

## Article

**Genome-Wide Association Studies Combined with Genomic Selection as a Tool to Increase Fusarium Head Blight Resistance in Wheat**Virginia L. Verges<sup>1</sup>, Gina L. Brown-Guedira<sup>2</sup>, David A. Van Sanford<sup>1,\*</sup><sup>1</sup> Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY 40546-0312, USA<sup>2</sup> USDA-ARS Plant Science Research Unit, Raleigh, NC 27695, USA

\* Correspondence: David A. Van Sanford, Email: dvs@uky.edu; Tel.: +11-859-257-7125.

**ABSTRACT**

Genomic selection (GS) has shown successful results as a tool to increase Fusarium Head Blight (FHB) resistance in wheat (*Triticum aestivum* L.). In this study we performed a genome-wide association study (GWAS) on regional FHB nurseries to select significant SNPs for deoxynivalenol (DON) and DSK, an index of DON, FHB rating and Fusarium damaged kernels (FDK). The objective was to determine whether a reduced number of markers could improve predictions of FHB traits compared to the full set of markers for three populations of 306, 281 and 198 lines that were evaluated in 2017, 2018, 2019 respectively at Lexington, Kentucky. Under a forward GS scheme, using regional nurseries as training populations (TP) of sizes 100 and 400, there was a substantial positive increase in prediction accuracy (PA) of 21% for DON (0.28 vs 0.22) and 12% for DSK (0.32 vs 0.28) using a reduced marker set at the smallest TP size. With cross validation, moderate PA was obtained consistently among populations and marker sets for both traits. While the full marker set showed the best performance, PA with reduced marker sets was only slightly lower, (0.55 vs 0.54) for DON and (0.60 vs 0.57) for DSK. Our results confirm first, that GWAS offers an excellent tool to select significant markers for traits like DON and DSK, which reduces the number of markers considerably. Secondly, under a forward GS scheme, using only SNPs significant at  $P < 0.1$  was the most effective strategy in that PA was highest. With these results we move a step forward in selecting lines with good resistance to DON accumulation and other FHB traits before evaluating them in the field, reducing the costs of phenotyping and genotyping.

**Open Access**

Received: 13 January 2021

Accepted: 02 June 2021

Published: 20 August 2021

Copyright © 2021 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:** wheat; genomic selection; genome wide association studies; forward genomic selection; fusarium head blight; disease resistance; DON content

## ABBREVIATIONS

GS, Genomic selection; GWAS, genome wide association study; FHB, Fusarium head blight; FDK, Fusarium damaged kernels; DON, deoxynivalenol; DSK, index of DON, FHB rating, FDK; TP, training population; VP, validation population; PA, prediction accuracy

## INTRODUCTION

Fusarium head blight (FHB) is one of the most devastating diseases of bread wheat (*Triticum aestivum* L.) worldwide, which leads to significant losses in grain yield and quality. FHB is particularly aggressive in regions with cropping systems in rotation with maize and high humidity and moisture through heading and maturity. The disease is primarily caused by *Fusarium graminearum* Schwabe, which infects spikes of wheat leading to the discoloration and deterioration of grain, and the contamination with mycotoxins, mainly deoxynivalenol (DON) [1–3].

Control of FHB is difficult because of the complexity of the disease, and the need to use different management strategies has been proven [3]. Breeding resistant cultivars should be a major part of an integrated approach to reduce the damage from FHB. In this sense, FHB adds complexity to the objective, because resistance is quantitatively inherited with many QTLs involved [4]. Breeding for resistance to a quantitatively inherited disease is a difficult task that requires multiple cycles of breeding, leading to a gradual improvement of resistance over time [5]. The use of an optical sorter based selection has shown promising results, increasing the percentage of individuals with higher levels of FHB resistance [6,7]. Together with marker-assisted selection (MAS), which has been used for improving FHB resistance [8–15], these approaches have become useful strategies in this fight against FHB. However, attempts to improve complex quantitative traits by using QTL-associated markers is not completely successful because of the difficulty of finding the same QTL across multiple environments (due to QTL x environment interactions) or variable effectiveness in different genetic backgrounds [16,17]. Moreover, the existence of multiple minor QTLs responsible for FHB resistance in different backgrounds has been addressed [18,19]. Genome wide association studies (GWAS), through the use of high-density SNPs maps, have been successful at detecting a high number of significant marker-trait associations for FHB traits [20–25].

Genomic selection (GS) is a form of MAS that simultaneously estimates all locus, haplotype or marker effects across the entire genome to calculate genomic estimated breeding values (GEBVs) [26]. Since its inception, there have been many studies that demonstrate the utility of GS in breeding for disease resistance in crops [5,16,27–29]. In wheat, FHB resistance is a challenging breeding target due to the combination of quantitatively inherited resistance and a challenging phenotype that is not easy to reproduce artificially. Thus, GS provides a great opportunity to breed FHB-

resistant wheat cultivars. Research evaluating the performance of GS on the prediction of FHB traits in wheat and barley (*Hordeum vulgare* L.) has produced some promising results. Some studies have predicted GEBVs under a cross validation scheme [30–35], while others have investigated the application of GS models under a forward selection scheme [22,36–40].

In a recently published study, Verges et al. [40] found prediction accuracies (PA) of 0.5 when predicting scab traits for populations of the University of Kentucky (UK) wheat breeding program in a forward GS scheme, with regional scab nurseries serving as the TP. In this study, three different optimization methods and four TP sizes were tested, at a constant number of markers; a high density SNP set with 20,929 SNPs. Some research has been done regarding the effect of marker number, with an agreement that a plateau in PA is reached with low to medium size marker sets [27,31,41]. All of these studies were done under a cross validation scheme. We are not aware of studies that evaluate the effect of marker number when training population (TP) and validation population (VP) are independent samples when predicting scab traits.

One way to evaluate different marker sets that vary in size is to define them based on the magnitude of their effect as indicated by *P* value. This would mean a strategy that begins with a GWAS to find marker-trait significant associations, followed by genomic prediction to calculate the GEBVs. Some studies have taken this approach and evaluated it in rice [42,43], maize [44] and wheat [34,45,46], with mixed results in terms of success. Positive results were achieved by Hoffstetter et al. [34], where they predicted FHB resistance and other traits with reduced marker sets with different levels of significance, finding increases in PA of 50% as an average. Conversely, Larkin et al. [46] found a reduction in PA compared to a GS model when GWAS-derived significant markers were added as fixed effects to the GS model to predict FHB traits.

For this study, we tried to go one step further in investigating the application of GS to predict FHB traits. Our first objective was to investigate the effectiveness of using GWAS to establish marker-trait relationships that could provide statistically significant SNPs associated with each individual trait under evaluation (DON and DSK, an index of DON, FHB rating and Fusarium damaged kernels (FDK)) for different UK populations and the regional scab nurseries. Secondly, we proposed to evaluate the impact of these reduced marker sets in predicting both traits under a cross validation scheme for all populations and under a forward GS scheme, where the regional scab nurseries become the TP. Finally, we investigated the impact of predicting FDK and FHB rating, using marker subsets defined for DON and DSK.

## MATERIALS AND METHODS

The plant material in this study comprised lines from the University of Kentucky (UK) soft red winter wheat breeding program, and the 2014–

2018 Uniform Northern and Uniform Southern soft red winter wheat scab nurseries (NUS and SUS respectively; Supplementary Table S1).

