Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

GENOME-WIDE ASSOCIATION MAPPING OF FUSARIUM HEAD BLIGHT RESISTANCE IN WHEAT (*Triticum aestivum* L.) USING GENOTYPING-BY-SEQUENCING

Marcio P. Arruda, Patrick Brown, Gina Brown-Guedira, Allison M. Krill, Carrie Thurber, Keith R. Merrill, Bradley J. Foresman, and Frederic L Kolb*

M.P. Arruda, P. Brown, A.M. Krill, B.J. Foresman, and F.L. Kolb, Dep. of Crop Sciences, Univ. of Illinois, 1102 S. Goodwin Ave., Urbana, IL 61801; G. Brown-Guedira, USDA-ARS Plant Science Research Unit, 4114 Williams Hall, Raleigh, NC 27695; C. Thurber, School of Science and Mathematics, Abraham Baldwin Agricultural College, 2802 Moore Highway, Tifton, GA 31793; K.R. Merrill, Department of Crop Science, North Carolina State University, Raleigh, NC 27695. Received _____. *Corresponding author (f-kolb@uiuc.edu).

Abreviations:

cMLM = compressed mixed linear model

DON = deoxynivalenol

FDK = Fusarium diseased kernels

FHB = Fusarium head blight

FHBdx = Fusarium head blight index

GBS = Genotyping-by-sequencing

GWAS = Genome-wide association study

INC = Incidence

ISK = Incidence, severity, and kernel quality index

PCA = Principal component analysis

QTL = Quantitative trait loci

SEV = Severity

SNP = Single nucleotide polymorphism

SSR = Simple sequence repeat

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Abstract

Fusarium head blight (FHB) is one of the most important wheat diseases worldwide and host resistance displays complex genetic control. A genome-wide association study (GWAS) was performed on 273 winter wheat breeding lines from the midwestern and eastern regions of the United States to identify chromosomal regions associated with FHB resistance. Genotyping-by-sequencing (GBS) was used to identify 19,992 single nucleotide polymorphisms (SNPs), covering all 21 wheat chromosomes. Marker-trait associations were performed with different statistical models, the most appropriate being a compressed mixed linear model (cMLM) controlling for relatedness and population structure. Ten significant SNP-trait associations were detected on chromosomes 4A, 6A, 7A, 1D, 4D, 7D, and multiple SNPs were associated with *Fhb1* on chromosome 3B. Although combination of favorable alleles of these SNPs resulted in lower levels of severity (SEV), incidence (INC), and deoxynivalenol (DON) concentration, lines carrying multiple beneficial alleles were in very low frequency for most traits.These SNPs can now be used for creating new breeding lines with different combinations of favorable alleles. This is one of the first GWAS using genomic resources from the International Wheat Genome Sequencing Consortium.

Introduction

Fusarium head blight (FHB), also known as head scab or ear blight, is a destructive disease of wheat (*Triticum aestivum* L.), and is considered the most important plant disease to hit the United States since the stem rust epidemics of the 1950's (Wood et al., 1999). In North America it is primarily caused by *Fusarium graminearum* Schwabe [telemorph: *Gibberella zeae* Schw. (Petch)]. In addition to wheat, the pathogen causes disease on barley, oat, rye, and corn.

The economic losses associated with FHB are due to grain yield reduction, decrease of grain quality and mycotoxin-contaminated grain. Decreased grain quality results from low test weight and damaged kernels. Some damaged kernels are also frequently lost during harvest due to low grain weight (Adams, 2010). In addition, *F. graminearum* produces potent animal toxins such as trichothecenes and estrogenic metabolites. Deoxynivalenol (DON), the most common trichothecene in FHB infected wheat, is especially harmful to monogastric animals, causing vomiting, diarrhea, nausea and feed refusal by swine (Bennet and Klich, 2003). In humans, *F. graminearum* mycotoxins have been linked with alimentary disorders such as Akakabi toxicosis, which is characterized by vomiting, anorexia and convulsions (Bennett and Klich, 2003). Lastly, trichotecenes can survive the production processes employed by the food industry (Hazel and Patel, 2004), imposing technical challenges.

FHB resistant wheat cultivars play an important role in FHB management and prevention of mycotoxin contamination. Fortunately, sources of genetic resistance are available in the wheat gene pool (Buerstmayr et al. 2009). Resistance to *F. graminearum* in small grains was first classified by Schroeder and Christensen (1963) into two major components: resistance to initial infection by the pathogen (type I) and resistance to fungal spread along the rachis (type II). In wheat lines both types of resistance may be present individually or in combination. In addition,

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

other physiological resistances have been described such as resistance to toxin accumulation (type III), resistance to kernel infection (type IV) and tolerance (type V) (Mesterházy, 1995; Mesterházy, 1999).

During the last two decades, many QTL mapping studies have been conducted for multiple types of FHB resistance in wheat. The great majority of these studies were performed with bi-parental populations, some of them with limited population size. Buerstmayr et al. (2009) and Liu et al. (2009) both reviewed these QTL in an attempt to find stable and useful QTL for breeding purposes. Some chromosomal regions, such as the short arm of chromosome 3B, were shown to harbor stable QTL. While linkage mapping in bi-parental population has been successful in detecting QTL for FHB resistance, the strategy has drawbacks including the large amount of time and/or resources needed for developing RIL or DH populations. In addition, limited genetic variation is explored. In fact, only the alleles differing in the parents are considered, resulting in low mapping resolution in most cases. Once mapping is done in the bi-parental population, it is then necessary to validate the results in different breeding populations. In this context, genome-wide association studies (GWAS) can be seen as a promising strategy for identifying QTL for traits of interest. First developed for human genetics, GWAS exploit the recombination events present in a group of unrelated individuals, usually resulting in higher mapping resolution. A key strategy in genome-wide association is to have enough genome coverage so that functional alleles will be in linkage disequilibrium (LD) with at least one marker (Myles et al., 2009). To date, a few GWAS have been conducted for FHB resistance traits in wheat, but marker availability has substantially improved recently, suggesting that newer studies having improved genome coverage are warranted. For instance, Miedaner et al. (2011) applied 115 single sequence repeat (SSR) markers in a GWAS involving European breeding

lines. Kollers et al. (2013), also working with European germplasm, used 732 SSRs to detect QTL for FHB resistance. Ghavami et al. (2011) used 2300 diversity array technology (DArT) markers in a durum wheat (*Triticum durum* L. var. *durum* Desf) association study. Although DArT markers are more abundant than SSRs, the authors found some chromosomes to be poorly covered. More recently, Gurung et al. (2014) used 4781 single nucleotide polymorphisms (SNPs) to identify QTL associated with five diseases in wheat: *Stagonospora nodorum* blotch, Tan spot (caused by *Pyrenophora tritici-repentis*) races 1 and 5, bacterial leaf streak (caused by *Xanthomonas translucens* pv. *Undulosa)*, spot blotch [caused by *Cochliobolus sativus* (Ito & Kuribayashi) Drechsler ex Dastur)], and *Septoria tritici* blotch. More recently, Jiang et al. (2015) genotyped 372 European wheat varieties using both the Illumina iSelect assays for wheat having 9,000 SNPs (Cavanagh et al., 2012) and 90,000 SNPs (Wang et al., 2014), in addition to 732 SSR markers. Jiang et al. (2015) detected multiple significant marker-trait associations, which where then used for predicting phenotypes of individual breeding lines. Genotyping-by-sequencing (GBS) is a high throughput method for SNP genotyping that combines genome complexity reduction with restriction enzymes and second generation sequencing technology (Elshire et al., 2011). Methylation-sensitive restriction enzymes can be used to target euchromatic, gene-rich regions, and hundreds of samples can be pooled and processed together following ligation with unique barcodes. The resulting pooled libraries are then polymerase chain reaction (PCR) amplified and Illumina sequenced. Poland et al. (2012a) modified the original protocol to accommodate a combination of enzymes. The striking advantages of GBS over other platforms are: i) identification of abundant SNPs at low cost; ii) simultaneously discovery and genotyping; iii) reduced ascertainment bias when compared to array based markers; iv) relatively easy automation (Elshire et al., 2011; Poland and Rife., 2012).

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Poland et al. (2012b) applied GBS in a wheat panel consisting of 254 lines from CIMMYT and were able to identify 41,371 SNPs. These SNPs were then used for genomic selection for grain yield, thousand kernel weight, and heading date. Rutkoski et al. (2013) used GBS to identify 130,000 GBS polymorphisms from 360 elite spring lines belonging to CIMMYT. Gemomic selection analyses were performed for quantitative resistance to stem rust, caused by *Puccinia graminis* f.sp. *tritici*. The number of SNPs was reduced to 2014 after applying a per-marker percent missing data threshold of 20%. Both studies successfully applied the GBS-SNP markers for genomic selection (GS). Langer et al. (2014) used GBS to identify 23,371 SNPs in a collection of 410 European winter wheat lines in a GWAS for flowering time.

In this study GBS-SNP markers were identified from a wheat panel of elite lines composed for genome-wide marker-trait association purposes. SNPs associated with resistance can accelerate the breeding process through marker assisted selection (MAS) or can be incorporated into GS strategies. In addition, significant SNPs can give insights into the biological function of the polymorphism and how it relates to resistance. The objective of this study was to establish marker-trait associations for the different types of resistance to FHB using GBS-SNP markers.

