM) panel developed by Triticeae Coordinated Agricultural Project (T-CAP). The Illumina 90K genotypic markers for these 143 lines were downloaded from the Triticeae Toolbox (T3) website<sup>1</sup> operated by T-CAP. The remaining 27 lines were also genotyped with the Illumina 90K SNPs assay at the USDA-ARS Cereal Crops Research Unit, Fargo, ND and the allele calls were performed with GenomeStudio v2011.1 (Illumina Inc., Hayward, CA, United States). Finally, a total of 11,523 common SNP markers were selected by combining the two sets of genotyping data. Genetic positions for the selected SNP markers were retrieved from the consensus map for 90K SNP markers developed by Wang et al. (2014). After filtering out SNP markers with a missing rate of more than 10% or with a minor allele frequency (MAF) less than 0.05, 10,101 SNPs were retained for further analysis. The missing genotypes were imputed using the java package “LinkImpute” (Money et al., 2015), based on a *k*-nearest neighbor genotype imputation method, LD-kNNi (Troynskaya et al., 2001), which is designed for unordered markers.

## Principal Component (PC) Analysis

Principal Component analysis was conducted with all genetic markers by using the prcomp function in R (R Core Team, 2017). We used the default (correlation) option to derive eigen values and eigen vectors, which implied that all genetic markers were weighted the same. The markers with large MAFs contributed the same as markers with small MAFs. The first three PCs were used to present the population structure categorized by market classes and origins. The first three PCs were also used as the fixed effects for estimation of heritability and genomic prediction.

## Estimation of Heritability

The observations in each environment, means and BLUPs across environments, were the response variables in a fixed and random

<sup>1</sup><https://triticeaetoolbox.org/wheat/>

effects mixed linear model. The fixed effects were the first three PCs. The random effects were the total additive genetic effects of individuals in addition to the residuals. The variance and covariance matrix of the total additive genetic effects was defined as the product of the additive relationship matrix derived from markers (Zhang et al., 2007) and the genetic variance. The statistical model can be written as follows:

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

where  $y$  is a vector ( $n \times 1$ ) of observations with  $n$  as number of lines;  $\mu$  is the overall mean;  $\beta$  is a vector ( $p \times 1$ ) of fixed effects with  $p$  as the number of fixed effects, specifically, the first three PCs derived from all markers ( $p = 3$ );  $X$  is a design matrix ( $n \times p$ ) for fixed effects;  $u$  is a vector ( $n \times 1$ ) of random effects of the total additive genetic effects of individuals;  $Z$  is a design matrix ( $n \times n$ ) for random effects  $u$ ; and  $\varepsilon$  is the residuals. The random effects follow normal distributions:  $u \sim N(0, K\sigma_u^2)$ ,  $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$ , where  $I$  is the identity matrix and  $K$  is the additive relationship matrix calculated from all genetic markers by using VanRaden algorithm (VanRaden, 2008). The matrix was rescaled to pedigree-like relationship matrix, implemented in GAPIT (Lipka et al., 2012; Tang et al., 2016). The rescaled relationship matrix had the maximum elements of 2 on the diagonals and the minimum element of 0 off diagonal for the two individuals that were the least related;  $\sigma_u^2$  and  $\sigma_\varepsilon^2$ , are the variance of individual additive genetic effects and variance of residuals, respectively. The estimations of the genetic and residual variances were conducted by using the EMMA algorithm (Kang et al., 2008).

## Genomic Prediction

Ridge regression was employed to perform genomic prediction by using mixed.solve function implemented in R package “rrBLUP v4.5” (Endelman, 2011). This statistical model has the same format as the model to estimate heritability (1). However, matrix  $Z$  and vector  $u$  are defined differently.  $Z$  is a design matrix ( $n \times m$ ) for random effects, specifically, the matrix for genotypes;  $u$  is a vector ( $m \times 1$ ) of random effects of markers with  $m$  as number of markers. The random effects of markers follow normal distributions:  $u \sim N(0, I\sigma_u^2)$ , where  $I$  is the identity matrix and  $\sigma_u^2$  is the variance of individual marker genetic effects.

## Evaluation of Prediction Accuracy

Monte Carlo cross-validations (Xu and Liang, 2001) were employed to evaluate accuracy of prediction by using rrBLUP. A proportion, varying from 10 to 90%, of lines were randomly selected as the testing population and the remaining lines as the training population. Both the genotypes and the observed phenotypes in the training population were used to estimate the effects of genetic markers. The estimated marker effects and the genotypes of the testing population were used to calculate the predicted phenotypes, which also included the estimated fixed effects of the first three PCs. The observed phenotypes of the testing population were only used to calculate the Pearson correlation coefficient between the observed and the predicted phenotypes. The stochastic process was replicated 1,000 times

and the prediction accuracy distributions, means, and standard errors were reported.

Selection of testing populations and training populations was also restricted to specific cohorts of the spring wheat lines. The first scenario had no restriction; all 170 lines were treated homogeneously. The second scenario restricted the testing population to either FHB-resistant lines or FHB-susceptible lines. The third scenario restricted the testing population to the lines from a specific market class. We also created some scenarios that restricted both the training and testing populations. For example, when we restricted the testing population to SWS, we also restricted the training population to SWS.

## Data Availability

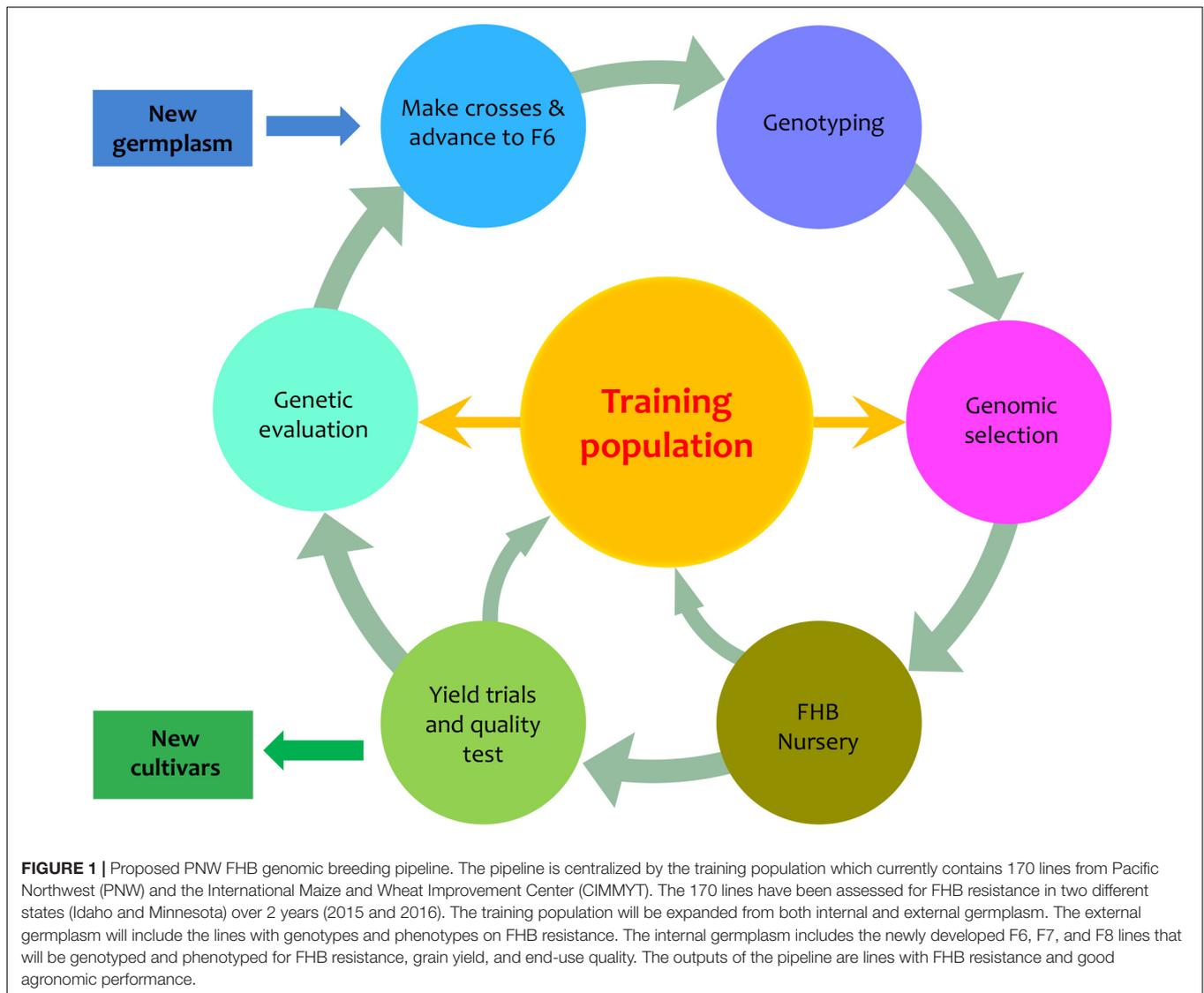
Genotypes, phenotypes and the classification information (market classes, origins, and FHB susceptibility) of 170 wheat lines are available in Supplementary Table S1.