Lines belonging to the UK wheat breeding program derived from multiple F<sub>4:5</sub> and F<sub>4:6</sub> families and were evaluated in yield trials as part of the testing program. They come from crosses made by the breeding program pursuing the program's goals, one of which is increased FHB resistance. The breeding lines may have in their pedigree a parent that was evaluated for FHB resistance in the uniform scab nursery, but it was not a condition of this study. Three populations of 306, 281 and 198 lines were evaluated in 2017, 2018, 2019 respectively at Lexington, Kentucky. In the three growing seasons the genotypes were planted in 1.2 m rows long, spaced 30 cm apart in the UK mist-irrigated, inoculated FHB nursery. The soil type at the site is a Maury silt loam (fine, mixed, semiactive, mesic typic Paleudalfs). The experiment was planted in a randomized complete block design with two replications. Two checks, a resistant line (KY02C-3005-25) and a susceptible cultivar (Pioneer Brand 2555) were planted across the experiment.

In all seasons, the FHB Nursery had an overhead mist irrigation system on an automatic timer that started three weeks before heading. The irrigation schedule was as follows: 5 min periods every 15 min from 2000 to 2045 h, 2100 to 2145 h, 0200 to 0245 h, 0500 to 0530 h, and 0830 h [13]. The experiment was inoculated with *Fusarium graminearum*—infected corn (*Zea mays* L.). Inoculum comprised 27 isolates taken from scabby wheat seeds collected over the years 2007–2010 from multiple locations across Kentucky [47]. The inoculum was prepared by allowing corn to imbibe water for approximately 16 h before autoclaving. After autoclaving, a solution of 0.2 g streptomycin in 150 mL sterile water was mixed in the corn to avoid the growth of other microorganisms. The corn was inoculated with potato dextrose agar (PDA) plugs containing *Fusarium graminearum*, covered and incubated for 2 weeks until fully colonized by the fungus. After that, the corn was spread on the floor until dry, and put in storage bags in a freezer until use. Approximately 3 weeks prior to heading, the scabby corn was spread in the rows at a rate of 11.86 gm<sup>-2</sup> [13].

For the NUS and SUS each nursery cooperator submits his or her breeding materials for evaluation and conducts an inoculated FHB trial at his or her location following the protocols developed by the US Wheat and Barley Scab Initiative (<https://scabusa.org/>) whose aim is to develop control measures against FHB. Two hundred twenty-nine lines belonging to the NUS that represented elite germplasm from public and private breeding programs were evaluated in field environments from 2014 to 2018. The NUS was evaluated at one or two locations in up to nine states per year from 2014 to 2018: Indiana, Illinois, Kentucky, Michigan, Missouri, Nebraska, Ohio, Virginia. The data set was balanced for individual years where the same set of genotypes was evaluated across different locations and unbalanced between years. Another set of 223 lines was evaluated in field environments from 2014 to 2018; these experiments were part of the SUS and represented elite germplasm from public and private breeding

programs. The SUS was evaluated at one or two locations in up to 10 states per year from 2014 to 2018: South Carolina, Georgia, Louisiana, Arkansas, North Carolina, Virginia, Illinois, Kentucky, Missouri and Indiana. The data was balanced for individual years where the same set of genotypes was evaluated across different locations and unbalanced between years. A list of location/year combinations for each regional nursery is shown in Supplementary Table S1.

### Phenotypic Evaluation

At 24 days after heading, FHB rating was recorded using a 0–9 scale. FHB rating is a visual estimate of the incidence and severity of the disease ranging from 0 (absence of FHB symptoms) to 9 ( $\geq 90\%$  of FHB blighted spikelets). Heading date (HD) was recorded when 50% of the spikes in a row had emerged from the flag leaf sheath (in Julian dates; data not shown). Plant height (cm) was measured from the soil surface to the top of the spike, excluding awns (data not shown). Lines were manually harvested using a sickle, mechanically threshed and cleaned. After cleaning, a grain sample of approximately 15 g from each row was further cleaned by hand and evaluated for Fusarium damaged kernels (FDK). The percentage of FDK was estimated by visually comparing samples with known levels of FDK ranging from 5 to 90%. The same sample (15 g) was subsequently sent to the University of Minnesota DON testing laboratory for DON analysis. DON concentration was determined by gas chromatography with mass spectrometry [48,49]. An index was created [40] combining FHB rating, FDK percentage and DON content with the formula:

$$\text{DSK index} = \text{FHB} \times 0.2 + \text{FDK} \times 0.3 + \text{DON} \times 0.5, \quad (1)$$

DSK index was created to emphasize the importance of kernels traits (FDK, DON) in breeding for FHB resistance.

The regional nurseries data were obtained for every genotype, location, year combination. Lines were planted in a 1.2 m row spaced 30 cm with two blocks. A common check cultivar (Ernie) was planted in the NUS and SUS across years and locations. Historical data consisted of entry mean data for FHB rating, FDK and DON concentration for each combination of genotype/location/year.

### Data Analysis

The following linear mixed model was utilized for the analysis of the FHB traits for which individual row-level data were available:

$$Y_{lk} = \mu + B_k + G_l + \varepsilon_{kl}, \quad (2)$$

where  $\mu$  was the mean,  $Y_{lk}$  was phenotypic observation of the  $l$ th genotype at the  $k$ th block,  $B_k$  was effect of the block,  $G_l$  was the effect of the genotype, and  $\varepsilon_{kl}$  represented the residual term. The overall mean and the genotypic effects were considered fixed, and the block term was treated as a random

effect. Best Linear Unbiased Estimators (BLUES) were derived from the model above.

For the historical data of the NUS and SUS nurseries, a single value of each line-environment combination was available for the different traits (FHB, FDK, DON). Therefore, the following linear mixed model was used for this data:

$$Y_{ijl} = \mu + Y_i + L_j + YL_{ij} + G_l + YG_{il} + LG_{jl} + \varepsilon_{ijl}, \quad (3)$$

where  $\mu$  was the mean,  $Y_{ijl}$  was phenotypic observation of the  $l$ th genotype at the  $i$ th year in the  $j$ th location,  $Y_i$  was the effect of the year,  $L_j$  was the effect of the location,  $G_l$  was the effect of the genotype and  $YG_{il}$  and  $LG_{jl}$  were the interaction terms year by genotype and location by genotype respectively, while  $\varepsilon_{ijl}$  represented the residual term. The overall mean and the genotypic effects were considered fixed and all the remaining terms random. The model above is the one from which BLUES were derived.

### Genotyping

For the 785 breeding lines from the University of Kentucky wheat breeding program, DNA was extracted using the Sbeadex plant kit from BioSearch Technologies; using leaf samples from the  $F_{4:5}$  or  $F_{4:6}$  lines that were collected by sampling a minimum of eight 7–10 day old seedlings. Genotyping by sequencing (GBS) [50] using the protocol described by Poland et al. [51] was conducted for the 785 lines that were phenotyped. Single nucleotide polymorphism (SNP) calling on raw sequence data for UK breeding lines and regional scab nurseries was done with Tassel-5GBSv2 pipeline version 5.2.35. SNPs with  $\leq 50\%$  missing data,  $\geq 5\%$  minor allele frequency and  $\leq 10\%$  of heterozygous calls per marker locus were retained and imputation performed using Beagle v4.0. The final number of SNPs defined was 20,929.

### Design of the Training Populations and Validating Populations

For this study we used the same training populations (TPs) with 100 and 400 individuals selected at random by Verges et al. [40]. In this study, four TP sizes were created (100,200,300,400) based on three different optimization methods to select lines for the TP: at random, based on the two tails distributions of lines for a specific trait and based on PEV (Prediction Error Variance) algorithm. As a summary, the NUS and SUS were combined together and we randomly selected 100 and 400 lines to constitute two different TP sizes to estimate GEBVs for the UK breeding lines in the three consecutive years. The validating populations (VP) were created by selecting 50 genotypes randomly from the total breeding lines for each year independently, creating a total of 20 validation sets for 2017, 2018 and 2019. The validating populations created for 2017 and 2018 sets were used previously in the study mentioned above and the VP created for the 2019 sets were created for this study.

