

GENOMIC SELECTION AND QUANTITATIVE TRAIT LOCI MAPPING FOR
FUSARIUM HEAD BLIGHT RESISTANCE IN WHEAT (TRITICUM AESTIVUM L.)

BY

MARCIO PAIS DE ARRUDA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Frederic Kolb, Chair
Associate Head Professor Brian Diers
Associate Professor Carl Bradley
Associate Professor Martin Bohn
Assistant Professor Patrick Brown

ABSTRACT

Fusarium head blight (FHB) is an important disease of wheat (*Triticum aestivum* L.), occurring in most growing areas. It is primarily caused by *Fusarium graminearum* Schwabe [teleomorph: *Gibberella zeae* Schw. (Petch)], in North America, and the majority of current wheat cultivars are susceptible to it. Significant economic losses are associated with FHB since it results in yield reduction, poor kernel quality, and grain contamination by mycotoxins. Resistance to FHB has been identified in the wheat gene pool, but breeding for it remains a challenge for several reasons, including the complex genetic control of resistance and poor adaptability of the traditional sources. In this context, molecular markers could contribute to the identification of genomic regions associated with FHB resistance. In addition, markers could be used to calculate breeding values for wheat lines, for traits related to resistance disease. In the first study of this dissertation, genomic selection (GS) models were compared for predicting traits associated with resistance to FHB resistance, using 273 breeding lines in use at the University of Illinois' soft red winter wheat breeding program. Genotyping-by-sequencing (GBS) was used to identify 5,054 single nucleotide polymorphisms (SNPs) which were then treated as predictor variables in GS analysis. Different parameters affecting the prediction accuracy of the genomic estimated breeding values (GEBVs) were tested, including: i) five genotypic imputation methods (random forest imputation – RFI, expectation maximization imputation – EMI, k-nearest neighbor imputation – KNNI, singular value decomposition imputation – SVDI and the mean imputation – MNI); ii) three statistical models (ridge regression best linear unbiased predictor – RR-BLUP, least absolute shrinkage and operator selector – LASSO, and elastic net); iii) marker density ($p = 500, 1500, 3000, \text{ and } 4500$ SNPs); and iv) training population size ($n_{TP} = 96, 144, 192, \text{ and } 218$). No discernable differences in

prediction accuracy were observed among imputation methods. For five of six traits, RR-BLUP outperformed other statistical models (LASSO and elastic-net), and a significant reduction in prediction accuracy was observed when marker number decreased to 3000 or 1500 SNPs, depending on the trait. Lastly, prediction accuracies decreased significantly when the sample size of the training set was less than 192. The second study consisted in a genome-wide association study (GWAS) performed on the same panel used in the first study. A total of 19,992 SNPs were obtained with GBS and ten significant SNP-trait associations were detected for multiple parameters associated with FHB resistance on chromosomes 4A, 6A, 7A, 1D, 4D, 7D, and multiple SNPs were associated with *Fhb-1* on chromosome 3B. *Fhb-1* is a major effect QTL identified in China, and it is very popular among wheat breeders worldwide. The genomic region on chromosome 6A appears to be new, as no other study reported QTL for that region. In addition, combination of favorable alleles of these SNPs resulted in lower levels of disease. The third study compared marker-assisted selection (MAS) with GS using different sets of genotypic data, including the QTL identified in the second study and *Fhb-1*. GS greatly outperformed MAS, with cross-validated prediction accuracy varying from 0.24 to 0.74 and from 0.59 to 0.98 for MAS and GS, respectively. Treating QTL as fixed effects in GS models resulted in higher prediction accuracy when compared with a GS model with only random effects. For the same selection intensity, GS resulted in higher selection differentials than MAS for all traits. This study indicates that GS is a more appropriate strategy than MAS for FHB resistance. The last study of this dissertation was concerned with a linkage mapping study using a population of 233 recombinant inbred lines obtained from IL97-181 (resistant) and Clark (susceptible). Neither parent possesses the traditional Asian sources of resistance to FHB in their pedigree. A total of 2275 single nucleotide polymorphisms (SNPs) were detected using genotyping-by-sequencing

(GBS) and a genetic map was built covering all 21 wheat chromosomes. Inclusive composite interval mapping (ICIM) analysis identified four genomic regions associated with multiple FHB parameters, across all environments. Four QTL were detected for FHB resistance under field conditions on chromosomes 1B, 2D, 6D, and 7B. Two QTL were associated with type I resistance (6D and 7B), and two were associated with type II resistance (1B and 2D). The percentage of the phenotypic variation explained by these QTL varied between 6.7 and 12.5%. For QTL on sub-genome B, intervals smaller than 2 cM were obtained. The results show that elite germplasm can contribute to FHB resistance.

ACKNOWLEDGMENTS

Above all, I would like to thank my wife for her unconditional love and support. Thank you for making my dreams yours; my brothers for their friendship, and my parents for praying for me.

I would like to thank my advisor Dr. Fred Kolb for his guidance and friendship. Thank you for making my doctoral studies possible.

I acknowledge my committee members Drs. Martin Bohn, Carl Bradley, Patrick Brown, and Brian Diers for their thoughtful criticism.

I am grateful to Monsanto for my fellowship and the US Wheat and Barley Scab Initiative for supporting our research.

I would like to thank Brad Foresman, Allison Krill, Norman Smith, Elias Handal, Eric Brucker, and all students in our group for their friendship and assistance.

I would like to thank Dr. Gina Brown-Guedira and Dr. Alex Lipka for their technical support.

I am grateful to Dianne Carson, Megan Ward, Wendy White, and Tracey Malkovich for making my life easier at the university.

To my friends Marcio, Michele, and Nicole Megda, David Walker, Steve Clough, Abraham, Ophelia, Samuel and Edward Akpertey, Dmitry, Elena, and Alex Tartakovsky, Debora Pfeiffer, Emerson Sebastião, Adriano Mastrodomenico, Luciano Santos, Narjara Catelmo, Samuel Fernandes, Vitor Freitas, Jair Unfried, Ezequiel de Oliveira, Rodrigo Sampaio, Felipe Zambonato and family, Ismael Duarte and family, and the Zimbres family.

Thank you Marcelo Pacheco, Luiz Carlos Federizzi, and José Baldin Pinheiro for helping me at the beginning of this journey.

TABLE OF CONTENTS

OBJECTIVES	1
CHAPTER 1: LITERATURE REVIEW	2
Wheat and the importance of Fusarium head blight	2
Disease cycle	5
FHB management	7
Wheat resistance to FHB	8
Breeding for FHB resistance	9
Quantitative trait loci mapping for FHB resistance	11
Genomic selection	13
Genotyping-by-sequencing	18
CHAPTER 2: GENOMIC SELECTION FOR PREDICTING FUSARIUM HEAD BLIGHT RESISTANCE IN A WHEAT BREEDING PROGRAM	21
Abstract	21
Introduction	22
Materials and methods	25
<i>Plant material and phenotypic data</i>	25
<i>Genotypic data</i>	27
<i>Imputation methods</i>	27

<i>Assessing the population structure level</i>	28
<i>Calculation of genomic-estimated breeding values</i>	28
<i>Genomic selection prediction accuracies</i>	30
<i>Number of SNP markers</i>	30
<i>Training and validation populations</i>	31
<i>GEBVs and PEBVs</i>	31
Results	32
<i>Genotypes, structure and imputation methods</i>	32
<i>GS models and SNP number</i>	33
<i>Training population size</i>	33
<i>GEBVs and PEBVs</i>	34
Discussion	35
<i>Genotypes and imputation methods</i>	35
<i>GS models and SNP number</i>	36
<i>Training population size</i>	37
<i>GEBVs and PEBVs</i>	38
Conclusions	39
CHAPTER 3: GENOME-WIDE ASSOCIATION MAPPING OF FUSARIUM HEAD BLIGHT RESISTANCE IN WHEAT (<i>Triticum aestivum</i> L.) USING GENOTYPING-BY-SEQUENCING	40

Abstract	40
Introduction	41
Materials and methods	44
<i>Plant material and disease assessment</i>	44
<i>Phenotypic data analysis</i>	45
<i>Genotypic data</i>	46
<i>Genome-wide association analysis</i>	47
Results	48
<i>Phenotypic data</i>	48
<i>Genotypic data and population structure</i>	49
<i>Marker-trait associations</i>	49
<i>FHB-related traits</i>	50
<u><i>Severity</i></u>	50
<u><i>Incidence</i></u>	51
<u><i>DON</i></u>	52
<u><i>Disease indexes</i></u>	52
Discussion	52
Conclusion	57

**CHAPTER 4: COMPARING GENOME-WIDE ASSOCIATION MAPPING AND
MARKER-ASSISTED SELECTION FOR FUSARIUM HEAD BLIGHT IN WHEAT**

<i>(Triticum aestivum L.)</i>	59
Abstract	59
Introduction	60
Materials and methods	62
<i>Plant material and phenotypes</i>	62
<i>Genome-wide association analysis</i>	63
<i>Comparison of models</i>	64
<i>Prediction accuracy</i>	65
<i>Selection differential</i>	65
Results	66
Discussion	67
Conclusion	69

**CHAPTER 5: MAPPING QTL ASSOCIATED WITH NATIVE RESISTANCE TO
FUSARIUM HEAD BLIGHT IN WHEAT (*Triticum aestivum L*)**

Abstract	70
Introduction	71
Materials and methods	73
<i>Plant material</i>	73

<i>Phenotypes</i>	74
<i>Genotyping-by-sequencing</i>	75
<i>Linkage map and QTL detection</i>	76
Results	77
<i>Phenotypes</i>	77
<i>Linkage map and QTL identification</i>	77
Discussion	78
Conclusion	80
FIGURES	82
TABLES	98
REFERENCES	109

OBJECTIVES

My objectives in this dissertation were to:

- i. Assess the accuracy of genomic predictions for parameters associated with FHB resistance and how accuracy is impacted by different factors;
- ii. Perform a genome-wide association study (GWAS) for detecting QTL associated with FHB resistance;
- iii. Compare marker-assisted selection (MAS) and genomic selection (GS);
- iv. Perform linkage mapping for native resistance to FHB using a bi-parental population.

CHAPTER 1: LITERATURE REVIEW

Wheat and the importance of Fusarium head blight

Wheat (*Triticum aestivum* L) is a self-pollinating, allohexaploid plant with chromosome number equal to $2n = 6X = 42$. The 17-gigabase pair genome is divided in three sets of seven homeologous chromosomes, with each set belonging to subgenomes A, B, and D (IWGSC, 2014). The crop is one of the most important staple grains worldwide. The economic and social importance of wheat is derived from its magnitude of production and ability to generate diverse foods (McFall and Fowler, 2009). In 2013-2014, the world wheat production was estimated to be 716.1 million metric tons, and the United States was responsible for 58.1 million metric tons, the third largest production for a single country. China and India were responsible for 121.9 and 93.5 million metric tons, respectively. Considering exportation, in the same period the US ranked number one with 32 million metric tons exported, which corresponded to 19% of the total wheat exported in the world. In Illinois, winter wheat was cultivated on 271,000 hectares, mainly in the southern area of the state (USDA, 2015). Wheat is an important crop for Illinois, especially for the southern area where it can be used for double-cropping with soybeans.

Fusarium head blight (FHB), also known as head scab or ear blight, is one of most destructive diseases of wheat. It can be caused by many different species of *Fusarium* such as *Fusarium graminearum* Schwabe [telemorph: *Gibberella zeae* Schw. (Petch)], *F. culmorum*, *F. avenacum*, *F. poae* and *Microdochium nivale* (formerly classified as *F. nivale*) (Parry et al., 1995). *Fusarium graminearum*, the prevailing species in North America, causes FHB on wheat, barley, oats, rye, and rice.

In the US, FHB was first described by Chester (1890) in Delaware. One year later, the disease was reported by Arthur (1891) in Indiana. By the first decades of the 19th century, the disease was already reported across Europe, Asia and Australia (Stack, 2003). During the early 1900's, several outbreaks of the disease were reported in the Midwest US, as in 1905 and 1907 in Minnesota (MacInnes and Fogelman, 1923). In 1919, several states reported significant losses, especially Illinois and Iowa (Dickson and Mains, 1929). Epidemics of FHB continued to take place for decades in the US as well as in other wheat growing countries (Stack, 2003). From 2003 to 2005, FHB outbreaks were reported in many wheat producing areas of the US again (McMullen et al., 2008), and FHB continues to be one of the most important diseases of wheat in both temperate and sub-tropical wheat growing areas.

Recent epidemics of FHB caused unprecedented economic losses and prolonged hardships for producers and rural communities (McMullen, 2003). According to Wood et al. (1999), FHB is the most important plant disease to hit the nation since the stem rust epidemics of the 1950's. From 1992 to 1993, yield losses of 95, 43 and 18 million bushels were estimated for North Dakota, Minnesota and South Dakota, respectively (McMullen et al., 1997). In northeastern North Dakota, an average yield reduction of 45% was reported for the 1993 epidemic. During the 1990's, American farmers lost 500 million bushels of wheat due to FHB, an amount equivalent to \$2.5 billion (Windels, 2000). In a meta-analysis involving 77 studies, Madden and Paul (2009) estimated a yield reduction of 0.038MT/ha per unit of disease index. Yield losses can be attributed to failed kernel development or to shriveled kernels with lower test weight. In addition, *F. graminearum*-damaged kernels, low test weight and mycotoxin contaminated grain severely reduce wheat market price (McMullen et al. 2012).

Severely damaged kernels are called “tombstones”, which are lightweight and frequently lost during harvest (Adams, 2010). Besides that, *F. graminearum* infection leads to poor quality wheat flour and low baking performance. Dexter et al. (1996) showed that FHB negatively impacted flour refinement (ash and color). Some cultivars showed reduced baking performance as a result of the disease, and test weight and kernel weight were inversely related to FHB. Boyacioğlu and Hettiarachcy (1995) demonstrated compositional changes as a consequence of FHB. Moderately diseased kernels were found to have reduced cellulose, amylase and hemicelluloses content and higher total protein content (but less glutenin and albumin) when compared to an uninfected wheat control.

Mycotoxin contaminated grain is another important concern in *F. graminearum*-infected wheat. The pathogen produces potent animal toxins such as trichothecenes and estrogenic metabolites. Deoxynivalenol (DON) is the most common trichothecene found in FHB-affected wheat grain. DON is especially harmful to monogastric animals, and its effects include vomiting, diarrhea, nausea and feed refusal by swine (Bennett and Klich, 2003), and decreased performance of cattle and other animals. The effects of zearalenone (ZEA) include reduced litter size, abortion and disfigured sexual organs in animals. Tiemann et al. (2008) found evidence of spleen dysfunction (hemosiderosis) in sows fed ZEA and DON contaminated wheat. In humans, ZEA may be associated with endometrial adenocarcinoma (Tomaszewski et al. 1998). Also, *F. graminearum* mycotoxins have been linked with alimentary disorders such as Akakabi toxicosis, a disease characterized by vomiting, anorexia and convulsions (Bennett and Klich, 2003).

Although some industrial treatments have been found to reduce the levels of mycotoxins in the grain, no single method has been developed to equally reduce the wide range of mycotoxins that may occur simultaneously in the commodities (Shapira, 2004). Since

trichothecenes such as DON are heat-stable molecules, they may survive the food production process. Consequently food cereals and beverages are frequently found to contain trichothecenes (Hazel and Patel, 2004). In order to protect consumers, many countries have defined maximum tolerable levels that are unlikely to be of health concern (Kuiper-Goodman, 2004). In the US, the Food and Drug Administration recommends DON maximum tolerable level of 1 ppm for finished wheat products for human consumption, 10 ppm on grains and grain by-products and 30 ppm in distillers and brewer grains.

Disease cycle

Fusarium head blight is primarily a monocyclic disease. The fungus can survive on living plants of many different crops such as wheat, barley, oats and corn (Bai and Shaner, 2004). According to Farr et al. (1989), species of several genera serve as hosts to *F. graminearum*, including: *Avena*, *Bromus*, *Cucumis*, *Hordeum*, *Secale*, *Sorghum*, *Trifolium*, *Triticum*, *Zea* and others. Crops such as soybean (Baird et al., 1997), potato (Estrada Jr et al., 2010), and sugar beet (Christ et al., 2011) have been shown to host *F. graminearum*. Overwintering structures of the pathogen includes perithecia and sporodochia. Ascospores, macroconidia, chlamydospores, and hyphae all can serve as primary inoculum (Bai and Shaner, 1994); however, ascospores are considered the principal inoculum. During spring, ascospores are discharged from perithecia and are carried out by splashing rain and air currents to wheat heads. Similarly, macroconidia produced from sporodochia are transported to the heads.

Wheat is susceptible to infection from head emergence to harvest; however, it is during anthesis that infections have the greatest destructive potential (Adams, 2010). Strange and Smith (1971) showed that wheat anthers contain fungal growth stimulants that are not present on other flower structures. After spores are dispersed to the wheat spike, germination occurs if conditions

are warm and wet. The optimum temperature range for infection is 20 – 30°C (Sutton, 1982). Germ tubes grow rapidly over floret and glume surfaces forming a mycelia network. The fungus penetrates the plant through the opening between lemma and palea of the spikelet; through the base of glumes; or through stomata. No appressorium formation occurs for direct penetration (Bushnell et al., 2003).

At this early stage of infection, the fungus establishes a biotrophic relationship with the host and starts growing in the apoplast. Floral tissues remain alive and functional, and no trichothecene production occurs. Visual symptoms such as brown or purplish brown lesions on the first floret to flower can be seen at this stage. Then, a subsequent shift from biotrophic to necrotrophic interaction takes place. This shift is probably mediated by DON and other trichothecene mycotoxins (Bushnell et al., 2003). These mycotoxins inhibit protein synthesis by interfering with important plant enzymes (Desjardins, 2006). Although not necessary for the initial infection, DON is a virulence factor and is important for fungal spread. Bai et al. (2001) demonstrated that *F. graminearum* mutants not producing DON cause less severe disease. At this phase, floret discoloration spreads apically and basally in the head (Astanasoff, 1920), and one or more spikelets or even the entire head become prematurely bleached. The fungus spreads from spikelet to spikelet through vascular bundles in the rachis and rachilla. When the rachis is infected, all tissues above that point may be killed (Adams, 2010). The last stage of the fungus infection is characterized by trichothecene accumulation and the death of the floral structure. At this truly necrotic stage, black perithecia can be identified on infected spikelets, nodes and crowns. The grains usually are shriveled, with a “scabby” or chalky appearance.

FHB management

FHB management in the field includes a wide range of strategies. The use of genetic resistance, cultural practices, and fungicides represent important measures for managing the disease (Edwards et al., 2004; Willyerd et al., 2012). Cultural practices such as crop rotation and tillage both reduce the amount of initial inoculum of the pathogen. Guo et al. (2010) demonstrated that reduced tillage and no till both had significantly higher amounts of initial inoculum of *F. graminearum* in spring wheat when compared to conventional tillage. Also, the authors quantitatively demonstrated a reduction in initial inoculum when rotating spring wheat with non-host crops.

The use of fungicides is becoming more common for FHB management; however, relying only on fungicides for FHB control may lead to mixed success since in most cases the level of control is inadequate when susceptible cultivars are used. Poor efficacy of the products, incorrect timing of application, variation among tillers in flowering, timing and severity of infection and virulence of isolates are possible reasons for the lack of control (Mesterházy et al., 2003). Some fungicides such as strobilurins can even increase the level of DON in the grains when compared to untreated control (Oldenbur et al., 2001). Tebuconazole was previously the best fungicide in terms of reducing FHB severity and DON concentration and increasing yield, as demonstrated by Mesterházy et al. (2003). Paul et al. (2008) showed prothioconazole + tebuconazole to be the most effective fungicide for controlling FHB when compared to metconazole, prothioconazole, tebuconazole, and propiconazole. For DON, however, metconazole was the most effective treatment. According to Wegulo et al. (2011), in years with high FHB incidence the disease is better controlled with fungicide combined with moderately resistant cultivars; whereas in years with low to moderate incidence it may not be necessary to

apply fungicide on moderately resistant cultivars. Willyerd (2012) reported that a single tebuconazole + prothioconazole application combined with cultivar resistance is the most effective and stable FHB management practice. More recently, Salgado et al. (2014) reported the efficacy and cost-benefit of different FHB management strategies involving chemical control, cultivar resistance, and combine configuration. In order to assist American farmers on their decision about fungicide application for FHB control, a collaborative research group from multiple US universities developed the Fusarium head blight risk assessment tool (http://www.wheatscab.psu.edu/riskTool_2011.html), which is a FHB forecasting system.

Wheat resistance to FHB

FHB resistant wheat varieties play an important role in FHB management and prevention of mycotoxin contamination. Fortunately, large genetic variability for resistance to FHB is present in the wheat gene pool (Buerstmayr et al. 2009). In contrast to the wheat rusts, complete resistance to FHB is not known. Resistance to *F. graminearum* in small grain crops 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, other physiological resistance types have been described, which include resistance to toxin accumulation (type III), resistance to kernel infection (type IV) and tolerance (type V) (Mesterházy, 1995; Mesterházy, 1999). Also, morphological traits may be associated with FHB resistance such as flower opening (Gilsinger et al., 2005).

Progress has been made towards understanding the biological basis of resistance to FHB in wheat. Ribichich et al. (2000) reported increased callose deposition between the rachilla and rachis in Sumai-3 (resistant) when compared to a susceptible cultivar. This callose deposition

was associated with less spread of the pathogen, generally referred to as type II resistance. The plant defense-related callose biosynthesis can be disrupted by lipase FGL1, a virulence factor of *F. graminearum* (Blümke et al., 2014). Proteins with antioxidant function were found to be induced or upregulated in wheat spikelets of the resistant cultivar Ning7840 only when inoculated with the pathogen (Zhou et al., 2005). Using a metabolon-proteomic approach, Gunnaiah et al. (2012) reported that resistance controlled by *Fhb-1* is mainly associated with cell wall thickening due to phenolic glucosides, hydroxycinnamic acid amides, and flavonoids. Xiao et al. (2013) reported the proteins *PR4*, *PR5*, *ABC* transporter, and the jasmonic acid (JA) pathway were crucial in FHB resistance, especially resistance mediated by the quantitative trait loci (QTL) *Fhb-1*. Kugler et al. (2013) reported G-protein coupled receptor kinases and biosynthesis genes for jasmonate and ethylene to be earlier induced by *Fhb-1*. Resistance to mycotoxin accumulation has been regarded as an important resistance mechanism, which can occur independently from resistance to spread and penetration. It has been shown that some wheat lines can convert deoxynivalenol into the less phytotoxic DON-3-O-glycoside. This conversion is mediated by DON-glucosyl-transferase, and a QTL on chromosome 3B either encodes it or regulates its expression (Lemmens et al., 2005).

Breeding for FHB resistance

Conventional plant breeding has made significant progress in introgressing FHB resistance into wheat genotypes since the first report of the disease by Chester (1890). Arthur (1891) was the first to observe intervarietal variability for FHB susceptibility. Later, MacInnes and Fogelman (1923) reported the prevalence of the disease on the susceptible cultivar Marquis. According to Stack (2003) the “classical” era of research on FHB started in 1908 with the experimental reproduction of the disease by inoculation. Two important problems in screening

for FHB resistance were resolved during the 1930's and the 1940's. One was the development of statistical tools, allowing analysis of experiments with complex effects. The second problem was related to the control of the environment. Devices such as muslin tents were developed in order to keep high relative humidity, which created a favorable microclimate for the pathogen. Despite these advances and the significant amount of research regarding FHB during the last century, the development of high yielding, FHB resistant cultivars is still a challenge for the wheat breeding community. The polygenic inheritance of resistance, poor agronomic performance of most sources of resistance, and the environmental effect on resistance are some of the factors pointed out by Bai and Shaner (2004) as constraining the development of commercial cultivars.

In wheat breeding programs, the worldwide spring wheat sources of resistance to FHB include the Chinese cultivars and lines Sumai 3, Nobeoka Bozu, Shangai 7-31B, Nyubai, Fan 1, Ning 8343, Ning 7840, Pekin 8, the Brazilian cultivars Frontana and Encruzilhada and others (Mesterházy, 2003). These genotypes have shown to have poor agronomic performance (Kohli, 1989). Compared to spring wheat, winter wheat received less emphasis and investment in regard to FHB resistance studies. However, resistance sources for winter wheat were also identified, such as the North American cultivars Freedom, Roane (Liu et al., 2005), Bess, Ernie (McKendry et al., 1995), Truman (McKendry et al., 2005), and the European cultivars Praag 8, Sorbas, Arina and others (Buerstmayr et al. 1996). According to Mesterházy (2003), many successful breeding institutions have demonstrated that developing FHB resistant wheat lines is no longer a scientific problem. As mentioned earlier, the challenge is combining resistance to the disease with other desirable agronomic traits.