## RESULTS

We designed a PNW FHB genomic breeding pipeline that includes six elements in addition to input and output (Figure 1). Central to the pipeline are the 170 PNW and CIMMYT spring wheat lines, which are used as the primary training population to develop new varieties with FHB resistance. The input is the newly introduced germplasm. The output is the newly developed varieties. The six surrounding elements of the pipeline are advanced breeding lines (F6), genotyping, GS, FHB nursery, field trials, and genetic evaluation. GS is used to assess new genotypes and genetic evaluation is used to assess new phenotypes. Both the FHB nursery and field trials focus on phenotyping, but the FHB nursery evaluates FHB resistance and the field trials evaluate agronomic traits such as grain yield and quality. Our primary interest was to determine how accurately we could predict the newly developed or the newly introduced varieties to our existing training population. We investigated the phenotypic correlation across environments, trait heritability under each environment and heritability of mean across environments, and prediction accuracy through validation.

## Phenotypic Correlation Across Locations and Years

Three FHB-related traits, INC, SEV, and DON, were measured in two to four locations during a 2-year period. The distributions of the three traits under different environments and their BLUPs were demonstrated in our previous study (Wang et al., 2017). In this study, we detailed the phenotypic correlation across environments with scatter plots and Pearson correlations (Figure 2). As shown by the distributions on the diagonals, DON and SEV exhibit close to a normal distribution. However, the INC distribution is close to the Poisson distribution. The side of tails varied across environments for the Poisson distribution. Two of them were on the left and one on the right. Nevertheless, INC had the highest correlation, ranging from 0.24 to 0.5 with mean of 0.33. SEV had the lowest correlations across environments, ranging from 0.07 to 0.52 with mean of 0.25. In general, the



correlations were low for all three traits. This finding suggests that FHB must be evaluated with multiple locations and multiple years to achieve reliable mean values.

### Similarity Between Mean and BLUP

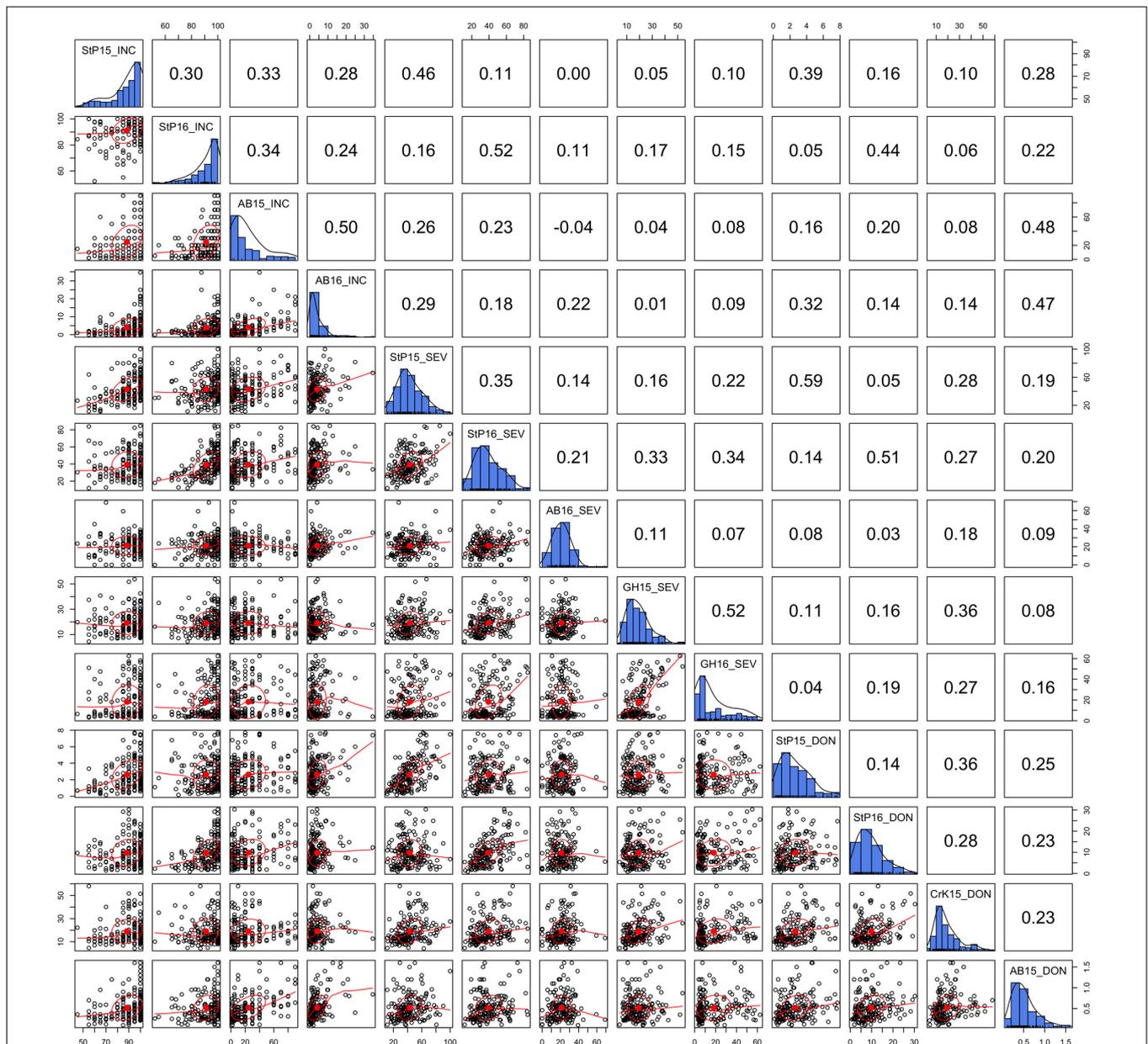
To examine prediction accuracy through validation, the training population and testing population should be completely separate so that the phenotypes of the testing population are not used as the training data. Means across the environments straightly satisfies this requirement. Although BLUP is derived in a mixed linear model with individual lines treated as uncorrelated, BLUPs are potentially dependent among lines due to adjustments for the effects of location, year, and replicate. In our previous gene mapping study (Wang et al., 2017), BLUPs were used as the response variable, instead of means. In this study, we wanted to know how similar the mean is to BLUP.

We demonstrated that the BLUP was almost identical to the mean, especially for INC and SEV, where the data were

nearly balanced and missing data were minimal (Figure 3). In the case of lines treated as unrelated, the BLUP is identical to the mean. Even for DON, BLUPs were almost identical to the means except for a few lines with missing values. The BLUPs of these lines equal their means adjusted by the means of the other lines. Using the BLUP of the training population involves the phenotypes of the testing population. Therefore, we chose to use means to evaluate prediction accuracy through validation.