### Genome Wide Association Study (GWAS)

Marker-trait associations were tested in the Genome Association and Prediction Integrated Tool (GAPIT) [52] using a mixed linear model (MLM) for the regional scab nurseries (NUS + SUS). A mixed linear model (MLM) includes both fixed and random effects. Individuals are included as random effects and these gives a MLM the ability to incorporate information about relationships among individuals. This information about relationships is conveyed through the kinship (K) matrix, which is used in an MLM as the variance-covariance matrix between individuals [52]. GAPIT produces a series of output files, including Manhattan plots, Q-Q plots and an association table with GWAS results for all SNPs analyzed, including *P*-values. First, GWAS was performed for each one of the 10 sets of lines becoming the TPs for cross validation, separately. This was done to prevent “inside trading” effect, described by Arruda et al. [29]; SNPs significantly associated with all traits were identified and specifically for DON content and DSK index, significant SNPs were selected to create three marker sets at different *P* value levels (0.01, 0.05 and 0.1). Secondly, GWAS was performed to the complete set of regional nursery lines (442 lines) to identify and select SNPs significantly associated with DON and DSK and create markers sets at different *P* value levels (0.01, 0.05 and 0.1) which were used on the forwards GS approach to predict UK breeding lines in the three consecutive years (2017,2018,2019).

### Genomic Prediction

GEBVs for FHB rating, FDK, DON, and DSK were estimated using ridge regression best linear unbiased prediction (RR-BLUP) [26] with the model.

$$y = X\beta + Zu + e, \quad (4)$$

where *y* is a vector of BLUEs for one trait for each wheat genotype,  $\beta$  is a vector of fixed effects which includes the overall mean and fixed covariates (major QTL and association mapping markers), *u* is a vector of random marker effects, *X* and *Z* are the design matrices for fixed and random effects, respectively, and *e* is a vector representing residual terms. The variance–covariance structure associated with the random term was  $u \sim N(0, I\sigma_u^2)$  and for the residual term was  $e \sim N(0, I\sigma_e^2)$ . The estimates of *u* were obtained from the `mixed.solve` function using the package RR-BLUP in R [53]. Prediction accuracy was defined as the Pearson correlation between the phenotypic values (BLUEs) and the GEBVs (predicted) values.

### Cross Validation

For cross validation, a total of 10 different TP (*N* = 351 lines) and VP (*N* = 91 lines) sets were created with the set of lines from the regional nurseries. We investigated the predictive ability of the genomic selection model for each of the two traits and calculated the prediction accuracy (Pearson correlation between phenotypic values and GEBVs) across 10

iterations of cross validation. A random sampling cross validation was conducted, training the model with 10 different TPs and VPs that were created to avoid the possible overestimated PA produced when the GS model is trained with markers selected by GWAS on the same lines that will become the VP “inside trading effect” (Table 1). Each TP was created randomly, had 351 lines, a 80% of the total regional nurseries lines and the other 20% become the VP.

**Table 1.** Number of SNPs for DON and DSK that constituted the marker sets for Cross Validation. Scenario 1 (GS) included the full set of SNPs. Scenarios 2–4 comprised SNPs with a significant effect for DON/DSK obtained after GWAS for the regional nurseries.

| TRAIT | Scenarios  | Marker Sets   | Number of SNPs |        |        |
|-------|------------|---------------|----------------|--------|--------|
|       |            |               | AV             | Min    | Max    |
| DON   | Scenario 1 | GS (Full Set) | 20,939         | 20,929 | 20,929 |
|       | Scenario 2 | GWAS (0.01)   | 218            | 205    | 238    |
|       | Scenario 3 | GWAS (0.05)   | 941            | 877    | 1078   |
|       | Scenario 4 | GWAS (0.1)    | 1878           | 1756   | 1990   |
| DSK   | Scenario 1 | GS (Full Set) | 20,939         | 20,929 | 20,929 |
|       | Scenario 2 | GWAS (0.01)   | 171            | 143    | 201    |
|       | Scenario 3 | GWAS (0.05)   | 875            | 829    | 914    |
|       | Scenario 4 | GWAS (0.1)    | 1852           | 1724   | 1935   |

### Forward Prediction Study

The forward prediction study included the use of two different training population sizes (100,400) comprising the combination of NUS and SUS to estimate GEBVs of 20 prediction sets with 50 breeding lines each, for each of the three years (2017,2018,2019).

**Table 2.** Different marker scenarios applied for forward genomic selection. Significant SNPs obtained after GWAS for the two traits are shown in this order: DON/DSK.

| Scenarios  | Marker Sets   | Number of SNPs |
|------------|---------------|----------------|
| Scenario 1 | GS (Full Set) | 20,929         |
| Scenario 2 | GWM (0.01)    | 234/176        |
| Scenario 3 | GWM (0.05)    | 899/831        |
| Scenario 4 | GWM (0.1)     | 1780/1756      |

Regional nurseries were defined as TPs and GWAS was used to define marker set scenarios for GS (Table 2). Scenario 1 (GS-full set) used all the SNPs that passed the filtering and imputation process (20,929). Scenario 2 (GWM-0.01) used only a set of significant ( $P < 0.01$ ) SNPs for DON content and DSK index. Scenario 3 (GWM-0.05) used only a set of significant ( $P < 0.05$ ) SNPs for DON content and DSK index. Scenario 4 (GWM-0.1) used only a set of significant ( $P < 0.1$ ) SNPs for DON and DSK index. Therefore, the study included five (4) different marker sets, two (2) TP sizes (100,400), and

three (3) different sets of lines as validating populations (2017, 2018 and 2019 populations).

## RESULTS

### Genome Wide Association Study

Manhattan and QQ plots from GWAS for DON content and DSK for the regional FHB nurseries are shown in Supplementary Figure S1. The Manhattan plot is a scatter plot where the X-axis is the genomic position of each SNP, and the Y-axis is the negative logarithm of the  $P$ -value obtained. The Quantile-Quantile (Q-Q) plot is a tool for assessing how well the model used in GWAS accounts for population structure and family relatedness. In this study, the Q-Q plot for each analysis showed that the observed- $\log_{10}$  ( $P$  value) was close to the expected- $\log_{10}$  ( $P$  value), but in the tail of the distribution, deviations from observed values in most cases indicated that significant marker effects were found. GWAS provided the significant markers at levels 0.01, 0.05 and 0.1 for the different populations and scenarios in this study.