Materials and methods

Plant material and disease assessment

The germplasm used in this study and the disease assessment are described in Arruda et al. (2015). In short, a total of 273 breeding lines were evaluated, with 185 lines belonging to the University of Illinois soft red winter wheat breeding program, and the remaining lines selected from 17 different land grant universities and private companies across the midwestern and

eastern regions of the United States. Phenotypic data were obtained from multiple experiments conducted in 2011, 2013, and 2014 in Urbana, Illinois. No symptoms were obtained in 2012 due to extreme drought conditions. Each year, the experiment was set up as a complete randomized block design (RCBD) with two replications. Since not all lines were present in 2011, the experiment was analyzed as an unbalanced design. The experimental unit consisted of a 1-meter long single-row plot. The field was inoculated with grain spawn, which was prepared from autoclaved maize kernels and isolates collected throughout Illinois. The infested kernels were spread at a rate of approximately 287 $kg.ha^{-1}$, starting two to four weeks before anthesis. Mist irrigation was applied three times per 24 hour period prior to, during, and after anthesis. FHB resistance was assessed by measuring the following parameters: disease incidence (INC), severity (SEV), FHB index (FHBdx), Fusarium-damaged kernels (FDK), incidence-severity-kernel index (ISK), and deoxynivalenol concentration (DON). INC and SEV are referred to as type I and type II resistances (Schroeder and Christensen, 1963), and are used to quantify resistance to penetration and spread of the pathogen, respectively. INC and SEV were measured approximately 21 days after anthesis. INC is measured as a visual estimate of the percentage of infected heads from a sample of 20 heads per plot. SEV was an estimate of infected spikelets in an infected head. FHBdx is calculated as (INC x SEV / 100). FDK is referred to as type III resistance (Mesterhazy 1995, Merterhazy et al., 1999), and is a visual estimate of the percentage of Fusarium-damaged kernels. For each breeding line, one sample of kernels was taken and compared against a set of known FDK standards. The ISK index is often used for making selections in the University of Illinois breeding program, and is calculated as (0.3 x INC + 0.3 x SEV + 0.4 x FDK). Lastly, DON was quantified by gas chromatography-mass spectrometry at the Department of Plant Pathology at the University of Minnesota. Resistance to mycotoxins has

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

been classified as type IV resistance (Miller et al., 1985), and DON is the most abundant mycotoxin in FHB damaged kernels in wheat.

Phenotypic data analysis

Best linear unbiased predictors (BLUPs) for each trait were calculated using a mixed model approach, as described by Arruda et al. (2015):

$$Y_{ijkl} = \mu + year_i + block(year)_{ij} + line_k + heading_{ijkl} + (year\ x\ line)_{ik} + \varepsilon_{ijkl} \quad \text{(i)}$$

where Y_{ijkl} is the observed phenotype, μ is the overall mean, $year_i$ is the random effect of the *i*th year, $block(year)_{ij}$ is the random effect of *j*th block within the *i*th year, $line_k$ is the random effect of the *k*th line, $heading_{ijk}$ is a quantitative covariate trait treated as fixed, consisting of the Julian day when the heading note was taken for the *l*th replicate of the *k*th line in the *j*th block within the *i*th year, $year\ x\ line_{ik}$ is the random effect of the interaction between the *i*th year and the *k*th line, and ε_{ijkl} is the random error term. The plot mean-based broad-sense heritability (H^2) was calculated for each trait using the variance components estimated from equation (i).

Genotypic data

DNA extraction was performed using a cetyltrimethyl ammonium bromide (CTAB)/chloroform protocol and then diluted to a concentration of 25ng*ul^{-1}. GBS libraries were prepared according to Poland et at. (2012b), with modifications. Three restriction-ligation reactions were performed on each genomic DNA sample, using three two-enzyme combinations for genome complexity reduction: *PstI*-HF-*MspI*, *PstI*-HF-*HinP1I*, and *PstI*-HF-*BfaI*. The enzyme *PstI*-HF (CTGCAG) is a rare cutter, whereas *MspI* (CCGG), *HinP1I* (GCGC), and *BfaI* (CTAG) are common cutters. We used these enzyme combinations hoping to have good coverage of the genome. Ninety six DNA samples were sequenced per lane using a set of 288 barcodes (96 samples x 3 enzyme

combination) with the *PstI* overhang, and a total of 288 DNA samples were sequenced across 3 lanes on the Illumina HiSeq2000 at the W. M. Keck Center for Comparative and Functional Genomics.

SNP calling was performed using the TASSEL 4 GBS pipeline (Glaubitz et al., 2013) by aligning reads against a pseudo-reference genome developed from the *T. aestivum* Chinese Spring chromosome survey sequence (here after referred to as the WCSS1 reference). The pseudo reference consisted of 41 molecules, one for each chromosome arm expect for chromosome 3B, which was a single molecule. Contigs longer than 200 bp originating from the chromosome arm specific libraries were downloaded at https://urgi.versailles.inra.fr/download/iwgsc/. Each pseudo molecule consisted of concatenated sequences for chromosome specific contigs with a string of 64 Ns inserted between contigs. The WCSS1 reference was indexed and alignment was done using the Burrows-Wheeler aligner (BWA) version 0.6.2. A text file having the start and end position of each contig within the pseudo molecule was created and used to identify SNP containing contigs and to determine location of SNPs on contigs. SNPs were named in the following format: IWGSC_CSS_ChromArm_ scaff _contig number_ SNP position within the contig. SNP in tags aligning to contigs having map positions from the POPSEQ map (IWGSC, 2014) were retained for analysis. Map positions are available at http://wheat-urgi.versailles.inra.fr/Seq-Repository/Publication. After obtaining their location, SNPs were excluded from the analysis if: i) per-marker missing data level > 50%; ii) minor allele frequency < 5%; iii) percent of heterozygotes > 10%. The three replicate libraries (three two-enzyme combinations) for each DNA sample were merged prior to imputation. Missing data were imputed using the expectation maximization (EM) imputation method as described by Rutkoski et al. (2013). The panel was

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

also genotyped with PCR-based markers reported to be associated with FHB resistance, as well as specific KASP assays for the reduced plant height genes *Rht-B1* and *Rht-D1* (Table 1).

Wheat is an allohexaploid species with 2n = 6X = 42 chromosomes distributed over sub-genomes A, B, and D. Since these sub-genomes share similarities, the same SNP may map to homoeologous chromosomes (1A, 1B, 1D) or even to non-homoeologous chromosomes (1A and 2B, for example). For these reasons, SNPs that mapped to multiple locations were excluded from the analysis. Lastly, we removed non-informative SNPs in order to reduce the number of markers and make the analysis more efficient. SNPs were eliminated based on LD ($r^2 > 0.8$) using the LD tagSNP selection option in JMP Genomics 7 (SAS Institute, 2015). This procedure uses an algorithm that keeps the maximally informative SNPs in the analysis, eliminating redundant information (Carlson et al., 2004).

Genome-wide association analysis

Principal component analysis (PCA) was first performed to access the level of genetic structure of the panel. The analysis was performed in JMP Genomics v.7, PCA for Population Stratification option (SAS Institute, 2015). Marker-trait associations were tested in the Genome Association and Prediction Integrated Tool – GAPIT (Lipka et al., 2012), using five linear models: 1) "naïve" model, with no control for population structure and relatedness, implemented in GAPIT by setting the parameters group.from and group.to equal to 1 and using no covariates. 2) Q model, similar to model 1 but using 4 principal components from the PCA as fixed effects (Zhao et al., 2007); 3) K model, with a variance-covariance matrix (K) between individuals treated as random, implemented in GAPIT by setting the parameters group.from and group.to equal to the number of taxa; 4) Mixed Linear Model "MLM", incorporating both the fixed effects from model 2 and the random effect from model 3 (Yu et al., 2006); 5) compressed MLM

(cMLM), in which individuals are clustered in groups and a reduced kinship matrix is used in the analysis (Zhang et al., 2010). The same number of principal components were used in models 2, 4, and 5.

The "naïve" and Q models can be expressed as:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e} \qquad \text{(iii)},$$

in which Y represent the phenotypes, $\boldsymbol{X}$ is the design matrix, $\boldsymbol{\beta}$ is a vector containing fixed effects, and $\boldsymbol{e}$ is the random error. In the "naïve" model $\boldsymbol{\beta}$ contains only the markers, whereas in the Q model it contains both the markers and the eigenvalues from the PCA. The other three models (K, MLM, and cMLM) can be expressed as:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} \qquad \text{(iv)},$$

in which the $\boldsymbol{Z}$ is a design matrix and $\boldsymbol{u}$ is a vector of random additive genetic effects. In the K model, $\boldsymbol{\beta}$ contains only markers and $\boldsymbol{u}$ contains the K matrix. In both MLM and compressed MLM, $\boldsymbol{\beta}$ has both markers and eigenvalues, and $\boldsymbol{u}$ has the K matrix. The significance of marker-trait associations was based on a false discovery rate (FDR)-adjusted P-value of 0.10. In GAPIT, the adjusted P-value follows the FDR-controlling procedure proposed by Benjamini and Hochberg (1995), the so called "BH procedure". The FDR is defined as the proportion of false positives among all discoveries, and the BH procedure controls the FDR at FDR $\leq \alpha$.