### Market Classes and Origin of Lines in Relation to PC Analysis

Market class is a major characteristic in wheat breeding. Three major classes of spring wheat are grown in the PNW. HRS and HWS wheat are used for different end-use products compared to SWS wheat. Therefore, HRS and HWS differ from SWS in important requirements such as protein content, flour hardness, and other quality traits. Among the 170 lines we studied, 79 were

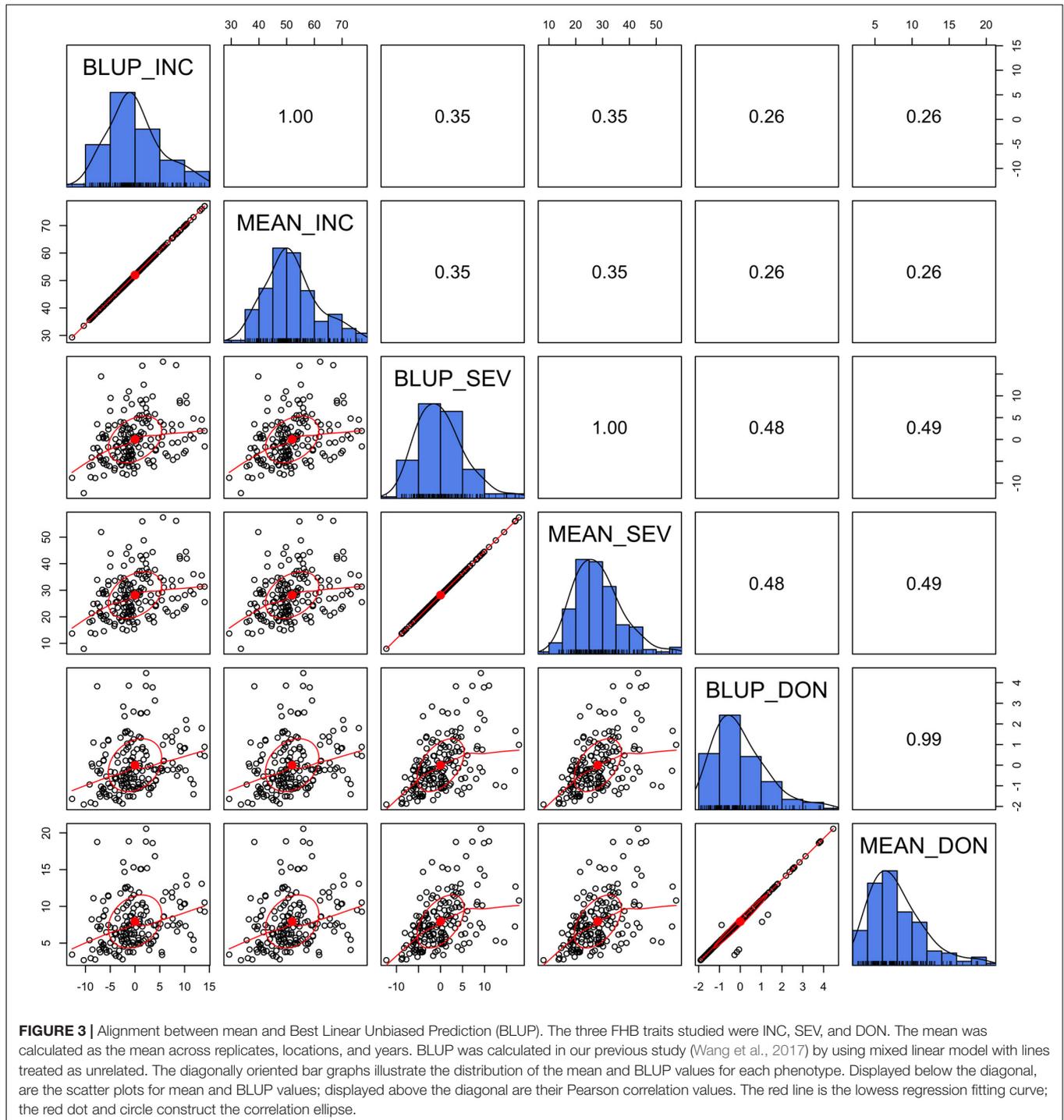


**FIGURE 2 |** Correlation and distribution of FHB phenotypes at different environments. The three FHB traits studied were incidence (INC), severity (SEV), and deoxynivalenol concentration (DON). The environments were defined as combinations of year and location. The study locations were StP (Saint Paul, MN, United States), CrK (Crookston, MN, United States), AB (Aberdeen, ID, United States), and greenhouse (GH). Years included 2015 (15) and 2016 (16). Environments are labeled on the diagonals. The diagonals also illustrate the distribution of the means across replicates within each environment. The dots on the scatterplots off the diagonals represent the mean across replicates. The red line is the robust fitting using lowess regression, the red dot and the circle represent the correlation ellipse.

HWS lines, 65 were HRS, and 24 were SWS. The remaining two lines were soft red spring (SRS) wheat. The 170 lines were from seven origins. The top five major origins (UCD, UI, WSU, MSU, and CIMMYT) contributed 26–49 lines each and 167 in total. The remaining lines consisted of two from LCS and one from SSK (Supplementary Table S1).

To explore the relationship among lines, origins, and market classes, PCs were derived from all available SNPs. The first three PCs explained around 20% of the total variance. Pairwise

relationships for the 170 lines are illustrated in **Figure 4** with market classes and origins indicated by colors and shapes, respectively. The SWS market class stands out from the other two classes on all three plots. Substantial overlap occurs between HRS and HWS. Most of the lines with the same origin were clustered. For example, the lines from CIMMYT (circles) are centered in the upper half of the PC1 vs. PC2 plot. The lines from MSU (diamonds) are concentrated in the top left of the PC2 vs. PC3 plot. We found a strong association between origins and