### *Phenotypic summary*

The training populations used in this study consisted of a set or lines entered in the US regional scab nurseries: Uniform Northern (NUS) and Uniform Southern Scab Nursery (SUS), and three set of breeding lines from the University of Kentucky wheat breeding program. This historical TP data comprised five years that were evaluated and curated ([https://scabusa.org/publications#pubs\\_uniform-reports](https://scabusa.org/publications#pubs_uniform-reports); verified January 13 2021). It is important to note that the regional nursery entries have been selected by breeders on the basis of their scab resistance, while the breeding lines had been advanced on the basis of agronomic performance and had not yet been screened for scab resistance. The phenotypic information (Table 3) for the nurseries, and the populations evaluated in Lexington, KY in 2017, 2018 and 2019, showed that good levels of infection were achieved, so that we were able to score genotypes and differentiate resistant and susceptible reactions for the different traits. The means for FHB rating, ranged from 3.30 in the regional nurseries to 5.35 in the 2018 population, with a minimum rating of 1–1.25 and a maximum rating score of 8.5 in the four sets. The mean FDK percentage ranged from 12.56% for the 2019 population and 48.62% for the 2017 population. The mean FDK for the 2017 population is higher than the one obtained for the regional nurseries (29.84%). FDK ranged from minimum values between 3.5 and 12% to higher values between 40–90%. The highest FDK was achieved with the 2017 population, and the lowest with 2019 population, both planted in Lexington KY. Regarding DON levels, the mean DON content ranged from 1.74 ppm for 2019 population to 24.92 ppm for 2017 population. Regional nurseries and 2018 population had mean values intermediate between these two contrasting values for DON content. The DON values ranged

from 0.16 to 5.12 ppm for 2019 population, 3.3 to 36.55 ppm for 2018 population, 11.1 to 51.4 ppm for 2017 population and 2.27 to 24.46 for the regional nurseries. DON levels in Lexington, KY in 2017 were higher than generally occurs. Despite these high values, we still could observe phenotypic variance among the evaluated lines. There was a range of 40 ppm between the lowest and highest values for the 2017 population and a very low range (5 ppm) between DON content values for the 2019 population, indicating that data from 2019 the lowest DON content accumulation. DSK index was calculated based on these traits.

**Table 3.** Summary of the phenotypic information for FHB rating, FDK (Fusarium damaged kernels) and DON (Deoxynivalenol content) for the regional scab nurseries and the Kentucky breeding populations evaluated in 2017, 2018 and 2019.

|                       | <b>FHB rating<br/>(0–9)</b> | <b>FDK<br/>(%)</b> | <b>DON<br/>(ppm)</b> |                       | <b>FHB rating<br/>(0–9)</b> | <b>FDK<br/>(%)</b> | <b>DON<br/>(ppm)</b> |
|-----------------------|-----------------------------|--------------------|----------------------|-----------------------|-----------------------------|--------------------|----------------------|
| Reg Nur* (2014-2018)  |                             |                    |                      | Breeding lines (2017) |                             |                    |                      |
| Mean                  | 3.30                        | 29.84              | 9.01                 | Mean                  | 4.10                        | 48.62              | 24.92                |
| Min                   | 1.25                        | 9.70               | 2.27                 | Min                   | 1.0                         | 12.00              | 11.10                |
| Max                   | 7.50                        | 61.47              | 24.46                | Max                   | 8.50                        | 90.00              | 51.40                |
| Breeding lines (2018) |                             |                    |                      | Breeding lines (2019) |                             |                    |                      |
| Mean                  | 5.35                        | 32.70              | 11.97                | Mean                  | 4.16                        | 12.56              | 1.74                 |
| Min                   | 2.50                        | 5.00               | 3.30                 | Min                   | 2.0                         | 3.50               | 0.16                 |
| Max                   | 8.50                        | 85.00              | 36.55                | Max                   | 7.0                         | 40.00              | 5.12                 |

\*Regional nurseries.

*Effect of the different marker sets with cross validation*

We evaluated the effect of the different marker set scenarios shown in Table 1 to predict DON and DSK for the regional nurseries with cross validation. Table 4 showed for DON a moderate PA with all scenarios. The highest PA was obtained with scenario 1, the full set of markers, and a slight reduction is observed with the three scenarios of marker subsets. Using scenario 3 (GWM-0.05) PA was 0.54, a 2% reduction compared to the full marker set (PA = 0.55). With scenario 4 (GWM-0.1) PA was 0.53, a 4% reduction compared to scenario 1. The lowest PA was obtained with scenario 2 (GWM-0.01), the smaller marker set. We observed the same trend for DSK index, with scenario 1 obtaining the highest PA, 0.6 and scenario 3 and 4 (GWM-0.05; GWM-0.1) having a slight reduction of a 5–8% in PA compared to scenario 1.

**Table 4.** Mean prediction accuracy and standard deviation for DON and DSK index with Cross Validation.

| <b>Marker set/Trait</b> | <b>GS</b>   | <b>GWM (0.01)</b> | <b>GWM (0.05)</b> | <b>GWM (0.1)</b> |
|-------------------------|-------------|-------------------|-------------------|------------------|
| DON                     | 0.55 ± 0.07 | 0.51 ± 0.07       | 0.54 ± 0.05       | 0.53 ± 0.06      |
| DSK                     | 0.6 ± 0.05  | 0.49 ± 0.06       | 0.55 ± 0.06       | 0.57 ± 0.05      |

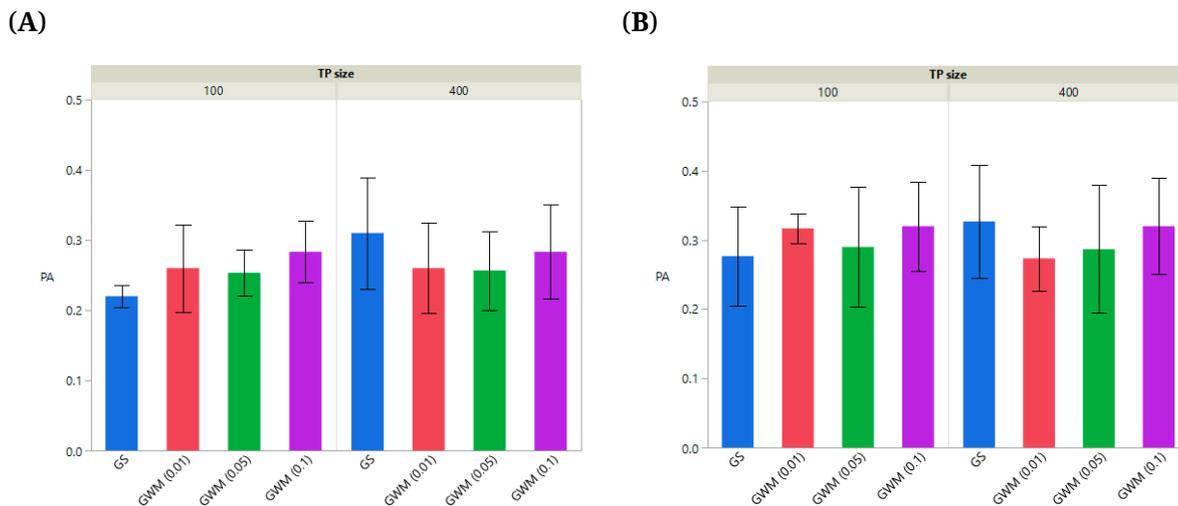
*Effect of the different marker set scenarios and TP size on predicting DON content and DSK index in a forward selection scheme*

Table 5 shows the prediction accuracies obtained for the two traits under all marker scenarios (Table 2) and for the three different set of lines evaluated, with a TP size of 100. As a general conclusion, at TP = 100 the highest PA obtained for DON (0.38) and for DSK (0.44) was achieved, predicting the 2017 population under scenarios 2 (GWM-0.01) for DON and Scenario 3 (GWM-0.05) for DSK. If we look at the PA obtained by year, for the 2017 population, the highest prediction accuracy (PA = 0.38) for DON content was achieved with Scenario 2 (GWM-GWM-0.01), and the lowest PA (0.24) was obtained with Scenario 1, (GS-full set). For DSK, also for 2017 population, the highest PA (0.44) was achieved with Scenario 3 and the lowest PA (0.36) was achieved with the Scenario 2 (GWM-0.01). For the 2018 population, the highest PA for DON content was obtained with scenario 4 (GWM-0.1) (PA = 0.25) and the lowest PA was obtained with Scenario 3, showing for this trait/year combination similar PA ranging from 0.22–0.25 with all scenarios. For DSK, also for 2018 population, the highest PA (0.29) was achieved with scenario 2 (GWM-0.01) and the lowest PA (0.14) was achieved with Scenario 3 ( $P < 0.05$ ). For the 2019 population, the highest PA (0.23) for DON content was obtained with Scenario 4 (GWM-0.1), while the lowest PA (0.17) achieved with Scenario 2 (GWM-0.01). For DSK, also for the 2019 population, the highest PA (0.34) was obtained with scenarios 4 (GWM-0.1) and the lowest PA (0.21) was obtained with scenario 1 (GS-full set).