Results

Phenotypic data

Substantial phenotypic variation was observed for all traits, with SEV and FHBdx showing the largest range (Table 2). For DON, the mean value of the untransformed BLUPs was 13.48 ppm,

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

which is considerably higher than the maximum allowed in wheat grains in the United States (10 ppm). Also, values of SEV as high as 100% for individual plots were observed in all years, indicating adequate conditions for disease development. Medium to high values of broad-sense heritabilities were observed in this experiment. The lowest value was observed for DON (0.43) and the highest value observed for SEV (0.67). These heritability estimates indicate that the panel contains adequate levels of genetic variation for the traits considered in this study.

Genotypic data and population structure

A total of 32,483 SNPs were identified with the GBS protocol after applying the filtering criteria. The observed SNP missing level was 0.31 and minor allele frequency was 0.17. After imputation, the number of SNP was reduced to 19,992 after the LD-SNP analysis, in which a single tagSNPs was used to represent markers showing linkage disequilibrium (r^2) higher than 0.8. The B-genome showed the highest number of SNPs (9084), followed by the A-genome (6992), and D-genome (3916) (Table 3). On average 1297, 998, and 559 SNPs were found for each chromosome of genomes B, A, and D, respectively. The overall polymorphism information content (PIC) of the panel was 0.2029, and the SNP diversity (DIV) was 0.2405. The values of PIC and DIV were similar among sub-genomes A and B, but lower for D. Not surprisingly, the chromosome with the highest number of SNPs was 3B, which is the largest of the wheat chromosomes, and for which a complete draft sequence is available for alignment of tags (Choulet et al., 2014). Although the PCA showed lines from Illinois clustering together, separated from lines of different origin (Fig. 1), the level of population structure across the panel was considered low, with the first four PCs accounting for 5.0, 3.0, 2.7, and 2.4% of the total genetic variation, respectively. No other grouping pattern could be detected with this analysis.

Marker-trait associations

Five statistical models were compared in their ability to detect marker-trait associations using 19,992 SNPs and markers associated with FHB resistance in previous studies. The models that did not include the K matrix - "naïve" and Q model - revealed a large number of significant SNPs for most traits (Table 4). For instance, the "naïve" model detected 160 QTL for INC whereas the Q model detected 208 for the same trait. These numerous associations probably arose from the fact that the breeding lines used in this study have a high degree of relatedness and the K matrix was not included in those models. In fact, the number of significant associations reduced considerably with the inclusion of the K matrix. Very similar numbers of associations were detected for the same trait across the K, MLM, and cMLM models.

Ideally, the *p*-value distribution would follow a uniform distribution around the expected values. A considerable deviance from the expected distribution was observed, especially for the "naïve" and Q models (data not shown). At the same time, the models including the K matrix showed good agreement with the expected distribution, with the outliers representing the significant SNPs based on the raw p-value. Models K, MLM, and cMLM performed very similarly and the cMLM was selected as the model of choice for subsequent analysis for all traits due to its improved statistical power (Zhang et al., 2010).

FHB-related traits

Severity

SNP IWGSC_CSS_3B_scaff_10676713_7175 was significantly associated with SEV (Fig. 2a). This SNP accounted for 8% of the variability and its estimated effect was -9.54 (Table 5). Chromosome 3B harbors a major effect QTL, *Fhb1*, which has been extensively studied. The

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

significant SNP was found to be in linkage disequilibrium with the PCR markers flanking *Fhb1* (D' = 0.39 and 0.34 for *umn10* and *gwm533*, respectively); however, these markers did not meet the threshold to be declared a QTL in this study, perhaps due to genotyping errors or recombination between *Fhb1* and the linked markers during the development of the breeding lines. Altough not significant, three additional SNPs on chromosome 3B, located at similar genetic position of *Fhb1*, were among the top markers when SNPs were ranked based on P values. Two of these three SNPs (IWGSC_CSS_3B_ scaff _10352272_5482 and IWGSC_CSS_3B_ scaff _10699215_3620) were also in linkage disequilibrium with the PCR markers flanking *Fhb1* (data not shown). Due to the importance of this genomic region for FHB resistance, it was decided to further analyze the relationship between these four SNPs on chromosome 3B and their impact on SEV. Thirteen haplotypes involving these four SNPs were found in the panel. The haplotypes were compared using a series of contrasts (Table 6). Haplotypes with fewer than 10 individuals were not used in the analysis. In general, a tendency towards lower levels of SEV was observed as favorable alleles accumulated (Fig. 3a). The lowest mean SEV was observed for the group with lines carrying all four favorable alleles ("+ + + +"); however, this group did not differ from the haplotype ("+ + - +").

Incidence

Five SNPs on chromosomes 4D, 7D, 4A, 6A, and 7A, were significantly associated with INC (Fig. 2b). IWGSC_CSS_7DS_ scaff _3876750_2023 accounted for 16% of the variance, and its effect was -6.74 (Table 5). The other SNPs showed effects varying from -2.62 to -4.56. Unfortunately, lines carrying more than one favorable allele were very rare, and a meaningful statistical comparison could not be performed. The only two contrasts performed were "- - - - -" versus "- - - - +" and "- - - - -" versus "- - - + -" (Table 6). The first was significant ($p = 0.017$),

but the second was not (p = 0.57). In general, a pattern similar to SEV was observed for INC, with lower disease levels as more favorable alleles accumulated (Fig. 3b).

DON

One SNPs associated with DON was located on chromosomes 1D, and three were located on chromosome 3B (one on the short arm that was also detected for SEV and two SNPs located in the long arm at 73.67 cM) (Table 5). With the exception of the SNP on 3BS that was also associated with SEV, the SNPs significantly associated with DON had the major allele as the favorable one. In other words, the lines showing the minor allele had the highest DON values, especially when multiple favorable alleles were absent. Eight different haplotypes were present in the panel (Fig. 3c), with two having more than 10 individuals. The first group consisted of lines carrying four "+" alleles, and the second comprised lines with "+" alleles for IWGSC_CSS_3B_ scaff _10676713_7175, IWGSC_CSS_1DS_ scaff _1879930_3352, and IWGSC_CSS_3B_ scaff _10764618_2168. These groups were not significantly different from each other (Table 6).

Disease indexes

Marker-trait association were tested for disease indexes, which incorporate multiple disease measurements all at once. One SNP was detected for FHBdx (IWGSC_CSS_3B_ scaff _10676713_7175) and two for ISK (IWGSC_CSS_3B_ scaff _10676713_7175 and IWGSC_CSS_7DS_ scaff _3876750_2023). These are the same SNPs detected for SEV and INC (Table 5). Both indexes were based on values of SEV and INC (plus FDK in the case of ISK).

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Discussion

Fusarium head blight is the most important wheat disease in the Midwest, United States, and cultivars with higher levels of resistance are urgently needed. The identification of QTL associated with resistance can potentially facilitate the incorporation of resistance into elite germplasm. In this study GBS was used to generate SNPs for a panel consisting of 273 winter wheat lines. This germplasm is mainly from soft winter wheat breeding programs across the midwestern and eastern regions of the United States. A total of 19,992 SNPs were used for genome-wide marker-trait association for FHB-resistance traits. The largest number of SNPs was detected for the B-genome. The levels of PIC and DIV were comparable between A- and B-genomes, but considerably lower for D. The relative lower diversity of this genome has been reported in other studies (Poland et al., 2012a, Nielsen et al., 2014). Akhunov et al. (2010) point to the low effective recombination and prevention of homoeologous chromosome paring as an explanation for the difference in genetic diversity among the wheat genomes. In addition, it is known that D-genome was latest to be added to the cultivated bread wheat and that this polyploidization event represented a genetic bottleneck (Peng et al., 2011). Chromosome 3B showed the largest number of SNPs. This is not surprising since 3B is the largest wheat chromosome and has a completed draft sequence available for aligning tags. In fact, wheat chromosome 3B is approximately 1 GB in size, more than twice the size of the entire rice genome (Itoh et al., 2007; Paux et al., 2008).

Population structure can result in spurious associations between markers and traits (Matthies et al., 2012), and association studies that do not account for it should be viewed with skepticism (Flint-Garcia et al. 2003). The level of stratification of the panel was assessed via PCA using all 19,992 SNPs. Only a low level of structure was detected, as revealed by the

modest contribution of the first four PCs to the total genetic variance (13.1%). This reduced population structure is most likely due to the fact that 86% of the lines used in the panel were breeding lines from the University of Illinois' breeding program, and/or due to the extensive germplasm exchange among the breeding programs from which the lines originated. This exchange of breeding lines is facilitated by the US Wheat and Barley Scab Initiative through a cooperative system of FHB screening nurseries.