Classic genetic studies showed that FHB resistance in wheat is under polygenic (Van Ginkel et al., 1996; Liu et al., 2005) or oligogenic (Bai et al., 2000) control. Several factors may

influence the number of genes found to be controlling FHB resistance, such as the source of resistance, phenotyping procedures and isolates of the pathogen (Kolb et al., 2001). Additive effects have been shown to play a major effect on the trait (Bai et al. 2000; Snijders, 1990), indicating that accumulation of resistance genes may be possible. Although less frequent, dominance and epistatic effects can also be present (Bai et al. 2000). FHB resistance is strongly influenced by the environment and, consequently, heritability estimates are usually low and/or inconsistent. For example, using $F_{2.4}$ families Verges et al. (2006) reported broad-sense heritability estimates varying between 0.3-0.33 for FHB severity and between 0.16-0.20 for Fusarium-diseased kernels (FDK). Snijders (1990) studied 39 F_2 populations segregating for FHB resistance and found broad-sense heritability estimates varying between 0.05 and 0.89, with average of 0.39.

Quantitative trait loci mapping for FHB resistance

Selection for FHB resistance has been performed mainly based on phenotypic data. Since FHB resistance is strongly influenced by the environment, biased measurements of true genetic resistance of an individual frequently occur. Selection based on genotypic data instead of phenotypic data can improve the breeder's ability to select specific genomic segments responsible for improved performance. In this context, marker assisted selection (MAS) has been regarded as a promising breeding strategy. During the last two decades several mapping studies were carried out to identify QTL associated with wheat resistance to FHB, and the most commonly used molecular markers have been: restriction fragment length polymorphisms – RFLP (Waldron et al., 1999), amplified fragment length polymorphisms – AFLP (Bai et al., 1999; Guo et al., 2003), random amplified polymorphic DNA – RAPD (Ban, 2000) and single sequence repeats – SSR (Bonin and Kolb, 2009; Mardi et al., 2005; Jia et al., 2005; Zhou et al.,

2004; Yu et al., 2008). Methods for detecting single nucleotide polymorphism (SNP) have been developed for wheat (Somers et al., 2003) and are now being used in several genetics/breeding studies.

More than one hundred QTL for FHB resistance in wheat were reviewed by Buerstmayr et al. (2009). These QTL were identified through linkage mapping and were localized across all 21 chromosomes, except on chromosome 7D; however, Cativelli et al. (2013) identified QTL for type II resistance for this chromosome. Twenty two QTL reviewed by Buerstmayr et al. (2009) were found independently in different studies, indicating that these QTL may be stable. They are distributed on chromosomes 1B (two regions), 1D, 2A (2), 2B (2), 2D (2), 3A, 3B (2), 3D, 4B, 4D, 5A, 5B, 6A, 7A and 7B (2). Liu et al (2009) conducted a meta-analysis with 245 QTL for FHB resistance published in 45 studies. Confidence intervals were calculated for 43 clusters of QTL on 21 chromosomes, and marker information surrounding these clusters is provided in their work. Löffler et al. (2009) conducted a similar study and the mean location of each cluster is provided. The review by Buerstmayr et al. (2009) and these two meta-analyses represent a rich source of information for QTL associated with the disease.

The great majority of the QTL associated with FHB resistance were identified through linkage mapping, most of the time exploring the polymorphism of only two parents. Genome wide association studies (GWAS) is another method to detect QTL, and it has already been implemented for FHB resistance in wheat. While linkage mapping relies on experimental populations such as recombinant inbred line population from biparental crosses, GWAS are based on diverse panel of unrelated lines, taking advantage of the recombination history of the panel. As consequence, higher map resolutions are expected in GWAS (Myles et al., 2009). In a study with 455 soft winter wheat European lines and 115 SSRs, Miedaner et al. (2011) detected

nine QTL on multiple chromosomes, two of them on genomic regions (1D and 3A) not previously reported. Also working with European germplasm, Kollers et al. (2013) evaluated 358 winter wheat and 14 spring wheat cultivars genotyped with 732 SSRs. QTL were identified in all chromosomes except 6B, the highest number of QTL being detected on chromosome 5B.

Despite the fact that hundreds of QTL for FHB resistance have been identified in wheat, selection of highly resistance individuals remains a challenge for many reasons. First, in QTL studies, only the statistically significant marker/trait associations are considered, leaving small effect QTL out. For this reason, the effect of the significant markers could very well be inflated. In addition, choosing the statistical threshold to determine which markers are significant is arbitrary. Second, considering that FHB resistance is a quantitatively inherited trait and major QTL are not common, selection based on few significant QTL may not be the best strategy for developing resistant lines.

Genomic selection

Genomic selection (GS) or genome wide selection may represent an advance compared to marker-assisted selection (MAS). The GS concept was first proposed by Meuwissen et al. (2001) in animal breeding, and it consists of using all marker information to calculate genomic estimated breeding values (GEBV). Instead of considering only the significant QTL, GS incorporates all loci, regardless of the effect of each locus. This approach is particularly interesting when high-density genome-wide molecular markers are available. The first step in GS is defining the training population, which will consist of individuals having both phenotypic and genotypic information. This group is used to estimate model parameters that will be necessary to calculate the GEBV of selection candidates. Having phenotypic and genotypic data, one can proceed to the second step in GS, which involves estimating marker effects using an appropriate

GS model. Lastly, marker effects are used to obtain the GEBVs in a validation population, for which only genotypic information is available. Lorenz et al. (2011) provides a detailed, comprehensive description of each step.

Genomic selection has traditionally been used on dairy cattle (*Bos taurus*) where there is great availability of SNP markers (Van Raden et al., 2009). In crop plants, the use of genomic selection has been delayed by the number of molecular markers available and the size and complexity of some polyploidy genomes, as is the case of the wheat genome. Nakaya and Isobe (2012) reviewed seven studies involving GS in plants. Among the annual crop plants, the number of markers ranged from 69 to 1339; whereas, for perennial crops, the number ranged from 3120 to 3938 markers. More recently, high density genome-wide markers are becoming available for multiple crops due to the significant reduction in sequencing costs. The pace of this recent plant genomic boom is illustrated by Michael and Jackson (2013), who reviewed the publication of 55 plant genomes, 37 of them published between 2011 and 2013. As plant genomes are becoming available, breeders can have access to a whole new level of genotypic information, which can be incorporated in breeding strategies such as genomic selection.

In genomic selection, markers are used as predictor variables (p) whereas phenotypes are used as the response variable (n). With high density genome-wide markers becoming more available, the number of markers frequently surpasses the number of phenotypes. In other words, the number of predictor variables gets larger than the number of observations, creating the so called “large p small n problem”. In this case, models relying on shrinkage factors or variable selection are needed in order to estimate the effect of each marker. There is an abundance of statistical models being used in the context of genomic selection, with the most popular being ridge regression best unbiased linear predictor (RR-BLUP), which is equivalent to the genomic

best unbiased linear predictor (G-BLUP), least absolute shrinkage and selection operator (LASSO), elastic-net, and Bayesian models (BayesA, BayesB, BayesC, wBSR, BL), (Lorenz et al., 2011; Heslot et al., 2012).

An infinitesimal model is assumed in RR-BLUP, in which markers are assumed to have a common variance (homocedastic) and are equally shrunken towards zero (Meuwissen et al. 2011). That does not mean the marker effects are equal to zero, but the amount of shrinkage is the same for all markers (Bernardo and Yu, 2007; Heffner et al., 2009). The assumption of homocedastic genetic variance for all markers is unrealistic in animal and plant breeding for the great majority of traits. Despite that, RR-BLUP has been extensively used in plant breeding with promising results (Riedelsheimer et al., 2012; Lipka et al., 2014). The LASSO allows some predictor variables to have variance equal to zero and performs continuous shrinkage of the remaining variables (Tibshirani, 1996). For this reason, the LASSO has been regarded as a suitable model when major genes are present; however, the LASSO only selects n predictor variables at most and does not perform well when the variables are correlated. The elastic-net is a combination of both RR-BLUP and LASSO, allowing $p \gg n$ and encouraging grouping of correlated variables (Zou and Hastie, 2005). In a recent study with 540 switchgrass (*Panicum virgatum*) accessions, Lipka et al. (2014) compared these three statistical models and obtained numerically higher prediction accuracies for RR-BLUP. In a simulation study, however, Ogotu et al. (2012) observed higher prediction accuracy for LASSO, elastic net and their extension when compared to RR-BLUP.

Bayesian models differ from the models above by allowing predictor variables to have different variances. In BayesA and BayesB, a marker or predictor variable has effect equal to zero with probability π , and normally distributed around mean equal to zero and a locus-specific

variance with probability $1 - \pi$. In BayesA $\pi = 0$, meaning that no marker can have effect equal to zero. BayesB has a more realistic prior, with $\pi > 0$. The locus-specific variance has a scaled inverted chi-square prior (Meuwissen et al., 2011; Habier et al., 2011). In BayesC, some markers will have common variances whereas others will have variance set to zero. Different from BayesB, π is not fixed, but estimated from the data. Heffner et al. (2011) compared these three Bayesian models with RR-BLUP and other breeding strategies for 13 agronomic traits in wheat. Despite different assumptions, the four models performed equally well, with the mean accuracy for all traits ranging from 0.58 to 0.60. One disadvantage of Bayesian models is that they are computationally demanding.

Recently, new models have been proposed in order to overcome the homocedastic variance requirement of infinitesimal models and still be computationally efficient (Shen et al., 2013; Hofheinz and Frisch, 2014). Another strategy that allows markers to display a non-common variance is by treating them as fixed effects, as successfully demonstrated by Bernardo (2014), Owens et al. (2014), and Zhao et al. (2014). In addition, methods that specifically model marker by environment interaction have been proposed (Jarquín et al., 2014) and tested (Zhang et al., 2015).

In genomic selection, a model that produces high accuracy (r) is desired since accuracy impacts the response to selection, $R = ir\sigma_A$, where i is the selection intensity and σ_A is the additive variance. Accuracy is calculated as $cor(GEBV, EBV)$, Pearson's correlation between GEBV and the true breeding value (EBV), which is not known; however, this correlation is equivalent to $cor(GEBV, PEBV)/\sqrt{h^2}$, in which the phenotypically estimated breeding value (PEBV) and the heritability of the trait are used (Lorenz et al., 2011). Numerous parameters have been shown to affect the estimates of prediction accuracy. Zhong et al. (2009) studied how

estimates of prediction accuracy are affected by marker density, level of linkage disequilibrium (LD), number of QTL affecting the trait, sample size, and level of replication in populations generated by inbred lines. Bernardo and Lorenzana (2009) also showed how accuracy is affected by the size of the training population, marker density, and GS models. Daetwyler et al., (2013) discussed the drivers of genomic prediction accuracy using a framework of deterministic predictions. When several QTL control a specific trait, the prediction accuracy can be calculated with $\rho = \sqrt{N_p h^2 (N_p h^2 + M_e)^{-1}}$, in which N_p is the number of individuals in the training population, h^2 is the heritability of the trait, and M_e is the number of independent chromosome segments. Another factor that influences accuracy is the degree of relatedness between the training and validation population. The more related the two sets are, the higher the accuracies (Riedelsheimer et al., 2013). Over time, new recombination events will affect the linkage disequilibrium between markers and QTL, reducing the predictive ability of the model. Then, the GS model can be updated as more lines are advanced in the breeding pipeline and more phenotypic and genotypic information are obtained, feeding the training set (Heffner et al., 2009).

There are several opportunities where GS can be applied in a breeding pipeline. Heslot et al. (2015) present a simple doubled-haploid based scheme of one breeding cycle and highlight how GS can benefit the breeding process. At the early stages, parent selection and prediction of cross variances could be supported by GS. At the later stages, the method can improve accuracy of expensive phenotype traits and predict genotype by environment interactions. Other schemes for incorporating GS in plant breeding pipelines are presented by Heffner et al. (2009), Rutkoski et al. (2011), and Desta and Ortiz (2014).

Genotyping-by-sequencing

Recently, a new approach has been developed for generating genome-wide markers using next generation sequencing (NGS) technology. Genotyping by sequencing (GBS) consists of a highly multiplexed system for developing reduced representation libraries for the Illumina next-generation platform. By using methylation-sensitive restriction enzymes, the genome of interest is cut in such a way that repetitive regions can be avoided and lower copy regions targeted (Elshire et al., 2011). This genome reduction is followed by a DNA barcoding step, where adapters are attached to the ends of the restriction fragments. These fragments are then sequenced in parallel multiplex samples on a single run of next-generation sequencing platforms (Poland and Rife, 2012). GBS can generate tens of thousands to hundreds of thousands of tags that can be used as dominant markers (presence/absence of each tag) or can also be used for SNP calls. Considering the high complexity and large size of the wheat genome, the GBS methodology would be highly desirable for generating high-density genome-wide molecular markers.

GBS has already been applied to small grains, including wheat and barley (Poland et al., 2012a). The authors developed a high density map for each for these crops using a two-enzyme GBS protocol (a rare cut-site and a common cut-site enzyme) and bi-parental populations. Approximately 34,000 and 20,000 SNPs were identified for barley and wheat, respectively. Interestingly, the authors successfully developed a high density map for both a diploid and an allopolyploid crop species. Now, these maps can be used for genomic assisted selection and precise mapping of agronomically important genes (Poland et al., 2012a). In a panel of 254 lines from CIMMYT, Poland et al. (2012b) identified 41K SNPs using GBS. The authors calculated GEBV for each line and observed an increase in performance of 10-20% when compared to an

established marker platform. Lado et al. (2013) applied GBS to call 102,324 SNPs in a panel consisting of wheat lines from Chile, Uruguay, and CYMMIT. The authors developed GS models for agronomic traits such as heading date, grain yield, thousand-kernel weight, and number of kernels per spike.

In a recent study, Forrest et al. (2014) applied three different genotyping platforms in bi-parental populations and a wheat panel to identify markers linked to *Lr67*, a gene that confers resistance to leaf rust (*Puccinia triticina*). The authors reported that SNP-based markers tightly linked to *Lr67* were obtained with GBS, but not with exhaustive SSR genotyping and the 9K Infinum iSelect array. The poor performance of the 9K SNP chip was attributed to difficulties in distinguishing duplicated gene copies or ascertainment bias caused by pre-selection of SNPs in populations of reduced size.

Talukder et al. (2014) performed a QTL linkage mapping study for heat tolerance in a wheat bi-parental population that had been previously genotyping with AFLPs, SSRs, and ESTs. By adding GBS SNPs, the authors were able to resolve conflicting chromosomal assignment of the linkage groups. More specifically, two linkage groups could not be differentiated based on SSR markers (*Xgwm356* and *Xbarc353*) since the markers mapped to multiple chromosomes, 2A and 6A. The extra GBS SNPs allowed the identification of a single linkage group. In addition, the authors were able to successfully detect multiple QTL associated with heat tolerance.

Langer et al. (2014) conducted a GWAS in a wheat panel to investigate the genetic architecture of flowering time. A total of 410 European winter wheat lines were genotyped for candidate genes (*Ppd-D1*, *Ppd-B1*, and others) and genotyping-by-sequencing was used for SNP calling. Although the gene *Ppd-D1* accounted for most of the variation in the panel, GBS allowed the detection of small effect QTL for flowering time.

Considering the recent advances in DNA marker technology and its application on wheat, the way wheat breeders use markers is expected to change dramatically. The great availability of high throughput markers at decreasing cost per sample will allow precise mapping of agronomically important genes by using high-density marker maps as well as the development of robust GS models in order to calculate GEBVs. These recently developed approaches represent a great opportunity for wheat breeders in dealing with quantitatively controlled traits.

CHAPTER 2: GENOMIC SELECTION FOR PREDICTING FUSARIUM HEAD BLIGHT RESISTANCE IN A WHEAT BREEDING PROGRAM

Abstract

Genomic selection (GS) is a breeding method that uses marker-trait models to predict unobserved phenotypes. This study developed GS models for predicting traits associated with resistance to Fusarium head blight (FHB) in wheat (*Triticum aestivum* L.), using 273 breeding lines in use at the University of Illinois' soft red winter wheat breeding program. Genotyping-by-sequencing (GBS) was used to identify 5,054 single nucleotide polymorphisms (SNPs) which were then treated as predictor variables in GS analysis. In this study, the following parameters affecting the prediction accuracy of the genomic estimated breeding values (GEBVs) were tested: i) five genotypic imputation methods (random forest imputation – RFI, expectation maximization imputation – EMI, k-nearest neighbor imputation – KNNI, singular value decomposition imputation – SVDI and the mean imputation – MNI); ii) three statistical models (ridge regression best linear unbiased predictor – RR-BLUP, least absolute shrinkage and operator selector – LASSO, and elastic net); iii) marker density ($p = 500, 1500, 3000, \text{ and } 4500$ SNPs); and iv) training population size ($n_{TP} = 96, 144, 192, \text{ and } 218$). No discernable differences in prediction accuracy were observed among imputation methods. For five of six traits, RR-BLUP outperformed other statistical models, and a significant reduction in prediction accuracy was observed when marker number decreased to 3000 or 1500 SNPs, depending on the trait. Lastly, prediction accuracies decreased significantly when the sample size of the training set was less than 192. Overall, the moderate to high prediction accuracies observed in this study suggest GS as a very promising breeding strategy for FHB resistance in wheat.

Introduction

Originally proposed by Meuwissen et al. (2001) for animal breeding, genomic selection (GS) predicts breeding values of individuals based on genome-wide molecular markers. It can be considered as a form of marker-assisted selection in which all markers are used to calculate genomic estimated breeding values (GEBVs). It is assumed that all quantitative trait loci (QTL) are in linkage disequilibrium with at least one marker and that all the genetic variance can be explained by markers (Goddard and Hayes, 2007). For this reason, GS is particularly promising for predicting quantitative, complex traits where many small effect loci contribute to phenotypic variation. In GS, marker effects are estimated from a training population (TP), for which phenotypes and genotypes are available. Marker effects are then used to predict phenotypes in a set of individuals that will only be genotyped, the breeding population. Based on GEBVs, individuals can be selected before being tested in field experiments, potentially speeding up the breeding cycle (Jannink et al., 2010; Heffner et al., 2009). Several factors can impact the accuracy of GS models, including marker density, QTL number, sample size, and genotypic imputation methods (Zhong et al., 2009; Heffner et al., 2011; Rutkoski et al., 2013).

Fusarium head blight (FHB) is the single most important wheat disease in the US Midwest, causing yield losses, reduction of grain quality (Dexter et al., 1996), and grain contamination due to mycotoxins. *Fusarium graminearum* (teleomorph: *Gibberella zea*) Schw. (Petch)] is the predominant pathogen that causes FHB of wheat and barley in North America. Classic genetic studies showed that wheat resistance to FHB is quantitatively inherited (Van Ginkel et al., 1996; Liu et al., 2005). Previous studies suggest that the sources of genetic variation for FHB resistance are predominantly additive (Bai et al. 2000; Snijders, 1990), indicating that accumulation of resistance genes may be possible. Selection for FHB resistance is

performed mainly using phenotypic data, and because FHB resistance is strongly influenced by the environment, inaccurate measurements of true genetic resistance of an individual frequently occur. Selection based on GEBVs instead of phenotypes could improve the breeder's ability to select individuals with superior FHB resistance. Although over 100 QTL for FHB have been reported in the literature (Buerstmayr et al. 2009), breeding for FHB resistance remains an important challenge as most QTL detected to date explain less than 15% of the variation.

Although the availability of genotypic information has greatly improved over the past few decades, phenotyping capabilities have not kept pace. Phenotyping for FHB resistance is usually a laborious task, requiring mist irrigation systems in the field or greenhouse capabilities. Some measurements, like mycotoxin analysis, are only obtained after harvest and are time consuming. As a consequence, breeders are now facing the so-called "large p , small n problem" (i.e., $p \gg n$) when applying markers to predict phenotypes. In this situation the number of predictor variables is larger than the number of observations, resulting in an infinite number of marker effect estimates (Gianola, 2013). To address this problem, several penalized regression approaches have been proposed for GS models. Ridge regression best linear unbiased predictor (RR-BLUP) is based on an infinitesimal model, in which all markers are equally shrunken towards zero. This model sets markers to be random effects with a common variance (Meuwissen et al., 2001). The least absolute shrinkage and selection operator (LASSO) performs continuous shrinkage and variable selection simultaneously. In the $p \gg n$ setting, LASSO will select n variables at most and set the effects of the remaining predictors equal to zero (Tibshirani, 1996). The elastic net is a combination of both RR-BLUP and LASSO, where the penalty is a weighted average of the penalties from these two approaches (Zou and Hastie, 2005).

In a previous publication, Rutkoski et al. (2012) compared GS models for FHB-related traits in wheat using 2402 DArT and 38 single sequence repeat (SSR) markers. This study involved germplasm from 18 different breeding programs at the USA and Canada. Heffner et al. (2011) used 1158 Diversity Array Technology (DArT) markers to compare genomic selection models, marker assisted selection, and phenotypic selection for agronomic traits, using germplasm from the Cornell University wheat breeding program (Ithaca, NY). Both studies showed GS as a promising strategy for wheat breeding.

With the advent of high-throughput single nucleotide polymorphism (SNP) genotyping methods, high-density genome-wide marker sets are becoming viable in several crop species, making GS a practical approach for predicting complex traits. A robust, high throughput genotyping method called genotyping-by-sequencing (GBS) has been applied to wheat (*Triticum aestivum* L.) with promising results (Poland et al., 2012a). Briefly, this technique consists of reducing genome complexity using restriction enzymes that target gene-rich regions (Elshire et al., 2011). The targeted segments are then polymerase chain reaction (PCR) amplified, barcoded and sequenced in a multiplexed reaction. The protocol can be modified to accommodate a combination of enzymes, as performed in wheat and barley by Poland et al. (2012b). GBS typically generates many thousands of SNPs with minimum of ascertainment bias, a common problem of SNP-chips.

Resistance to FHB is an important breeding objective in most programs in the Corn Belt, including the soft red winter wheat program of the University of Illinois at Urbana-Champaign. This study aimed to develop genomic models for FHB-related traits, using GBS. Three penalized models were compared: RR-BLUP, LASSO, and elastic net. This study also investigated how the

GS accuracies are impacted by imputation methods, number of SNP markers, the size of the training population, and the number of folds used in cross validation.

Materials and methods

Plant material and phenotypic data

The germplasm used in this study consisted of 273 breeding lines involving 233 different crosses. A total of 185 lines came from the University of Illinois soft red winter wheat program, and 88 lines belonging to 17 programs across the midwestern and eastern US. Of these 88 lines, 50 have been used as parents at the University of Illinois. Hence, this collection is representative of the genetic diversity currently being explored at the University of Illinois. Measurements associated with FHB resistance were collected from one-meter long single rows, cultivated in a field nursery during 2011, 2013 and 2014 field seasons. Due to drought conditions no phenotypic data were obtained in 2012. Each experiment was set up as a randomized complete block design with two replicates. Since not all lines were present in 2011, the experiment was analyzed as an unbalanced design. Grain spawn inoculum was prepared with a mixture of ten different isolates of *F. graminearum* collected throughout Illinois over several years. Agar plugs containing the isolates were placed in carboxymethyl-cellulose liquid media, and allowed to grow and sporulate (Tuite, 1969). After seven days, a conidial suspension was obtained and mixed to autoclaved maize grain. Colonized grains were then spread in the field at a rate of approximately 287kg.ha⁻¹, two to four weeks before anthesis. The field was mist irrigated three times per day during six weeks, starting two weeks before the first wheat line headed. The following traits were used to evaluate FHB resistance: incidence (INC), severity (SEV), FHB index (FHBdx; [(incidence x severity)/100], Fusarium-damaged kernels (FDK), incidence-severity-kernel index (ISK; [(0.3 x incidence + 0.3 x severity + 0.4 x FDK)] and deoxynivalenol concentration (DON).

INC was estimated as the percentage of infected heads, irrespective of the severity of disease on each head. SEV was visually estimated as the percentage of infected spikelets in a wheat head. INC and SEV are used to quantify the resistance to penetration and spread of the disease, respectively, and they are also known as type I and type II resistance (Schroeder and Christensen, 1963). FDK is a visual estimate of the percentage of *F. graminearum*-damaged, tombstone kernels using known standards, and it is often classified as type III resistance (Mesterházy 1995, Merterházy et al., 1999). FHBnx and ISK are disease indexes routinely used by breeders when performing selection. SEV and INC were recorded one month after the heading date of each breeding line, whereas FDK and DON were recorded after harvest. Resistance to toxin accumulation has been classified as type IV resistance (Miller et al., 1985) and deoxynivalenol (DON) is by far the most important mycotoxin produced by *F. graminearum* in wheat. Its concentration was determined by gas chromatography-mass spectrometry at the Department of Plant Pathology at the University of Minnesota by Dr. Yanhong Dong.

For each line and each trait, best linear unbiased predictions (BLUPs) were calculated using PROC MIXED SAS version 9.4 (SAS Institute 2013), according to equation (i):

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

in which Y_{ijk} 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 for the k th line in the j th block within the i th year, $year \times line_{ik}$ is the random effect of the interaction between the i th year and k th line, and ε_{ijkl} is the random error term.

Genotypic data

DNA was extracted from 5-day-old leaves using a cetyltrimethyl ammonium bromide (CTAB)/chloroform protocol. After quantification, the GBS libraries were constructed according to Poland et al. (2012b), with a few modifications. For genomic complexity reduction, three two-enzyme combinations were used: *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. A different set of barcodes were used for each enzyme combination. Sequence data were obtained from 96-plex Illumina HiSeq2000 runs (W.M. Keck Center for Comparative and Functional Genomics).