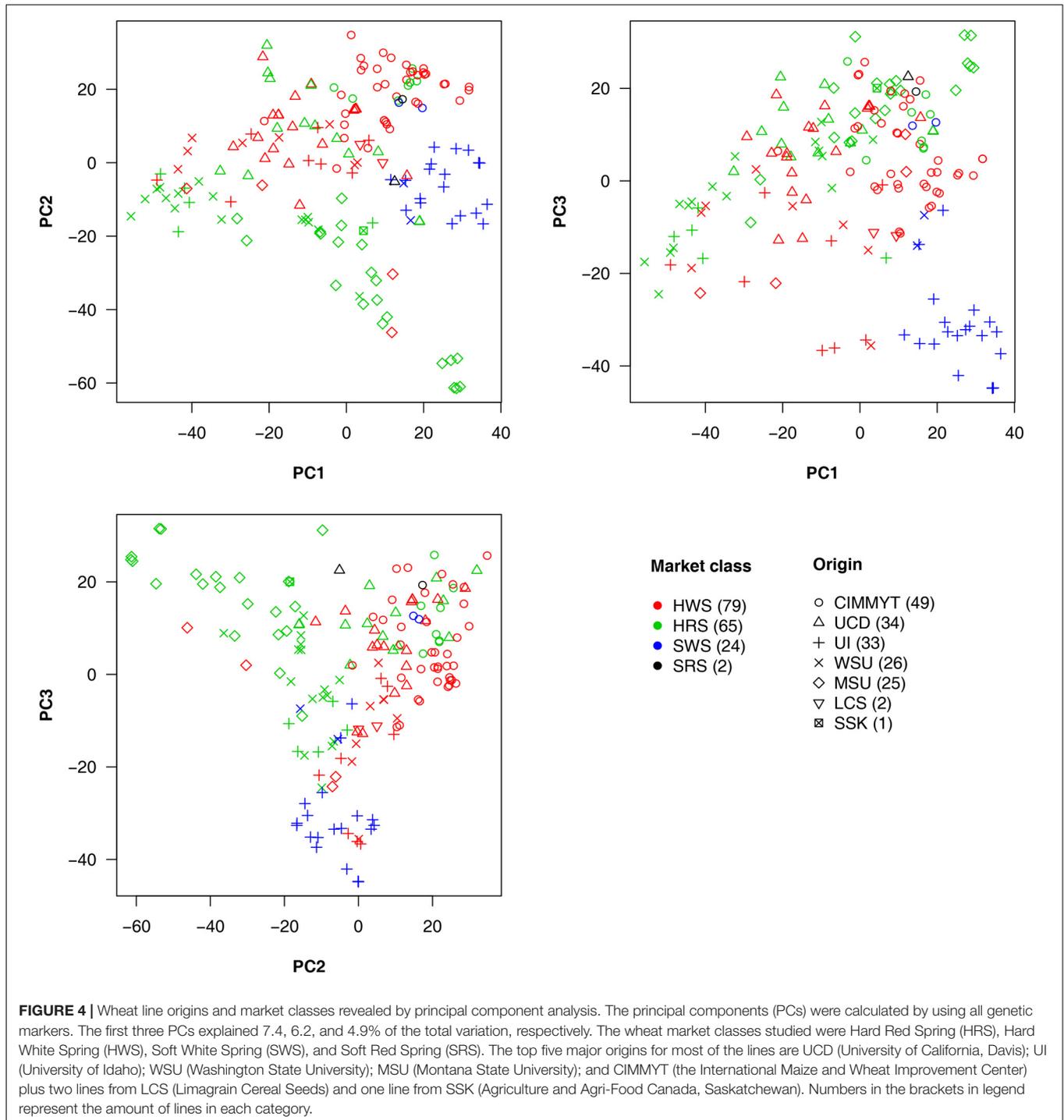


market classes. Almost all the individuals from UI are SWS, those from MSU are mostly HRS, and those from CIMMYT are mostly HWS.

### Heritability Estimation

We estimated the narrow sense heritability by using a fixed effect and random effect mixed linear model. Besides the residual effects, the random effects are the individual total additive genetic

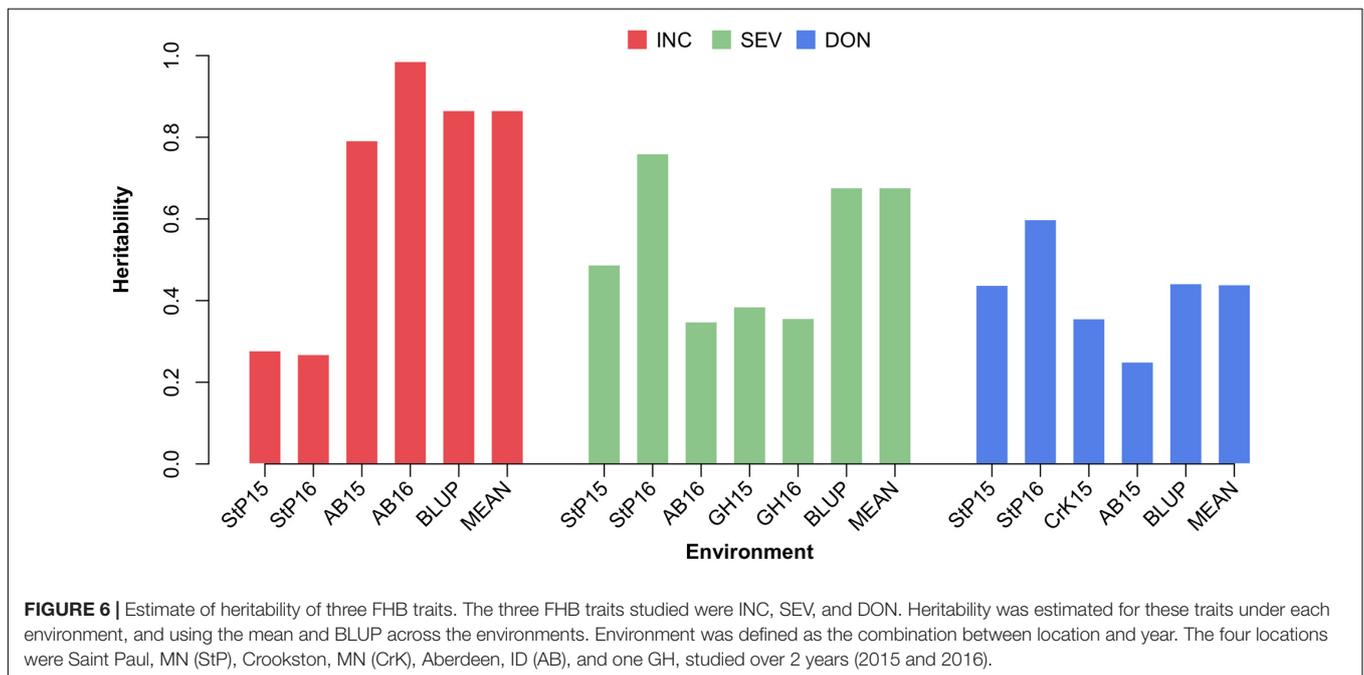
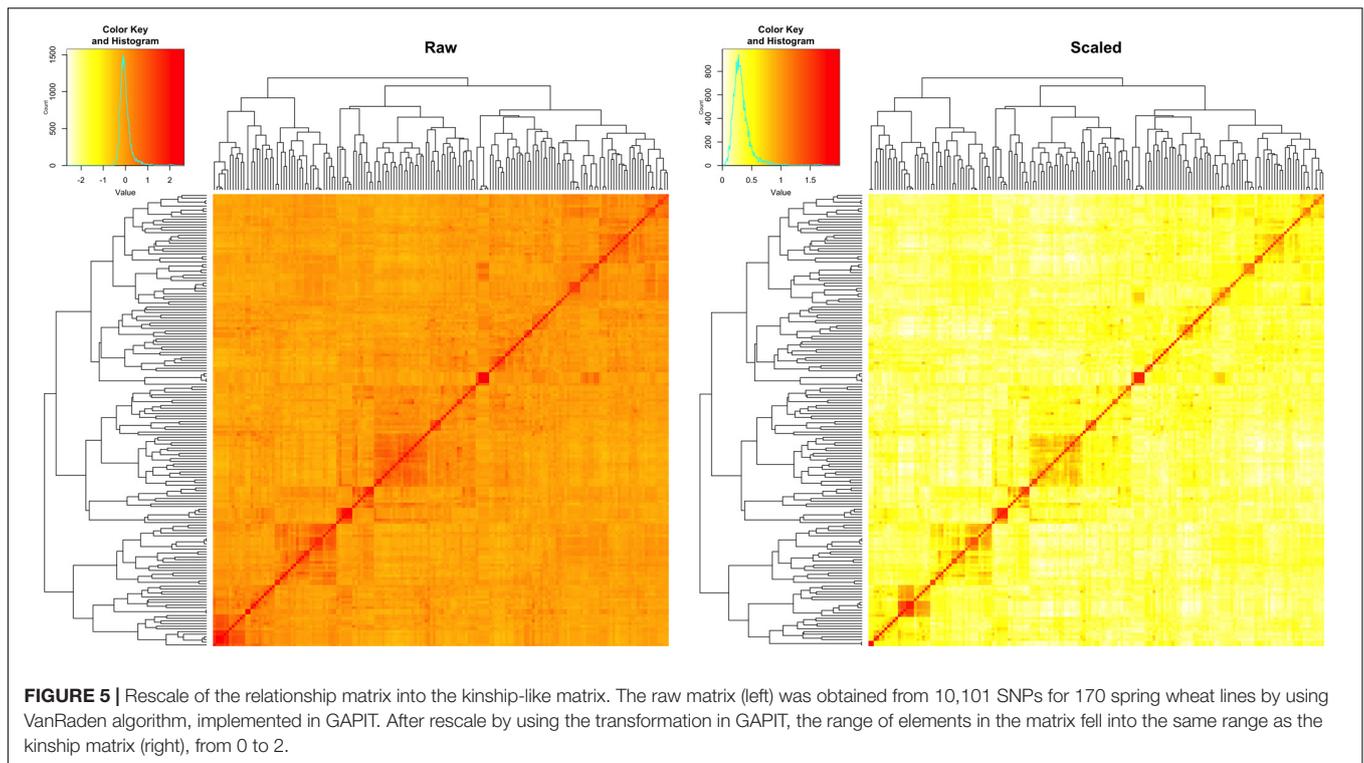
effects with variance structure defined by an additive relationship matrix. The raw additive relationship matrix was derived by using the VanRaden algorithm implemented in GAPIT. Most of the elements in the matrix were close to zero, ranging from  $-3$  to  $3$  (Figure 5). The estimation of heritability requires the minimum number of elements to be zero for uncorrelated individuals and the maximum number of elements to be two on the diagonals for inbred individuals. Thus, we conducted the



pedigree kinship-like transformation, implemented in GAPIT. The scaled relationship matrix had desirable features to estimate heritability (Figure 5).