**Table 5.** Accuracy for DON and DSK index under different marker sets and two TP sizes (100 and 400). GS whole genotype marker data (~ 20,929 SNPs), GWM (0.01) marker subset based on significance level  $P < 0.01$  (234 and 176 SNPs), GWM (0.05) marker subset based on significance level  $P < 0.05$  (899 and 831 SNPs), GWM (0.1) marker subset based on significance level  $P < 0.1$  (1780 and 1756 SNPs).

| Marker set | Year | DON  |      | DSK  |      |
|------------|------|------|------|------|------|
|            |      | 100  | 400  | 100  | 400  |
| GS         | 2017 | 0.24 | 0.43 | 0.42 | 0.49 |
| GS         | 2018 | 0.23 | 0.34 | 0.20 | 0.25 |
| GS         | 2019 | 0.19 | 0.16 | 0.21 | 0.24 |
| GWM (0.01) | 2017 | 0.38 | 0.38 | 0.36 | 0.36 |
| GWM (0.01) | 2018 | 0.23 | 0.24 | 0.29 | 0.20 |
| GWM (0.01) | 2019 | 0.17 | 0.16 | 0.30 | 0.26 |
| GWM (0.05) | 2017 | 0.32 | 0.34 | 0.44 | 0.45 |
| GWM (0.05) | 2018 | 0.22 | 0.28 | 0.14 | 0.13 |
| GWM (0.05) | 2019 | 0.22 | 0.15 | 0.29 | 0.28 |
| GWM (0.1)  | 2017 | 0.37 | 0.39 | 0.42 | 0.45 |
| GWM (0.1)  | 2018 | 0.25 | 0.30 | 0.20 | 0.21 |
| GWM (0.1)  | 2019 | 0.23 | 0.16 | 0.34 | 0.30 |

At TP size of 400 individuals, as a general conclusion, the highest PA obtained for DON (0.43) and for DSK (0.49) was achieved predicting the 2017 population under scenario 1 (GS). With this TP size, PA for DON in 2017 varied from 0.34 to 0.43; the highest PA was obtained with Scenario 1 (GS) and the lowest PA with Scenario 3 (GWM-0.05). For DSK, the PA ranged from 0.36 to 0.49, with the highest PA obtained with Scenario 1 (GS) and the lowest with Scenario 2 (GWM-0.01). With 2018 population set, PA for DON varied from 0.24 to 0.34, being the highest PA obtained with scenario 1 (GS-full set) and the lowest with scenario 2 (GWM-0.01). For DSK, the PA ranged from 0.13 to 0.25, with the highest PA obtained with scenario 1 (GS) and the lowest with scenario 3 (GWM-0.05). With respect to DON, PA in the 2019 population varied from 0.15 to 0.16. For DSK, the PA ranged from 0.24 to 0.30 with the highest PA obtained with scenario 4 and the lowest PA obtained with scenario 1 (GS-full set).



**Figure 1.** Prediction Accuracy (PA) for DON (A) and DSK index (B) under different marker sets for genomic selection. Average between 2017-18-19. GS: full marker set (~20,929 SNPs), GWM (0.01), GWM (0.05) and GWM (0.01), marker subsets based on different significance level.

Figure 1A,B showed PA for both traits and four marker scenarios, where PA is the average of the three populations (2017, 2018 and 2019). As a general conclusion, under scenario 4 (GWM-0.1) we observed the highest PA for both traits at TP size = 100 and under scenario 1 (GS) at TP size = 400. At TP = 100 (Figure 1A), scenario 4 (GWM-0.1) increased PA for DON (0.28) by 21% compared to scenario 1 (GS full set) where we obtained the lowest PA (0.22). Similarly, for DSK (Figure 1B), scenario 4 (GWM-0.1) increased the PA by 12% compared to scenario 1, that obtained the lowest PA. At TP = 400 (Figure 1A), for DON, scenario 1 (GS) reached an average PA of 0.21, 10% higher than PA obtained with scenario 4 (GWM-0.1) the second highest value. For DSK, scenario 1 reached an average PA of 0.33 that is a 3% higher than PA obtained with scenario 4 (GWM-0.1), the second highest value. We observed for DON, that the difference between the size of the TP did not have an impact on the average PA under the marker

subsets, but it did have an impact with the full set of markers where PA at TP = 400 outperformed by a 21% PA at TP size = 100. We observed a similar trend for DSK with scenarios 3 and 4, but we observed an impact of TP size with scenarios 1 and 2. When looking at scenarios 2–4, where SNPs were selected based on GWAS  $P$  values on the regional nurseries only, scenario 4 (GWM-0.1), showed the highest PA compared to scenario 2 and 3, for both DON and DSK and at both TP sizes.

#### *Effect of the different marker set scenarios on prediction of FHB rating and FDK*

We also investigated the effect of genomic selection under these marker scenarios when predicting other scab traits like FHB rating and FDK (Table 6). These traits are known to be correlated with DON, and along with DON, they constitute the DSK index. These traits were predicted based on the markers selected with GWAS for DON and DSK; evaluation was carried out under different marker scenarios, the effects of markers coming from GWAS for DON or DSK, and under two TP sizes. The results showed that PA for FHB rating was, on average, lower (0.2) than PA obtained for DON (0.28) or DSK (0.31) and PA for FDK was of a similar magnitude (0.31) when compared to DON (0.28) and DSK (0.31).

**Table 6.** Average prediction accuracy for FHB rating and FDK under different marker set scenarios, 2017–19. GS whole genotype marker data (~ 20,929 SNPs), GWM (0.01) marker subset based on significance level  $P < 0.01$  (234 and 176 SNPs), GWM (0.05) marker subset based on significance level  $P < 0.05$  (899 and 831 SNPs), GWM (0.1) marker subset based on significance level  $P < 0.1$  (1780 and 1756 SNPs) and GWM (corr) marker subset based on correlation. Training population sizes = 100, 400.

| Trait      | N     | FHB Rating |      |      | FDK  |      |      |
|------------|-------|------------|------|------|------|------|------|
|            |       | 100        | 400  | AV   | 100  | 400  | AV   |
| <b>DON</b> |       |            |      |      |      |      |      |
| GS         | 20929 | 0.15       | 0.15 | 0.15 | 0.29 | 0.35 | 0.32 |
| GWM (0.01) | 234   | 0.18       | 0.22 | 0.20 | 0.29 | 0.25 | 0.27 |
| GWM (0.05) | 899   | 0.17       | 0.19 | 0.18 | 0.30 | 0.33 | 0.32 |
| GWM (0.1)  | 1780  | 0.15       | 0.18 | 0.16 | 0.30 | 0.36 | 0.33 |
| <b>DSK</b> |       |            |      |      |      |      |      |
| GS         | 20929 | 0.15       | 0.15 | 0.15 | 0.29 | 0.35 | 0.32 |
| GWM (0.01) | 176   | 0.17       | 0.16 | 0.17 | 0.34 | 0.26 | 0.30 |
| GWM (0.05) | 831   | 0.20       | 0.20 | 0.20 | 0.30 | 0.29 | 0.30 |
| GWM (0.1)  | 1756  | 0.20       | 0.21 | 0.21 | 0.34 | 0.32 | 0.33 |

N: SNPs number. FHB rating (0–9), FDK: Fusarium damaged kernels, (%).