SNPs were called using the UNEAK pipeline (Lu et al., 2013) using a population-based approach similar to the one described by Poland et al. (2012b). A mismatch of 1 bp in a 64 bp tag was allowed when aligning the tag sequences using BWA (Li and Durbin, 2009). Putative SNPs were identified from tags that differed by 1 bp, and then filtered with the Fisher's exact test ($p < 0.001$), which tests whether two alleles are independent in a population of inbred lines. The minor allele frequency cut off was set to 5% and SNPs with more than 20% missing data were eliminated from the analysis.

Imputation methods

Genotypic missing values were imputed using five different methods: mean imputation (MNI), k -nearest neighbors imputation (KNNI) (Troyanskaya et al. 2001), expectation maximization imputation (EMI) (Dempster et al. 1977), random forest regression imputation (RFI) (Stekhoven and Bühlmann 2011), and singular value decomposition (SVDI). These methods are described in detail by Rutkoski et al (2013). The imputations were performed using the R scripts provided by the same authors. The methods were compared in terms of GS

accuracy for all traits, using the RR-BLUP model. All mean comparison tests were performed in SAS PROC GLM using the Ryan-Einot-Gabriel-Welch (REGWQ) multiple testing correction at $\alpha = 0.05$ level. The REGWQ correction conservatively controls the familywise error.

Assessing the population structure level

Population structure is an important factor in genomic selection as it can affect the prediction of breeding values (Lipka et al., 2014; Isidro et al., 2015). Biased estimates of genomic selection accuracy can be obtained when population structure is not taken into account (Riedelsheimer et al., 2013). For these reasons, the level of structure of the germplasm was assessed before running any GS analyses. A relationship matrix was built in GAPIT (Lipka et al., 2012) using 5054 SNPs. Principal component analysis (PCA) was performed with the same SNPs in order to detect genetic structure in this collection of lines. The PCA was performed in JMP Genomics 7 (SAS Institute 2015).

Calculation of genomic-estimated breeding values

GEBVs were calculated by the general equation (Lorenz et al., 2011):

$$g(x_i) = \sum_{k=1}^p x_{ik} \beta_k, \quad (\text{ii})$$

where $g(x_i)$ is a function relating phenotypes to genotypes, x_{ik} is the score of the k^{th} SNP (coded additively as -1, 0, or 1) for the i^{th} individual, and β_k is the effect of the k^{th} SNP. For each trait, three models were tested: ridge regression best unbiased prediction (RR-BLUP) (Hoerl and Kennard, 1970; Meuwissen et al., 2001), least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), and elastic net (Zou and Hastie, 2005). Shrinkage models differ from ordinary least square (OLS) regression by adding a penalty to the cost function. If in OLS

regression the cost function is represented by the sum of squared residuals (i.e., $\sum e_i^2$), in shrinkage models the general cost function is given by equation (iii):

$$\frac{1}{2} \sum_i e_i^2 + \frac{\lambda}{2} \sum_k |\beta_k|^q, \quad (\text{iii})$$

where λ is the regularization parameter, and the term q is equal to 2 for RR-BLUP, 1 for LASSO, and $0 < q < 1$ for elastic net (Lorenz et al. 2011). While in OLS regression the parameter estimators are calculated with $\underset{\beta}{\operatorname{argmin}} \|Y - X\beta\|_{ln}^2$, the parameter estimators for RR-BLUP,

LASSO, and elastic net are given by equations (iv), (v), and (vi), respectively:

$$\underset{\beta}{\operatorname{argmin}} \left(\frac{\|Y - X\beta\|_2^2}{2\sigma^2} + \lambda \|\beta\|_2^2 \right), \quad (\text{iv})$$

$$\underset{\beta}{\operatorname{argmin}} \left(\frac{\|Y - X\beta\|_2^2}{2\sigma^2} + \lambda \|\beta\|_1 \right), \quad (\text{v})$$

$$(1 + \lambda_2) \underset{\beta}{\operatorname{argmin}} \left(\frac{\|Y - X\beta\|_2^2}{2\sigma^2} + \lambda_2 \|\beta\|_2^2 + \lambda_1 \|\beta\|_1 \right), \quad (\text{vi})$$

where the notation $\|\cdot\|_2$ is used for the L_2 - norm loss function (Euclidian norm) $\|\beta\|_2 = (\sum_i \beta_i^2)^{1/2}$, $\|\cdot\|_1$ is used for the L_1 norm (Manhattan norm) $\|\beta\|_1 = \sum_i |\beta_i|$, and $\lambda \geq 0$ is the tuning (penalty or regularization) parameter which regulates the strength of shrinkage of the estimates. The L_1 part is responsible for automatic variable selection, whereas the L_2 part does grouped selection, encouraging grouping of highly correlated variables. The expression “ $\underset{\beta}{\operatorname{argmin}}$ ” refers to the solution of coefficients β that minimizes the equation inside the brackets.

The terms λ_1 and λ_2 refer to the tuning parameters associated with the L_1 and L_2 norm penalties (Heslot et al., 2012; Ogutu, 2012). The “rr-BLUP package (Endelman, 2011)” in the R

programming language was used to conduct RR-BLUP, while LASSO and elastic net were conducted in the “glmnet R package (Friedman et al., 2010)”.

Genomic selection prediction accuracies

For each GS model and trait, the prediction accuracy was calculated as follows:

$$\frac{r(GEBV:PEBV)}{\sqrt{h^2}}, \quad (\text{vii})$$

where r is the Pearson’s correlation between the GEBVs and the phenotypically estimated breeding values (PEBVs) in the validation population (Dekkers 2007; Albrecht et al., 2011; Zhao et al., 2012), and h^2 is the broad-sense heritability on a line mean basis (Fehr, 1991). The variance components estimated from PROC MIXED in SAS (described in “Plant Material and Phenotypic data”) were used to calculate these heritabilities.

Number of SNP markers

The number of markers used in GS has a strong impact on the time required to run each analysis. Thus, it was assessed how a reduction on the number of markers would impact the GS accuracies. Sixty randomly sampled marker datasets, each consisting of either 500, 1500, 3000, or 4500 SNPs, were drawn from the original genotypic data. For each combination of GS model (RR, LASSO, and elastic net), marker set, and trait, GS was performed using 5-fold cross validation, resulting in 5 values of prediction accuracy for each of the 60 runs. For each trait, the mean prediction accuracy was compared among GS models for the marker sets containing the same number of SNPs, and between numbers of SNPs within the same GS model. The exact same folds were used to compare the prediction accuracy of different GS models.

Training and validation populations

Marker effects estimated from the training population (TP) are used to calculate GEBVs for individuals in the validation population (VP). For a fixed number of breeding lines, one needs to decide how many lines will be part of the TP and VP. Therefore, the impact of the TP size ($n_{TP} = 96, 144, 192, \text{ and } 218$) on prediction accuracy was assessed in this study. Also, different proportions of the TP and VP were by evaluated by comparing different number of folds in cross validation. In 2-fold cross validation the population is equally divided in two groups (VP and TP), yielding a proportion of 0.5 of the TP in relation to the total. When 3-fold cross validation is used, this proportion is 0.66, as two thirds of the lines are assigned to the training set. Prediction accuracy was compared for 2-fold, 3-fold, 5-fold, and 10-fold cross validation. For each scenario, 150 analyses were performed, resulting in different number of accuracy values, depending upon the number of folds. For instance, 2-fold cross validation resulted in 300 values. Then, 300 values of accuracy were randomly selected from the 3, 5, and 10-fold schemes and their means were compared in SAS PROC GLM using the REGWQ procedure at $\alpha = 0.05$ level.

GEBVs and PEBVs

After determining the best imputation method, GS model, SNP number, TP size and proportion of TP/total, marker effects were estimated using cross validation, and a total of 300 estimated were obtained for each marker. Then, the mean effect of each marker of obtained and used to calculate GEBVs for all 273 lines, for each trait. As a final step, the bias of the predictions was estimated as the slope (b) of the linear regression of PEBVs on GEBVs, as performed by Zhang et al. (2014).

Results

Genotypes, structure and imputation methods

More than 30,000 SNPs were called this collection of breeding lines using the UNEAK pipeline. After applying a Fisher's exact test cutoff ($p < 0.001$), a cut off no more than 20% missing data and a minor allele frequency cutoff of $> 5\%$, the number of SNPs was reduced to 5054 with 12.57% overall missing data.

No clear genetic structure was detected in this population, as revealed by Figure 1. This finding is not surprising as 86% of the lines belong to the University of Illinois breeding program or have been used in its crossing block. This low level of population structure was also detected by the principal component analysis, with the first and second eigenvectors explaining only 4.6% and 2.8% of the total variability, respectively.

Four imputation methods were used to generate different genotypic data sets, which were then compared in terms of accuracy of GS modeling. As shown in Figure 2, the methods performed equally well for all traits. The EMI method was numerically superior for two traits (FDK and INC), KNNI was superior for two traits (DON and ISK) and SDVI resulted in numerically higher accuracies for FHBdx and SEV. Although no statistically significant difference was detected among the imputation methods for most traits, the methods differed considerably in terms of computation time required to impute missing genotypic data, with RFI being by far most time demanding (over 8 hours), followed by EMI (1min and 10s), SVDI (< 1min), KNNI (<1min) and MNI (<0.5min). These analyses were performed in a Dell Precision with CORE i7 vPro 3.00GHz, 8GB RAM, 700GB HD. In a study involving five populations and a super computer set, Rutkoski et al. (2013) also found the RFI method to be substantially more computationally demanding than the other methods. The EMI approach was selected for

subsequent analysis because it gave numerically higher accuracies for two traits, it was not very computationally demanding, and it has been used in other wheat genomic selection studies (Poland et al., 2012b; Lado et al., 2013, Rutkoski et al., 2013).

GS models and SNP number

Moderate to high values of cross-validated accuracies were observed for all traits, with the highest accuracies observed for FDK, ranging from 0.73 to 0.88, and the lowest for SEV, ranging from 0.46 to 0.59. As a general trend, the accuracies increased with the number of SNPs, albeit at different rates depending on the GS model and trait (Figure 3). For instance, a plateau was observed for ridge regression after 1500 (FDK and DON) or 3000 SNPs (SE, INC, and ISK), but not for FHBdx. The LASSO and the elastic net behaved very similarly to each other, showing a different pattern from the RR-BLUP. No plateau was observed when they were used for FHBdx, FDK, and ISK. For all three models, the trait with the least increase in accuracy was DON.

For the same number of SNPs, RR-BLUPs outperformed LASSO and elastic net for all traits, except INC. For this trait, the GS models provided non-significantly different results when SNPs were equal to 3000 or higher. According to Zou and Hastie (2005), elastic net is expected to outperform the LASSO when $p \gg n$. In this study, that was indeed the case for all traits, but the differences among these models were negligible. Considering the overall performance of the GS models, RR-BLUP was selected as the model of choice for subsequent analyses.

Training population size

Increasing the size of the training population (TP) resulted in higher accuracies for all traits (Figure 4). For most traits, the greatest increase occurred when TP changed from 96 to 144 breeding lines. On average, accuracies were boosted by 16.6% across all traits. The rate of gain

decreases as the TP size increases, reaching a plateau between 192 and 218 for all traits, except FHBdx. Among all traits, FDK showed the highest accuracy values, followed by ISK and DON. Even with only 96 individuals in the TP, FDK showed mean accuracy of 0.76. The lowest values were obtained for FHBdx, ranging from 0.35 ($n_{TP} = 96$) to 0.61 ($n_{TP} = 218$). In a study with agronomic and quality traits in wheat, Heffner et al. (2011) reported a decrease of the mean accuracy across all traits by 30% when n_{TP} decreased from 288 to 96, which is an average reduction of 0.156% per individual removed from the TP. In this study, the mean decreased by 20.51% by changing n_{TP} from 218 to 96, resulting in an average reduction of 0.168% per individual.

In another series of analyses, the effect of the proportion of the TP in relation to the total number of lines was assessed by changing the number of folds used in cross-validation. Accuracies increased linearly with the proportion of TP until 0.8 or 0.9, except for DON (Figure 5); Marker effects are better estimated as the TP increases; however, as the proportion of the TP reached 0.9 the variance of the accuracy values increased substantially. In fact, whereas the mean accuracy increased linearly with TP, the variance increased exponentially, as demonstrated in Figure 6 for FDK.

GEBVs and PEBVs

For each trait, GEBVs were calculated using the same imputation method (EMI), the best GS model (RR-BLUP), the largest training population size ($n_{TP} = 218$), and 5-fold cross validation. For all traits, the values were distributed around the reference straight line reasonably well, with higher agreement for the mid values. The fitted linear regression of PEBVs on GEBVs all obtained slope estimated of approximately $b = 1$, suggesting little to no bias in the GEBVs (Table 2). Routinely, lines showing phenotypic values of $ISK \leq 45$ are selected for further

testing. In this study, 65 or 23.81% of the lines showed PEBVs below that threshold. Using $GEBVs \leq 45$ as the criterion, 43 or 15.75% of the lines would have been selected, with 42 selections shared between PEBVs and GEBVs, and three unique to PEBVs. Twenty five lines with PEBVs below 45 would have been missed since as their GEBVs were above 45. For DON, using a $PEBV \leq 8$ ppm for DON, 51 or 18.68% of the lines would have been selected. At the same time, using $GEBV \leq 8$ ppm as the threshold instead, 36 lines would have been chosen, with 31 being shared between PEBVs and GEBVs, and five being unique to PEBVs. In addition, 20 lines with PEBVs below 8 ppm would have been missed ($GEBVs > 8$ ppm). Similar trends were observed for the other traits when comparing selection based on GEBVs instead of PEBVs.

Discussion

Genotypes and imputation methods

A collection of 273 breeding lines in use at the University of Illinois' soft red winter wheat breeding program was used to develop genomic selection models for traits associated with Fusarium head blight resistance. Cultivars with higher levels of resistance to this disease are urgently needed across the United States. Five thousand and fifty four SNPs were called by genotyping-by-sequencing and the UNEAK pipeline (Lu et al., 2013). Allelic effects of each marker were estimated and used to calculate genomic estimated breeding values for the lines.

Five different imputation methods were compared in terms of GS accuracy. The EMI, SVDI, KNNI, RFI and MNI methods performed equally well for all traits. At the same level of missing data (20%), Rutkoski et al. (2013) found similar results. Moreover, they demonstrated that these imputation methods tend to show differences with respect to GS accuracy as the level of missing data increases. Poland et al. (2012b) found no difference in GS accuracy for wheat traits when imputation methods were compared, even working with up to 80% missing data per

marker. In this study, EMI was numerically superior for two traits (FDK and INC) and required only 1 minute and 10 seconds to run a complete set of imputation for one data set. At the same time, RFI required over 8 hours and did not result in higher GS accuracy for any traits. Two other methods required almost 1 minute to run one set of imputation, and MNI required less than 30 seconds, but this method resulted in the lowest accuracies.

A low level missing data was selected for two reasons. First, the GS analysis gets computationally intensive as the number of SNP increases. Second, a total of 5054 SNPs were called with 20% missing data per markers. GS analyses with a larger number of SNPs were performed, but the increments in accuracy were not substantial enough to justify include the larger marker sets (data not shown).

GS models and SNP number

In this study, RR-BLUP, LASSO, and elastic net were compared for six traits under a range of SNP numbers. RR-BLUP is based on an infinitesimal model where all predictors are maintained in the analysis, resulting in a non-parsimonious model. In contrast, the LASSO performs variable selection and keeps a subset of predictors in the model, which is a reasonable assumption in plant breeding, as some SNPs are expected to not be associated with the response variable. This method, however, has its shortcomings: it can be unstable with high dimension data. In addition, the LASSO tends to be problematic when the predictor variables are highly correlated, but that should not be the case in this study as SNPs exhibiting pairwise correlations exceeding $r^2 \geq 0.8$ were not used. The elastic net can be seen as a combination of both methods. For most traits, higher values of accuracy were obtained when using RR-BLUP than with the other methods. In only one situation (trait = INC and SNP = 4500) did the LASSO and elastic-net provide numerically higher accuracies, although they were not significantly different from

RR-BLUP. This led to the conclusion that, for the traits used in this study, a model with a very large number of small effect SNPs is more appropriate than a reduced model. It has been suggested that the LASSO is better suited for situations when large effect genes are present. It is possible that such genes are present in the University of Illinois' germplasm for INC, and possibly for other traits. Even so, the superiority of RR-BLUP was demonstrated.

Training population size

A significant increase in accuracy was observed when more breeding lines were included in the TP. The overall rate of gain averaged over the six traits was comparable to a study in wheat conducted by Heffner et al. (2011), although they worked with different traits. In contrast to their results, a plateau in accuracy was observed when $n_{TP} > 192$ for all traits, except FHBdx. The impact of the training population size on accuracy was assessed in other crops such as barley with empirical (Lorenz et al., 2012) and simulated data (Iwata and Jannink, 2012), oats (Asoro et al., 2011), wheat (Heffner et al., 2011), sugar beet (Würschum et al., 2013), and maize (Crossa et al., 2014). Lorenz et al. (2012) reported diminishing returns for prediction accuracies of DON and FHB resistance when $n_{TP} > 200$, with no real gain when n_{TP} changed from 200 to 300 barley lines. Considering that expected accuracy depends on n_{TP} , h^2 , and the number of loci involved (Daetwyler et al., 2008), it would be necessary to increase the genetic variability of the germplasm and the number of testing environments in order to obtain higher accuracies.

The influence of the ratio TP/total number of lines was also assessed in terms of accuracies. Different number of folds used in k-fold cross validation schemes were tested: 2, 3, 5, and 10-fold. For a fixed number of breeding lines, one needs to decide how many lines will be part of the TP and how many will be used for validation. Due to a very low level of structure in the group of lines used in this study, no particular sampling technique (Lipka et al., 2014; Isidro

et al., 2015) was used for choosing the lines to be part of the TP. As the ratio TP/total number of lines increased, higher values of accuracies were obtained. Similar results were obtained in an analysis conducted in maize by Zhao et al. (2012). In this study it was observed that the values of accuracy became excessively variable when the proportion is larger than 0.8.

GEBVs and PEBVs

A high agreement between PEBVs and GEBVs was found all for parameters associated with FHB resistance. Consequently, wheat lines with higher levels of resistance can be selected even before being planted. At the University of Illinois wheat breeding program, great emphasis is placed on ISK and DON when making selections. For ISK, it was observed that some lines would have been misclassified and missed by using the same value of GEBV as the reference PEBV value. That is because, although the agreement is sufficient, it is not perfect (Figure 7). Had the GEBV threshold been reduced by 14% (using 51.3 instead of 45), not a single line would have been missed. It is true that 11 extra lines would be selected inappropriately, but 100% of the right lines would be carried to the selected pool. For DON, using 9.48 instead of 8 ppm (relaxing the threshold by 18.5%) would guarantee all the right lines would have been chosen. In this case, 51 additional lines would have been selected. It is important to note that phenotyping for FDK and DON requires significant amount of time since grains must be harvested, samples must be cleaned, and then evaluated. In this case of DON, samples are shipped to a specific laboratory where mycotoxin quantification is performed. Given this scenario, genomic selection should have a significant impact in reducing resources associated with FDK and DON quantification.

Conclusions

Most of the germplasm used in this study is primarily represented by breeding lines from the University of Illinois soft red wheat program. This study showed that moderate to high prediction accuracies can be achieved for FHB resistance-related traits within the context of a breeding program. Ridge regression-BLUP outperformed the other models for all traits. It was also demonstrated that moderate to high accuracies can be obtained even with a reduced set of SNPs and a few hundred lines in the training set. These results are encouraging and show that genomic selection can indeed be a promising strategy when breeding for FHB resistance.

Historically, breeders have relied on phenotypic selection and, to some extent, on marker assisted selection when breeding for FHB resistance. The cooperative system of nurseries and research projects supported by the Agricultural Research Service/USDA under the US Wheat and Barley Scab Initiative (USWBSI), has greatly improved the knowledge about the disease and has been important for incorporating resistance into cultivars. The results obtained in this study support GS as a feasible breeding strategy for FHB resistance, aiding to the strategies already implemented by breeders and the USWBSI. A next step would be using GS to calculate GEBVs in the $F_{4.5}$ generation in order to reduce the number of lines tested in replicated field trials. Additionally, GEBVs can be used to select which lines will compose the crossing blocks. Heffner et al. (2010) provide a framework for incorporating GS in a winter wheat breeding program based on single seed descent. Doubled haploid-based programs could benefit even more from GS as the number of lines to be field-evaluated could be drastically reduced as soon as the doubled haploid lines are available.

CHAPTER 3: GENOME-WIDE ASSOCIATION MAPPING OF FUSARIUM HEAD BLIGHT RESISTANCE IN WHEAT (*Triticum aestivum* L.) USING GENOTYPING-BY- SEQUENCING

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 US 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 adequate 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 *Fhb-1* on chromosome 3B. The genomic region on chromosome 6A appears to be new, as no other study reported QTL for that region. In addition, combination of favorable alleles of these SNPs resulted in lower levels of disease; however, lines carrying multiple beneficial alleles were in very low frequency for most traits. Predicted protein for identified SNPs included receptor-like kinase, reverse transcriptase, retrotransposon, and others, indicating a complex network of genes is involved in FHB resistance.

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 USA 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, trichothecenes 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, large genetic variability for resistance to FHB is present 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, 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 linkage mapping studies in bi-parental populations have been published, identifying QTL associated with the different types of resistance to FHB on nearly every chromosome. Buerstmayr et al. (2009) and Liu et al. (2009) both reviewed these QTL (including over 100 and 249 QTL, respectively) in an attempt to find stable and useful QTL for breeding purposes. Genome-wide association studies (GWAS), first developed for human genetics, 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 changed since then. 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. (2013) 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.

Genotyping-by-sequencing (GBS) is a high throughput method for SNP genotyping that relies on genome complexity reduction with restriction enzymes (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 et al., 2012).

Poland et al. (2012b) applied GBS in a wheat panel consisting of 254 lines from CIMMYT and were able to identify 41,371 SNPs. Rutkoski et al. (2013) used GBS to identify 130,000 GBS polymorphisms from 360 elite spring lines belonging to CIMMYT. 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. (2012) studied a panel consisting of 410 European winter wheat lines, and identified a total of 23,371 SNPs for use in GWAS.

In this study GBS-SNP markers were developed 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

A total of 273 breeding lines were evaluated in this study, with 185 lines belonging to the University of Illinois soft red winter wheat breeding program, and the remaining lines selected from different land grant universities and private companies across the midwestern and eastern regions of the US. 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 224kg.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; [(severity x incidence)/100], Fusarium-damaged kernels (FDK), incidence-severity-kernel index (ISK; [(0.3 x incidence + 0.3 x severity + 0.4 x FDK)], 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 $(INC \times SEV / 100)$. 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 $(0.3 \times incidence + 0.3 \times severity + 0.4 \times FDK)$. Lastly, DON was quantified by gas chromatography-mass spectrometry at the Department of Plant Pathology at the University of Minnesota. Resistance to mycotoxins has 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:

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

where Y_{ijk} 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 for the k th line in the j th block within the i th year, $year \times line_{ik}$ is the random effect of the interaction between the i th year and the k th line, and ε_{ijk} 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 al. (2012b), with modifications. Three two-enzyme combinations for genome complexity reduction were used: *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. For each enzyme combination a different set of barcodes was used. A set of 384 barcodes with the *PstI* overhang was used, and a total of three 96-plex libraries were sequenced 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 except 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: CSS1_ChromArm_contig number_SNP position within the contig. Map positions of SNP

containing contigs were obtained from the POPSEQ map (IWGSC, 2014) 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% ; iv) SNP mapped into multiple chromosomes. Missing data were imputed using the expectation maximization (EM) imputation method as described by Rutkoski et al. (2013). The panel was also genotyped with single sequence repeat (SSR) markers known to be associated with FHB resistance at the Eastern Regional Small Grains Genotyping Lab, USDA-ARS, Raleigh, NC.

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). In this study, SNPs that mapped to multiple locations were excluded from the analysis. Lastly, redundant SNPs were eliminated based on LD ($r^2 > 0.8$) using the LD tagSNP selection option in JMP Genomics 7 (Carlson et al., 2004; SAS Institute, 2015).

Genome-wide association analysis

Principal component analysis (PCA) was first performed to assess 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, was 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:

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

in which Y represent the phenotypes, X is the design matrix, β is a vector containing fixed effects, and e is the random error. In the “naïve” model β 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 compressed MLM) can be expressed as:

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

in which the Z is a design matrix and u is a vector of random additive genetic effects. In the K model, β contains only markers and u contains the K matrix. In both MLM and compressed MLM, β has both markers and eigenvalues, and u has the K matrix. The significance of marker-trait associations was based on a false discovery rate-adjusted P-value of 0.10.