The variances of individuals' total additive genetic effects and residual effects were estimated by using the EMMA algorithm. The total phenotypic variance was defined as the sum of the additive genetic variance and the residual variance. The proportion of the additive genetic variance over the total

variance was defined as the narrow sense heritability. The heritability estimates for phenotypes measured under different environments, the means across environments, and BLUP are illustrated in Supplementary Table S2, Supplementary Figure S1, and Figure 6. Overall, the estimated heritabilities of INC were much higher than SEV and DON. This finding agrees with the finding that the phenotypes of INC were more correlated among environments.



## Validation With a Homogenous Cohort

We were particularly interested in prediction accuracy when crosses among our current 170-line training population were advanced to higher generations (e.g., F6–F8) or when similar new external lines were introduced. To fill this knowledge gap, we randomly masked the phenotypes (i.e., means) of a certain proportion of the 170 lines and treated the masked lines as the

future testing population (inference). The remaining lines were treated as the training population (reference). The genotypes of individuals in both the training and testing populations were used to calculate the Genomic Estimated Breeding Values (GEBVs) for all individuals, including those in the training and testing populations. However, only the phenotypes of individuals in the training population were used to estimate the effects of all

markers by using the rrBLUP package in R. The prediction accuracy was calculated as the Pearson correlation between the predicted and the observed phenotypes. The predicted phenotypes were the sum of GEBV and the fixed effects, the first principal components. The randomization was replicated 1,000 times and the distributions, means, and standard deviations of the prediction accuracies were reported (Figure 7 and Supplementary Table S3).

Consistent to the higher correlation across environments and higher heritability for INC compared to the other traits, prediction accuracy was also higher for INC. Prediction accuracy also depended on training population size. The scenario with 90% of 170 lines as the testing population had the lowest prediction accuracy. The scenario with 10% of the lines remaining as the testing population had the highest prediction accuracy. In practice, this latter scenario is the most useful because it uses almost all of the currently available resources if the number of introduced new lines is less than 17 and they have similar relationships with the 170 lines. In this case, the prediction accuracy is about 0.65 for INC and 0.45 for both SEV and DON.

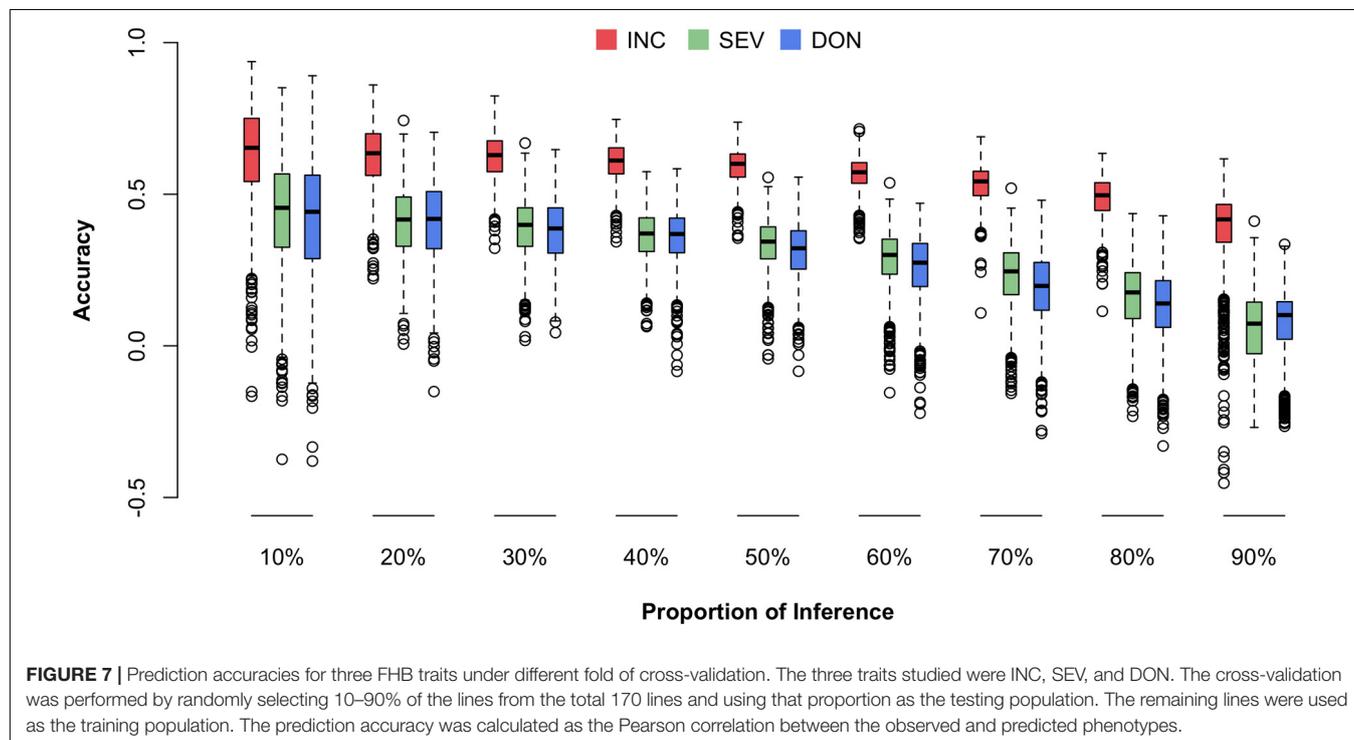
### Validation With a Resistant Cohort

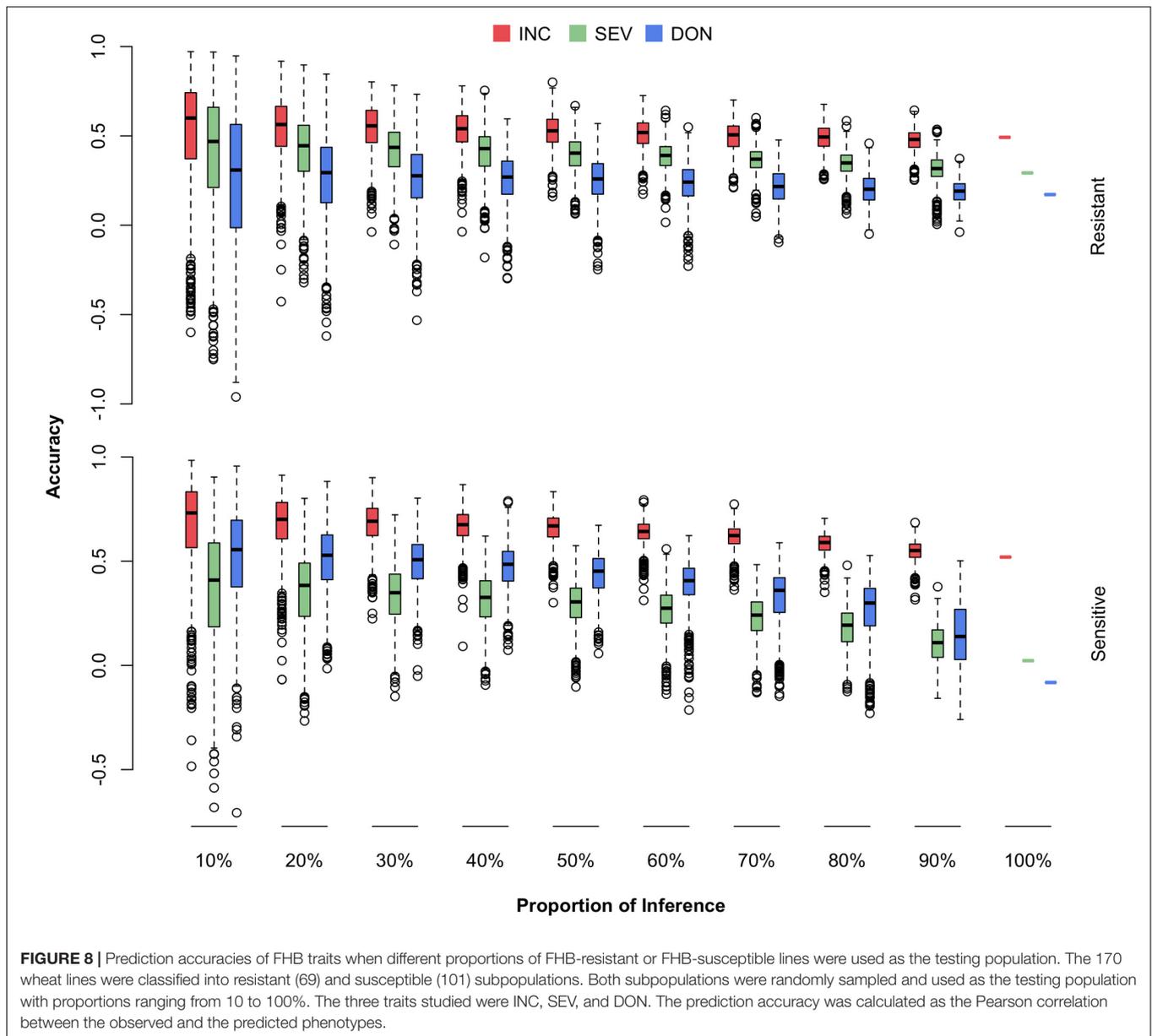
The newly developed varieties, or the new lines introduced to the current breeding population, are most likely to be resistant to FHB. The most critical question is how well a certain amount of such lines can be predicted. Among the 170 lines in the current breeding population, 69 were identified by the breeders (Dr. Jianli Chen and Dr. Michael Pumphrey) as resistant lines and the rest as susceptible lines (Supplementary Table S1 and Supplementary Figure S2). For each subpopulation,