Five statistical models were tested for detecting marker-trait associations. The best model can vary with the trait, as reported by Gurung et al. (2014). By comparing the agreement between predicted and observed p-values, the authors found the K model to be the best for *Stagnosporam nodorum* blotch and tan spot caused by *Pyrenophora tritici-repentis* races 1 and 5. The MLM (Q + K matrices) was the most suitable for bacterial leaf streak, spot blotch, and *Septoria tritici* blotch. In this study, models lacking control for relatedness detected a large number of (most likely) spurious associations. The lines used in this study belong to breeding programs that extensively exchange germplasm. For this reason, high degree of relatedness was expected. The marker-based kinship (K) matrix was shown to be important for controlling false-positive associations.

Although several QTL have been identified for FHB resistance, few of them are considered stable, being detected by independent studies. To date, the best studied and characterized QTL is *Fhb1*, on the short arm of chromosome 3B, originally identified in spring wheat germplasm from China. Independent studies detected QTL for multiple traits associated with FHB resistance on the *Fhb1* region. In a meta-analysis with 249 QTL from 45 studies, Liu et al. (2009) reported 27 QTL associated with SEV, INC, FDK, and DON on 3BS between 0 and 21.6 cM. This cluster, or meta-QTL (MQTL), was flanked by the SSR markers *Xfba311* and

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Xgwm493. Working with 30 mapping populations of mostly European germplasm, Löffler et al. (2009) performed a meta-analysis and found 13 QTL for SEV and INC, clustered around 16.1cM. In a comprehensive review, Buerstmayr et al. (2009) compiled information from 52 mapping studies and found 26 QTL for multiple FHB resistance traits falling within positions 0 to 20 cM on chromosome 3B, between *Xgwm533* and *Xgwm493.* The mapping populations used in these meta-analyses and review show *Fhb1* QTL region spanning over loci 0 and 20 cM distally on 3BS.

This major effect QTL has been introgressed into several adapted winter wheat lines and cultivars from the United States (Jin et al., 2013) and worldwide. In the germplasm used in this study, 97 breeding lines (35%) are known to have Ning7840 or Sumai-3 in their pedigree, and we deteremined that 15 of those carried the alleles of *gwm533* and *umn10* associated with *Fhb1*. Interestingly, the two markers linked to *Fhb1* did not reach significance in our GWAS. However, our analysis was able to detect a highly significant GBS-SNP marker associated with resistance at the the *Fhb1* locus. The SNP IWGSC_CSS_3B_ scaff _10676713_7175 was associated with SEV, FHBdx, and DON on chromosome 3B at 18.32 cM, and was in linkage disequilibrium with markers *umn10* and *gwm533.* In addition, it was shown that a combination of favorable alleles of four GBS SNPs in the region, albeit not all significant when considered individually, was associated with lower levels of SEV.

Lack of significant marker-trait association for *umn10* and *gwm533* may be explained by recombination between the markers and *Fhb1* during the course of line development and/or lack of sufficient LD between the markers and the resistance gene. Thirty-five lines in this panel carry the favorable IWGSC_CSS_3B_ scaff _10676713_7175 allele and 24 of those had the most favorable haplotypes ("+ + + +" and "+ + - +"). A total of 32 lines having alleles associated with

the *Fhb1* gene at *umn10* and *gwm533* were identified, of which only 24 and 18 carried the favorable IWGSC_CSS_3B_ scaff _10676713_7175 allele or the most favorable haplotypes of GBS markers, respectively. Marker *umn10* developed by Liu et al. (2008) is closely linked with *Fhb1* and is widely used for MAS. However, some winter wheat lines in the eastern United States that are considered FHB susceptible and do not have Asian sources of resistance in their pedigrees amplify a DNA fragment the same size as that amplified from Ning7840 and Sumai-3 using the UMN10 primer pair (Brown-Guedira, unpublished data). The alleles of SSR marker *gwm533* associated with *Fhb1* are not generally found in eastern United States winter wheats that do no have these resistance sources in their pedigrees. However, this locus is located approximately 10 cM distal to *Fhb1* and recombination may have occurred during development of SRWW breeding lines selected for FHB resistance in disease screening nurseries.

The marker association for INC on chromosome 4D in our study could potentially be unique. Several studies have associated the *Rht-D1b* dwarfing gene on the short arm of 4D with increased FHB susceptibility (Srinivasachary et al., 2009). The SNP IWGSC_CSS_4DS_ scaff _2300354_4482 was not in signficant LD with *Rht-D1* in our study and no significant marker trait association was identified for *Rht-D1*. Liu et al. (2009) reported 10 QTL on chromosome 4D clustered around locus 12 cM, most of them associated with SEV and all overlapped with the major plant height locus *Rht-D1*. Jiang et al. (2015) also found significant associations of the *Rht-D1* locus with FHB reaction in a GWAS of 372 European wheat varieties. The *Rht-D1b* allele was present at high frequency in their germplasm and was also determined to be associated with relatedness among lines. Based on assays for the functional polymorphisms, approximately 15 % of lines in our panel were determined to have the *Rht-D1b* dwarfing gene while 80 % had the *RhtB1b* gene and 5% of lines did not have either dwarfing gene. Almost half of lines having

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

RhtD1b were derived from programs in the outheast and mid-atlantic regions and selected for inclusion in the study based on their moderate level of resistance to FHB. The selection of moderately FHB resistant *Rht-D1b* lines for inclusion in the study may have prevented detection of deleterious effects associated with the locus.

Additional SNPs associated with INC were located on chromosomes 4A (locus 78.35 cM), 6A (134.15 cM), 7A (22.82 cM), and 7D (70.84 cM), all of which may correspond with previously reported QTL. Liu et al. (2009) reported a MQTL in a similar region of chromosome 4A between loci 75.7 to 77.1 cM associated with SEV. Previous studies reported QTL for FHB resistance on proximal (Schmolke et al., 2005; Holzapfel et al., 2008) and distal (Paillard et al., 2004; Kollers et al., 2013) positions of chromosome 6A. The significant SNP we detected on the long arm of chromosome 6A could potentially be associated with the QTL detected by Paillard et al. (2004) and Kollers et al. (2013) in the distal region of 6AL. Petersen et al. (2015) recently reported a QTL on 6A in the soft red winter wheat NC-Neuse. It is known that two regions on chromosome 7A harbor QTL for SEV and one region on 7D for multiple FHB resistance traits (Liu et al., 2009). In this study, SNPs detected for INC were relatively close to those regions. SNP IWGSC_CSS_7DS_ scaff _3876750_2023 that was present in 7% of the lines in our panel explained the largest amount of variance (16%) observed for any FHB related traits and had effects ranging from -6.74% to -7.47% for INC and ISK, respectively. As was found for SEV, lines carrying multiple favorable ("+") alleles showed lower levels of disease.

Mycotoxin accumulation in harvested grain is an important problem for growers, the food industry, and consumers. DON is the most important mycotoxin in wheat, and QTL associated with lower levels of DON have been reported on multiple chromosomes. This study detected a significant DON effect for the SNP in the *Fhb1* region on the short arm of chromosome 3B

associated with SEV and the disease indexes. The SNPs we detected proximally on 3B may be associated with QTL reported previously. For instance, Liu et al. (2009) detected a cluster of QTL between loci 49 and 57 cM, and Löffler et al. (2009) reported a cluster around locus 45.5 cM. Interestingly, QTL in the proximal region of chromosome 3B are reported from eastern soft winter wheat cultivars, including NC-Neuse (Petersen et al., 2015), Truman (Islam et al., 2015), Ernie and Massey (Liu et al., 2007; Liu et al., 2009). In this study, the SNP detected on the short arm of chromosome 1D falls within the same cluster detected by Liu et al. (2009) on that chromosome.

Combining favorable alleles for significant SNPs also resulted in a reduction on DON level. A numerical difference was observed when compared with the classes contrasting for the presence of the marker associated with *Fhb1* ("+ + + +") and ("+ - + +"), although the classes were not significantly different. Miedaner et al. (2006) compared eight combinations of QTL on chromosomes 3A, 3B, and 5A in terms of DON and FHBdx. For both traits, the combination 3B and 5A did not differ from three QTL stacked together. In this study, most lines in the panel (90.4%) already have favorable alleles for SNP in the proximal region of chromosome 3B (IWGSC_CSS_3B_ scaff _10413672_4839, and IWGSC_CSS_3B_ scaff _10764618_2168), and the short arm of 1D (IWGSC_CSS_1DS_ scaff _1879930_3352). When favorable alleles at these three SNPs are simultaneously absent, DON levels increase substantially (> 17 ppm) Interestingly, there were only two lines with no favorable allele (- - - -), both of which were derived from hard wheat winter breeding programs. It is possible to speculate that the combination of QTL associated with these SNPs may play a role in DON accumulation and/or detoxification that have been selected by soft winter wheat breeders in screening nurseries.

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Hypothetical protein predictions were obtained for eight out of twelve contigs (Table 7), many of which seem to be involved in transcription. For the contig on chromosome 1D harboring IWGSC_CSS_1DS_ scaff _1879930_3352, the predicted protein is a receptor-like kinase (RLK). The RLKs are known to be involved in a wide range of plant responses including development, growth, and response to pathogen (Goff et al., 2007). Other predicted proteins contain a reverse transcriptase, protein phosphatase inhibitor, transcription factors, and other ubiquitin protein. Based on the wide range of protein functions observed, we speculate that FHB resistance in wheat may involve intricate gene x gene interactions. In fact, some SNP x SNP interactions were significant in this study (data not shown). Miedaner et al. (2011) also found digenic epistatic interactions to be involved in FHB resistance. Although our results did not lead to a clear conclusion about protein function associated with FHB resistance, they point to directions for future research and investigation on this topic.