Results

Phenotypic data

Substantial phenotypic variation was observed for all traits, with SEV and FHBdx showing the largest range (Table 3). For DON, the mean value of the untransformed BLUPs was 13.48 ppm, which is considerably higher than the maximum allowed in wheat grains in the US

(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. This number was then reduced to 19,992 SNPs after the LD-SNP analysis, in which SNPs showing linkage disequilibrium (r^2) higher than 0.8 were eliminated. The observed SNP missing level was 0.31 and MAF was 0.17. Genome B showed the highest number of SNPs (9084), followed by A (6992), and D (3916) (Table 4). 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 draft genome sequence is available (Choulet et al., 2014). Although the PCA showed lines from Illinois clustering together, separated from lines of different origin (Figure 8), 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. The models that did not include the K matrix - “naïve” and Q model - revealed a large number of

significant SNPs for most traits (Table 5). 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 WCSS1_contig10676713_3B_7175 was significantly associated with SEV (Fig. 9a). This SNP accounted for 8% of the variability and its estimated effect was -9.54 (Table 6). Chromosome 3B harbors a major effect QTL, *Fhb-1*, which has been extensively studied. The significant SNP was found to be in linkage disequilibrium with SSR markers flanking *Fhb1* ($D' = 0.39$ and 0.34 for *umn10Fd* and *Xgwm533*, respectively); however, these SSRs did not meet the threshold to be declared a QTL in this study, probably due to genotyping errors. When SNPs were ranked based on FDR-adjusted p values, three additional SNPs on chromosome 3B, located at similar genetic position of *Fhb1*, were among the top seven markers. Two of these three SNPs

were also in linkage disequilibrium with the SSRs flanking *Fhb-1* (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 7). 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 accumulate (Fig. 10a). The lowest mean SEV was observed for the group with lines carrying all four favorable alleles (“+ + + +”); however, this group did not differ from the second best haplotype (“+ + - +”), suggesting SNPs WCSS1_contig10676713_3B_7175, WCSS1_contig10352272_3B_5482, and WCSS1_contig10699215_3B_3620 are sufficient for providing lower SEV levels.

Incidence

Five SNPs on chromosomes 4D, 7D, 4A, 6A, and 7A, were significantly associated with INC (Fig. 9b). WCSS1_contig3876750_7DS_2023 accounted for 16% of the variance, and its effect was 6.74 (Table 6). The other SNPs showed effects varying from -4.56 to 3.60. A positive effect means that the presence that allele is associated with higher levels of disease. In these cases, the alternate allele is the one to be selected. 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 “- - - + -”). 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 with more favorable alleles (Fig. 10b).

DON

Four SNPs were associated with DON on chromosomes 1D and 3B (two on the short arm and one on the long arm) (Table 6). Although in the same locus, SNPs WCSS1_contig10413672_3B_4839 and WCSS1_contig10764618_3B_2168 were not in linkage disequilibrium. For this reason, they were both considered in further analysis. Differently from other traits, some of the SNPs detected for DON had the major allele as being 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. 10c), with two having more than 10 individuals. The first group consisted of lines carrying four “+” alleles, and the second comprised lines with “+” alleles for WCSS1_contig10676713_3B_7175, WCSS1_contig1879930_1DS_3352, and WCSS1_contig10764618_3B_2168. These groups were not significantly different from each other (Table 7).

Disease indexes

Marker-trait association were tested for disease indexes, which incorporate multiple disease measurements all at once. One SNP was detected for FHBdx (WCSS1_contig10676713_3B_7175) and two for ISK (WCSS1_contig10676713_3B_7175 and WCSS1_contig3876750_7DS_2023). These are the same SNPs detected for SEV and INC (Table 4). Both indexes were based on values of SEV and INC (plus FDK in the case of ISK).

Discussion

Fusarium head blight is the most important wheat disease in the Midwest US, 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 US. A total of 19,992 SNPs were used for genome-wide marker-trait association for FHB-resistance traits. A larger number of SNPs was detected for sub-genome B. The levels of PIC and DIV were comparable between sub-genomes A and B, 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 pairing as an explanation for the difference in genetic diversity among the wheat genomes. In addition, it is known that genome D was latest to be added to the cultivated bread wheat (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 more sequence available for aligning tags. In fact, wheat chromosome 3B is more than twice 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 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). 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.

Several QTL have been identified for FHB resistance and few of them are considered stable, being detected by independent studies. To date, the best studied and characterized QTL is *Fhb-1*, on the short arm of chromosome 3B. This major effect QTL has been introgressed into several adapted lines and cultivars worldwide. In this study, one SNP on the short arm of chromosome 3B (locus 18.32 cM) was associated with SEV. Independent studies detected QTL for multiple traits associated with FHB resistance on that same 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 within loci between 0 and 21.6 cM. This cluster, or meta-QTL (MQTL) was flanked by the SSR markers *Xfba311* and *Xgwm493*. Working with 30 mapping populations of mostly European germplasm, Loffler 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*. Although some overlap exists between the mapping populations used in these meta-analyses and

review, they all show *Fhb-1* spanning over loci 0 and 20 cM. In this study, a SNP associated with SEV, FHBdx, and DON on chromosome 3B locus 18.32 cM is reported. This SNP was in linkage disequilibrium with SSRs flanking *Fhb-1*, suggesting that this GWAS was able to detect the same chromosomal region associated with *Fhb-1*. In addition, it was shown that a combination of SNPs in that region, albeit not all significant when considered individually, can confer lower levels of SEV. Interestingly, the combination without SNP 3B-3 did not differ from the best combination (“+ + +”).

Five SNPs were associated with INC, on chromosomes 4A (locus 78.35 cM), 4D (0), 6A (134.15), 7A (22.82), and 7D (70.84). Liu et al. (2009) reported a MQTL consisting of two QTL on chromosome 4A between loci 75.7 to 77.1 cM. Both QTL were associated with SEV. They also found 10 QTL on chromosome 4D clustered around locus 12 cM, most of them associated with SEV. Interestingly, these 10 QTL all overlapped with *Rht-D1*, a plant height gene. The 273 breeding lines used in this study were genotyped for markers linked to *Rht-D1*, and they showed no association with FHB resistance. The SNP on chromosome 6A seems to be a unique QTL, as no other mapping study has reported associations with FHB resistant traits around locus 134.15 cM. 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. As it was found for SEV, lines carrying multiple (“+”) 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 four SNPs significantly associated with DON on chromosomes 3B (one on the short arm and two

on the long arm) and 1D. The SNP on short arm of chromosome 3B is the same found for SEV and the disease indexes. The SNPs on the long arm of 3B could potentially be associated with QTL reported previously. For instance, Liu et al. (2009) detected a cluster of QTL between loci 49 and 57 cM, and Loffler et al. (2009) reported a cluster around locus 45.5 cM. In this study, the SNP detected on chromosome 1D falls within the same cluster detected by Liu et al. (2009) on that chromosome. The combination of SNPs had a numerical reduction on DON level.

Unfortunately the group of lines with no positive allele (“- - -”) consisted of only two lines and were not included in the contrasts; however, a substantial numerical difference was observed when compared with the class (“+ + +”) and (“+ - +”). These two 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, it was observed that most lines in the panel (90.4%) already have at least three favorable alleles (WCSS1_contig10676713_3B_7175, WCSS1_contig1879930_1DS_3352, and WCSS1_contig10764618_3B_2168). When this combination is absent, DON levels increase substantially. It is possible to speculate that the combination of QTL associated with these SNPs may play a role in DON accumulation and/or detoxification. Future research is necessary to clarify the function of the QTL associated with these SNPs and DON levels.

In a study with recombinant inbred lines from a cross involving Ning 7840, the main source of *Fhb-1* in North America, Bernardo et al. (2012) reported seven SNPs mapped within the *Fhb-1* region flanked by *Xgwm533* and *Xgwm493*. The nucleotide sequences of the contigs containing the significant SNPs found in this study were compared with the sequences reported

by Bernardo et al. (2012). No correspondence was found, indicating the high level of polymorphism within the *Fhb-1* region.

Hypothetical protein predictions were obtained for eight out of twelve contigs (Table 8), many of which seem to be involved in epigenetic regulation. For the contig on chromosome 1D harboring WCSS1_contig1879930_1DS_3352, the predicted protein is a receptor-like kinase (RLK), which 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. These results suggest resistance to FHB 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.

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 a new genomic region as well as in regions previously reported as harboring QTL, especially on chromosome 3B. This study also provides evidence that QTL accumulation can result in higher levels of resistance, but certain haplotypes occur in rare combinations. The SNPs reported in this study can be used immediately for marker-assisted selection and/or genomic selection. In addition, functional characterization of the underlying QTL could be performed for elucidating their biological role in FHB resistance. This is the first GWAS for FHB resistance using the contig information from the International Wheat Genome Sequencing Consortium. The

sequences associated with the SNPs found in this study can be obtained in the consortium's website for future research.

**CHAPTER 4: COMPARING GENOME-WIDE ASSOCIATION MAPPING AND
MARKER-ASSISTED SELECTION FOR FUSARIUM HEAD BLIGHT IN WHEAT
(*Triticum aestivum* L.)**

Abstract

Genomic selection (GS) and markers-assisted selection (MAS) rely on marker-trait association and are routinely used for breeding purposes. Although similar, these strategies differ regarding how markers are used to estimate breeding values. In this study, GS and MAS were compared for predicting six parameters associated with resistance to a destructive wheat disease, Fusarium head blight (FHB). A panel consisting of 273 soft red winter wheat lines from the US midwestern and eastern regions was used in this study. Quantitative trait loci (QTL) identified in a previous genome-wide study and *Fhb-1*, the best studied QTL for FHB resistance, were used for MAS using multiple linear regression (MLR). GS relied on 19,992 SNPs, QTL and *Fhb-1*, and marker effects were estimated with ridge regression-best unbiased linear predictor (RR-BLUP). GS greatly outperformed MAS, with cross-validated prediction accuracy varying from 0.24 to 0.74 and from 0.59 to 0.98 for MAS and GS, respectively. Treating quantitative trait loci (QTL) as fixed effects in GS models resulted in higher prediction accuracy when compared with a GS model with only random effects. For the same selection intensity, GS resulted in higher selection differentials than MAS for all traits. This study indicates that GS is a more appropriate strategy than MAS for FHB resistance.

Introduction

Several marker-based strategies are being applied in modern plant breeding with the objective of selecting individuals with superior performance; marker-assisted selection (MAS) and genomic selection (GS) are two examples. In MAS, individual lines are selected based on quantitative trait loci (QTL), which are detected through linkage mapping (LM) or genome-wide association studies (GWAS). While LM relies on experimental populations with limited recombination events, GWAS is performed on panels or collections of lines, taking advantage of all recombination events that occurred throughout the history of the panel, usually resulting in higher mapping resolution when compared to LM (Myles et al., 2009). In both LM and GWAS, only QTL meeting a certain threshold will be declared as significant, all the remaining marker-trait associations being ignored for further analysis. Thus, the number of markers per trait used in traditional MAS is generally low. For traits under complex genetic control, with multiple small effect genes playing a role, MAS can be of limited applicability (Bernardo et al., 2008). GS can be viewed as an extension of MAS. Instead of defining a statistical threshold for declaring significant QTL, all markers are used in GS for modeling the performance of an individual, regardless of the magnitude of their effect (Meuwissen et al., 2001). Therefore, for traits with complex inheritance, GS is expected to perform better because the variation due to small effect genes is potentially captured.

During the last decade, an exponential reduction in sequencing cost has occurred (Wetterstrand et al., 2015), and numerous next generation sequencing platforms are available to the plant research community (Patel et al., 2015). As a consequence, high density genome-wide markers are becoming available for most crops, making MAS and GS even more attractive. Assessing how these strategies compare is important for making more rational and efficient use

of markers in a breeding pipeline. Heffner et al. (2011) compared GS and MAS for 13 agronomic traits in wheat and reported a 28% advantage of GS over MAS for prediction accuracies. Wang et al. (2014) found GS to outperform MAS for agronomic and quality traits in inbred and hybrid rye. In the case of grain yield, the prediction accuracy, defined as the correlation between phenotypically estimated breeding values (PEBVs) and genomic estimated breeding values (GEBVs), increased from 0.12 (MAS) to 0.59 (GS) in one population. Zhao et al. (2014) compared both methods for heading time and plant height in rye, and found the best method to be trait specific. When major genes are present, GS models such as ridge regression-best unbiased linear prediction (RR-BLUP) will underestimate the variance associated with these genes, negatively impacting accuracies. By treating major genes as fixed effects in the GS model even higher accuracies can sometimes be achieved (Bernardo, 2014; Owens et al., 2014; Zhao et al.; 2014), but not always (Rutkoski et al., 2012).

Fusarium head blight (FHB), caused by *Fusarium graminearum* Schwabe [telemorph: *Gibberella zeae* Schw. (Petch)], is a major wheat disease in most growing areas, including the US midwestern region, where the pathogen can overwinter on maize debris. The pathogen causes damage due to significant grain yield reduction (Madden and Paul, 2009) and mycotoxin contamination in infected grain. Breeding for FHB resistance has been challenging due to the complex, polygenic nature of resistance, labor intensive phenotyping, and the delayed availability of high density, genome-wide markers compared to other economically important crops. One major QTL, *Fhb-1* has been identified by independent studies and multiple small effect QTL have been detected on nearly all wheat chromosomes, revealing the complex genetic architecture of FHB resistance (Buerstmayer et al., 2009; Liu et al., 2009; Loffler et al., 2009).

The objective of this study was to compare marker-based breeding strategies (MAS and GS) in their ability to predict phenotypes associated with FHB resistance.

Materials and methods

Plant material and phenotypes

A panel consisting of 273 breeding lines from multiple institutions was used in this study. Most of the lines originated from the University of Illinois soft red winter wheat breeding program, the remaining lines came from breeding programs in the US midwestern and eastern regions. The panel was phenotyped as described in Chapters 2 and 3. Briefly, six parameters associated with FHB resistance were recorded: severity (SEV), incidence (INC), FHB index (FHBdx; $[(SEV \times INC)/100]$), Fusarium-damaged kernel (FDK), incidence-severity-kernel index (ISK; $[0.3 \times INC + 0.3 \times SEV + 0.4 \times FDK]$), and mycotoxin accumulation (DON). SEV was recorded as the percentage of infected spikelets within a wheat head. INC is the percentage of infected heads in an experimental unit. FDK was recorded as the visual estimate of the percentage of Fusarium-damaged kernels in a sample of kernels. DON was quantified using chromatography-mass spectrometry at the Department of Plant Pathology at the University of Minnesota by Dr. Yanhong Dong. The experiment was grown in Urbana, Illinois in 2011, 2013, and 2014. Each year, the experiment was set up as a randomized complete block design (RCBD), with two replications. Since not all lines were planted in 2011, the data was analyzed as an unbalanced design. The experimental unit consisted of one-meter long single rows cultivated in a scab nursery with mist irrigation and grain spawn inoculation. Maize kernels were infested with *F. graminearum* from isolates collected throughout Illinois over several years, and then spread in the field at a rate of approximately $224\text{kg}\cdot\text{ha}^{-1}$.

For each trait, best linear unbiased predictor (BLUP) was calculated for each line using PROC MIXED SAS version 9.4 (SAS Institute 2013), according to equation (i):

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

where Y_{ijk} 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 heading date in Julian days for the k th line in the j th block within the i th year, $year \times line_{ik}$ is the random effect of the interaction between the i th year and the k th line, and ε_{ijk} is the random error term. The BLUPs were used as phenotypically estimated breeding values (PEBVs) for model comparisons.

Genome-wide association analysis

The procedures for single nucleotide polymorphisms (SNPs) identification and QTL detection were described in Chapter 3. In short, genotyping-by-sequencing (GBS) was employed using three two-enzyme combinations, where a rare cutter (*PstI*-HF) was combined with three common cutters (*MspI*, *HinPII*, and *BfaI*). TASSEL GBS version 4 was used against a pseudo reference genome developed from the Spring Chinese chromosome survey. A total of 19,992 SNPs were identified after applying the following filtering criteria: i) maximum per-marker missing data level of 50%; ii) minimum allele frequency of 5%; iii) maximum heterozygosity level of 10%; iv) SNPs mapped to single chromosomes. Missing data was imputed using the expectation maximization method (Rutkoski et al., 2013). The panel was also genotyped with single sequence repeat (SSR) markers known to be associated with FHB resistance. Two SSRs (*Xgwm533* and *umn10Fd*) flanking *Fhb-1* were used to determine whether or not a particular line

had this major effect QTL. Lines carrying the resistant allele for both SSRs simultaneously was scored as “2” (*Fhb-1*-resistant), whereas lines carrying the susceptible allele for both SSR was scored as “0” (*fhb-1*-susceptible). Other combinations were classified as heterozygote (“1”).

Genome-wide marker-trait associations were tested using the Genome Association and Prediction Integrated Tool – GAPIT (Lipka et al., 2012). A compressed mixed linear model was used having a marker-based kinship (K) matrix treated as random effect and matrix (Q) with the first four eigenvalues from a principal component analysis, treated as fixed effects. A false-discovery rate (FDR)-adjusted *p*-value of 0.10 was used for detecting significant marker-trait associations. The number of QTL identified with this procedure varied according to the trait: SEV(1), INC (5), FHBdx (1), DON (4), ISK (2), and FDK (0).

Comparison of models

Genomic selection and marker-assisted selection were compared in this study using different genotypic data sets. The GS model was based on RR-BLUP under three different scenarios: all markers treated as random effects (GS1); all markers as random effects and *Fhb-1* treated as fixed effect (GS2); all markers as random effects and QTL treated as fixed effects (GS3); For marker-assisted selection, multiple linear regression (MLR) was used with QTL treated as fixed (MAS1); and MLR with *Fhb-1* treated as fixed (MAS2). Although *Fhb-1* was not associated with some traits such as INC and DON, this QTL was included for these traits due to its importance for FHB resistance and popularity among wheat breeders worldwide. Since no QTL were identified for FDK, the five SNPs with the lowest FDR-adjusted *p*-values were selected from GAPIT and treated as fixed effects in GS1 and MAS1. In addition, three extra SNPs were considered for SEV because they were ranked among the top eight SNPs with the

lowest FDR-adjusted p -values, and they targeted the *Fhb-1* region. For more information about the SNPs used in this study, see Table 6.

Prediction accuracy

A 5-fold cross validation scheme was used to estimate SNP effects. Initially, the 273 lines were divided into five groups, four groups consisting of the training population (TP) and one group consisting of the validation population (VP). SNP effects were estimated in the TP and genomic estimated breeding values (GEBVs) were calculated for the lines in the VP. Prediction accuracy was calculated as:

$$\frac{r(GEBV:PEBV)}{\sqrt{h^2}}$$

where r is the Pearson's correlation between the GEBVs and the PEBVs in the validation population, and h^2 is the broad-sense heritability, estimated in Chapter 2 (Dekkers 2007; Albrecht et al., 2011). For each model and trait, 60 TPs were randomly obtained, resulting in 300 values of accuracy. The models were compared in SAS PROC GLM using the Ryan-Einot-Gabriel-Welch test at $\alpha = 0.05$ level.

Selection differential

In a second step, GEBVs were calculated for the entire panel using the averaged SNP effect across the 300 runs. For each trait, the 273 lines were then ranked based on their GEBVs (provided by GS1, GS2, GS3, MAS1, and MAS2). In other words, five different ranks were obtained for each trait. Then, the selection differential was calculated as the difference between the mean PEBV of the top 5, 10, 15, 20, and 25% best lines and the mean PEBV of the reference, unselected group of 273 lines.

Results

Prediction accuracies for six parameters associated with FHB resistance are presented in Figure 11. For SEV, genomic selection models GS1, GS2, and GS3, provided the highest mean accuracy, varying from 0.58 to 0.67, whereas MAS1 and MAS2 resulted in accuracies of 0.26 and 0.29, respectively. Interestingly, model GS3 outperformed GS2, even though the QTL considered in GS3 is localized on the same region as *Fhb1*, on chromosome 3BS. This result suggests that using the four SNPs around the *Fhb-1* region is more advantageous than using *Fhb-1* itself. Whereas MAS2 performed poorly for SEV (accuracy = 0.26), this model resulted in the second best mean prediction accuracy for INC (0.74), after GS3 (0.84). The best two models for INC both included five significant QTL, indicating these QTL may play an important role for this trait. The results for FHBdx and ISK are similar to SEV, the only difference being that GS1 and GS2 were not significantly different. For FHBdx it was expected that GS2 and GS3 would not differ since the QTL for FHBdx shares the same location as *Fhb-1*. A striking difference was observed between the MAS models and GS models for FDK. Accuracies varied from 0.24 to 0.39 for MAS1 and MAS2 and varied from 0.87 for GS1 to as high as 0.98 for GS3. Even though no QTL has been detected for this trait, treating five SNPs as fixed effect resulted in a substantial increase in accuracy. The GS models showed high accuracies for DON, but there were no significant differences among them.

The selection differential varied greatly between MAS and GS models but not within these groups (Figure 2). In fact, GS1, GS2, and GS3 overlapped for all traits, except FHBdx. At the same time, MAS2 outperformed MAS1 for SEV, INC, and ISK when at least 10% of the individuals were selected. For FDK, the selection differential barely changed with the selection intensity when MAS1 and MAS2 were applied, indicating the models are not appropriate for

ranking the breeding lines for that trait. This is not surprising since *Fhb-1* does not seem to be involved with FDK, and no other QTL was detected for this trait in the GWAS. The GS models also performed better for DON when compared to MAS, with no discernable difference among GS models. MAS2 was consistently superior than MAS1 across all selection intensities.

Discussion

Breeding for FHB resistance remains a major challenge among wheat breeders for several reasons, the complex genetic control of resistance being one of them. Since the introgression of *Fhb-1* from Sumai3 and its derivatives into adapted germplasm in North America, significant emphasis has been put on MAS, in which *Fhb-1* and a few other QTL are selected for. In this study, GS and MAS were compared using different genotypic data sets, including SSRs markers flanking *Fhb-1* and SNPs significantly associated with FHB resistance. The results showed that MAS with only *Fhb1* (MAS1) and QTL (MAS2) can lead to poor accuracy for nearly all traits. In addition, the values of accuracy showed much higher variability associated with these two models when compared to the GS models. GS greatly outperformed MAS for all traits, particularly for FDK, which seemed to not be controlled by large effect QTL. For one trait, INC, MAS2 was able to provide accuracy values higher than GS1 and GS2, suggesting the presence of medium to large effect QTL; however, incorporating these QTL in genomic selection (GS3) resulted on even greater accuracy. Five QTL have been shown to be associated with INC and this study suggests that they should be included in any of the marker-based breeding strategies tested in this study. Zhao et al. (2014) found MAS outperformed GS for heading time in rye, but not for plant height. In their study, the photoperiod insensitivity gene *Ppd-D1* was found to be associated with a single SNP whereas plant height was associated with 16 SNPs.

Treating QTL as fixed effects significantly improved prediction accuracy for all traits except DON. In fact, neither the QTL detected in chapter 3 nor *Fhb-1* treated as fixed effects resulted in higher accuracy for this trait. Rutkoski et al. (2012) compared several GS models and MAS for FHB resistance and concluded that GS outperformed MAS, but treating QTL as fixed effects did not improve accuracy; however, in a recent study with stem rust (caused by *Puccinia graminis* f.sp. *tritici*) resistance in wheat, Rutkoski et al. (2014) showed that treating major genes as fixed effects can lead to higher GS prediction accuracy. Bernardo (2014) observed that having a single gene treated as fixed was never disadvantageous, except when the variability explained by the major gene was lower than 10%. In this study, considering only the GS models, having a single major gene treated as fixed effect (*Fhb-1*) led to higher accuracies when compared to a model with all markers as random (GS1), except for one trait; however, treating all QTL as fixed (GS3) resulted in even better performance.

Although the GS models differed in terms of prediction accuracy, they performed equally well for selection differential. One possible explanation is that the differences in prediction from these GS models are not enough to substantially change the order of the breeding lines when they are ranked for a specific trait. The same cannot be said for MAS1 and MAS2. For INC, MAS2 was not superior than the GS models for selection differential, although it provided higher prediction accuracies than GS1 and GS2. One possible explanation is that MAS2 was more appropriated for predicting susceptible lines (and they are in larger number than resistant ones); however, when ranking the most resistant lines, the GS models were able to discriminate small differences better than MAS2.

Conclusion

GS greatly outperformed MAS in both prediction accuracy and selection differential for traits associated with FHB resistance. Treating significant QTL as fixed effects in GS models resulted in even higher accuracies, but this increase did not substantially change the order of lines when ranked according to their breeding values. These results indicate that GS is a more appropriate marker-based strategy when breeding for a complex trait such as FHB resistance.