For FHB rating, predictions based on GWAS for DON or DSK gave similar accuracies (0.18 and 0.19; Table 6). The highest prediction accuracy

for this trait was found with Scenario 2 (GWM-0.01) at TP = 400, PA = 0.22. At TP = 400, PA for FHB rating under all reduced marker sets, was 6 to 32% higher than scenario 1 (GS-full set). At TP = 100 the results were similar; highest PA was obtained with scenario 3 and 4 (0.2), the PA obtained under these scenarios was 12–25% higher than scenario 1, and the lowest PA obtained under scenario 1 (GS-full set). For FDK, Table 6 shows that predictions based on GWAS for DON and DSK had similar accuracies (0.31–0.32) and the scenario that showed the best results was scenario 4 (GWM-0.1), PA = 0.33 as an average between two TP sizes. Scenario 4 outperformed scenario 1 (GS-full set) by a 3% and scenarios 2 and 3 by a 3 to 18%. The highest prediction accuracy for this trait was found with scenario 4 (GWM-0.1) at TP = 400, PA = 0.36. The lowest PA was obtained with scenario 2 (GWM-0.01), at both TP sizes and with markers selected based on GWAS for DON and DSK.

## DISCUSSION

Genomic selection has become a primary technology for plant breeders looking to accelerate the breeding process. Some of the benefits of GS include increasing genetic gain per unit time, reducing phenotyping costs, reducing field testing and more accurate selection of parents for crosses. In this study we established several scenarios with different marker subsets based on level of significance obtained with GWAS in the regional nurseries, and afterwards evaluated them with both cross validation and forward GS, predicting independent sets of UK breeding lines.

Overall, results from our study showed positive and promising results regarding the use of a subset of markers based on GWAS. We established trait specific genomic relationship matrices, and defined different marker sets that include specific SNPs that were significant for DON or DSK. It is, to our knowledge, the first study reporting positive results with GWAS-GS for DON and DSK under a forward GS scheme, where a set of regional lines with known DON values becomes the TP to calculate GEBVs for UK breeding lines that don't have FHB phenotyping evaluation yet. It is known that FHB is a very complex disease and the traits evaluated to quantify disease resistance are explained by many genes with small effects. Therefore, GWAS can be used to identify trait-marker associations in order to improve and validate GS; in addition, this step reduces the number of markers used for the analysis significantly, with the cost reduction than implies. Alternatively, it may be possible to reduce costs using a targeted genotyping approach such as amplicon sequencing [54].

Under a forward selection approach, our results using regional nurseries as TPs (Figure 1A,B) over three years, showed a substantial positive increase in PA of 21% for DON (0.28 vs 0.22) and 12% for DSK (0.32 vs 0.28) under scenario 4 compared to scenario 1 at the smallest TP size. On the other hand, with the largest TP, the highest accuracies were obtained with the full marker set for both traits, being 10% and 3% higher for DON and DSK, respectively than scenario 4, the second best scenario.

Based on these results, we conclude that the association of GWAS and GS is a successful strategy that allows one to reduce the marker set with minimal effect on PA but with a great impact on the marker number; our results showed an average 93.6% reduction in marker's number (20,932 vs 1900) and only an average 6% reduction in PA for both traits (TP = 400). We validated in a forward GS scheme that significant SNPs for regional nurseries were also significant SNPs for the UK material, showing association with QTLs that are mainly responsible for the resistance or susceptibility of lines. Rutkoski et al. [30] suggested that fewer loci were involved in DON resistance compared to other FHB traits like Severity, Incidence and FDK; our results agree with this concept as we observed that a reduced number (~ 1800) of SNPs were enough to estimate GEBVs in an accurate and consistent way for DON accumulation and DSK index.

In studies applying a forward GS scheme with independent samples of related material [27,38], investigators found prediction accuracies for DON in barley ranging from 0.14 to 0.67 and for FHB ranging from 0.58 to 0.77. In wheat, using an independent sample for TP and VP, Jiang et al. [22] found prediction accuracies of 0.58 for FHB rating using a TP and VP evaluated in different years, for sets of European wheat populations. In another study, Schulthess et al. [37] found prediction accuracies ranging from 0.4 (lower relatedness between TP and VP) to 0.8 (higher relatedness between TP and VP) when predicting severity in hybrid wheat. In our study, the highest PA obtained, with scenario 4, for DON (0.39) and DSK (0.45) was found when predicting the 2017 population. Lower prediction accuracies were observed with the 2018 and 2019 populations.

Year to year variability is a normal phenomenon in a breeding program: every year new families are evaluated, and the environmental conditions are unpredictable. In our results, GxE interaction was manifest in the differences among 2017 (high PA), 2018 and 2019 (moderate-low PA). In a forward GS scheme, prediction accuracies are affected by degrees of relatedness between TP and VP, and the year effect, when TP and VP are evaluated in different years. Therefore, we see our results as promising in that we are using regional nurseries composed of lines from different breeding programs to predict UK breeding lines; nursery entries may not be closely related to the UK material, adding more complexity to GS. Something else to consider is the phenotypic expression showed in the three different sets of validating populations for DON. The expression of DON content varied greatly among UK populations (Table 3). In 2017 and 2018, a range was observed of 40.3 and 33.25 ppm between the minimum and the maximum values, respectively, whereas in 2019 we observed little variability with 4.96 ppm between the lowest and highest values. This year-to-year difference in phenotypic expression for the different lines affected the PA.

In order to place the PAs obtained with forward GS in context, we can look at previous studies from our group [40,55]. Verges et al. [40] found that with a 0.40–0.43 PA for DON, using a 30–40% selection intensity,

common in early generation selection, 50 to 60% of the lines were correctly selected for low DON based on GEBVs. We also found that with a 0.49 PA for DSK, up to 68% of lines were correctly selected for low DON, using the same selection intensity. Therefore, we think that a prediction accuracy of 0.4 would be acceptable to most breeders when selecting for DON resistance in lines not yet tested in a scab nursery.

Under cross validation (Table 4), our results agreed with the values found in the literature, for DON and DSK, and with consistency among all marker subsets. The GS model trained with the full marker sets showed the highest PA for DON (0.55). In scenario 3 with a reduced markers set, using SNPs significant at  $P < 0.05$ , we observed the highest PA for DON (0.54) compared to the 0.55 observed with the full set marker set. Different authors investigated this trait in wheat [30,31,35,40] and barley [27], reporting moderate PA for DON with cross validation, less than or equal to 0.6 on average. Therefore, our results provide strong evidence about the model's predictive ability with a 95% reduction of marker number used for this trait through building trait-specific genomic relationship matrices that exploit GWAS via rrBLUP [42].

The DSK index was proposed by Verges et al. [40] with the objective of weighting the values of FDK and DON, traits that affect grain quality, food safety and economic return to the farmer. Under cross validation (Table 4), this trait showed a moderate PA ranging from 0.49–0.57 for scenarios 2–4. The highest PA with the marker subsets was obtained under scenario 4 (PA = 0.57), SNPs selected with GWAS for DSK with a  $P < 0.1$  level of significance. This value represents a slight reduction in accuracy of 5% compared to scenario 1 (PA = 0.6). DSK is a novel index and therefore difficult to compare with other indices from the literature, but in another study Arruda et al. [31] evaluated two different indices (FHB index and ISK), finding prediction accuracies with cross validation of around 0.5 for FHB index and 0.7 for ISK. Rutkoski et al. [30] found prediction accuracies ranging from 0.44 to 0.54 for the same index.