we randomly masked a certain proportion of the lines (10–100%) as the testing population. The remaining lines, including all the lines from the opposite subpopulation, were treated as the training population. The randomization was also replicated 1,000 times. The distributions, means, and standard deviations of the prediction accuracies were reported (Figure 8 and Supplementary Table S4).

We observed that prediction accuracies for the resistant lines were less sensitive to the proportion of resistant lines removed from the training population, compared with prediction accuracies for the susceptible lines as the proportion of susceptible lines were removed. This finding suggests that the current FHB-susceptible lines are valuable for predicting the resistant lines. Especially for the two low heritability traits, SEV and DON, the prediction accuracies for susceptible lines dropped almost to zero when the training population contained only the resistant lines. In contrast, the prediction accuracies for the resistant lines remained at 50, 35, and 20% for INC, SEV, and DON, respectively, even when the training population contained susceptible lines only.

We also observed that prediction accuracy was much lower for the scenario that restricted the testing population to resistant lines than the scenario with a homogeneous cohort. For example, a homogeneous cohort with 90% of lines as the training population resulted in a prediction accuracy of 0.63 for INC. In contrast, a resistant cohort that retained all susceptible lines and 90% of resistant lines as the training population, resulted in a prediction accuracy of 0.53 for INC on resistant lines. This finding suggests that to prevent the overestimate of prediction accuracy as future testing population contains resistant lines only, not a mixture of both resistant and susceptible lines, breeding



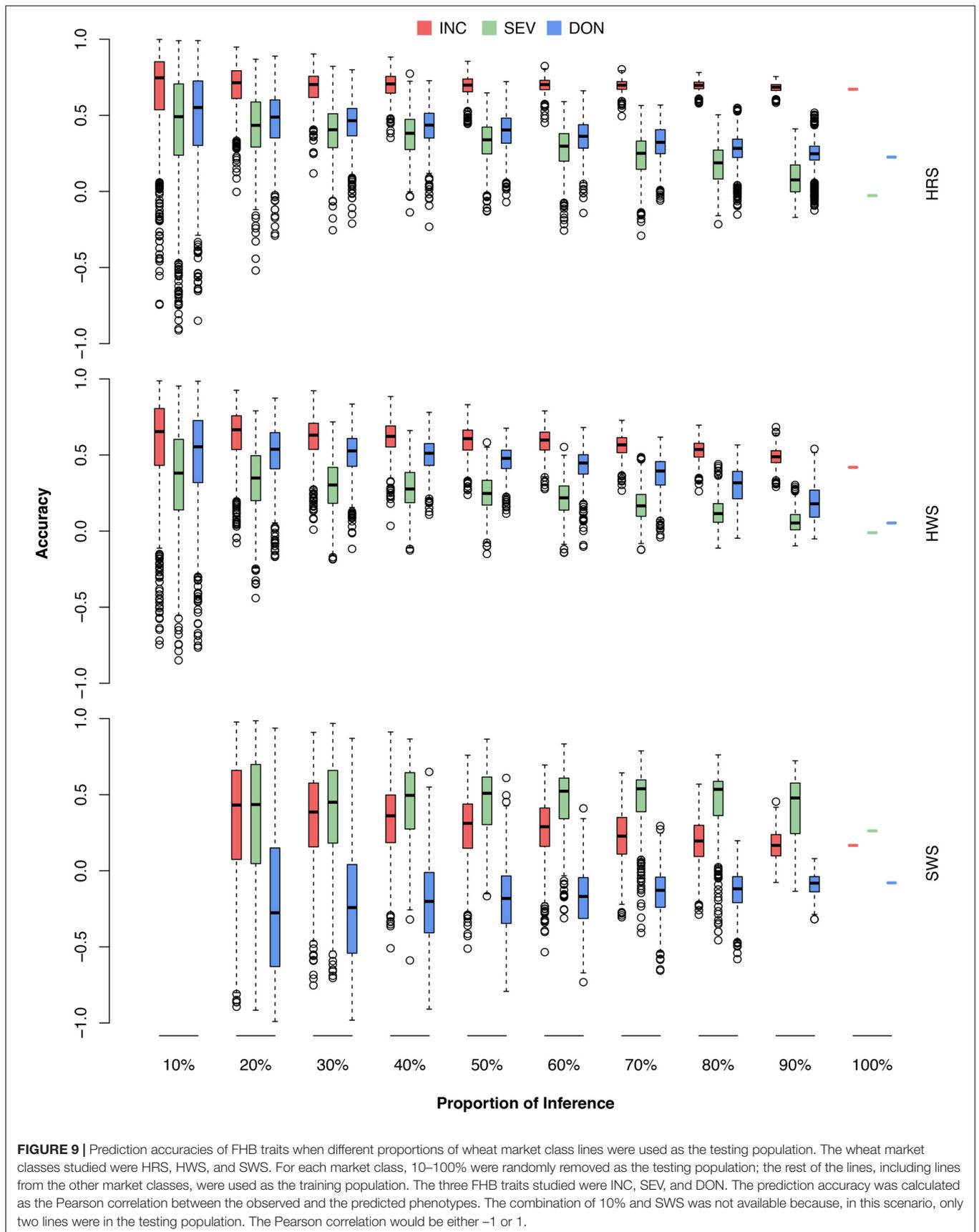


plans should be based on the scenarios that validate a resistant cohort, not scenarios with a homogeneous cohort.

### Validation With a Market Class Cohort

Our wheat breeding program involves three major market classes (HRS, HWS, and SWS) for different end-use products (Supplementary Table S1). The knowledge gap is how the prediction accuracy relates to the amount of new lines from a particular market class. To fill this gap, we randomly masked a certain proportion (10–100%) of lines from a particular market class into the testing population and used the remaining lines, mainly from other market classes, as the training population. The randomization was also replicated 1,000 times. The distributions, means, and standard deviations of the prediction accuracies were reported (Figure 9 and Supplementary Table S5).