CHAPTER 5: MAPPING QTL ASSOCIATED WITH NATIVE RESISTANCE TO FUSARIUM HEAD BLIGHT IN WHEAT (*Triticum aestivum* L)

Abstract

Fusarium head blight (FHB) is a destructive disease of wheat, and most cultivars are susceptible to it. Host resistance plays an important role in FHB management and native sources of resistance have been identified in elite background, adding to the traditional Sumai-3 derived resistance sources. In this study, a population of 233 recombinant inbred lines derived from a cross between IL97-1828 (resistant) and Clark (susceptible) was used to identify quantitative trait loci (QTL) associated with FHB resistance. Neither parent possesses the traditional Asian sources of resistance to FHB in their pedigree. A total of 2275 single nucleotide polymorphisms (SNPs) were detected using genotyping-by-sequencing (GBS) and a genetic map was built covering all 21 wheat chromosomes. Inclusive composite interval mapping (ICIM) analysis identified four genomic regions associated with multiple FHB parameters, across all environments. Four QTL were detected for FHB resistance under field conditions on chromosomes 1B, 2D, 6D, and 7B. Two QTL were associated with type I resistance (6D and 7B), and two were associated with type II resistance (1B and 2D). Type I and type II resistances are quantified by severity and incidence, respectively. The percentage of the phenotypic variation explained by these QTL varied between 6.7 and 12.5%, and they seem to co-localize with regions previously identified with FHB resistance. For QTL on sub-genome B, intervals smaller than 2 cM were obtained. The results show that elite germplasm can contribute to FHB resistance.

Introduction

Fusarium head blight (FHB) is one of the most important wheat diseases, occurring in most wheat producing countries. The disease is caused by many different *Fusarium* species (Parry et al., 1995), and in North America it is primarily caused by *Fusarium graminearum* Schwabe [teleomorph: *Gibberella zeae* Schw. (Petch)]. The effects of the disease on wheat include reduction in yield and grain weight, poor quality, and contamination by mycotoxins such as deoxinivalenol. Multiple components of FHB resistance have been identified in wheat: resistance to infection (type I) and spread (type II) (Schroeder and Christensen, 1963), resistance to mycotoxin accumulation (type III), resistance to kernel infection (type IV), and tolerance (type V) (Mesterházy, 1995; Mesterházy, 1999). The development of high yield potential cultivars with acceptable levels of FHB resistance remains a challenge for several reasons: the complex inheritance of the trait, difficulties in obtaining reproducible phenotypes, and the lack of adaptation of the main sources of resistance. Many quantitative trait loci (QTL) associated with FHB resistance have been reported in the literature (Buerstmayer et al., 2009) and a major QTL (*Fhb1*) on chromosome 3B has been identified in Sumai-3 and its derivatives by a number of independent studies (Bai et al., 1999; Anderson et al., 2001; Shen et al., 2003; Zhou et al., 2004; Chen et al., 2006; Liu et al., 2009; Loffler et al., 2009). Sumai-3 is a Chinese spring cultivar and it is arguably the most widely used source of FHB resistance in the world, although it shows undesirable agronomic traits in most wheat growing areas. Despite that, *Fhb1* has been successfully introgressed into multiple lines and cultivars around the world and the QTL itself does not seem to affect agronomic and quality traits (Salameh et al., 2011; Suzuki et al., 2012, Bakhsh et al., 2013); however, Von der Ohe et al. (2010) showed that, depending on the genetic

background, *Fhb1* can have a small negative effect on grain yield when introgressed into elite germplasm.

Numerous sources of resistance unrelated to Sumai-3 have been identified in elite germplasm. In Europe, QTL for FHB resistance have been detected on multiple chromosomes from numerous resistance sources including Renan (Gervais et al., 2003), Arina (Paillard et al., 2004), Cansas (Klahr et al., 2007; Häberle et al., 2010), Dream (Schmolke et al., 2008), Apache, Rubens and Romanus (Hopzapfel et al., 2008), and Frontana (Szabó-Hevér et al., 2014). The majority of the QTL identified in these studies can be considered minor effect QTL, some being stable across environments. In the US, many cultivars and breeding lines have been classified as moderately to highly resistant such as Ernie (McKendry et al., 1995), Freedom (Gooding et al., 1997), Massey, Roane (Griffey et al., 2001), Truman (McKendry et al., 2005), Bess (McKendry et al., 2007) and Goldfield (Ohm et al., 2000). Gilsinger et al. (2005) identified two regions of chromosomes 2B and 7B associated with FHB incidence and narrow flower opening in a population derived from Goldfield. Liu et al. (2007) reported four QTL from Ernie on chromosomes 2B, 3B, 4B, and 5A. Bonin and Kolb (2009) detected three QTL for FDK and DON accumulation in the line IL94-1653. A study with multiple native sources of resistance including Ernie, Freedom, Lyman, and Overland identified 15 QTL for multiple FHB resistance traits on 12 chromosomes (Eckard et al., 2015).

Until very recently, mapping studies in wheat relied mostly on a limited number of manually scored markers, the most popular being single sequence repeat (SSR), restriction fragment length polymorphism (RFLP), and amplified fragment length polymorphism (AFLP). The diversity array technology (DArT) represented an advance due to its automation, but its distribution across the wheat genome remains non-uniform. Recent advancements in next

generation sequencing (NGS) technology have dramatically increased the availability of SNP markers at decreasing costs (Wetterstrand, 2015). Akhunov et al., (2010) used the Illumina GoldenGate assay to identify SNPs in tetraploid and hexaploid wheat. Since then, SNPs have been increasing in popularity in mapping studies in wheat, including FHB resistance QTL identification. Bernardo et al., (2012) detected 7 SNPs between *Xgwm533* and *Xgwm493*, the SSRs flanking the *Fhb1* region. Petersen et al. (2014) used DArTs, SSRs and SNPs to detect QTL for FHB resistance in a population derived from Neuse, a native source of resistance. Genotyping-by-sequencing (GBS) is a next generation sequencing method that performs SNP discovery and genotyping simultaneously, generating many thousands of SNPs with lower ascertainment bias than SNP-chips (Elshire et al., 2011). GBS can be easily modified to accommodate multiple enzyme combinations, allowing detection of SNPs uniformly distributed across the genome (Poland et al., 2012b). The method has been successfully applied in wheat (Poland et al., 2012a; Poland et al., 2012b; Lado et al., 2013; Forrest et al., 2014; Talukder et al., 2014; Langer et al. 2014). The objective of this study was to identify QTL associated with FHB resistance in a population derived from the cross between the native source of resistance IL97-1828 and Clark (susceptible), using GBS-SNPs.

Materials and methods

Plant material

A recombinant inbred line (RIL) population consisting of 233 individuals was derived from the cross Clark (susceptible) x IL97-1828 (resistant). The line IL97-1828 displays both type I, II, and III resistances under high disease pressure. In addition, it does not have the traditional Asian source of resistance in its background (PioW549/PioW558//P79404G1-26-2/3/P81131-16-2-1-2-3-3). Clark (Beau//665256A1-8-1/67137B5-16/4/Sullivan/3/Beau5517B8-5-3-3/Logan)

(Ohm et al., 1988) is a winter wheat cultivar with high susceptibility to FHB. The cross was made in 2005 in greenhouse, and generations F₁ to F₅ were advanced by single seed descent.

Phenotypes

In 2008 Thompson (2010) evaluated F_{5:6} plants in the field at the Crop Science Research and Education Center, University of Illinois at Urbana-Champaign. The following year, F_{5:7} seeds were planted in two locations: Urbana, Illinois and Wooster, Ohio at the Snyder farm of the Ohio Agricultural Research and Development Center. In Urbana three resistant lines were used as FHB resistant checks (Ernie, IL97-6755 and IL00-8061) whereas in Wooster five resistant checks (Ernie, IL97-6755, IL00-8061, IL02-18228 and Truman) and one susceptible check (Pio25R47) were used. In both locations *F. graminearum* was inoculated using the grain spawn inoculation method, as described by Thompson (2010). Briefly, a conidial suspension was prepared with different isolates of *F. graminearum* and then used to infest autoclaved maize grain and allowed to colonize. In Urbana, grain spawn was spread in the field at a rate of 287 kg.ha⁻¹ two to four weeks before anthesis. Mist irrigation was used to guarantee high humidity, favoring disease development. In Wooster, Ohio grain spawn was disseminated at a rate of 360 kg.ha⁻¹ three weeks prior to anthesis. In addition, a suspension of macroconidia (25,000 conidia mL⁻¹) was applied every five days starting three days after heading of the earliest line.

FHB resistance was assessed by measuring the following parameters: incidence (INC), severity (SEV), FHB index (FHBdx; [(incidence x severity)/100], Fusarium-damaged kernels (FDK), incidence-severity-kernel index (ISK index; [(0.3 x incidence + 0.3 x severity + 0.4 x FDK)] and deoxynivalenol concentration (DON). INC and SEV were recorded one month after the heading date of each line, whereas FDK and DON were recorded after harvest. FDK is a visual estimate of the percentage of damaged kernel using known standards. DON concentration

was determined by Dr. Yanhong Dong in the Department of Plant Pathology at the University of Minnesota using gas chromatography-mass spectrometry.

For each trait and line, least-squares means were calculated in JMP v. 11 (SAS Institute, 2015) according to the equation (i):

$$Y_{ijkl} = \mu + year_i + location_j + line_k + (year \times line)_{ik} + (location \times line)_{jk} + block(year \times location)_{ijl} + \varepsilon_{ijkl} \quad (i)$$

where Y_{ijkl} is the observed phenotype, μ is the overall mean, $year_i$ is the random effect of the i th year, $location_j$ is the random effect of the j th location, $line_k$ is the fixed effect of the k th line, $(year \times line)_{ik}$ is the random effect of the interaction between the i th year and the k th line, $(location \times line)_{jk}$ is the random effect of the interaction between the j th location and the k th line, $block(year \times location)_{ijl}$ is the random effect of l th block within the i th year and j th location, and ε_{ijkl} is the random error term. For DON, the effects involving “location” were not included in the model since DON values were only obtained in Urbana, Illinois.

Genotyping-by-sequencing

Seeds of the 233 RILs and the parents were sent to the USDA Eastern Regional Small Grains Genotyping Laboratory, Raleigh, North Carolina for DNA extraction. A library was prepared with a two-enzyme restriction digest: *Pst*I-HF (CTGCAG) is a rare cutter whereas *Msp*I (CCGG) is a common cutter. Each restriction enzyme leaves an overhang on the DNA strand, to which the barcode is bound. The barcodes adapters were selected according to Poland et al. (2012). Only segments that were cut by both enzymes were sequenced. Two lanes of sequences were obtained from Illumina HiSeq 2000 runs and were then analyzed using the TASSEL GBS 4 against a pseudo reference genome developed from the *T. aestivum* Chinese Spring chromosome

survey (IWGSC, 2014) (as described in Chapter 3). After obtaining their location, SNPs were filtered based on: i) per-marker missing data level < 50%; ii) minor allele frequency > 5%; iii) percent of heterozygotes < 10%; iv) probability of chi-square test (1:1) > 0.05.

Linkage map and QTL detection

A linkage map was built using the QTL IciMapping v. 4 software (Li et al., 2008, Meng et al., 2015). Initially, non-redundant SNPs with missing data level < 20% were selected to build a preliminary genetic map. SNPs were considered redundant if they were completely correlated or identical. For two or more redundant SNPs, the one with lowest percentage of missing value was kept. The Chinese Spring chromosome survey provided the chromosome arm to which each SNP was assigned. In other words, for every SNP, its chromosome number and arm was known in advance, and this information was used for assigning SNPs into linkage groups (LG). Ordering within LG was performed with the “nnTwoOpt” option. The genetic distances obtained with this preliminary map were used as anchors for a second map including SNPs with up to 50% missing data level. The final map was aligned with the POPSEQ map (IWGSC, 2014) available at <http://wheat-urgi.versailles.inra.fr/Seq-Repository/Publication>. The inclusive composite interval mapping (ICIM) analysis performed for QTL detection in QTL IciMapping v. 4 (Meng et al., 2015). The method first identifies the most significant markers and marker-pairs multiplication using stepwise regression. Then, a two-dimension scan is performed to identify significant digenic epistasis based on adjusted phenotypic data (Li et al., 2008). Significant thresholds were based on 1000 permutations and $\alpha = 0.05$ and 0.10.

Results

Phenotypes

A detailed phenotypic description of the population used in this study is provided by Thompson (2010). Least-squares mean were newly calculated from the raw data and are presented in Table 9. The traits with the largest variation were SEV, FDK, and FHBdx, as indicated by their standard deviation. At the same time, DON and ISK showed the lowest variation. All traits were correlated with each other at $\alpha = 0.05$ level, with the Spearman correlation coefficient varying from 0.30 (SEV and DON) to 0.92 (FHBdx and SEV). The data indicate that this population shows adequate phenotypic variation for the parameters evaluated, and it is suitable for mapping purposes.

Linkage map and QTL identification

The GBS method identified 2275 non-redundant SNPs after passing the filtering criteria described in the materials and methods. The final genetic map spanned 2875 cM over the 21 wheat chromosomes, with an overall marker density of 1.26 cM per SNP (Table 10); however, the number of SNPs varied greatly among the sub-genomes, resulting in densities of 1.11, 1.41, and 6.46 cM/SNP for sub-genomes A, B, and D, respectively. Chromosome 3B displayed the highest number of markers (578), more than the entire D sub-genome (187). A satisfactory agreement was observed between the map developed in this study and the POPSEQ map (IWGSC, 2014).

Two QTL were detected for SEV on the short arms of chromosomes 1B and 2D (Table 11). The QTL on chromosome 1B was flanked by the SNPs WCSS1_1190535_1BS_1034 and WCSS1_3457475_1BS_7477, within a 1.09 cM region (Fig. 13). Its estimated effect was 2.10.

The QTL on chromosome 2D was flanked by WCSS1_5315750_2DS_5213 and WCSS1_5354063_2DS_3647, within an 11.5 cM region (Fig. 14), and estimated effect of -2.05. The percentage of the variation explained by these two QTL was 7 and 6.6%, respectively. Two QTL were detected for INC on chromosomes 6D and 7B. The first was flanked by WCSS1_3239902_6DL_273 and WCSS1_2895599_6DL_459, within 6.87cM region (Fig. 15), and the second was flanked by WCSS1_5085227_7BL_199 and WCSS1_3115996_7BL_172, spanning over 1.88 cM (Fig. 16). A larger percentage of the variation was explained by the QTL for INC than for SEV, accounting for 12.6% (6D) and 10.3% (7B). Their estimated effects were comparable to the QTL for SEV; however, comparing effect sizes between traits can be misleading since the traits were not normalized at first. One QTL was detected for FDK on chromosome 2A using data only from Urbana (2009 and 2010) and three extra QTL were found to be environment-specific, being significant in only one environment.

Discussion

In this study, a population of 233 RILs from a cross between IL97-1828 and Clark was used for mapping QTL associated with FHB resistance. The line IL97-1828 does not have the traditional Asian sources of FHB resistance on its pedigree, and it has been extensively used in the University of Illinois winter wheat breeding program as a parent. For these reasons, IL97-1828 represents an important source of native resistance to FHB.

In a previous study, Thompson (2010) studied the same population using 214 DArT and 16 SSR markers, and detected multiple QTL, but only three genomic regions (on chromosomes 1B, 2B, and 3B) were significant in more than one environments. In this study, a new genetic map was built using 2275 SNPs identified from GBS, covering all wheat chromosomes, with an

overall marker density of 1.26 cM per SNP. In addition, exact chromosome assignment of the markers was possible since the chromosome arm of each SNP was known in advance, except for chromosome 3B (both arms of this chromosome have been sequenced together). This new map, and the ICIM analysis allowed the identification of four genomic regions associated with FHB resistance using phenotypic data from all environments. These genomic regions belong to chromosomes 1B, 2D, 6D, and 7B.

The QTL on chromosomes 1B and 2D are associated with SEV. Several other QTL in these regions have been reported elsewhere. In a meta-analysis involving 249 QTL for FHB resistance, Liu et al. (2009) identified clusters of QTL or meta-QTL (MQTL) on all wheat chromosomes, with 3 MQTL on chromosomes 1B and 2A each. Loffler et al. (2009) also performed a meta-analysis with QTL for FHB resistance and detected the same number of MQTL on those chromosomes. In this study, the total genetic distance for chromosome 7B was 211 cM, which is much larger when compared to the distance mapped in the meta-analyses, most likely due to the number of markers used on each map. Using the genetic distance to infer QTL location in clusters can be misleading for chromosome 1B; however, the name of the SNPs flanking the QTL on 1B indicates it belongs to the short arm. Both meta-analyses showed Asian and non-Asian sources contributing to QTL in those regions.

Two QTL were detected for INC on the long arms of chromosomes 6D and 7B. The meta-analysis by Liu et al. (2009) also detected MQTL for these regions. Interestingly, only non-Asian sources contributed with QTL for these clusters. Loffler et al. (2009) also reported two MQTL for chromosome 7B, one on each arm. Lastly, one QTL was detected for FDK on the short arm of chromosome 2A. In this same region, Liu et al. (2009) detected two clusters. The cluster closer to the centromere contains a QTL for FDK from the Chinese landrace

Wangshuibai. Overall, narrower intervals were obtained for QTL on genomes A and B than for D. This is not surprising since far more SNPs were identified for these genomes. Genome D is known to bear a reduced level of polymorphism when compared to A and B (Poland et al., 2012a, Nielsen et al., 2014).

The parents of the RILs used in this study were present in the GWAS described in Chapter 3, but no similarity was found between the genomic regions associated with FHB resistance in both studies. In other words, the QTL identified in this study were not significant in the GWAS, and vice-versa. The reasons for this disparity are: i) the limited number of SNPs occurring in both studies (226 SNPs). Only two SNPs associated with QTL in the biparental population were present in the GWAS; ii) the linkage disequilibrium (LD) pattern in both studies is different; iii) the panel used in the GWAS bears polymorphisms not present in the biparental population; and iv) the effect of the environment. These studies were carried out in different years, and the RILs were evaluated in one extra location.

Conclusion

In this study, multiple QTL were detected for FHB resistance in a population from an elite by elite cross. These QTL were located in chromosomal regions not detected in a previous study with the same population, but in similar regions reported elsewhere. The SNP-based genetic map, and the ICIM analysis allowed the location of these QTL with relative precision. In three situations, the genetic distance between the flanking markers was less than 2 cM. These results are encouraging, especially considering that these QTL were identified in elite germplasm and no linkage drag is expected. Future research could explore the biological role played by these genomic regions on FHB resistance. In addition, crosses between full-sibs that differ for

these QTL could be used for fine-mapping. Lastly, the SNPs flanking the QTL detected in this study can be used in kompetitive allele specific PCR (KASP™) assays, which could facilitate the introgression of these QTL into newer, competitive wheat germplasm.

FIGURES

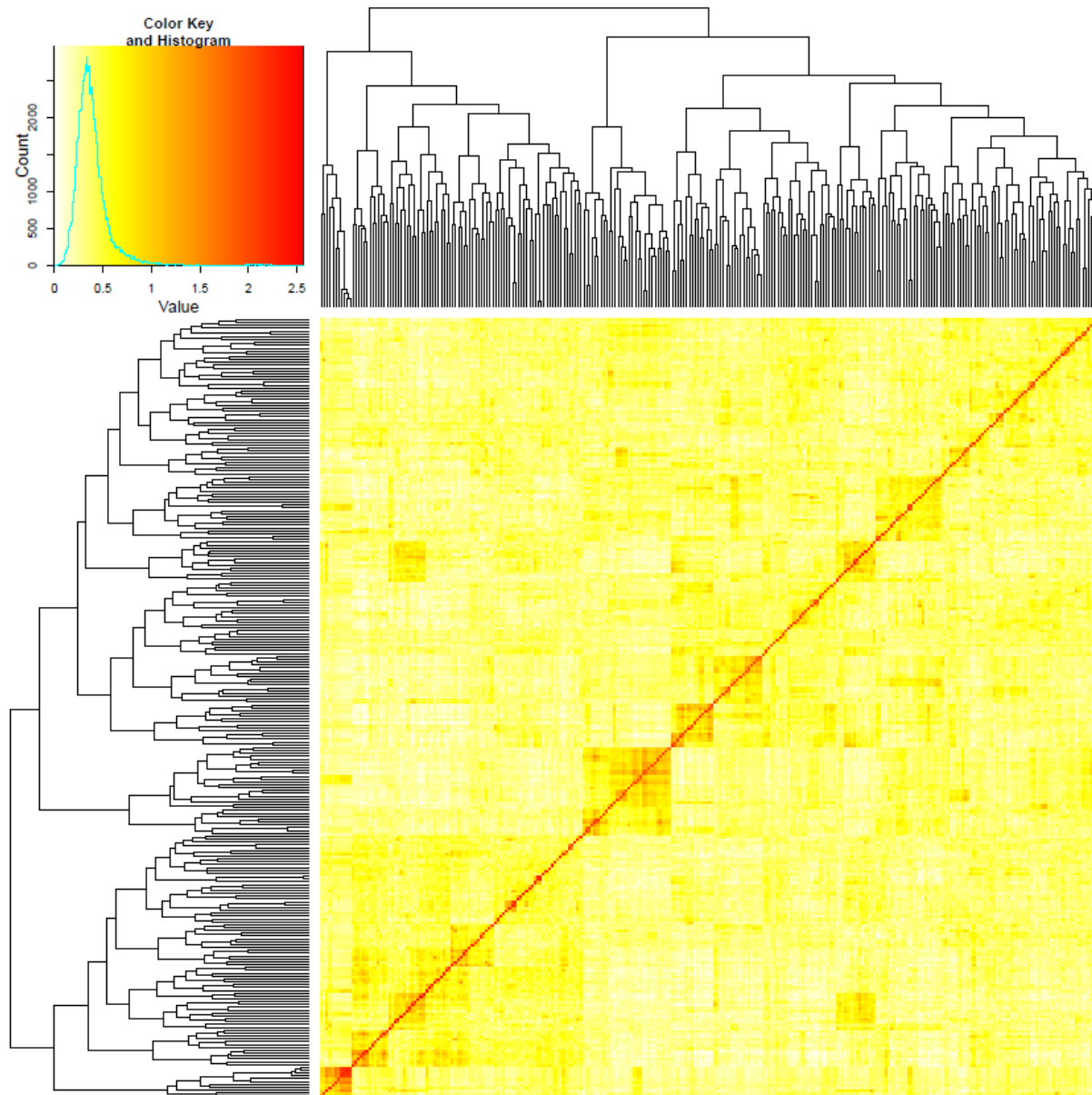


Figure 1. Heat map of the marker-based relationship matrix for 273 wheat breeding lines. Matrix obtained from 5054 SNPs. Rows and columns represent the breeding lines.

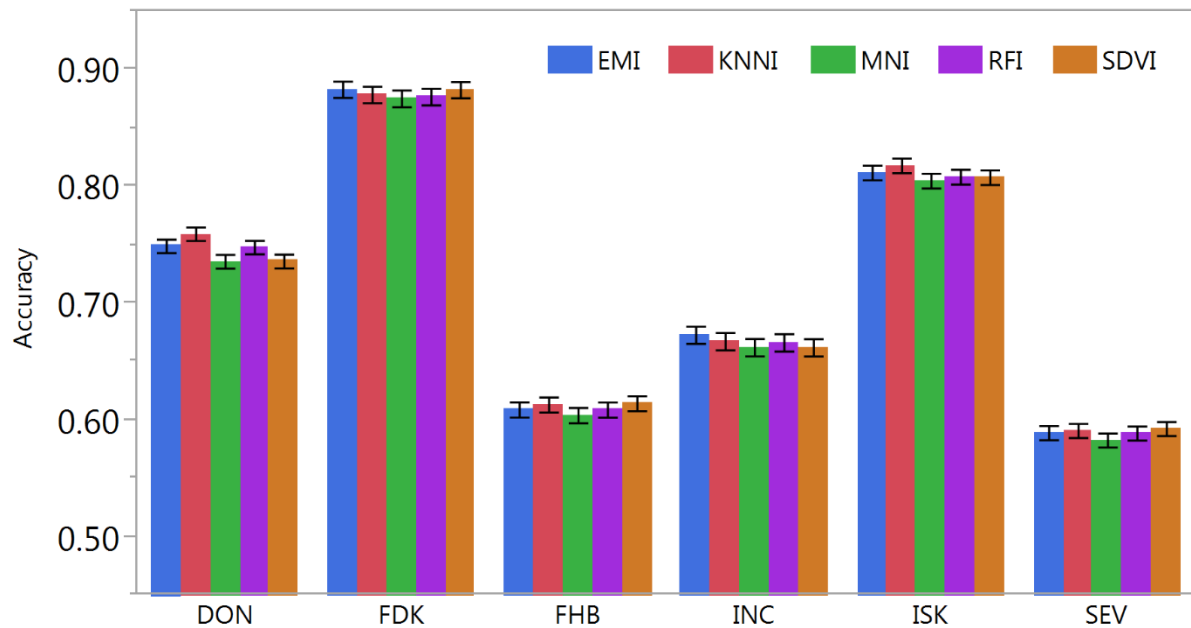


Figure 2. Five-fold cross validated genomic selection accuracies for six traits associated with FHB resistance and four imputation methods. Methods receiving the same letter do not differ according to the Ryan-Einot-Gabriel-Welch multiple comparison test at α 0.05 level. EMI = expectation maximization imputation, MNI = mean imputation, RFI = random forest imputation, and SDVI = singular value decomposition imputation. DON = deoxynivalenol concentration, FDK = Fusarium-damaged kernels, FHBdx = FHB index, INC = incidence, ISK = ISK index, SEV = severity. Error bars represent \pm one standard error of the mean.