We also investigated whether SNPs selected based on significant marker-trait associations for DON and DSK would be effective in estimating accurate GEBVs for FDK and FHB rating, two very important traits evaluated when breeding for FHB resistance. Our results (Table 6) showed for FDK, a similar average PA (0.3) to DON or DSK, based on all scenarios and the three different populations to which a forward GS scheme was applied. The prediction accuracies ranged from 0.25 to 0.35. As an average of two TP sizes, scenario 4 showed the highest PA for FDK (PA = 0.33) 5% higher than scenario 1. These results show consistency in PA and confirm the usefulness of GWAS identifying significant SNPs to target different scab traits when breeding to increase FHB resistance. Arruda et al. [31] found a PA of 0.8 for FDK; Rutkoski et al. [30] found prediction accuracies ranging from 0.35 to 0.46 and Larkin et al. [46] found PA of 0.53 all under a cross validation scheme. Even though the PA we obtained is lower, we should underscore that our results are based on a

forward GS scheme, where prediction accuracy is generally lower than is found with cross validation. Two aspects are important to address here: (1) the similar PA obtained for FDK is an average based on the three UK populations and all scenarios; (2) Scenario 4, with 1756–1780 SNPs outperformed scenario 1, which strongly supports the use of GWAS-GS approach in a forward GS strategy.

TP size has been extensively discussed in the literature, and there is agreement that highest PAs are achieved with 300–400 individuals and that at larger TPs, a plateau is achieved [27,56–59]. Our results, under a forward GS scheme, showed small differences between PA obtained with TP size 100 vs 400 using reduced marker numbers (scenarios 2–4). PA for DON, as an average of all scenarios, was 0.27 at TP size of 100 and 400. PA for DSK reached 0.31 at TP size of 100, and 0.29 at TP size of 400. With the full marker set, the results differed in that both traits had higher PA at TP = 400 compared to TP = 100.

In a simulation study, Hickey et al. [41] suggested that for related biparental populations, 300–500 SNPs are enough to get prediction accuracies of 0.6, with training populations ranging from 400–800 individuals. Numerous investigators have evaluated the effect of marker number on the PA for FHB traits and they observed a similar trend, where increases in marker number increase PA until a plateau is reached, in some studies sooner, with 250 to 380 markers [27,60,61] or later, at around 3000 SNPs [31]. Our results, under a forward GS approach, indicate that with 1700–1800 SNPs selected via GWAS, it is possible to obtain PA of 0.4, when TP and VP are independent sets of lines.

With this study we tried to improve our understanding of how GS and GWAS could improve breeding for a challenging disease like FHB, and very challenging and costly traits like DON accumulation. In a recent article, we stated that selections based on GEBVs could be done effectively in material that was not yet evaluated for FHB in the field [40]. Lines in earlier generations could be selected for resistance based on GEBVs, eliminating very susceptible material before testing it in the field. The results of this study reinforce this idea, given the usefulness of the regional nurseries to predict FHB traits, coupled with the use of GWAS for identifying a smaller number marker significantly associated with traits of interest. Reduced marker number decreases genotyping costs considerably, which is always good news for breeders.

#### **SUPPLEMENTARY MATERIALS**

The following supplementary materials are available online: <https://doi.org/10.20900/cbgg.20210007>. Supplementary Table S1: Uniform Northern and Uniform Southern Scab Nursery (NUS-SUS) for 2014–2018; Supplementary Figure S1: Manhattan and QQ plots from GWAS for DON content and DSK for the regional FHB nurseries and the three Kentucky populations.

### **AUTHOR CONTRIBUTIONS**

VLV designed the study, collected the data, did the data analysis and wrote the manuscript. GBG facilitated the genotyping by sequencing and provided input on marker set size. DVS provided the breeding material, obtained the grant funding, collaborated with VV on design of the study, and edited the manuscript.

### **CONFLICTS OF INTEREST**

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analysis or interpretation of data; in the writing of the manuscript and in the decision to publish the results.

### **FUNDING**

This work was funded by a grant from the US. Department of Agriculture, through the US Wheat and Barley Scab Initiative under agreement no. 59-0206-9-054.

### **ACKNOWLEDGMENTS**

We thank John Connelley and Sandy Swanson for their technical support.

### **REFERENCES**

1. Parry DW, Jenkinson P, McLeod L. Fusarium ear blight (scab) in small grain cereals—a review. *Plant Pathol.* 1995;44(2):207-38.
2. Dexter JE, Clear RM, Preston KR. Fusarium head blight: effect on the milling and baking of some Canadian wheats. *Cereal Chem.* 1996;73(6):695.
3. Bai G, Shaner G. Management and Resistance in Wheat and Barley To Fusarium Head Blight. *Annu Rev Phytopathol.* 2004;42(1):135-61.
4. Liu S, Abate ZA, McKendry AL. Inheritance of Fusarium head blight resistance in the soft red winter wheat Ernie. *Theor Appl Genet.* 2005;110(3):454-61.
5. Poland J, Rutkoski J. Advances and Challenges in Genomic Selection for Disease Resistance. *Annu Rev Phytopathol.* 2016;54(1):79-98.
6. Carmack WJ, Clark AJ, Dong Y, Van Sanford DA. Mass selection for reduced deoxynivalenol concentration using an optical sorter in SRW wheat. *Agronomy.* 2019;9(12):816.
7. Carmack WJ, Clark AJ, Dong Y, Brown-Guedira G, Van Sanford DA. Optical Sorter-Based Selection Effectively Identifies Soft Red Winter Wheat Breeding Lines With Fhb1 and Enhances FHB Resistance in Lines With and Without Fhb1. *Front Plant Sci.* 2020;11:1318.
8. Miedaner T, Wilde F, Steiner B, Buerstmayr H, Korzun V, Ebmeyer E. Stacking quantitative trait loci (QTL) for Fusarium head blight resistance from non-adapted sources in an European elite spring wheat background and assessing their effects on deoxynivalenol (DON) content and disease severity. *Theor Appl Genet.* 2006;112(3):562-69.

9. Anderson JA, Chao S, Liu S. Molecular Breeding Using a Major QTL for Fusarium Head Blight Resistance in Wheat. *Crop Sci.* 2007; 47(53):S-112-9.
10. Buerstmayr H, Ban T, Anderson JA. QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. *Plant Breed.* 2009;128(1):1-26.
11. Agostinelli AM, Clark AJ, Brown-Guedira G, Van Sanford DA. Optimizing phenotypic and genotypic selection for Fusarium head blight resistance in wheat. *Euphytica.* 2012;186(1):115-26.
12. Miedaner T, Korzun V. Marker-assisted selection for disease resistance in wheat and barley breeding. *Phytopathology.* 2012;102(6):560-6.
13. Balut AL, Clark AJ, Brown-Guedira G, Souza E, Van Sanford DA. Validation of Fhb1 and QFhs.nau-2DL in Several Soft Red Winter Wheat Populations. *Crop Sci.* 2013;53(3):934-45.
14. Liu S, Griffey CA, Hall MD, McKendry AL, Chen J, Brooks WS, et al. Molecular characterization of field resistance to Fusarium head blight in two US soft red winter wheat cultivars. *Theor Appl Genet.* 2013;126(10):2485-98.
15. Lemes da Silva C, Fritz A, Clinesmith M, Poland J, Dowell F, Peiris K. QTL mapping of Fusarium head blight resistance and deoxynivalenol accumulation in the Kansas wheat variety 'Everest'. *Mol Breed.* 2019;39(3):35.
16. Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. *Crop Sci.* 2009;49(1):1-12.
17. 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;22(11):961-75.
18. Cai J, Wang S, Li T, Zhang G, Bai G. Multiple minor QTLs are responsible for Fusarium head blight resistance in Chinese wheat landrace Haiyanzhong. *PLoS One.* 2016;11(9):1-15.
19. Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, et al. Whole Genome Association Mapping of Fusarium Head Blight Resistance in European Winter Wheat (*Triticum aestivum* L.). *PLoS One.* 2013;8(2):1-10.
20. Arruda MP, Brown P, Brown-Guedira G, Krill AM, Thurber C, Merrill KR, et al. Genome-wide association mapping of Fusarium head blight resistance in wheat using genotyping-by-sequencing. *Plant Genome.* 2016;9(1). doi: 10.3835/plantgenome2015.04.0028
21. Hoffstetter A, Cabrera A, Sneller C. Identifying quantitative trait loci for economic traits in an elite soft red winter wheat population. *Crop Sci.* 2016;56(2):547-58.
22. 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;130(3):471-82.
23. Wang R, Chen J, Anderson JA, Zhang J, Zhao W, Wheeler J, et al. Genome-wide association mapping of Fusarium head blight resistance in spring wheat lines developed in the Pacific Northwest and CIMMYT. *Phytopathology.* 2017;107(12):1486-95.