As demonstrated by the population structure from the PC plots (Figure 4), HRS and HWS are more connected to each other than to SWS. Consequently, the HRS lines can be predicted well from the HWS lines and vice versa. In contrast, the SWS lines cannot be predicted from the HRS or HWS lines; for example, the accuracies for DON were near or below zero. We narrowed down the problem by conducting the validation for each market class separately (Supplementary Figure S3 and Supplementary Table S6). In this scenario, all the lines in the training population were from the same market class as the lines in testing population. We observed a similar trend as the scenarios with training populations that contained lines from other market classes. The number of lines in SWS was almost one third of other two classes and almost all of them have the same origin. These are two possible reasons caused the low prediction in SWS.



## DISCUSSION

In this study, various types of analyses and validations were performed to evaluate the potential of genomic prediction to improve resistance to FHB for spring wheat in the PNW. The procedures we used included determination of response variables, estimation of heritability, and validation of prediction accuracy under different scenarios. Although some of the procedures are specific to our own breeding program, other wheat breeding programs and even other crop breeding programs can also benefit from our methods and findings.

### Mean vs. BLUP

The validation of GS requires that the phenotypes of the testing population be used only for comparison of the prediction accuracy. In this study, we chose to use the mean for two reasons. First, we wanted to satisfy the requirement of validation that prevents the use of phenotypes in the testing population to predict themselves. Second, the mean and BLUP were highly identical in this study. The correlation between mean and BLUP was 100% for INC and SEV and 99% for DON. Thus, the results inferred from mean are robust enough to presume similar results in future GS based on BLUP.

### Marker-Based Kinship and Heritability Estimation

Heritability is defined as the proportion of genetic variance over the total variance, which is the sum of the genetic variance and the residual variance. Genetic variance and residual variance are usually estimated in a mixed linear model with genetic effect and residual effect as the random effects. The variance structure of the genetic effect is the relationship matrix derived from either pedigree or genetic markers. The matrix is twice of co-ancestry, which is the probability of identical by descent. Consequently, the minimum element is zero indicating no relationship. The maximum element is two for identical twins or an inbred with itself. The genetic variance estimated with such a relationship matrix is adjusted to the base genetic variance among founder who are completely uncorrelated. When the matrix is derived from pedigree, it satisfies all the requirements.

Many algorithms can be used to calculate the relationship matrix among individuals based on genetic markers, including Loiselle (Loiselle et al., 1995) and VanRaden. Most of these relationships do not have the properties corresponding to pedigree-based kinship. Consequently, the estimated genetic variances do not correspond to the genetic variance among individuals of the base population. The proportion to the total variance is artificially defined based on the selection of algorithms.

The 170 lines used in this study were inbred wheat. The diagonals of the relationship matrix should be two. If we assume the farthest two inbred lines are uncorrelated, their corresponding element should be the minimum, zero. With these assumptions, we applied a transformation, implemented in GAPIT, for the raw relationship matrix calculated by the VanRaden algorithm. After the transformation, the maximum

element was 2 and the minimum element was 0. The maximum element was more likely to be correct than the minimum element. Therefore, interpretation of estimated heritability should be inferred with caution, even the pedigree-like relationship matrix used in this study.

### Breeding for FHB Resistance for SWS

Our primary interest is to develop and introduce new varieties with resistance to FHB, especially for SWS, a very important market class in PNW. Although the prediction accuracies were reasonably high under homogenous scenario, the scenario is far from practice. The future testing lines are more likely to be resistant lines from SWS market class. This study suggested that the current training population has reduced prediction accuracies when the testing population is restricted to either resistant lines, or the market class (Figures 8, 9). Introducing lines in these categories and phenotyping them on multiple locations across years are critical to ensure the success of GS in breeding to improve FHB resistance.

The actual prediction accuracy should be better than reported in this study for scenarios where the testing populations contained less than 15 lines, such as the testing cohort that was restricted to SWS. Pearson correlation coefficient is downward biased in such case (Zhou et al., 2017). It is expected that larger training population leads to higher prediction accuracy. This was true for all cases in this study except the scenarios with testing population restricted to SWS for SEV and DON (Figure 9). As there were only 24 SWS lines, 20–50% of them as inference corresponded to that there were only 5–12 lines in the testing population. The smaller the proportion, the more downward biased on the Pearson correlation coefficient. After applying the unbiased correction with the Olkin and Pratt's method (Olkin and Pratt, 1958), the reversed trend disappeared for SEV. For DON, all the averages of accuracies were below zero. The measurements of DON on the 24 SWS lines were not able to be predicted among themselves, or from the lines of the other two market classes.

## CONCLUSION

Correlation across environments were low for the three measurements of FHB: INC, SEV, and DON. Heritabilities of the means across the environments were 0.87, 0.68, and 0.44 for INC, SEV, and DON, respectively. Among the 170 lines, randomly selecting 90% as training population had accuracy of 0.63, 0.43, and 0.42 to predict the rest of 10% as testing population for INC, SEV, and DON, respectively. When testing population were restricted to resistant lines, or the wheat market class, substantial reductions of prediction accuracy were observed. Introducing new resistant soft white lines and phenotyping them in multiple environments would benefit GS for developing new wheat varieties with resistance to FHB.

## ETHICS STATEMENT

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not

necessarily reflect the views of the funding agencies. All datasets analyzed herein were published previously. This study did not involve samples from humans or animals.

## AUTHOR CONTRIBUTIONS

ZZ and JC conceptualized the study with assistance from MP. HD performed the data analyses with assistance from RW and advice from YY and MP. RW, JA, and JC conducted phenotypic assessment of FHB resistance. RW and JC generated genotypic data through service from two USDA-ARS genotyping centers. ZZ and JC wrote the manuscript. All the authors revised the manuscript.

## FUNDING

This research was partially supported by the USDA-ARS U.S. Wheat and Barley Scab Initiative and the National Research Initiative Competitive Grants (awards of 2011-68002-30029, 2017-67007-25939, 2015-05798, and 2016-68004-24770) from

the USDA National Institute of Food and Agriculture, the Idaho Agricultural Experimental Station (Project IDA01483), the Idaho Wheat Commission (Project BJKW32), the Washington Grain Commission (Endowment and award number 126593), and China Scholarship Council (CSC).

## ACKNOWLEDGMENTS

The authors thank spring wheat breeders J. Dubcovsky and L. Talbert for their contributions to wheat lines used in this study; S. Reynolds, W. Zhao, Y. Dong, D. See, and Sh. Chao for assistance in the phenotyping and genotyping processes. The authors also thank Dr. Linda R. Klein for valuable writing advice and editing the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00911/full#supplementary-material>