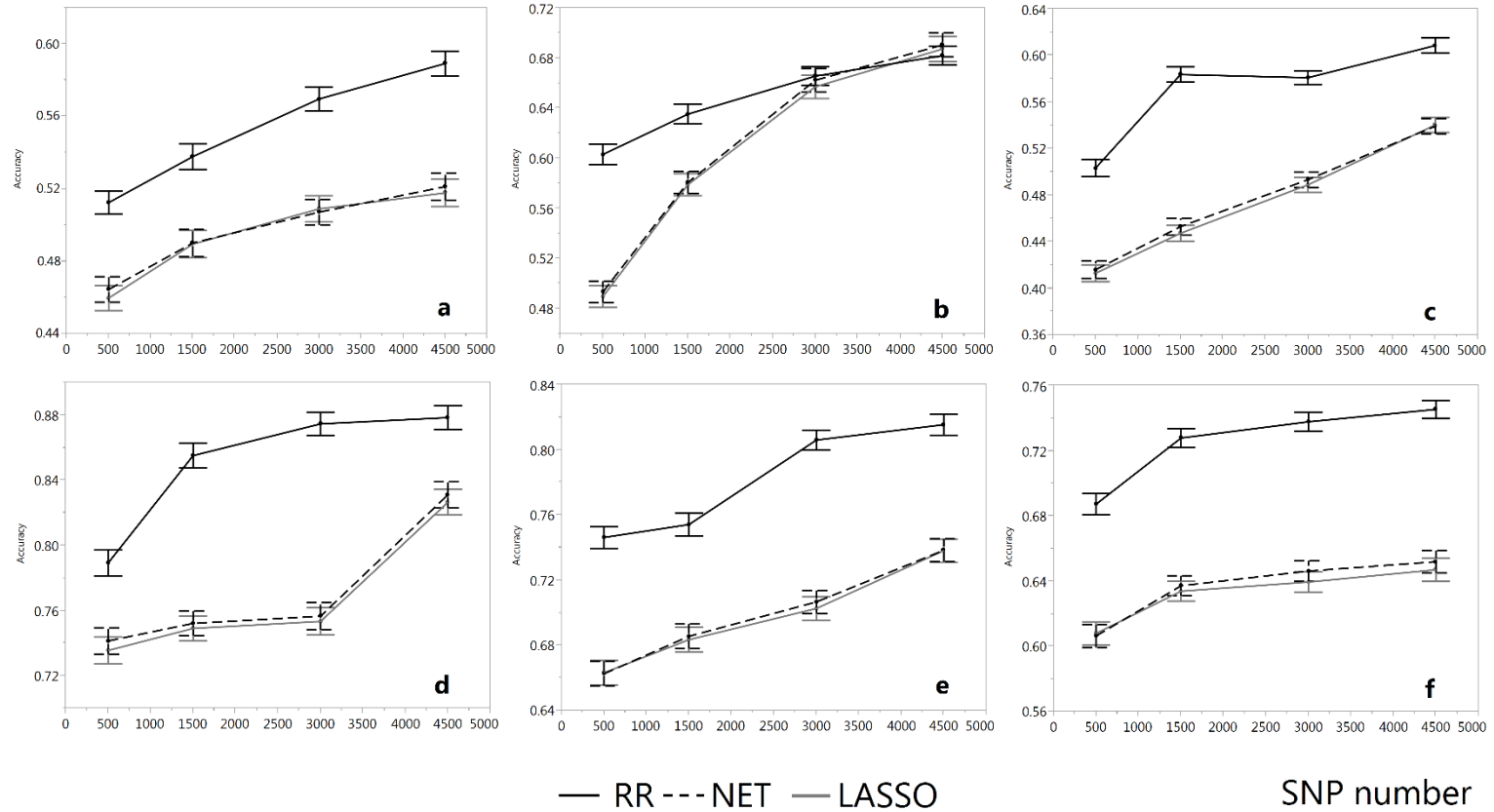


Figure 3. Five-fold cross validated genomic selection accuracies for six FHBrelated traits as a function of genomic selection models and SNP numbers. a = SEV (severity), b = INC (incidence), c = FHBdx (FHB index), d = FDK (Fusarium-damaged kernels), e = ISK (ISK index), f = DON (deoxynivalenol concentration). RR = ridge regression best unbiased linear predictor, NET = elastic net, LASSO = least absolute shrinkage and selection operator. Error bars represent \pm one standard error of the mean.

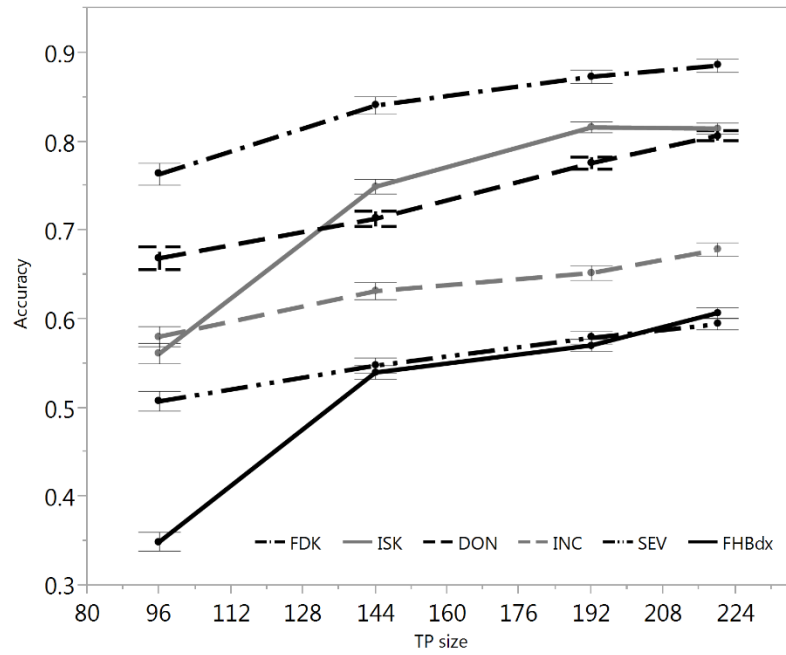


Figure 4. The effect of training population (TP) size on genomic selection accuracy for six FHB related traits. FDK = Fusarium-damaged kernels, ISK = ISK index, DON = deoxynivalenol concentration, INC = incidence, FHBdx = FHB index, SEV = severity. TP = training population. Error bars represent \pm one standard error of the mean.

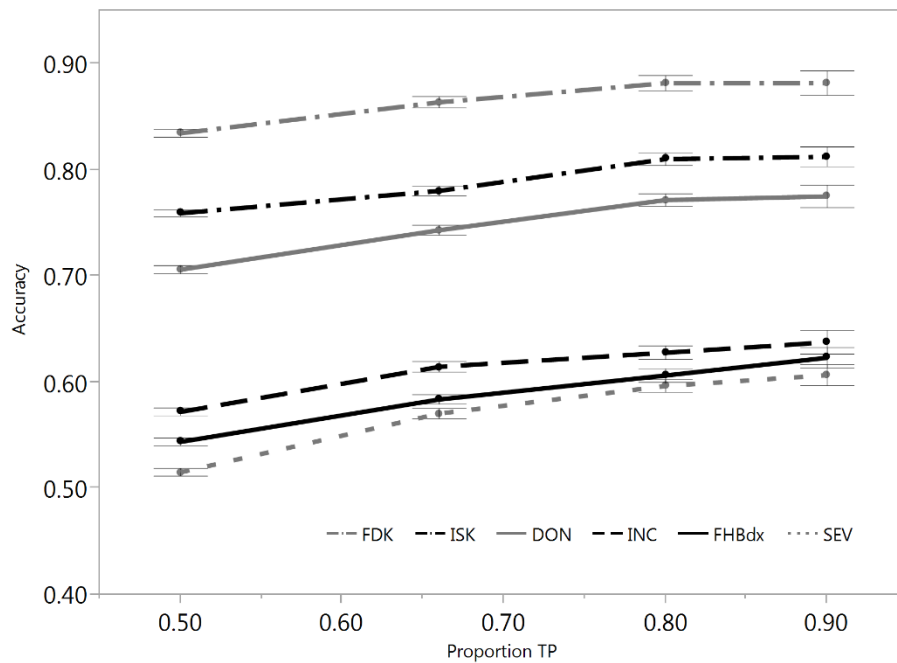


Figure 5. The effect of proportion of the training population/total number of lines on genomic selection accuracy for six FHB related traits. FDK = Fusarium-damaged kernels, ISK = ISK index, DON = deoxynivalenol concentration, INC = incidence, FHBdx = FHB index, SEV = severity. TP = training population. Error bars represent \pm one standard error of the mean.

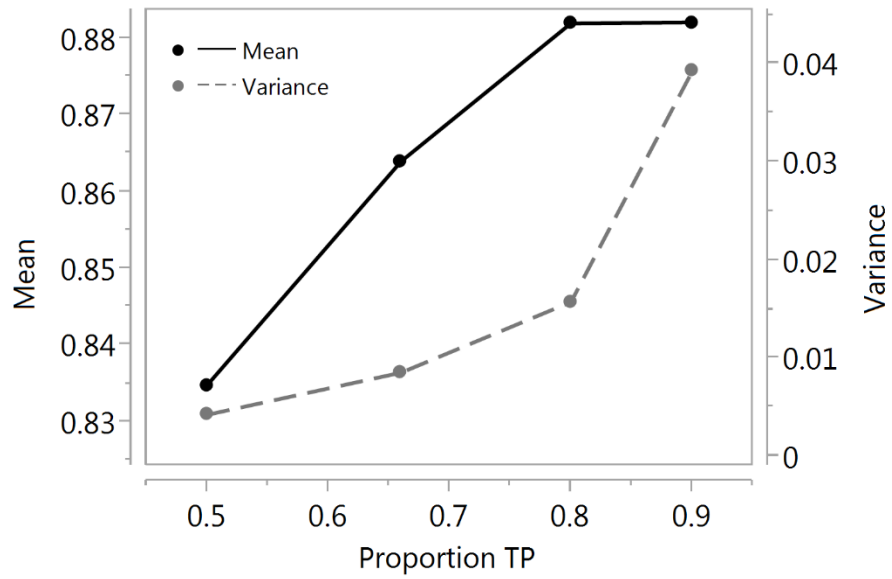


Figure 6. The effect of proportion of the training population/total number of lines on the mean and variance of genomic prediction accuracies for Fusarium-damaged kernels (FDK). Genomic estimated breeding values calculated using RR-BLUP with five-fold cross validation. TP = training population.

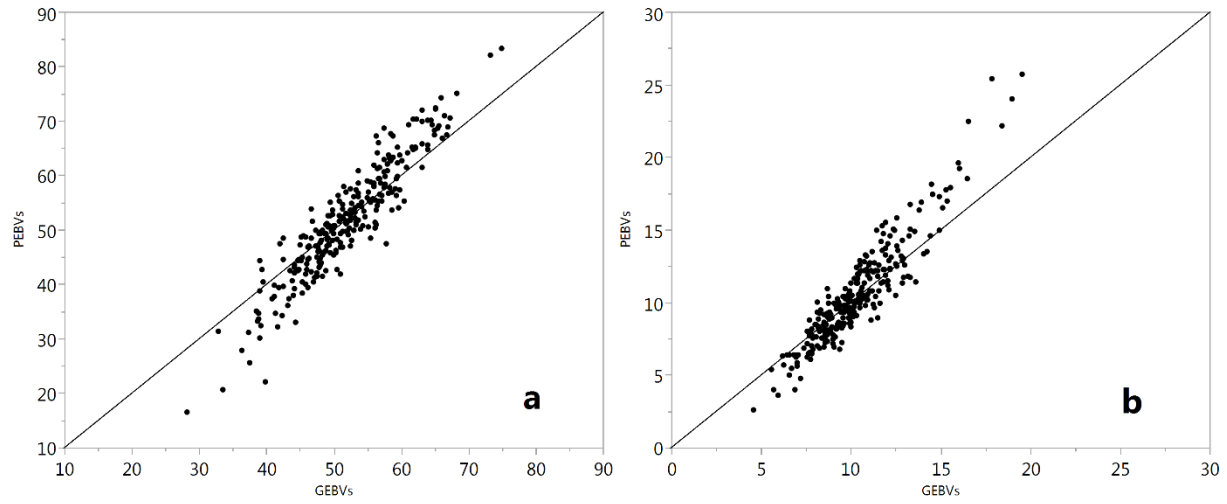


Figure 7. Distribution of phenotypically estimated breeding values (PEBVs) and genomic estimated breeding values (GEBVs) for ISK index (a) and deoxynivalenol concentration (DON) (b). GEBVs calculated using RR-BLUP with five-fold cross validation.

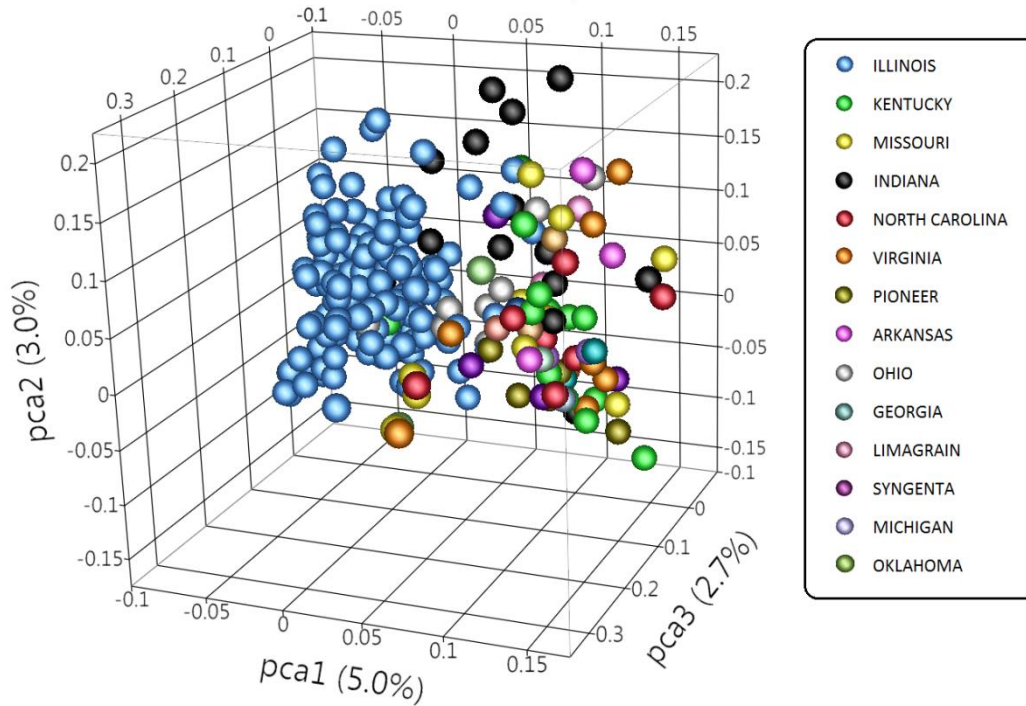


Figure 8. Principal component analysis of 273 winter wheat breeding lines using 19,992 SNPs. Colors represent the origin of the breeding lines. Illinois = University of Illinois at Urbana-Champaign; Kentucky = University of Kentucky; Missouri = University of Missouri; Indiana = Purdue University; North 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.

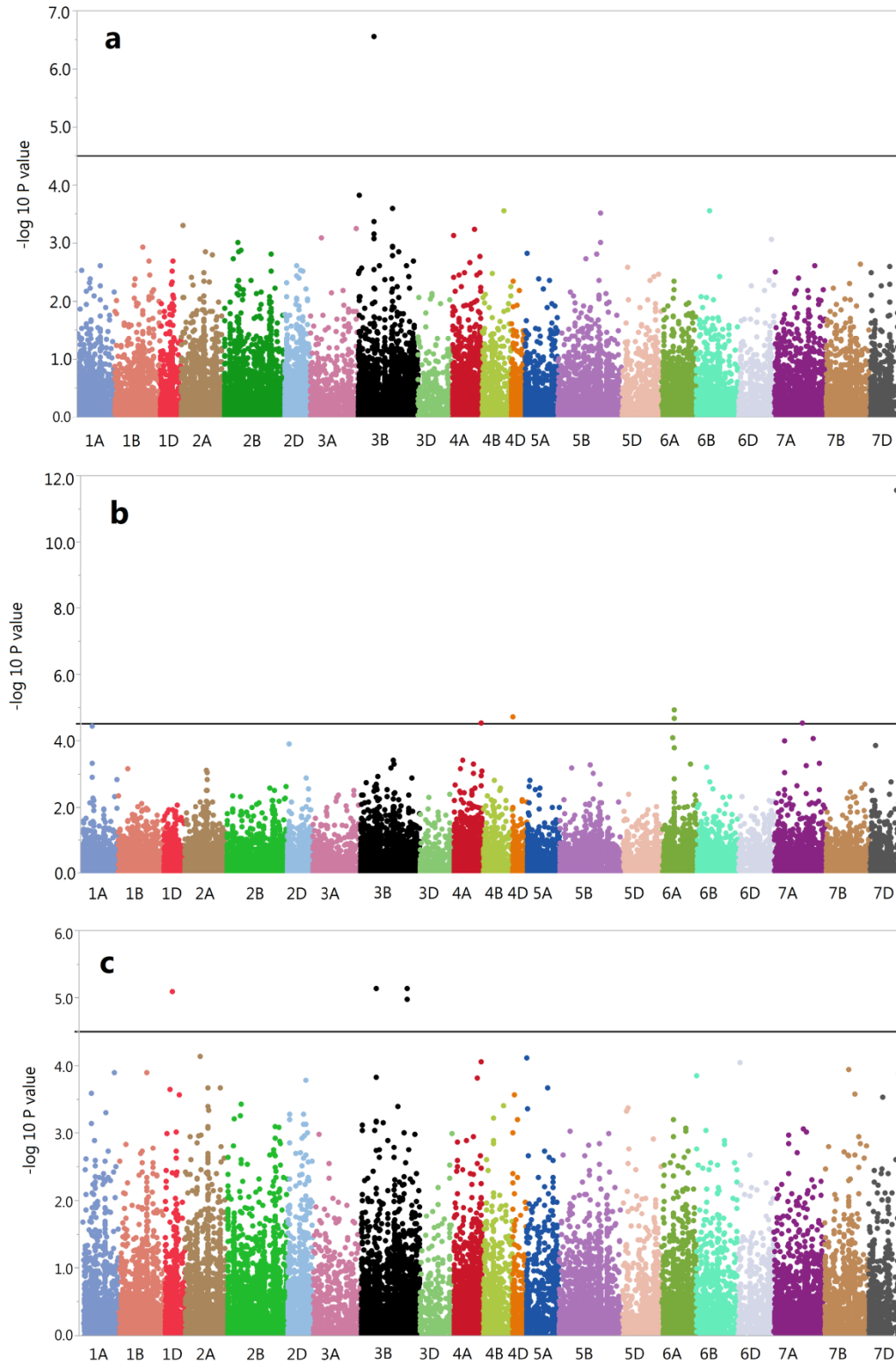


Figure 9. Genome-wide association scan for three parameters associated with FHB resistance: SEV (a), INC (b), and DON (c). The y axis represents the P value of the marker-trait association on a $-\log_{10}$ scale. The horizontal line represents the threshold for declaring a marker as significant (FDR-adjusted P-value ≤ 0.10).

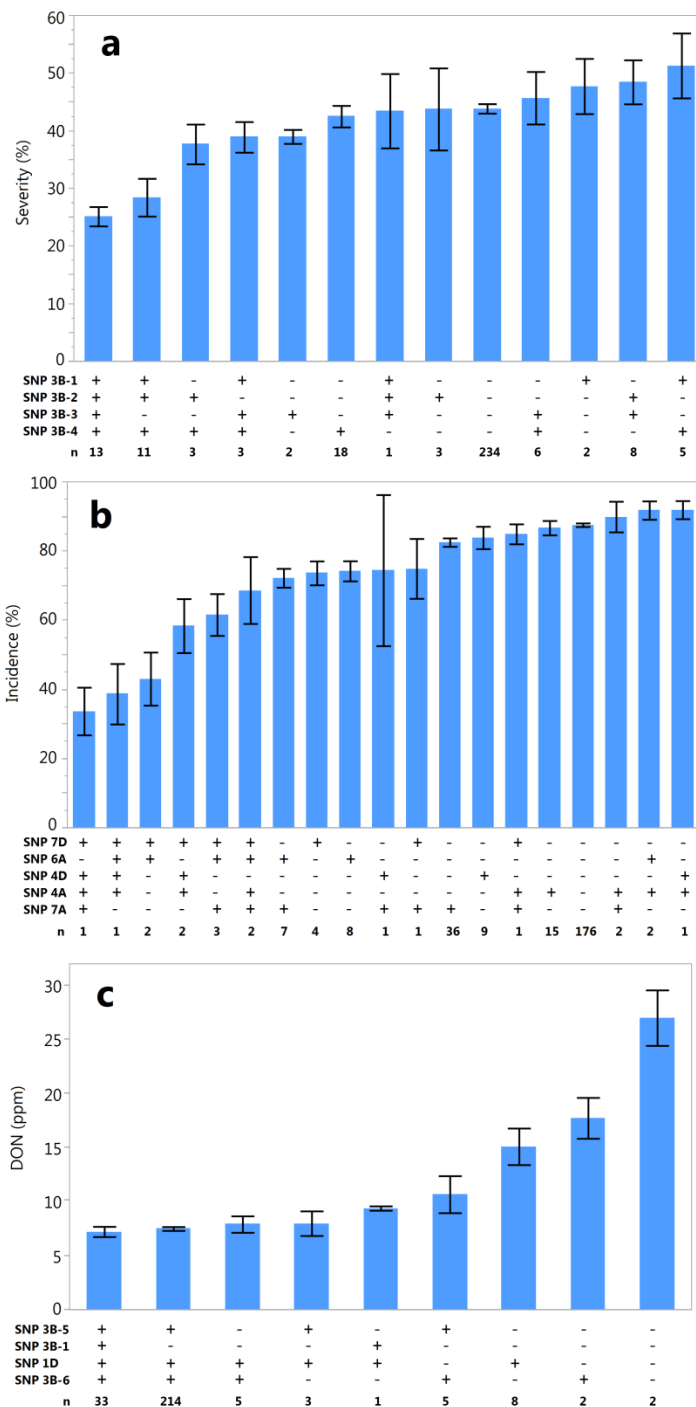


Figure 10. Mean phenotypic value of SEV (a), INC (b), and DON (c) for groups of wheat lines carrying different SNP combinations, with (“+”) signal representing the favorable allele. Lines were evaluated in 2011, 2013, and 2014 in Urbana-IL. SNP 3B-1 = WCSS1_contig10676713_3B_7175, SNP 3B-2 = WCSS1_contig10352272_3B_5482, SNP 3B-3 = WCSS1_contig10698462_3B_2332, SNP 3B-4 = WCSS1_contig10699215_3B_3620, SNP 7D = WCSS1_contig3876750_7DS_2023, SNP 6A = WCSS1_contig5780077_6AL_12152, SNP 4D = WCSS1_contig2300354_4DS_4482, SNP 4A = WCSS1_contig7146617_4AL_11335, SNP 7A = WCSS1_contig4132011_7AS_1400, SNP 3B-5 = WCSS1_contig10413672_3B_4839, SNP 1D = WCSS1_contig1879930_1DS_3352, SNP 3B-6 = WCSS1_contig10764618_3B_2168.