24. Tessmann EW, Dong Y, Van Sanford DA. GWAS for fusarium head blight traits in a soft red winter wheat mapping panel. *Crop Sci.* 2019;59(5):1823-37.
25. Hu W, Gao D, Wu H, Liu J, Zhang C, Wang J, et al. Genome-wide association mapping revealed syntenic loci QFhb-4AL and QFhb-5DL for Fusarium head blight resistance in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 2020;20(1):1-13.
26. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157(4):1819-29.
27. Lorenz AJ, Smith KP, Jannink JL. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Sci.* 2012;52(4):1609-21.
28. Rutkoski JE, Poland JA, Singh RP, Huerta-Espino J, Bhavani S, Barbier H, et al. Genomic selection for quantitative adult plant stem rust resistance in wheat. *Plant Genome.* 2014;7(3):1-10.
29. Arruda MP, Lipka AE, Brown PJ, Krill AM, Thurber C, Brown-Guedira G, et al. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol Breed.* 2016;36(7):1-11.
30. Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink JL, Sorrells M. Evaluation of genomic prediction methods for fusarium head blight resistance in wheat. *Plant Genome.* 2012;5(2):51-61.
31. Arruda MP, Brown PJ, Lipka AE, Krill AM, Thurber C, Kolb FL. Genomic selection for predicting fusarium head blight resistance in a wheat breeding program. *Plant Genome.* 2015;8(3):1-12.
32. Jiang Y, Zhao Y, Rodemann B, Plieske J, Kollers S, Korzun V, et al. Potential and limits to unravel the genetic architecture and predict the variation of Fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.). *Heredity.* 2015;114(3):318-26.
33. 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;128(12):2471-81.
34. Hoffstetter A, Cabrera A, Huang M, Sneller C. Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat. *G3.* 2016;6(9):2919-28.
35. Dong H, Wang R, Yuan Y, Anderson J, Pumphrey M, Zhang Z, et al. Evaluation of the potential for genomic selection to improve spring wheat resistance to fusarium head blight in the pacific northwest. *Front Plant Sci.* 2018;9:1-15.
36. Sallam AH, Smith KP. Genomic selection performs similarly to phenotypic selection in barley. *Crop Sci.* 2016;56(6):2871-81.
37. 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;131(3):685-701.
38. Tiede T, Smith KP. Evaluation and retrospective optimization of genomic selection for yield and disease resistance in spring barley. *Mol Breed.* 2018;38(5):55.

39. Herter CP, Ebmeyer E, Kollers S, Korzun V, Würschum T, Miedaner T. Accuracy of within- and among-family genomic prediction for Fusarium head blight and Septoria tritici blotch in winter wheat. *Theor Appl Genet.* 2019;132(4):1121-35.
40. Verges VL, Lyerly J, Dong Y, Van Sanford DA. Training Population Design With the Use of Regional Fusarium Head Blight Nurseries to Predict Independent Breeding Lines for FHB Traits. *Front Plant Sci.* 2020;11:1-13.
41. Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, et al. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* 2014;54(4):1476-88.
42. Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J, et al. Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS One.* 2014;9(3):1-12.
43. Spindel JE, Begum H, Akdemir D, Collard B, Redoña E, Jannink JL, et al. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity.* 2016;116(4):395-408.
44. Bian Y, Holland JB. Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Heredity.* 2017;118(6):585-93.
45. Lozada DN, Mason RE, Sarinelli JM, Brown-Guedira G. Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genet.* 2019;20(1):1-12.
46. Larkin DL, Holder AL, Mason RE, Moon DE, Brown-Guedira G, Price PP, et al. Genome-wide analysis and prediction of Fusarium head blight resistance in soft red winter wheat. *Crop Sci.* 2020;60(6):1-19.
47. Bec S, Ward T, Farman M, O'Donnell K, Hershman D, Van Sanford D, et al. Characterization of fusarium strains recovered from wheat with symptoms of head blight in Kentucky. *Plant Dis.* 2015;99(11):1622-32.
48. Mirocha CJ, Kolaczowski E, Xie W, Yu H, Jelen H. Analysis of Deoxynivalenol and Its Derivatives (Batch and Single Kernel) Using Gas Chromatography/Mass Spectrometry. *J Agric food Chem.* 1998;46(4):1414-8.
49. Dong Y, Steffenson BJ, Mirocha CJ. Analysis of Ergosterol in Single Kernel and Ground Grain by Gas Chromatography–Mass Spectrometry. *J Agric Food Chem.* 2006;54(12):4121-25.
50. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 2011;6(5):1-10.
51. 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.
52. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: Genome association and prediction integrated tool. *Bioinformatics.* 2012;28(18):2397-9.
53. Endelman JB, Jannink JL. Shrinkage estimation of the realized relationship matrix. *G3.* 2012;2(11):1405-13.

54. Bernardo A, St. Amand P, Le HQ, Su Z, Bai G. Multiplex restriction amplicon sequencing: a novel next-generation sequencing-based marker platform for high-throughput genotyping. *Plant Biotechnol. J.* 2020;18(1):254-65.
55. Verges VL, Van Sanford DA. Genomic selection at preliminary yield trial stage: Training population design to predict untested lines. *Agronomy.* 2020;10(1):1-16.
56. Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L, et al. Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics.* 2014;198(1):3-16.
57. Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME. Training set optimization under population structure in genomic selection. *Theor Appl Genet.* 2015;128(1):145-58.
58. Michel S, Ametz C, Gungor H, Akgöl B, Epure D, Grausgruber H, et al. Genomic assisted selection for enhancing line breeding: merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials. *Theor Appl Genet.* 2017;130(2):363-76.
59. Sarinelli JM, Murphy JP, Tyagi P, Holland JB, Johnson JW, Mergoum M, et al. Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theor Appl Genet.* 2019;132(4):1247-61.
60. Heffner EL, Jannink JL, Iwata H, Souza E, Sorrells ME. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci.* 2011;51(6):2597-606.
61. Sallam AH, Endelman JB, Jannink JL, Smith KP. Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *Plant Genome.* 2015;8(1). doi: 10.3835/plantgenome2014.05.0020

How to cite this article:

Verges VL, Brown-Guedira GL, Van Sanford DA. Genome-Wide Association Studies Combined with Genomic Selection as a Tool to Increase Fusarium Head Blight Resistance in Wheat. *Crop Breed Genet Genom.* 2021;3(4):e210007. <https://doi.org/10.20900/cbgg20210007>