Conclusion

Breeding wheat for FHB resistance has been a difficult task due to a number of reasons, including the complex nature of resistance. In this study QTL for FHB resistance were identified in regions previously reported as harboring QTL in bi-parental populations, especially on wheat chromosome 3B. Identification of marker-trait associations in our panel of soft winter wheat breeding lines suggests that the previously reported resistance QTL are contributing to resistance in this germplasm and are good targets for marker assisted selection. This study also provides evidence that QTL accumulation can result in higher levels of resistance, but certain haplotypes currently occur only rarely. The SNPs reported can be used to develop assays for marker-assisted selection for favorable multi-locus genotypes and/or inclusion in genomic selection approaches. It is possible that many of the SNP associated with resistance in this study are closely linked to

or are in high LD with the resistance QTL. Our GWAS for FHB resistance is one of the first to utilize sequence based genotyping with aligment to the scaffolds from the International Wheat Genome Sequencing Consortium (IWGSC, 2014), allowing integration with this significant genomic resource in wheat. Functional characterization of the underlying QTL could be performed for elucidating their biological role in FHB resistance.

Acknowledgments

The authors are grateful to Dr. Yanhong Dong, from the University of Minnesota for deoxynivalenol determination. This material is based upon work supported by the United States Department of Agriculture, under Agreement No. 59-0206-9-080. This is a cooperative project with the U.S. Wheat & Barley Scab Initiative. Marcio Arruda and Keith Merrill are grateful to Monsanto Company and the Triticeae-CAP (funded by AFRI grant No. 2011-68002-30029), for supporting their graduate studies, respectively.

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Figure 1. Principal component analysis of 273 winter wheat breeding lines using 19,992 SNPs. Colors represent the origin of the breeding lines. Illinois = University of Illinois at Urbana-Champaign; Kentucky = University of Kentucky; Missouri = University of Missouri; Indiana = Purdue University; North Carolina = North Carolina State University; Virginia = Virginia Tech University; Pioneer = Pioneer HiBred International; Arkansas = University of Arkansas; Ohio = Ohio State University; Georgia = University of Georgia; Limagrain = Limagrain Cereal Seeds; Syngenta = Syngenta/AgriPro Associates; Michigan = Michigan State University; and Oklahoma = Oklahoma State University. One line from Cornell University, one from KWS SAAT SE, and one from University of Nebraska-Lincoln can not be visualized in the graph.

Figure 2. Genome-wide association scan for three parameters associated with FHB resistance: SEV (a), INC (b), and DON (c). The y axis represents the P value of the marker-trait association on a -log10 scale. The horizontal line represents the threshold for declaring a marker as significant (FDR-adjusted P-value ≤ 0.10).

Figure 3. Mean phenotypic value of SEV (a), INC (b), and DON (c) for groups of wheat lines carrying different SNP combinations, with ("+") signal representing the favorable allele. Lines were evaluated in 2011, 2013, and 2014 in Urbana-IL. SNP 3B-1 = IWGSC_CSS_3B_ scaff _10676713_7175, SNP 3B-2 = IWGSC_CSS_3B_ scaff _10352272_5482, SNP 3B-3 = IWGSC_CSS_3B_ scaff _10698462_2332, SNP 3B-4 = IWGSC_CSS_3B_ scaff _10699215_3620, SNP 7D = IWGSC_CSS_7DS_ scaff _3876750_2023, SNP 6A = IWGSC_CSS_6AL_ scaff _5780077_12152, SNP 4D = IWGSC_CSS_4DS_ scaff _2300354_4482, SNP 4A = IWGSC_CSS_4AL_ scaff _7146617_11335, SNP 7A =

IWGSC_CSS_7AS_ scaff _4132011_1400, SNP 3B-5 = IWGSC_CSS_3B_ scaff _10413672_4839, SNP 1D = IWGSC_CSS_1DS_ scaff _1879930_3352, SNP 3B-6 = IWGSC_CSS_3B_ scaff _10764618_2168.

Table 1. Markers associated with quantitative trait loci (QTL) for FHB resistance and reduced height (*Rht*) genes in wheat.

Marker	Chromosome	Gene/QTL name	Reference
umn10	3B	*Fhb1*	Liu et al. 2008
gwm533	3B	*Fhb1*	Zhou et al., 2002
SNP3BS-8	3B	*Fhb1*	Bernado et al. 2012
cfd233	2D	*MQTL8*	Kollers et al., 2013, Löffler et al., 2009
gwm539	2D	*QFhs.nau-2DL/Qfhs.arc-2D*	Jiang et al., 2007a, Jian et al., 2007b
gwm304	5A	*Qfhs.ifa-5A/Qfhs.umc-5A*	Liu et al. 2007
wmc705	5A	*Qfhs.ifa-5A/Qfhs.umc-5A*	Buerstmayr et al., 2002
wMAS000001	4B	*RhtB1*	http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/kasp_download.php?URL
wMAS000002	4D	*RhtD1*	http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/kasp_download.php?URL

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Table 2. Descriptive statistics and Spearman correlation for untransformed BLUPs for 273 wheat lines, and broad sense heritabilities on a plot mean-basis.

Trait	BLUPs					h^2	correlations				
	Mean	Min	Max	Range	SD		SEV	INC	FHBdx	FDK	ISK
SEV	44.15	10.37	83.41	73.03	15.02	0.57					
INC	77.80	51.69	87.82	36.12	5.92	0.67	0.51**				
FHBdx	36.74	2.43	79.01	76.58	15.12	0.68	0.98**	0.62**			
FDK	38.56	15.93	69.00	53.07	8.82	0.55	0.55**	0.50**	0.58**		
ISK	51.99	16.57	83.43	66.86	10.55	0.53	0.85**	0.71**	0.89**	0.85**	
DON	13.48	6.21	27.86	21.65	2.69	0.43	0.47**	0.50**	0.52**	0.48**	0.57**

Mean, minimum, maximum, range, standard deviation, and Spearman correlation coefficient for untransformed best linear unbiased predictors (BLUP), and broad-sense heritabilities ($h^{2)}$ for six measurements associated with FHB resistance. The data was calculated for 273 winter wheat breeding lines in Urbana, Illinois, in 2011, 2013, and 2014. ** Significant at $\alpha = 0.05$ level. DON was measured in ppm, and the other parameters were measured in percentage.

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Table 3. SNP coverage and diversity obtained a wheat panel consisting in 273 breeding lines using genotyping-by-sequencing.

Chromosome		n	PIC	DIV
Genome A	1	916	0.217	0.26
	2	1521	0.213	0.254
	3	553	0.229	0.277
	4	1200	0.201	0.237
	5	590	0.215	0.258
	6	1028	0.223	0.268
	7	1184	0.222	0.267
		6992[a]	0.217[b]	0.26[b]
Genome B	1	991	0.214	0.256
	2	1962	0.231	0.28
	3	2154	0.204	0.241
	4	777	0.209	0.249
	5	1560	0.23	0.277
	6	796	0.221	0.265
	7	844	0.185	0.216
		9084	0.217	0.26
Genome D	1	610	0.179	0.208
	2	860	0.163	0.184
	3	286	0.172	0.197
	4	350	0.176	0.204
	5	662	0.176	0.204
	6	551	0.217	0.26
	7	597	0.213	0.254
		3916	0.175	0.202
	total	19992	0.203[c]	0.241[c]

[a] Number of SNPs on genome A. [b] Mean values for genome A. [c] Mean values across genomes A, B, and D.

Table 4. Number of QTL associated with FHB-related traits according to different statistical models.

	"naïve"	Q	K	MLM	cMLM	Chromosomes
SEV	80	25	1	1	1	3B
INC	160	208	2	3	5	7D, 6A, 4D, 4A, 7A
FHNdx	95	31	2	2	1	3B
FDK	38	162	0	0	0	-
ISK	132	59	2	2	2	3B, 7D
DON	60	57	6	3	4	1D, 3B (3)

Marker-trait association tests were performed on a panel consisting in 273 wheat breeding lines genotyped with 19,992 SNPs. "naïve" = statistical model with no control for population structure and relatedness; Q model = population structure controlled using four principal components (PCs), treated as fixed effects, from a principal component analysis (Q matrix); K model = relatedness controlled using a marker-based kinship (K) matrix, treated as random; MLM = mixed linear model having the Q and K matrices, with fixed and random effects, respectively; cMLM = compressed mixed linear model, similar to MLM, but with a compressed matrix of individual.

Table 5. SNPs associated with FHB resistance in a panel of 273 breeding lines, chromosomal position, *P* values, frequency, and effects of favorable alleles. Unit of effects are present except DON (ppm).