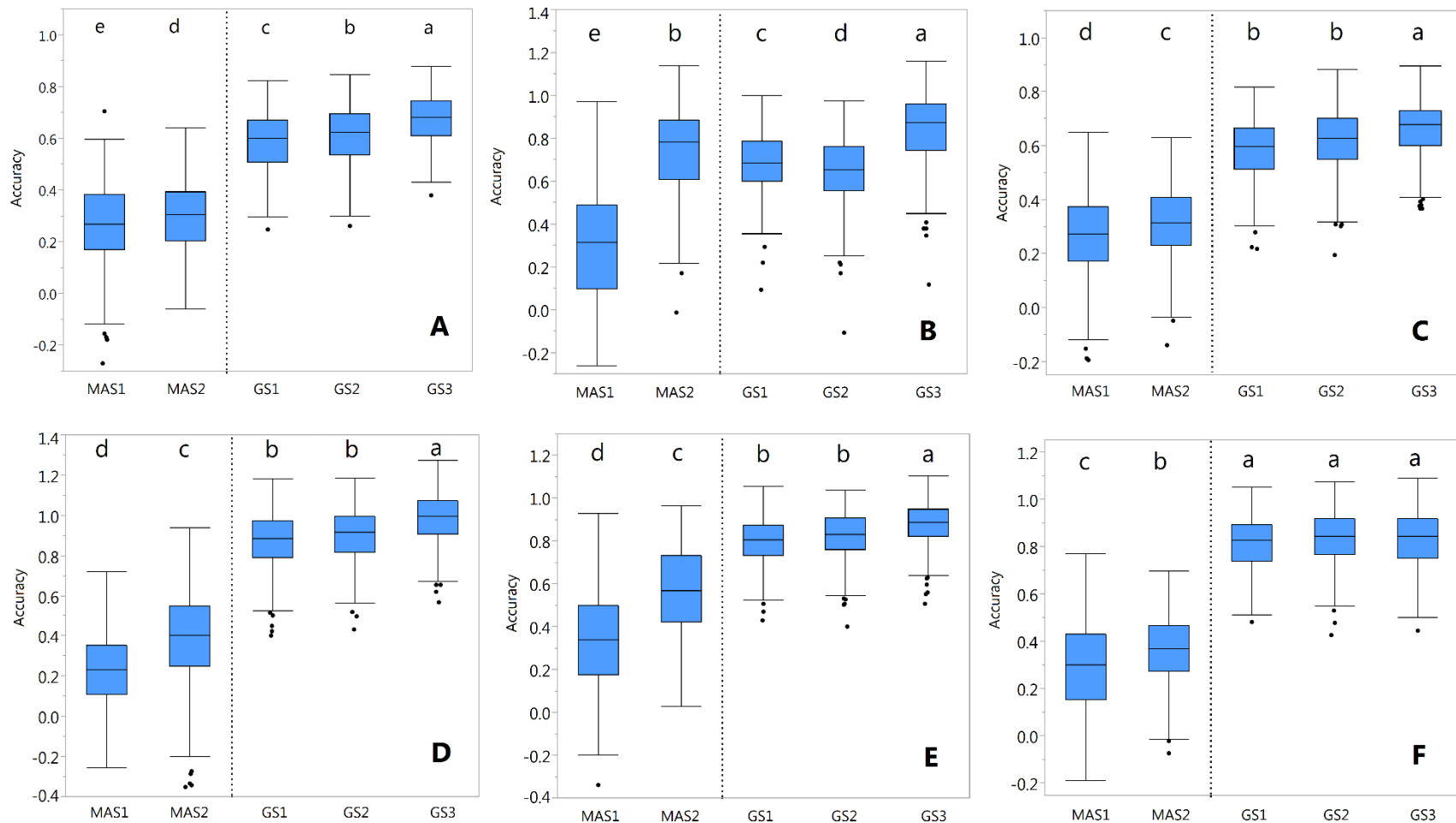


Figure 11. Cross validated prediction accuracies for six parameters associated with Fusarium head blight (FHB) resistance: severity (A), incidence (B), FHB index (C), Fusarium-damaged kernel (D), incidence-severity-kernel index (E), and deoxynivalenol concentration (F). The dotted line divides marker-assisted selection (left) and genomic selection (GS) (right). MAS1: Marker assisted selection with *Fhb-1* using linear regression. MAS2: Marker assisted selection with QTL using multiple linear regression. GS1: genomic selection with 19,992 SNPs using ridge regression-best linear unbiased predictor (RR-BLUP) model. GS2: genomic selection with 19,992 SNPs + *Fhb-1* using RR-BLUP. GS3: genomic selection with 19,992 SNPs + QTL using RR-BLUP. All marker estimates were obtained using a 5-fold cross validation scheme, 60 randomly selected training populations (TP) of size 218 and validation population (VP) of size 55. Treatments with the same letter at not significantly different according to the Ryan-Einot-Gabriel-Welch-q test at $\alpha = 0.05$ level.

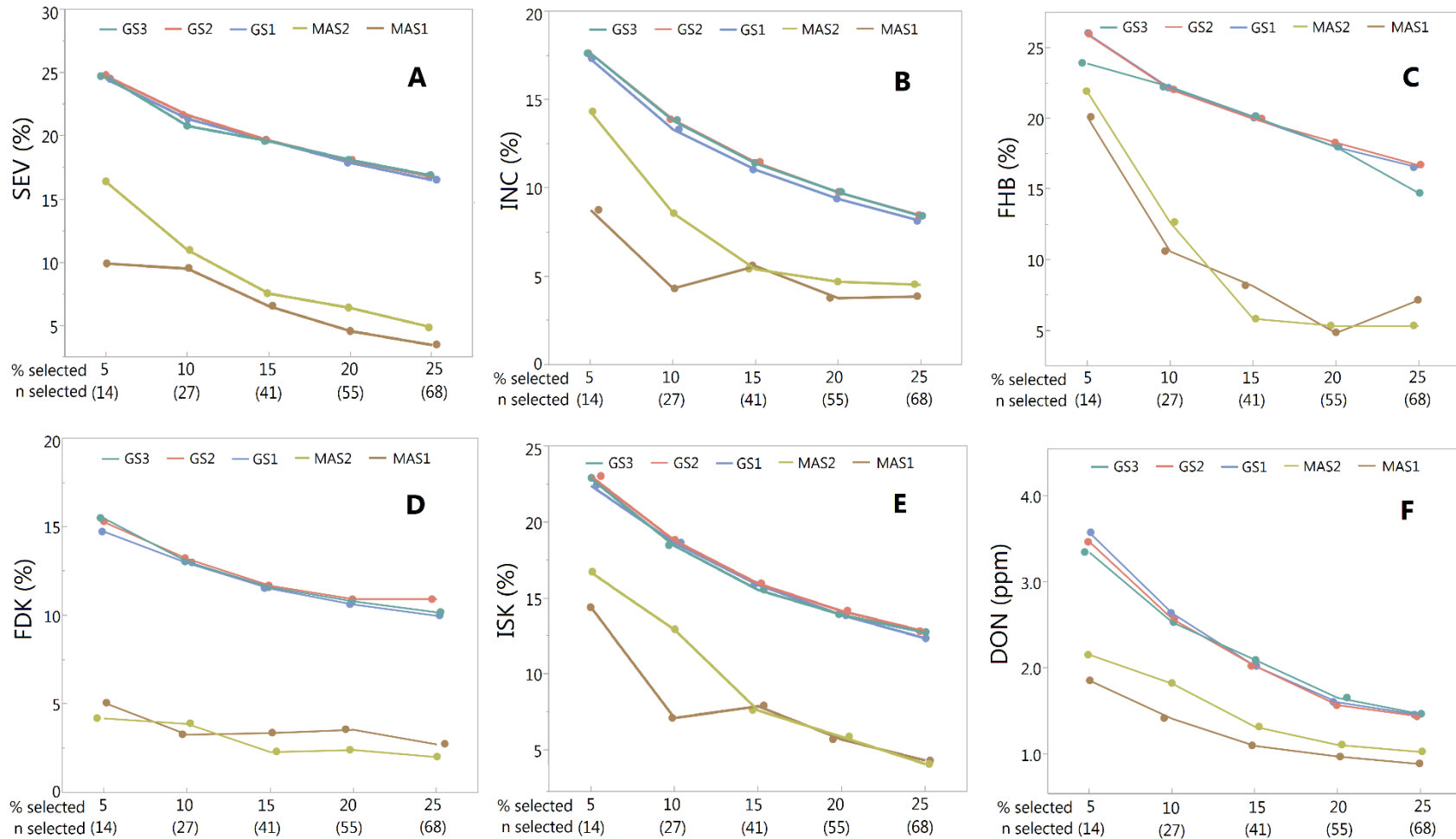


Figure 12. Selection differential for six parameters associated with Fusarium head blight (FHB) resistance: severity (A), incidence (B), FHB index (C), Fusarium-damaged kernel (D), incidence-severity-kernel index (E), and deoxynivalenol concentration (F). The numbers on the x-axis represent the percentage and number of selected individuals after ranking 273 breeding lines according to breeding values estimated through five different models: MAS1: Marker assisted selection with *Fhb-1* using linear regression, MAS2: Marker assisted selection with QTL using multiple linear regression, GS1: genomic selection with 19,992 SNPs using ridge regression-best linear unbiased predictor (RR-BLUP) model, GS2: genomic selection with 19,992 SNPs + *Fhb-1* using RR-BLUP, GS3: genomic selection with 19,992 SNPs + QTL using RR-BLUP.

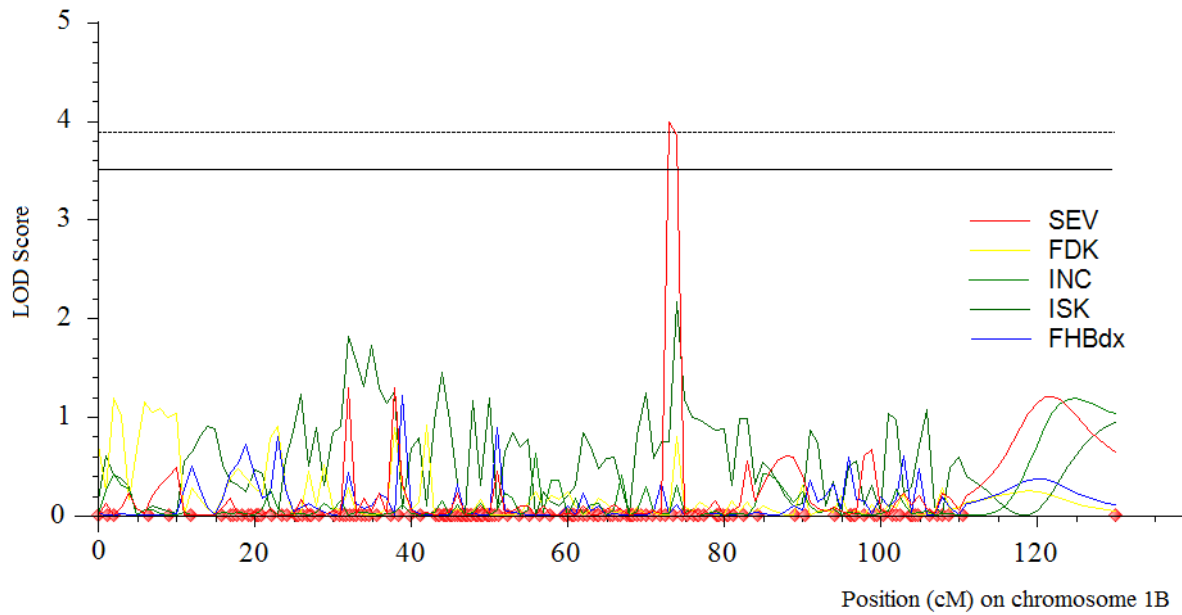


Figure 13. Linkage map and LOD scores after inclusive composite interval mapping (ICIM) analysis for FHB resistance for chromosome 1B. Dotted and solid lines represent significant thresholds obtained after 1000 permutations at type I error probability of 0.05 and 0.10, respectively.

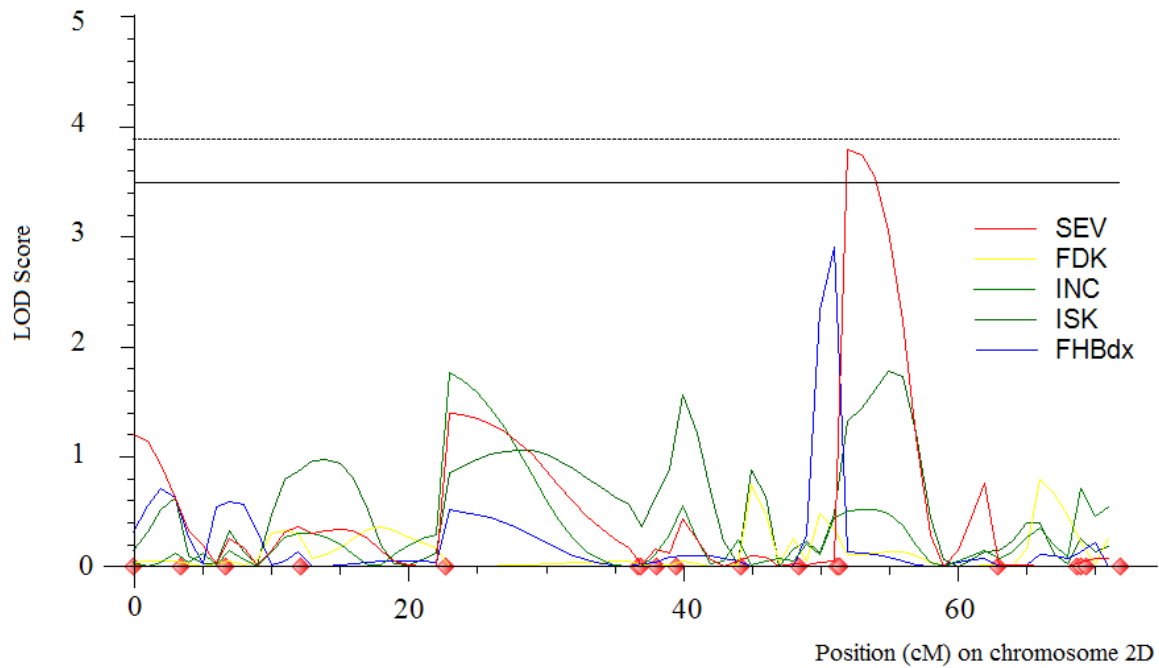


Figure 14. Linkage map and LOD scores after inclusive composite interval mapping (ICIM) analysis for FHB resistance for chromosome 2D. Dotted and solid lines represent significant thresholds obtained after 1000 permutations at type I error probability of 0.05 and 0.10, respectively.

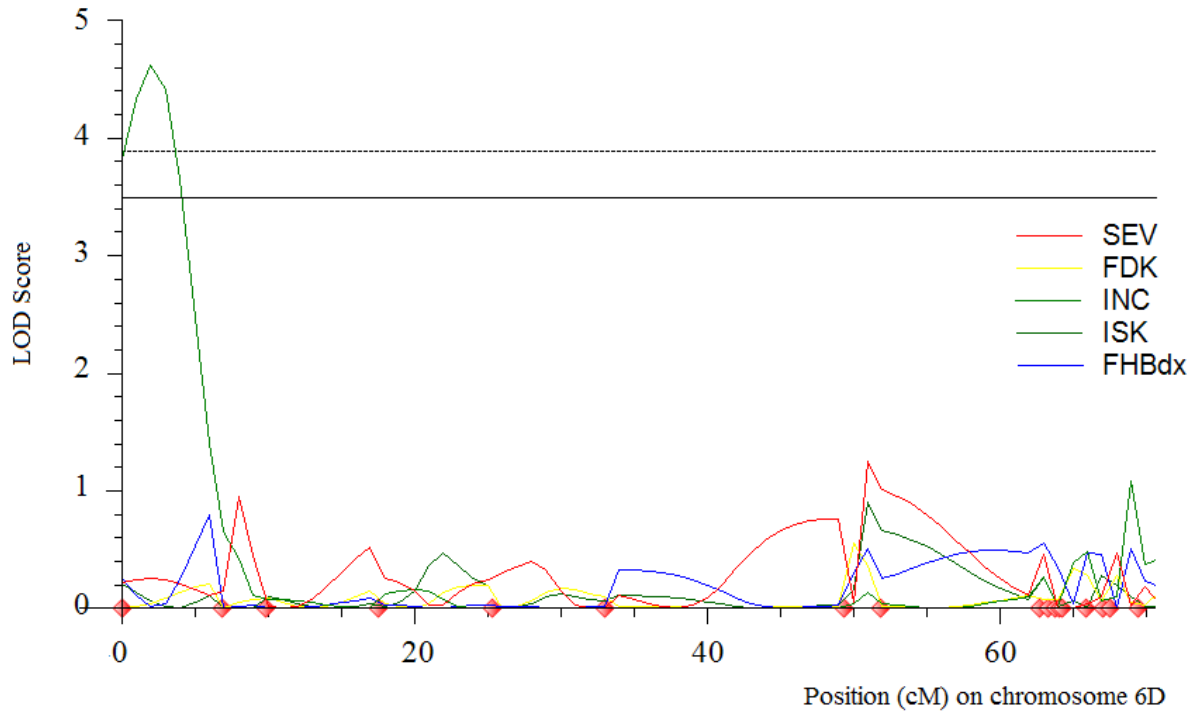


Figure 15. Linkage map and LOD scores after inclusive composite interval mapping (ICIM) analysis for FHB resistance for chromosome 6D. Dotted and solid lines represent significant thresholds obtained after 1000 permutations at type I error probability of 0.05 and 0.10, respectively.

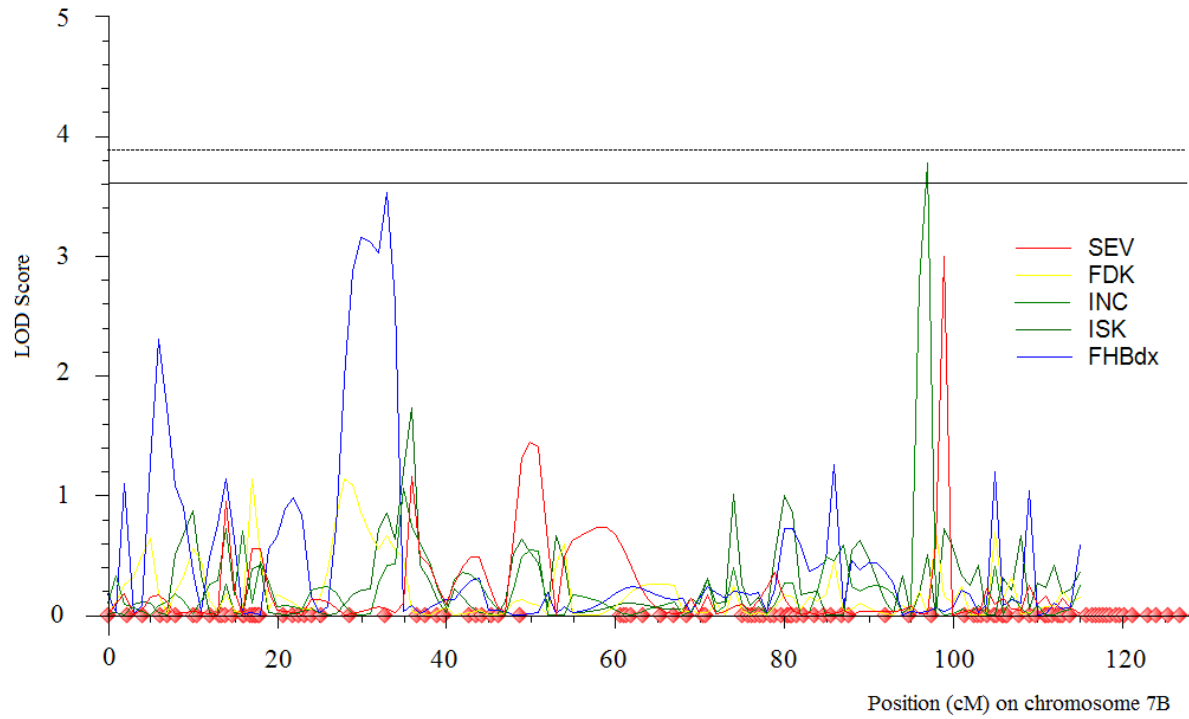


Figure 16. Linkage map and LOD scores after inclusive composite interval mapping (ICIM) analysis for FHB resistance for chromosome 7B. Dotted and solid lines represent significant thresholds obtained after 1000 permutations at type I error probability of 0.05 and 0.10, respectively.

TABLES

Table 1. Five-fold cross validated prediction accuracies (\pm standard error of the mean) for FHB related traits according to GS model and marker density. For this analysis, genotypes were imputed by expectation maximization imputation method. Training population size was equal to 218 and validation population size equal to 55. RR-BLUP = Ridge regression – best unbiased linear predictor, NET = ELASTI-NET, LASSO = least absolute shrinkage and selection operator, SEV = severity, INC = incidence, FHBdx = Fusarium head blight index, FDK = Fusarium-diseased kernels, ISK = incidence, severy and kernel quality index, DON = deoxynivalenol concentration.

SNP number	SEV		
	RR	NET	LASSO
4500	0.59 \pm 0.0067Aa ^{†‡}	0.52 \pm 0.0074Ba	0.52 \pm 0.0076Ba
3000	0.57 \pm 0.0061Aa	0.51 \pm 0.0071Bab	0.51 \pm 0.0073Bab
1500	0.54 \pm 0.0070Ab	0.49 \pm 0.0076Bb	0.49 \pm 0.0075Bb
500	0.51 \pm 0.0063Ac	0.46 \pm 0.0070Bc	0.46 \pm 0.0070Bc
INC			
	RR	NET	LASSO
4500	0.68 \pm 0.0103Aa	0.69 \pm 0.0138Aa	0.68 \pm 0.0143Aa
3000	0.66 \pm 0.0106Aab	0.65 \pm 0.0143Ab	0.65 \pm 0.0146Ab
1500	0.63 \pm 0.0078Ab	0.58 \pm 0.0089Bc	0.58 \pm 0.0088Bc
500	0.60 \pm 0.0081Ac	0.49 \pm 0.0086Bd	0.49 \pm 0.0086Bd
FHBdx			
	RR	NET	LASSO
4500	0.61 \pm 0.0065Aa	0.54 \pm 0.0065Ba	0.54 \pm 0.0065Ba
3000	0.58 \pm 0.0060Ab	0.49 \pm 0.0066Bb	0.49 \pm 0.0064Bb
1500	0.58 \pm 0.0065Ab	0.45 \pm 0.0071Bc	0.45 \pm 0.0068Bc
500	0.50 \pm 0.0073Ac	0.41 \pm 0.0075Bd	0.41 \pm 0.0072Bd
FDK			
	RR	NET	LASSO
4500	0.88 \pm 0.0074Aa	0.83 \pm 0.0079Ba	0.83 \pm 0.0078Ba
3000	0.87 \pm 0.0071Aa	0.76 \pm 0.0083Bb	0.75 \pm 0.0083Bb
1500	0.85 \pm 0.0076Aa	0.75 \pm 0.0077Bb	0.75 \pm 0.0076Bb
500	0.79 \pm 0.0079Ab	0.74 \pm 0.0081Bb	0.73 \pm 0.0082Bb
ISK			
	RR	NET	LASSO
4500	0.81 \pm 0.0065Aa	0.74 \pm 0.0070Ba	0.74 \pm 0.0071Ba
3000	0.81 \pm 0.0061Aa	0.71 \pm 0.0071Bb	0.70 \pm 0.0071Bb
1500	0.75 \pm 0.0069Ab	0.68 \pm 0.0075Bbc	0.68 \pm 0.0076Bbc
500	0.75 \pm 0.0068Ab	0.66 \pm 0.007B5c	0.66 \pm 0.0074Bc
DON			
	RR	NET	LASSO
4500	0.74 \pm 0.0054Aa	0.65 \pm 0.0069Ba	0.65 \pm 0.0072Ba
3000	0.74 \pm 0.0059Aa	0.65 \pm 0.0062Ba	0.64 \pm 0.0063Ba
1500	0.73 \pm 0.0057Aa	0.64 \pm 0.0062Ba	0.63 \pm 0.0061Ba
500	0.69 \pm 0.0065Ab	0.61 \pm 0.0070Bb	0.61 \pm 0.0071Bb

[†] Within rows, means followed by the same capital letter are not significantly different according to REGWQ (0.05)

[‡] Within columns, means followed by the same lower case letter are not significantly different according to REGWQ (0.05)

Table 2. Estimate (\pm standard error of the mean) and 99% confidence interval of unbiasedness for six FHB-related traits

Trait	$b(\text{PEBV}, \text{GEBV})^\dagger$	99% CI	
SEV	1.002 ± 0.0161	0.9598	1.0434
INC	1.048 ± 0.0195	0.9973	1.0987
FHBdx	1.009 ± 0.0163	0.9667	1.0510
FDK	1.003 ± 0.0142	0.9661	1.0402
ISK	1.014 ± 0.0124	0.9815	1.0461
DON	0.974 ± 0.0220	0.9169	1.031

[†] Unbiasedness estimated by the slope of the linear regression of PEBVs on GEBVs. The GEBVs were calculated using RR-BLUP with 5054 SNPs, genotypic missing data imputed by EMI, and TP size = 218. Confidence interval based on 300 values of the slope.

Table 3. 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.

Table 4. 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 5. 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, 3 on 3B

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 6. SNPs associated with FHB resistance in a panel of 273 breeding lines, chromosomal position, *P* values, and marker effects.