## REFERENCES

- Anderson, J. A., Stack, R. W., Liu, S., Waldron, B. L., Fjeld, A. D., Coyne, C., et al. (2001). DNA markers for Fusarium head blight resistance QTLs in two wheat populations. *Theor. Appl. Genet.* 102, 1164–1168. doi: 10.1007/s001220000509
- Arruda, M. P., Brown, P. J., Lipka, A. E., Krill, A. M., Thurber, C., Merrill, K. R., et al. (2016). Genome-wide association mapping of Fusarium head blight resistance in wheat using genotyping-by-sequencing. *Plant Genome* 9, 1–14. doi: 10.3835/plantgenome2015.04.0028
- Arruda, M. P., Brown, P. J., Lipka, A. E., Krill, A. M., Thurber, C., Kolb, F. L., et al. (2015). Genomic selection for predicting head blight resistance in a wheat breeding program. *Plant Genome* 8, 1–12. doi: 10.3835/plantgenome2015.01.0003
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci.* 34, 20–25. doi: 10.2135/cropsci1994.0011183X003400010003x
- Bernardo, A. N., Ma, H., Zhang, D., and Bai, G. (2012). Single nucleotide polymorphism in wheat chromosome region harboring *Fhb1* for Fusarium head blight resistance. *Mol. Breed.* 29, 477–488. doi: 10.1007/s11032-011-9565-y
- Buerstmayr, H., Steiner, B., Hatzgenbichler, E., Scholz, U., Mesterhazy, A., Lemmens, P., et al. (2003). “Resistance to Fusarium head blight in wheat: molecular mapping and validation of QTLs for horizontal resistance, implications for resistance breeding,” in *Proceedings of the 10th International Wheat Genetics Symposium*, Paestum, 113–116.
- Buerstmayr, H., Ban, T., and Anderson, J. A. (2009). QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. *Plant Breed.* 128, 1–26. doi: 10.1111/j.1439-0523.2008.01550.x
- Chen, J., Griffey, C. A., Maroof, M. A. S., Stromberg, E. L., Biyashev, R. M., Zhao, W., et al. (2006). Validation of two major quantitative trait loci for fusarium head blight resistance in Chinese wheat line W14. *Plant Breed.* 125, 99–101. doi: 10.1111/j.1439-0523.2006.01182.x
- Chen, J., and Marshall, J. (2012). Evaluation of FHB resistance in PNW spring wheat lines. *Idaho Grain* 1, 18–19.
- Cuthbert, P. A., Somers, D. J., and Brulé-Babel, A. (2007). Mapping of *Fhb2* on chromosome 6BS: a gene controlling Fusarium head blight field resistance in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 114, 429–437. doi: 10.1007/s00122-006-0439-3
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Fuentes, R. G., Mickelson, H. R., Busch, R. H., Dill-Macky, R., Evans, C. K., Thompson, W. G., et al. (2005). Resource allocation and cultivar stability in breeding for Fusarium head blight resistance in spring wheat. *Crop Sci.* 45, 1965–1972. doi: 10.2135/cropsci2004.0589
- Guo, G., Zhou, Z., Wang, Y., Zhao, K., Zhu, L., Lust, G., et al. (2011). Canine hip dysplasia is predictable by genotyping. *Osteoarthritis Cartilage* 19, 420–429. doi: 10.1016/j.joca.2010.12.011
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* 92, 433–443. doi: 10.3168/jds.2008-1646
- Heffner, E. L., Sorrells, M. E., and Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512
- Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* 9, 166–177. doi: 10.1093/bfpg/elq001
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723. doi: 10.1534/genetics.107.080101
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu, S., and Anderson, J. A. (2003). Targeted molecular mapping of a major wheat QTL for Fusarium head blight resistance using wheat ESTs and synteny with rice. *Genome* 46, 817–823. doi: 10.1139/g03-066
- Liu, S., Hall, M. D., Griffey, C. A., and McKendry, A. L. (2009). Meta-analysis of QTL associated with fusarium head blight resistance in wheat. *Crop Sci.* 49, 1955–1968. doi: 10.2135/cropsci2009.03.0115
- Loiselle, B. A., Sork, V. L., Nason, J., and Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am. J. Bot.* 82, 1420–1425. doi: 10.1002/j.1537-2197.1995.tb12679.x
- Marshall, J., Bissonnetter, K., and Chen, J. (2012). *The Re-Emergence of Fusarium Head Blight in Idaho*. Moscow, ID: University of Idaho.
- Marshall, J. (2014). *Fusarium Head Blight Forum*. Idaho Falls, ID: IBC/UI.

- McKendry, A. L., Liu, S., Abate, A., and Davis, G. L. (2004). "Inheritance of Fusarium head blight resistance in the US wheat cultivar 'Ernie,'" in *Proceedings of the 2nd International Symposium on Fusarium Head Blight Incorporating the 8th European Fusarium Seminar* (Orlando, FL: Wyndham Orlando Resort), 107–110.
- McMullen, M. P., Schatz, B., Stover, R., and Gregoire, T. (1997). Studies of fungicide efficacy, application timing, and application technologies to reduce Fusarium head blight and deoxynivalenol. *Cereal Res. Commun.* 25, 779–783.
- Mirocha, C. J., Kolaczowski, E., Xie, W., Yu, H., and Jelen, H. (1998). Analysis of deoxynivalenol and its derivatives (batch and single kernel) using gas chromatography/mass spectrometry. *J. Agric. Food Chem.* 46, 1414–1418. doi: 10.1021/jf970857o
- Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G. Y., and Myles, S. (2015). LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3* 5, 2383–2390. doi: 10.1534/g3.115.021667
- Olkin, I., and Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Ann. Math. Stat.* 29, 201–211. doi: 10.1214/aoms/1177706717
- Pumphrey, M. O., Bernardo, R., and Anderson, J. A. (2007). Validating the *Fhb1* QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. *Crop Sci.* 47, 200–206. doi: 10.2135/cropsci2006.03.0206
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Revelle, W. (2017). *Psych: Procedures for Personality and Psychological Research*. Evanston, IL: Northwestern University.
- Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., and Jannink, J. L. (2012). Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Genome* 5, 51–61. doi: 10.3835/plantgenome2012.02.0001
- Schroeder, H. W., and Christensen, J. J. (1963). Factors affecting resistance of wheat to scab caused by *Gibberella zeae*. *Phytopathology* 53, 831–838.
- Shen, X., Ittu, M., and Ohm, H. W. (2003a). Quantitative trait loci conditioning resistance to Fusarium head blight in wheat line F201R. *Crop Sci.* 43, 850–857. doi: 10.2135/cropsci2003.8500
- Shen, X., Zhou, M., Lu, W., and Ohm, H. (2003b). Detection of Fusarium head blight resistance QTL in a wheat population using bulked segregant analysis. *Theor. Appl. Genet.* 106, 1041–1047. doi: 10.1007/s00122-002-1133-8
- Sneller, C. H., Paul, P., and Guttieri, M. (2010). Characterization of resistance to Fusarium head blight in an eastern U.S. soft red winter wheat population. *Crop Sci.* 50, 123–133. doi: 10.2135/cropsci2008.09.0569
- Snijders, C. H. A. (1990). Fusarium head blight and mycotoxin contamination of wheat, a review. *Neth. J. Plant Pathol.* 96, 187–198. doi: 10.1007/BF01974256
- Steiner, B., Lemmens, M., Griesser, M., Scholz, U., and Schondelmaier, J. (2004). Molecular mapping of resistance to Fusarium head blight in the spring wheat cultivar frontana. *Theor. Appl. Genet.* 109, 215–224. doi: 10.1007/s00122-004-1620-1
- Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., et al. (2016). GAPIT version 2: an enhanced integrated tool for genomic association and prediction. *Plant Genome* 9, 1–9. doi: 10.3835/plantgenome2015.11.0120
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., et al. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17, 520–525. doi: 10.1093/bioinformatics/17.6.520
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Wang, Y., Yong, X., and Xiao, Q. (1982). The improvement of identification technique of scab resistance of wheat and the development of resistant sources. *Sci. Agric. Sin.* 5, 67–77.
- Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183
- Wang, R., Chen, J., Anderson, J. A., Zhang, J., Zhao, W., Wheeler, J., et al. (2017). Genome-wide association mapping of fusarium head blight resistance in spring wheat lines developed in the Pacific Northwest and CIMMYT. *Phytopathology* 10, 1486–1495. doi: 10.1094/PHTO-02-17-0073-R
- Xu, Q., and Liang, Y. (2001). Monte carlo cross validation. *Chemometr. Intell. Lab. Syst.* 56, 1–11. doi: 10.1016/S0169-7439(00)00122-2
- Yang, Z., Gilbert, J., Fedak, G., and Somers, D. J. (2005). Genetic characterization of QTL associated with resistance to Fusarium head blight in a doubled-haploid spring wheat population. *Genome* 48, 187–196. doi: 10.1139/g04-104
- Zhang, Z., Todhunter, R. J., Buckler, E. S., and Van Vleck, L. D. (2007). Technical note: use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *J. Anim. Sci.* 85, 881–885. doi: 10.2527/jas.2006-656
- Zhang, X., Rouse, M. N., Nava, I. C., Jin, Y., and Anderson, J. A. (2016). Development and verification of wheat germplasm containing both *Sr2* and *Fhb1*. *Mol. Breed.* 36:85. doi: 10.1007/s11032-016-0502-y
- Zhou, Y., Vales, M. I., Wang, A., and Zhang, Z. (2017). Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction. *Brief. Bioinform.* 18, 744–753. doi: 10.1093/bib/bbx133

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Dong, Wang, Yuan, Anderson, Pumphrey, Zhang and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.