Trait	SNP	C	F/U[a]	cM	*P*[b]	f[c]	r^2	Adj *p*[d]	effects
SEV	IWGSC_CSS_3B_ scaff _10676713_7175	3B	A/G	18.32	5.14	0.11	0.08	0.050	-9.54
	IWGSC_CSS_3B_ scaff _10352272_5482[e]	3B	G/A	10.19	3.83	0.14	0.05	0.980	-7.21
	IWGSC_CSS_3B_ scaff _10698462_2332[e]	3B	G/T	6.86	3.6	0.44	0.04	0.980	-6.02
	IWGSC_CSS_3B_ scaff _10699215_3620[e]	3B	T/C	18.32	3.37	0.20	0.04	0.980	-4.99
INC	IWGSC_CSS_7DS_ scaff _3876750_2023	7D	A/T	70.84	11.57	0.07	0.16	<0.001	-6.74
	IWGSC_CSS_6AL_ scaff _5780077_12152	6A	T/C	134.15	4.94	0.07	0.06	0.097	-3.60
	IWGSC_CSS_4DS_ scaff _2300354_4482	4D	G/T	0	4.72	0.06	0.06	0.097	-4.56
	IWGSC_CSS_4AL_ scaff _7146617_11335	4A	A/G	78.35	3.54	0.06	0.06	0.070	-3.02
	IWGSC_CSS_7AS_ scaff _4132011_1400	7A	C/G	22.82	4.54	0.16	0.05	0.097	-2.62
FHBdx	IWGSC_CSS_3B_ scaff _10676713_7175	3B	A/G	18.32	5.14	0.11	0.07	0.052	-8.96
ISK	IWGSC_CSS_3B_ scaff _10676713_7175	3B	A/G	18.32	5.14	0.11	0.07	0.052	-5.55
	IWGSC_CSS_7DS_ scaff _3876750_2023	7D	A/T	70.84	11.57	0.07	0.16	0.000	-7.47
DON	IWGSC_CSS_3B_ scaff _10413672_4839	3B	C/T	73.67	5.14	0.94	0.07	0.052	-1.75
	IWGSC_CSS_3B_ scaff _10676713_7175	3B	A/G	18.32	5.14	0.11	0.07	0.052	-1.36
	IWGSC_CSS_1DS_ scaff _1879930_3352	1D	A/T	19.04	5.10	0.95	0.06	0.052	-1.99
	IWGSC_CSS_3B_ scaff _10764618_2168	3B	T/A	73.67	4.99	0.93	0.06	0.052	-2.08

[a] F/U = favorable allele/unfavorable allele; [b] *P* value reported in a $-\log_{10}$ scale; [c] f = frequency of the favorable allele; [d] Adj p = FDR-ajusted *p* value; [e] SNP not significant according to the FDR-adjusted *p* value. Marker-trait associations were tested using a compressed mixed linear model with control for population structure and relatedness.

Table 6. Contrast between haplotypes for different traits associated with FHB resistance.

Trait	Haplotypes		Estimate	Std	F	Prob > F
SEV	"+ + + +"	" - - - -"	-19.68	5.05	15.13	< 0.001
	"+ + - +"	"- - - -"	-18.43	5.59	10.85	< 0.001
	"- - - +"	"- - - -"	-2.14	4.04	0.28	0.597
	"+ + + +"	"+ + - +"	-1.25	7.26	0.03	0.864
	"+ + - +"	"- - - +"	-16.29	6.59	6.12	0.014
INC	"- - - - +"	"- - - - -"	-2.50	1.05	5.73	0.017
	"- - - + -"	"- - - - -"	-1.20	2.18	0.33	0.567
DON	"+ + + +"	"+ - + +"	-0.28	0.36	0.57	0.98

The "+" signal represents the favorable allele of the SNP, and the "-" signal represents the unfavorable allele. For SEV, four SNPs were considered: IWGSC_CSS_3B_ scaff _10676713_7175, IWGSC_CSS_3B_ scaff _10352272_5482, IWGSC_CSS_3B_ scaff _10698462_2332, and IWGSC_CSS_3B_ scaff _10699215_3620. Haplotypes for INC were built with: IWGSC_CSS_7DS_ scaff _3876750_2023, IWGSC_CSS_6AL_ scaff _5780077_12152, IWGSC_CSS_4DS_ scaff _2300354_4482, IWGSC_CSS_4AL_ scaff _7146617_11335, and IWGSC_CSS_7AS_ scaff _4132011_1400. Haplotypes for DON were combinations of the following SNPs: IWGSC_CSS_3B_ scaff _10413672_4839, IWGSC_CSS_3B_ scaff _10676713_7175, IWGSC_CSS_1DS_ scaff _1879930_3352, and IWGSC_CSS_3B_ scaff _10764618_2168.

Table 7. Hypothetical protein prediction and conserved domains for contigs harboring SNPs associated with FHB resistance.

SNP	C	cM	hypothetical protein predition	conserved domains
IWGSC_CSS_1DS_ scaff _1879930_3352	1D	19.04	RLK protein (*Brachypodium*, *Setaria*, *Nicotiana*, *Solanum*)	none
IWGSC_CSS_3B_ scaff _10413672_4839	3B	73.67	chloroplast DNA / transcription factor	BSD superfamily
IWGSC_CSS_3B_ scaff F_10764618_2168	3B	73.67	uncharacterized hypothetical protein	none
IWGSC_CSS_4AL_ scaff _7146617_11335	4A	78.35	FAR-1-related (*Tritum urartu*)	FAR-1 DNA-binding
IWGSC_CSS_4DS_ scaff _2300354_4482	4D	0	PPI (*Brachypodium*, *Oryza*, *Zea*, *Elaeis*, etc)	Protein phosphatase inhibitor-2 superfamily
IWGSC_CSS_6AL_ scaff F_5780077_12152	6A	134.2	Rop guanine nucleotide exchange factor 2 (*Aegilops tauschii*)	FHA, PRONE, DNA-polymerase-viral-N-terminal,
IWGSC_CSS_7AS_ scaff _4132011_1400	7A	22.82	Putative reverse transcriptase (*Oryza sativa*)	Reverse-transcriptase superfamily and EEP family
IWGSC_CSS_7DS_ scaff _3876750_2023	7D	70.84	GDSL esterase/lipase (*Aegilops tauschii*)	SGNH_hydrolase superfamily, Lipase_GDSL, UBA2

RLK = Receptor-like kinase protein. BSD = BTF2-like transcription factors, synapse-associated proteins, and DOS2-like proteins. FAR = far-red-impaired response. PPI = Protein phosphatase inhibitor. Rop = Repressor of primer. FHA = Forkhead associated domain. PRONE = Plant-specific Rop nucleotide exchanger. EEP = Endonuclease/Exonuclease/phosphatase. GDSL = consensus amino acid sequence of glycine (G), aspaspartic acid (D), serine (S), and leucine (L). SGNH = consensus amino acid sequence of serine (S), glycine (G), asparagine (N), and histidine (H). Predicted amino acid sequences obtained with Fgenesh gene finder, and hypothetical protein prediction and conserved domain obtained with BLASTp.

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

References

Adams, M. J. 2010. Diseases caused by fungi and fungus-like organisms. pp.15-86. In W.W. Bockus, R.L. Bowden, and, R.M. Hunger, et al. (ed). Compendium of wheat diseases and pests. St Paul: The American Phytopathological Society

Akhunov, E., A. Akhunova, O. Anderson, J. Anderson, N. Blake, M. Clegg, et al. 2010. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11:702

Arruda, M.P., P.J. Brown, A.E. Lipka, A.M. Krill, C. Thurber, and F.L. Kolb. 2015. Genomic selection for predicting Fusarium head blight resistance in a wheat breeding program. Plant Gen., doi: 10.3835/plantgenome2015.01.0003.

Benjamini, Y., Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach for multiple testing. J. R. Stat. Soc. B 57:289-300

Bennett, J. W., and M. Klich. 2003. Mycotoxins. Clin. Microbiol. Rev. 16:497-516

Bernardo, A.N., H.X. Ma, D.D. Zhang, and G.H. Bai. 2012. Single nucleotide polymorphism in wheat chromosome region harboring *Fhb1* for Fusarium head blight resistance. Mol. Breeding 29:477-488

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Buerstmayr, H., T. Ban, and J.A. Anderson. 2009. QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: A review. Plant Breeding 128:1-26

Buerstmayr, H., M. Lemmens, L. Hartl, L. Doldi, B. Steiner, M. Stierschneider, et al. 2002. Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). Theor. Appl. Genet. 104:84-91

Carlson, C.S., M.A Eberle, M.J. Rieder, Q. Yi, L. Kruglyak, Nickerson, D.A. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am. J. Hum. Genet. 74:106-120.

Cavanagh, C.R., S. Chao, S. Wang, B.E. Huang, S. Stephen, S. Kiani, et al. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci U. S. A. 110:8057-8062

Choulet, F., A. Alberti, S. Theil, N. Glover, V. Barbe, J. Daron, et al. 2014. Structural and functional partitioning of bread wheat chromosome 3B. Science 345:1249721

Elshire, R.J., J.C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Flint-Garcia, S.A., J.M. Thornsberry, and E.S.I. Buckler. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357-374

Ghavami, F., E.M. Elias, S. Mamidi, O. Ansari, M. Sargolzaei, T. Adhikari, et al. 2011. Mixed model association mapping for Fusarium head blight resistance in tunisian-derived durum wheat populations. G3 1:209-218

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. *TASSEL-GBS*: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:e90346

Goff, K.E., and K.M. Ramonell. 2007. The role and regulation of receptor-like kinases in plant defense. Gene Regul. Syst. Bio 1:167-175

Gurung, S., S. Mamidi, J.M. Bonman, M. Xiong, G. Brown-Guedira, and T.B. Adhikari. 2014. Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. PLoS ONE 9:e108179

Hazel, C.M., and S. Patel. 2004. Influence of processing on tricothecene levels. Toxicol. Lett. 153:51-59

Holzapfel, J., H.-H Voss, T. Miedaner, V. Korzun, J. Häberle, G. Schweizer, et al. 2008. Inheritance of resistance to Fusarium head blight in three European winter wheat populations. Theor. Appl. Genet. 117:1119-1128

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Islam, M.S., G. Brown-Guedira, D.V. Sanford, H. Ohm, Y. Dong, A.L. McKendry. 2015. Novel QTL associated with the Fusarium head blight resistance in Truman soft red winter wheat. Euphytica 10.1007/s10681-015-1550-9

Itoh, T., T. Tanaka, R.A. Barrero, C. Yamasaki, Y. Fujii, P.B. Hilton, et al. 2007. Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. Genome Res. 17:175-183

IWGSC. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

Jiang, G.-L., Y. Dong, J. Shi, and R.W. Ward. 2007a. QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ 9306. II. Resistance to deoxynivalenol accumulation and grain yield loss. Theor. Appl. Genet. 115:1043-1052

Jiang, G.-L., J. Shi, and R.W. Ward. 2007b. QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ 9306. I. Resistance to fungal spread. Theor. Appl. Genet. 116:3–13. doi:10.1007/s00122-007-0641-y

Jiang, Y., Y. Zhao, B. Rodemann, J. Plieske, S. Kollers, V. Korzun, et al. 2015. 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 114:318-326

Jin, F., D. Zhang, W. Bockus, P.S. Baezinger, B. Carver, and G. Bai. 2013. Fusarium head blight resistance in US winter wheat cultivars and elite breeding lines. Crop Sci. 53:2006-2013

Kollers, S., B. Rodemann, J. Ling, V. Korzun, E. Ebmeyer, O. Argillier, et al. 2013. Whole genome association mapping of Fusarium head blight resistance in european winter wheat (*Triticum aestivum* L.). PLoS ONE 8:e57500

Langer, S.M., C.F.H. Longin, and T. Würschum. 2014. Flowering time control in European winter wheat. Front. Plant Sci. 5:537

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, et al. 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397-2399

Liu, S., M.D. Hall, C.A. Griffey, and A.L. McKendry. 2009. Meta-analysis of QTL associated with Fusarium head blight resistance. Crop Sci. 49:1955-1968

Liu S, M.O. Pumphrey, B.S. Gill, H.N. Trick, J.X. Zhang, J. Dolezelet, al. (2008) Toward positional cloning of *FHB1*, a major QTL for fusarium head blight resistance in wheat. Cereal Res. Commun. Suppl. B 36:195-2012

Liu, S., Z.A. Abate, H. Lu, T. Musket, G.L. Davis, and A.L. McKendry. 2007. QTL associated with Fusarium head blight resistance in the soft red winter wheat Ernie. Theor. Appl. Genet. 115:417-427

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Löffler, M., C.C. Schon, and T. Miedaner. 2009. Revealing the genetic architecture of FHB resistance in hexaploid wheat (*Triticum aestivum* L.) by QTL meta-analysis. Mol. Breeding 23:473-488

Matthies, I.E., T.V. Hintum, S. Weise, and M.S. Roder. 2012. Population structure revealed by different marker types (SSR or DArT) has an impact on the results of genome-wide association mapping in European barley cultivars. Mol. Breeding 30:951-966

Mesterhazy, Á. 1995. Types and components of resistance to Fusarium head blight of wheat. Plant Breeding 114:377-386

Mesterhazy, Á., T. Bartók, C.G. Mirocha, and R. Komoróczy. 1999. Nature of wheat resistance to Fusarium head blight and the role of deoxynivalenol for breeding. Plant Breeding 118:97-110

Miedaner, T., F. Wilde, B. Steiner, H. Buerstmayr, V. Korzun, and E. Ebmeyer. 2006. Stacking quantitative trait loci (QTL) for Fusarium head blight resistance from non-adapted sources in a European elite spring wheat background and assessing their effects on deoxynivalenol (DON) content and disease severity. Theor. Appl. Genet. 112:562-569

Miedaner, T., T. Würschum, H. Maurer, V. Korzun, E. Ebmeyer, and J. Reif. 2011. Association mapping for Fusarium head blight resistance in European soft winter wheat. Mol. Breeding 28:647-655

Miller, J.D., J.C. Young, and D.R. Sampson. 1985. Deoxynivalenol and Fusarium head blight resistance in spring cereals. J. Phytopathol. 113:359-367

Myles, S., J. Peiffer, P.J. Brown, E. Ersoz, Z. Zhang, D. Costich, et al. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell 21:2194-2202

Nielsen, N.H., G. Backes, J. Stougaard, S.U. Andersen, and A. Jahoor. 2014. Genetic diversity and population structure analysis of European hexaploid bread wheat (*Triticum aestivum* L.) varieties. PLoS ONE 9:e94000

Paillard, S., T. Schnurbusch, R. Tiwari, M. Messmer, M. Winzeler, B. Keller, et al. 2004. QTL analysis of resistance to Fusarium head blight in Swiss winter wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 109:323-332

Paux, E., P. Sourdille, J. Salse, C. Saintenac, F. Choulet, P. Leroy, et al. 2008. A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101-104

Peng, J., D. Sun, and E. Nevo. 2011. Domestication evolution, genetics and genomics in wheat. Mol. Breeding 28:281-301

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Petersen, S., J.H. Lyerly, P.V. Maloney, G. Brown-Guedira, C. Cowger, J. Costa, et al. 2015. Identification and mapping of QTL conferring resistance to Fusarium head blight resistance in the soft red winter wheat cultivar NC-Neuse. Crop Sciences /accepted for publication/

Poland, J.A., P.J. Brown, M. Sorrels, and J-L. Jannink. 2012a. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping sequencing approach. PLoS ONE 7:e32253

Poland, J.A., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, et al. 2012b. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Gen 5:103-113

Poland, J.A., and T.E. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. Plant Gen 5:92-102

Rutkoski, J.E., J.A. Poland, J.L. Jannink, and M.E. Sorrells. 2013. Imputation of unordered markers and the impact on genomic selection accuracy. G3 3:427-39

SAS Institute. 2015. JPM Genomics System for Windows. Version 7. SAS Institute Inc, Cary, NC

Schmolke, M., G. Zimmermann, H. Buerstmayr, G. Schweizer, T. Miedaner, V. Korzun, et al. 2005. Molecular mapping of Fusarium head blight resistance in the winter wheat population Dream/Lynx. Theor. Appl. Genet. 111:747-756

Schroeder, H.W., and J.J. Christensen. 1963. Factors affecting resistance of wheat to scab caused by *Gibberella zeae*. Phytopathology 53:831-838

Srinivasachary, A., N. Gosman, A. Steed, T. Hollins, R. Bayles, P. Jennings, P., et al. 2009. Semi-dwarfing*Rht-B1* and *Rht-D1* loci of wheat differ significantly in their influence on resistance to Fusarium head blight. Theor. Appl. Genet. 18:695-702

Wang, S.C, D. Wong, K. Forrest, A. Allen, S.M. Chao, B.E. Huang, et al. 2014. Characterization of polyploidy wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotech J. 12:787-796

Wood, M., D. Comis, D. Harden, L. McGraw, and K.B. Stelljes. 1999. Fighting *Fusarium*. Beltsville: United States Department of Agriculture, Agricultural Research Service.

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203-208

Zhang, Z., Ersoz, E., Lai, C.-Q, Todhunter, R.J., Tiwari, H.K., Gore, M.A., et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355-360.

Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, et al. 2007. An *Arabidopsis* example of association mapping in structured samples. PLoS Genetics 3:e4

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

Zhou, W., F.L. Kolb, G. Bai, G. Shaner, and L.L. Domier. 2002. Genetic analysis of scab resistance and QTL in wheat with microsatellite and AFLP markers. Genome 45:719-727

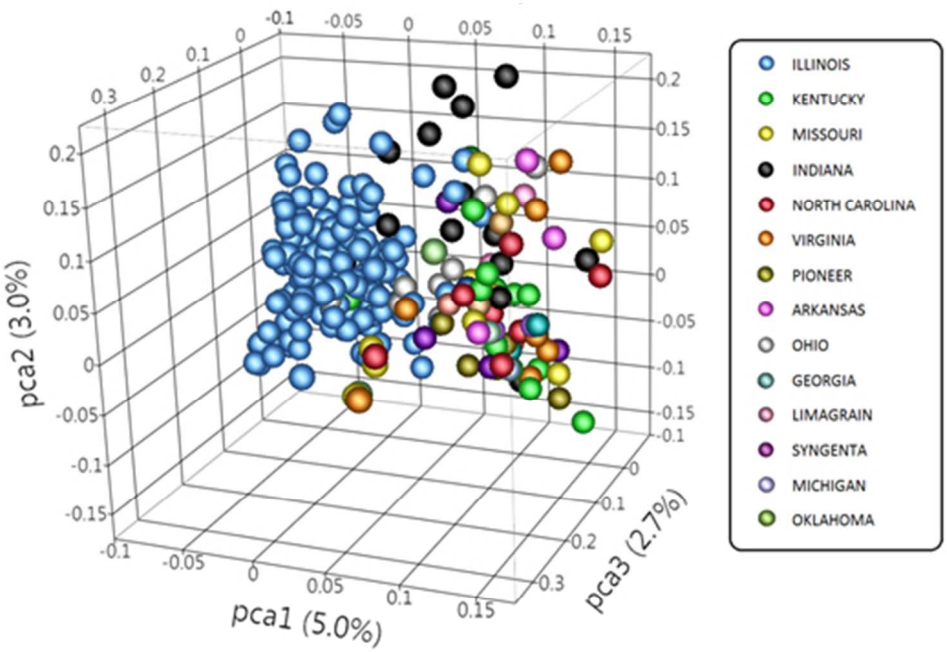

Figure 1. Principal component analysis of 273 winter wheat breeding lines using 19,992 SNPs. Colors represent the origin of the breeding lines. Illinois = University of Illinois at Urbana-Champaign; Kentucky = University of Kentucky; Missouri = University of Missouri; Indiana = Purdue University; North Carolina = North Carolina State University; Virginia = Virginia Tech University; Pioneer = Pioneer HiBred International; Arkansas = University of Arkansas; Ohio = Ohio State University; Georgia = University of Georgia; Limagrain = Limagrain Cereal Seeds; Syngenta = Syngenta/AgriPro Associates; Michigan = Michigan State University; and Oklahoma = Oklahoma State University. One line from Cornell University, one from KWS SAAT SE, and one from University of Nebraska-Lincoln can not be visualized in the graph.
144x100mm (96 x 96 DPI)

Plant Gen. Accepted Paper, posted 11/24/2015. doi:10.3835/plantgenome2015.04.0028

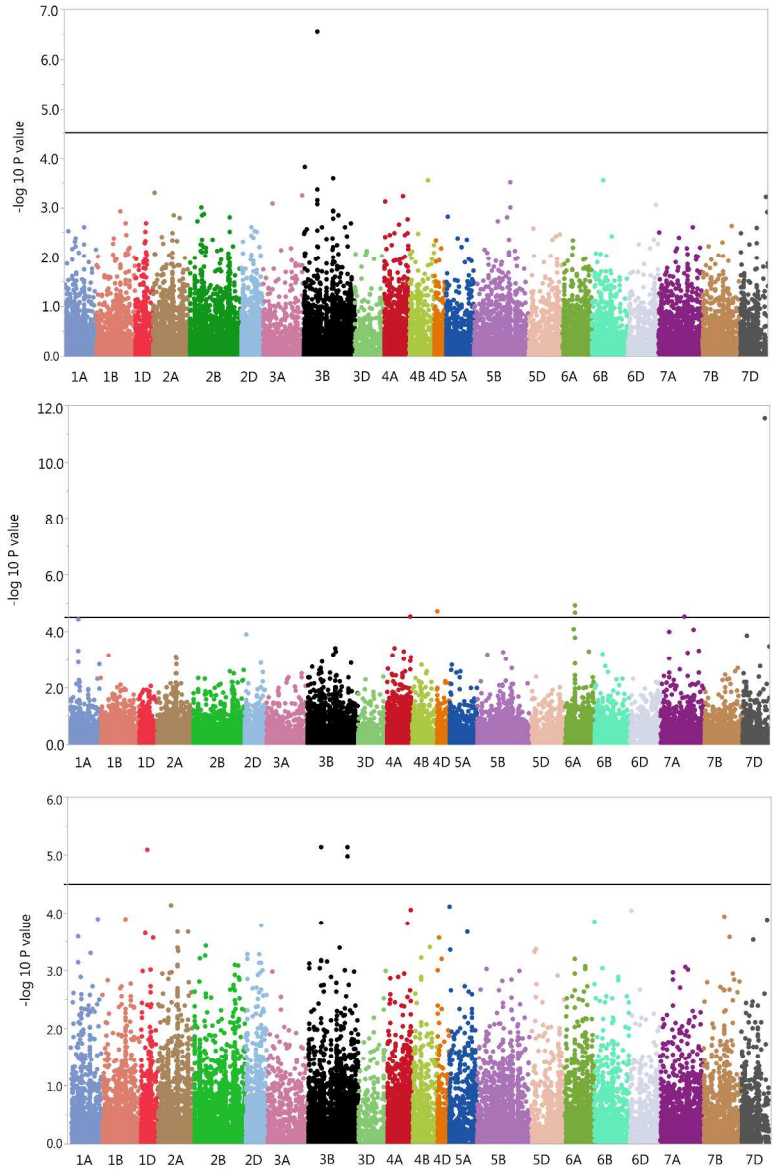

Figure 2. Genome-wide association scan for three parameters associated with FHB resistance: SEV (a), INC (b), and DON (c). The y axis represents the P value of the marker-trait association on a -log10 scale. The horizontal line represents the threshold for declaring a marker as significant (FDR-adjusted P-value ≤ 0.10). 1001x1507mm (96 x 96 DPI)

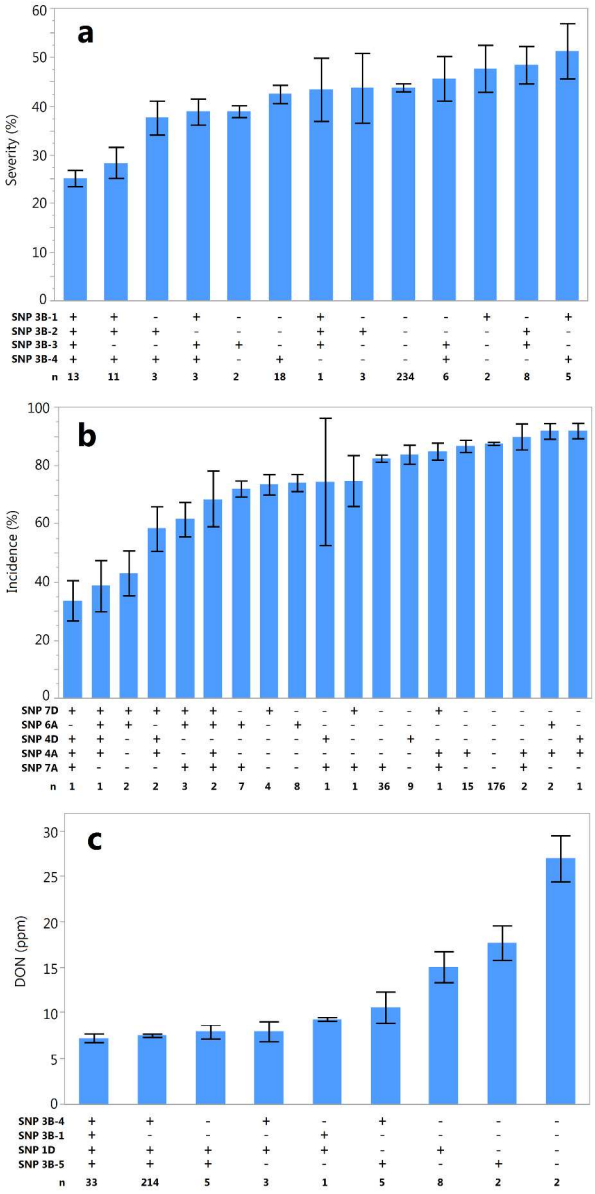

Figure 3. Mean phenotypic value of SEV (a), INC (b), and DON (c) for groups of wheat lines carrying different SNP combinations, with ("+") signal representing the favorable allele. Lines were evaluated in 2011, 2013, and 2014 in Urbana-IL. SNP 3B-1 = IWGSC_CSS_3B_ scaff _10676713_7175, SNP 3B-2 = IWGSC_CSS_3B_ scaff _10352272_5482, SNP 3B-3 = IWGSC_CSS_3B_ scaff _10698462_2332, SNP 3B-4 = IWGSC_CSS_3B_ scaff _10699215_3620, SNP 7D = IWGSC_CSS_7DS_ scaff _3876750_2023, SNP 6A = IWGSC_CSS_6AL_ scaff _5780077_12152, SNP 4D = IWGSC_CSS_4DS_ scaff _2300354_4482, SNP 4A = IWGSC_CSS_4AL_ scaff _7146617_11335, SNP 7A = IWGSC_CSS_7AS_ scaff _4132011_1400, SNP 3B-5 = IWGSC_CSS_3B_ scaff _10413672_4839, SNP 1D = IWGSC_CSS_1DS_ scaff _1879930_3352, SNP 3B-6 = IWGSC_CSS_3B_ scaff _10764618_2168.
545x1052mm (120 x 120 DPI)