Trait	SNP	C	cM	<i>p</i> ^a	maf ^b	<i>r</i> ²	Adj <i>p</i> ^c	effects
SEV	WCSS1_contig10676713_3B_7175	3B	18.32	5.14	0.11	0.08	0.050	-9.54
	WCSS1_contig10352272_3B_5482 ^d	3B	10.19	3.83	0.14	0.05	0.980	-7.21
	WCSS1_contig10698462_3B_2332 ^d	3B	6.86	3.6	0.44	0.04	0.980	6.02
	WCSS1_contig10699215_3B_3620 ^d	3B	18.32	3.37	0.20	0.04	0.980	-4.99
INC	WCSS1_contig3876750_7DS_2023	7D	70.84	11.57	0.07	0.16	<0.001	6.74
	WCSS1_contig5780077_6AL_12152	6A	134.15	4.94	0.07	0.06	0.097	3.60
	WCSS1_contig2300354_4DS_4482	4D	0	4.72	0.06	0.06	0.097	-4.56
	WCSS1_contig7146617_4AL_11335	4A	78.35	3.54	0.06	0.06	0.070	-3.02
	WCSS1_contig4132011_7AS_1400	7A	22.82	4.54	0.16	0.05	0.097	-2.62
FHBdx	WCSS1_contig10676713_3B_7175	3B	18.32	5.14	0.11	0.07	0.052	-8.96
ISK	WCSS1_contig10676713_3B_7175	3B	18.32	5.14	0.11	0.07	0.052	-5.55
	WCSS1_contig3876750_7DS_2023	7D	70.84	11.57	0.07	0.16	0.000	7.47
DON	WCSS1_contig10413672_3B_4839	3B	73.67	5.14	0.06	0.07	0.052	1.75
	WCSS1_contig10676713_3B_7175	3B	18.32	5.14	0.11	0.07	0.052	-1.36
	WCSS1_contig1879930_1DS_3352	1D	19.04	5.10	0.05	0.06	0.052	-1.99
	WCSS1_contig10764618_3B_2168	3B	73.67	4.99	0.07	0.06	0.052	2.08

^a *P* value reported in a $-\log_{10}$ scale; ^b maf = minor allele frequency; ^c Adj *p* = FDR-adjusted *p* value; ^d 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 7. 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: WCSS1_contig10676713_3B_7175, WCSS1_contig10352272_3B_5482, WCSS1_contig10698462_3B_2332, and WCSS1_contig10699215_3B_3620. Haplotypes for INC were built with: WCSS1_contig3876750_7DS_2023, WCSS1_contig5780077_6AL_12152, WCSS1_contig2300354_4DS_4482, WCSS1_contig7146617_4AL_11335, and WCSS1_contig4132011_7AS_1400. Haplotypes for DON were combinations of the following SNPs: WCSS1_contig10413672_3B_4839, WCSS1_contig10676713_3B_7175, WCSS1_contig1879930_1DS_3352, and WCSS1_contig10764618_3B_2168.

Table 8. Hypothetical protein prediction and conserved domains for contigs harboring SNPs associated with FHB resistance.

SNP	C	cM	hypothetical protein prediction	conserved domains
WCSS1_contig1879930_1DS_3352	1D	19.04	RLK protein (<i>Brachypodium</i> , <i>Setaria</i> , <i>Nicotiana</i> , <i>Solanum</i>)	none
WCSS1_contig10413672_3B_4839	3B	73.67	chloroplast DNA / transcription factor	BSD superfamily
WCSS1_contig10764618_3B_2168	3B	73.67	uncharacterized hypothetical protein	none
WCSS1_contig7146617_4AL_11335	4A	78.35	FAR-1-related (<i>Tritum urartu</i>)	FAR-1 DNA-binding
WCSS1_contig2300354_4DS_4482	4D	0	PPI (<i>Brachypodium</i> , <i>Oryza</i> , <i>Zea</i> , <i>Elaeis</i> , etc)	Protein phosphatase inhibitor-2 superfamily
WCSS1_contig5780077_6AL_12152	6A	134.2	Rop guanine nucleotide exchange factor 2 (<i>Aegilops tauschii</i>)	FHA, PRONE, DNA-polymerase-viral-N-terminal,
WCSS1_contig4132011_7AS_1400	7A	22.82	Putative reverse transcriptase (<i>Oryza sativa</i>)	Reverse-transcriptase superfamily and EEP family
WCSS1_contig3876750_7DS_2023	7D	70.84	GDSL esterase/lipase (<i>Aegilops tauschii</i>)	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), aspartic 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.

Table 9. Descriptive statistics and Spearman correlation coefficient for six measurements associated with Fusarium head blight (FHB) resistance.

Trait	Parents		233 RILs					Correlation ¹				
	IL97-1828 (R)	Clark (S)	Mean	Min	Max	Range	SD	SEV	INC	FHBdx	FDK	ISK
SEV	32.74	74.66	54.11	37.90	75.24	37.34	7.92					
INC	57.44	86.77	74.75	52.65	86.47	33.81	6.67	0.46*				
FHBdx	20.30	67.19	40.70	23.03	60.28	37.25	8.24	0.92*	0.75*			
FDK	21.25	63.75	38.99	22.50	62.50	40.00	7.56	0.33*	0.36*	0.39*		
ISK	36.07	72.50	54.79	40.00	70.73	30.73	5.78	0.67*	0.70*	0.79*	0.80*	
DON	8.90	11.90	10.19	4.55	22.30	17.75	3.05	0.30*	0.42*	0.40*	0.45*	0.47*

Summary statistics of least-squares means for six parameters associated with FHB resistance evaluated in 233 recombinant inbred lines (RILs) from a cross between IL97-1828 (resistant) and Clark (susceptible). Least-squares means were calculated from phenotypic data obtained in Urbana, Illinois in 2009 and 2010, and Wooster, Ohio in 2010. Deoxynivalenol (DON) concentration was not recorded in Wooster, Ohio in 2010. ¹Spearman correlation coefficient. * Significant at $\alpha = 0.05$ level. DON was measured in ppm, and the other parameters were measured in percentage.

Table 10. SNP coverage obtained in a population of 233 winter wheat recombinant inbred lines using genotyping-by-sequencing.

Chromosome	n	cM	D^a	
Genome A	1	97	86	0.88
	2	150	211	1.41
	3	57	96	1.69
	4	109	162	1.48
	5	47	117	2.49
	6	62	78	1.26
	7	175	194	1.11
	697^b	944^c	1.11^d	
Genome B	1	208	130	0.63
	2	313	232	0.74
	3	578	325	0.56
	4	48	104	2.17
	5	51	178	3.50
	6	100	100	1.00
	7	93	115	1.24
	1391	1185	1.41	
Genome D	1	70	154	2.20
	2	26	72	2.76
	3	8	59	7.36
	4	5	96	19.26
	5	34	178	5.24
	6	21	72	3.43
	7	23	115	4.99
	187	746	6.46	
total	2275	2875	1.26^e	

^a D: Density (genetic distance divided by the number of SNPs). ^b Number of SNPs on genome A. ^c Length of genome A in centimorgans (cM). ^d Average SNP density for genome A. ^e Average SNP density across genomes A, B, and D.

Table 11. Summary of inclusive composite interval mapping (ICIM) analysis for FHB resistance performed in 233 wheat recombinant inbred lines (RILs) from a cross between IL97-1828 (resistance) and Clark (susceptible).

Trait	C ^a	P ^b	Flanking markers		D ^c	LOD	r (%)	Effect	alpha	Env ^d
SEV	1B	73	WCSS1_contig_1190535_1BS_1034	WCSS1_contig_3457475_1BS_7477	1.09	3.99	7.03	2.10	0.05	3
SEV	2D	52	WCSS1_contig_5315750_2DS_5213	WCSS1_contig_5354063_2DS_3647	11.50	3.79	6.72	-2.05	0.10	3
INC	6D	72	WCSS1_contig_3239902_6DL_273	WCSS1_contig_2895599_6DL_459	6.87	4.61	12.49	2.36	0.05	3
INC	7B	97	WCSS1_contig_5085227_7BL_199	WCSS1_contig_3115996_7BL_172	1.88	3.78	10.66	-2.17	0.10	3
FDK	2A	54	WCSS1_contig_5308233_2AS_2588	WCSS1_contig_5237718_2AS_3323	1.55	3.48	5.91	1.84	0.10	2
SEV	7B	19	WCSS1_contig_3110112_7BS_2886	WCSS1_contig_3075459_7BS_802	0.44	5.22	8.85	3.61	0.05	1a
ISK	2A	143	WCSS1_contig_6344163_2AL_8696	WCSS1_contig_6430050_2AL_1026	1.10	7.54	11.66	-3.06	0.05	1a
ISK	2A	186	WCSS1_contig_5282575_2AS_7303	WCSS1_contig_5302874_2AS_3463	1.61	5.95	7.40	-1.76	0.05	1b

^a C: wheat chromosome. ^b P: genetic position (cM) of the QTL. ^c D: genetic distance (cM) of the interval between flanking markers. ^d Env: environments in which the quantitative trait loci was significant (3 = Urbana, Illinois – 2009 and 2010, and Wooster, Ohio - 2010; 2 = Urbana, Illinois – 2009 and 2010; 1a = Urbana, Illinois – 2009; 1b = Urbana, Illinois – 2010).

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
- Albrecht, T., V. Wimmer, H.J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, et al. 2011. Genome-based prediction of testcross values in maize. *Theoretical and Applied Genetics* 123:339-350
- Anderson, J.A., R.W. Stack, S. Liu, B.L. Waldron, A.D. Fjeld, C. Coyne, et al. 2001. DNA markers for Fusarium head blight resistance QTL its two wheat populations. *Theoretical and Applied Genetics* 102:1164-1168
- Arthur, J.C. 1891. Wheat scab. *Purdue University Agricultural Experiment Station Bulletin* 36: 129-132
- Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.L. Jannink. 2010. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *The Plant Genome* 4:132-144

- Astanasoff, D. 1920. Fusarium-blight (scab) of wheat and other cereals. *Journal of Agricultural Research* 20:1-40
- Bai, G.H., and G.E. Shaner. 1994. Wheat scab: Perspective and control. *Plant Disease* 78:760-766
- Bai, G.H., F.L. Kolb, G. Shaner, et al. 1999. Amplified fragment length polymorphism markers linked to a major quantitative trait locus controlling scab resistance in wheat. *Phytopathology* 89:343-348
- Bai, G.H., G. Shaner, and H. Ohm. 2000. Inheritance of resistance to *Fusarium graminearum* in wheat. *Theoretical and Applied Genetics* 100:1-8
- Bai, G.H., A.E. Desjardins, and R.D. Plattner. 2001. Deoxynivalenol-nonproducing *Fusarium graminearum* causes initial infection, but does not cause disease spread in wheat spikes. *Mycopathologia* 153:91-98
- Bai, G.H., and G.E. Shaner. 2004. Management and resistance in wheat and barley to Fusarium head blight. *Annual Review of Phytopathology* 42:135-161
- Bakhsh, A., N. Mengistu, P.S. Baenziger, I. Dweikat, S.N. Wegulo, D.J. Rose, et al. 2013. Effect of Fusarium head blight resistance gene *Fhb1* on agronomic and end-use quality traits of hard red winter wheat. *Crop Science* 53:793-801

- Ban, T. 2000. Analysis of quantitative trait loci associated with resistance to Fusarium head blight caused by *Fusarium graminearum* Schwabe and of resistance mechanisms in wheat (*Triticum aestivum* L.). *Breeding Science* 50:131-137
- Baird, R.E., Mullinix, B.G., Peery, A.B., and Lang, M.L. 1997. Diversity and longevity of the soybean debris mycobiota in a no-tillage system. *Plant Disease* 81: 530-534
- Bennett, J. W., and M. Klich. 2003. Mycotoxins. *Clinical Microbiological Reviews* 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. *Molecular Breeding* 29:477-488
- Bernardo, R., and J. Yu. 2007. Prospect for genomewide selection for quantitative traits in maize. *Crop Science* 47:1082-1090
- Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Sciences* 48:1649-1664
- Bernardo, R. 2014. Genomewide selection when major genes are known. *Crop Science* 54:68-75

Blümke, A., C. Falter, C. Herrfurth, B. Sode, R. Bode, W. Schäfer, et al. 2014. Secreted fungal effector lipase releases free fatty acids to inhibit innate immunity-related callose formation during wheat head infection. *Plant Physiology* 165:346-358

Bonin, C.M., and F.L. Kolb. 2009. Resistance to *Fusarium* head blight and kernel damage in a winter wheat recombinant inbred line population. *Crop Science* 49:1304-1312

Boyacıoğlu, D., and N.S. Hettiarachcy. 1995. Changes in some biochemical components of wheat grain that was infected with *Fusarium graminearum*. *Journal of Cereal Science* 21:57-62

Buerstmayr, H., M. Lemmens, H. Grausgruber, and P. Ruckebauer. 1996. Scab resistance of international wheat germplasm. *Cereal Research Communications* 24:195-202

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

Bushnell, W.R., B.E. Hazen, and C. Pritsch. 2003. Histology and physiology of *Fusarium* head blight. pp. 44-83. In Leonard, K.J. and W.R. Bushnell (ed). *Fusarium head blight of wheat and barley*. St Paul: The American Phytopathological Society.

Cativelli, M., S. Lewis, and M.L. Appendino. 2013. A *Fusarium* head blight resistance quantitative trait locus on chromosome 7D of the spring wheat cultivar Catbird. *Crop Science* 53:1464-1471

Chen, J., C.A. Griffey, M.A.S. Saghai-Marooif, E.L. Stromberg, R.M. Biyashev, W. Zhao, et al. 2006. Validation of two major quantitative trait loci for Fusarium head blight resistance in Chinese wheat line W14. *Plant Breeding* 125:99-101

Chester, F.D. 1890. The scab of wheat. *Delaware Agriculture Experiment Station Reports* 3:89-90

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

Christ, D.S., Gödecke, R., von Tiedemann, A., and Varrelmann, M. 2011. Pathogenicity, symptom development, and mycotoxin formation in wheat by *Fusarium* species frequently isolated from sugar beet. *Phytopathology* 101:1338-1345

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, J. Cerón-Rojas, et al. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112:48-60

Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* 3:e3395

Dekkers, J.C.M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics* 124:331-341

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39:1-38

Desjardins, A.E. 2006. *Fusarium mycotoxins chemistry, genetics and biology*. Sait Paul: The American Phytopathological Society.

Desta, Z.A., and R. Ortiz. 2014. Genomic selection: Genome-wide prediction in plant improvement. *Trends in Plant Science* 19:592-601

Dexter, J.E., R.M. Clear, and K.R. Preston. 1996. Fusarium head blight: Effect on the milling and baking of some Canadian wheats. *Cereal Chemistry* 73:695-701

Dickson, J.G., E.B. Mains. 1929. Scab of wheat and barley and its control. *USDA Farmers' Bulletins*. 1599. 18pp.

Eckard, J., J. Gonzalez-Hernandez, M. Caffè, W. Berzonsky, W. Bockus, G.F. Marais, et al. 2015. Native Fusarium head blight resistance from winter wheat cultivars 'Lyman,' 'Overland,' 'Ernie,' and 'Freedom' mapped and pyramided onto 'Wesley'-Fhb1 backgrounds. *Molecular Breeding* 35:1-16

Edwards, S.G. 2004. Influence of agricultural practices on fusarium infection of cereals and subsequent contamination of grain by trichothecene mycotoxins. *Toxicology Letters* 153:29-35

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

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* 4:250-255

Estrada Jr, R., Gudmestad, N.C., Rivera, V.V., and Secor, G.A. 2010. *Fusarium graminearum* as a dry rot pathogen of potato in the USA: prevalence, comparison of host isolates aggressiveness and factors affecting aetiology. *Plant Pathology* 59:1114-1120

Farr, D.F., G.F. Bills, G.P. Chamuris, et al. 1989. *Fungi on plants and plant products in the United States*. St. Paul: American Phytopathological Society Press. 1252pp.

Fehr, W.R. 1991. *Principles of cultivar development*. V1: Theory and technique. Ames: Iowa State Univ. Press.

Flint-Garcia, S.A., J.M. Thornsberry, and E.S.I. Buckler. 2003. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54:357-374

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33:1-22

Gervais, L., F. Dedryver, J.Y. Morlais, V. Bodusseau, S. Negre, M. Bilous, et al. 2003. Mapping of quantitative trait loci for field resistance to Fusarium head blight in a European winter wheat. *Theoretical and Applied Genetics* 106:961-970

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: Genes|Genomes|Genetics* 1:209-218

Gianola, D. 2013. Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* 194:573-596

Gilsinger, J., Kong, L., Shen, X., and H. Ohm. 2005. DNA markers associated with low Fusarium head blight incidence and narrow flower opening in wheat. *Theoretical and Applied Genetics* 110:1218-1225

Ginkel, M.V., W.V.D. Schaar, and Y. Zhuping. 1996. Inheritance of resistance to scab in two wheat cultivars from Brazil and China. *Plant Disease* 80:863-867

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

Goddard, M.E., and B.J. Hayes. 2007. Genomic selection. *Journal of Animal Breeding and Genetics* 124:323-330

- Goff, K.E., and K.M. Ramonell. 2007. The role and regulation of receptor-like kinases in plant defense. *Gene Regulation and Systems Biology* 1:167-175
- Gooding, R.W., H.N. Lafever, K.G. Campbell, and L.D Herald. 1997. Registration of Freedom wheat. *Crop Science* 37:1007
- Griffey, C.A., T.M. Starling, A.M. Price, W.L. Sisson, M.K. Das, T.H. Pridgen, et al. 2001. Registration of Roane wheat. *Crop Science* 41:1359
- Guo, P.G., G.H. Bai, and G.E. Shaner. 2003. AFLP and STS tagging of a major QTL for *Fusarium* head blight resistance in wheat. *Theoretical and Applied Genetics* 106:1011-1017
- Guo, X.W., X.G.D. Fernando, P. Bullock, and H. Sapistein. 2010. Quantifying cropping practices in relation to inoculums levels of *Fusarium graminearum* on crop stubble. *Plant Pathology* 59:1107-1113
- 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
- Häberle, J., J. Holzapfel, G. Schweizer, and L. Hartl. 2009. A major QTL for resistance against *Fusarium* head blight in European winter wheat. *Theoretical and Applied Genetics* 119:325-332

Habier, D., R.L. Fernando, K. Kizilkaya, and D.J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186

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

Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. *Crop Sciences* 49:1-12

Heffner, E.L., A.J. Lorenz, J.L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: Gain per unit time and cost. *Crop Sciences* 50:1681-1690

Heffner, E.L., J.L. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome* 4:65-75

Heslot, N., H.P. Yang, M.E. Sorrells, and J.L. Jannink. 2012. Genomic selection in plant breeding: a comparison of models. *Crop Sciences* 52:14-160

Hoerl, A.E., and R.W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55-67

Hofheinz, N., and M. Frisch. 2014. Heteroscedastic ridge regression approaches for genome-wide prediction with a focus on computational efficiency and accurate effect estimation. *G3: Genes|Genomes|Genetics* 4:539-546

Holzappel, J., H.H. Voss, T. Miedaner, V. Korzun, J. Haberle, G. Schweizer, et al. 2008. Inheritance of resistance to *Fusarium* head blight in three European winter wheat populations. *Theoretical and Applied Genetics* 117:1119-1128

Isidro, J., J.L. Jannink, D. Akdemir, J.A. Poland, N. Heslot, and M.E. Sorrels. 2015. Training set optimization under population structure in genomic selection. *Theoretical and Applied Genetics* 128:145-158

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

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 Research* 17:175-183

Iwata, H., and Jannink, J. L. 2011. Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines. *Crop Science* 51:1915-1927

Jannink, J.L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* 9:166-177

Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, J. Lorgeou, et al. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics* 127: 595-607

Jia, G., P. Chen, G. Qin, G. Bai, X. Wang, S. Wang, et al. 2005. QTL for Fusarium head blight response in a wheat DH population of Wangshuibai/Alondra's'. *Euphytica* 146:183-191

JMP, version 11. SAS Institute Inc., Cary, NC.

Klahr, A., G. Zimmermann, G. Wenzel, and V. Mohler. 2007. Effects of environment, disease progress, plant height and heading date on the detection of QTLs for resistance to Fusarium head blight in an European winter wheat cross. *Euphytica* 154:17-28

Kohli, M. M. 1989. Taller sobre fusariose de la espiga en America del Sur. Mexico City: CIMMYT

Kolb, F.L., G.H. Bai, G.J. Muehlbauer, et al. 2001. Host plant resistance genes for Fusarium head blight: Mapping and manipulation with molecular markers. *Crop Science* 41:611-619

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

Kugler, K.G., G. Siegwart, T. Nussbaumer, C. Ametz, M. Spannagl, B. Steiner, et al. 2013. Quantitative trait loci-dependent analysis of a gene co-expression network associated with Fusarium head blight resistance in bread wheat (*Triticum aestivum* L.). BMC Genomics 14:728

Kuiper-Goodman, T. 2004. Risk assessment and risk management of mycotoxins in food. In: Magan, N. and Olsen, M. (ed) Mycotoxins in food. Detection and control. Cambridge: Woodhead Publishing Limited. pp. 3-31

Lado, B., I. Matus, A. Rodríguez, L. Inostroza, J.A. Poland, F. Belzile, et al. Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. G3: Genes|Genomes|Genetics 3:2105-2114

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

Lemmens, M., U. Scholz, F. Berthiller, C. Dall'Asta, A. Koutnik, R. Schumacher, et al. 2005. The ability to detoxify the mycotoxin deoxynivalenol colocalizes with a major quantitative trait locus for Fusarium head blight resistance in wheat. Molecular Plant Microbe Interactions 18:1318-1324

Li, H., J.M. Ribaut, Z. Li, and J. Wang. 2008. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theoretical and Applied Genetics* 116:243-260

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760

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

Lipka, A.E., F. Lu, J.H. Cherney, E.S. Buckler, M.D. Casler, and D.E. Costich. 2014. Accelerating the switchgrass (*Panicum virgatum* L.) breeding cycle using genomic selection approaches. *PLoS ONE* 9:e112227

Liu, S., A. Abate, and A.L. McKendry. 2005. Inheritance of Fusarium head blight resistance in the soft red winter wheat Ernie. *Theoretical and Applied Genetics* 110:454-461

Liu, S., Z.A. Abate, H. Lu, G. Davis, and A. McKendry. 2007. QTL associated with Fusarium head blight resistance in the soft red winter wheat Ernie. *Theoretical and Applied Genetics* 115:417-427

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

Loffler, 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. *Molecular Breeding* 23:473-488

Lorenz, A.J., S. Chao, F.G. Asoro, et al. 2011. Genomic selection in plant breeding: Knowledge and prospects. *Advances in Agronomy* 110:77-123

Lorenz, A. J., K.P. Smith, J.L. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Science* 52:1609-1621

Lorenzana, R.E., R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics* 120:151-161

Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, et al. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet* 9: e1003215

MacInnes, J., and R. Fogelman. 1923. Wheat Scab in Minnesota. University of Minnesota Technical Bulletins 18.

Madden, L.V., and P.A. Paul. 2009. Assessing heterogeneity in the relationship between wheat yield and Fusarium head blight intensity using random-coefficient mixed models. *Phytopathology* 99:850-860

Mardi, M., H. Buerstmayr, B. Ghareyazie, M. Lemmens, S. Mohammadi, R. Nolz, et al. 2005. QTL analysis of resistance to Fusarium head blight in wheat using a 'Wangshuibai' - derived population. *Plant Breeding* 124:329-333

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. *Molecular Breeding* 30:951-966

McFall, K.L., and M.E. Fowler. 2009. Overview of wheat classification and trade. pp. 439-454. In B.F. Carver (ed). *Wheat science and trade*. Ames: Wiley-Blackwell

McKendry, A.L., J.E. Berg, D.N. Tague, and K.D. Kephart. 1995. Registration of 'Ernie' wheat. *Crop Science* 35:1513

McKendry, A.L., D.N. Tague, R.L. Wright, J.A. Tremain, and S.P. Conley. 2005. Registration of 'Truman' wheat. *Crop Science* 45:421-423

McKendry, A.L., D.N. Tague, R.L. Wright, and J.A. Tremain. 2007. Registration of 'Bess' Wheat. *Journal of Plant Registrations* 1:21-23

McMullen, M.P., R. Jones, and D. Gallenberg. 1997. Scab of wheat and barley - A re-emerging disease of devastating impact. *Plant Disease* 81:1340-1348

McMullen, M. 2003. Impacts of Fusarium head blight on the North American agriculture community: The power of one disease to catapult change. pp.484-503. In K.J. Leonard, and W.R. Bushnell (ed). Fusarium head blight of wheat and barley. St Paul: The American Phytopathological Society

McMullen, M., S. Halley, B. Schatz, S. Meyer, J. Jordahl, and J. Ransom. 2008. Integrated strategies for Fusarium head blight management in the United States. Cereal Research Communications 36:563-568

McMullen, M., G. Bergstrom, E. De Wolf, R. Dill-Macky, D. Hershman, G. Shaner, et al. 2012. A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Disease 96:1712-1728

Meng, L., H. Li, L. Zhang, and J. Wang. 2015. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal (in press)

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

Mesterházy, Á., T. Bartók, T., and C. Lamper. 2003. Influence of wheat cultivar, species of *Fusarium*, and isolates aggressiveness on the efficacy of fungicides for control of *Fusarium* head blight. *Plant Disease* 87:1107-1115

Mesterhazy, Á. 2003. Breeding wheat for *Fusarium* head blight resistance in Europe. pp.211-240. In K.J. Leonard, and W.R. Bushnell (eds.). St. Paul: The American Phytopathological Society

Meuwissen, T.H.E., B.J. Hayes, M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829

Michael, T.P., and S. Jackson. 2013. The first 50 plant genomes. *Plant Genome* 6(2):1-7

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. *Theoretical and Applied Genetics* 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. *Molecular Breeding* 28:647-655

Miller, J.D., J.C. Young, and D.R. Sampson. 1985. Deoxynivalenol and Fusarium head blight resistance in spring cereals. *Journal of Phytopathology* 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. *The Plant Cell* 21:2194-2202

Nakaya, A., and S.N. Isobe. 2012. Will genomic selection be a practical method for plant breeding? *Annals of Botany* 110:1303-1316

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

Ogutu, J.O., T. Schulz-Streeck, and H.P. Piepho. 2012. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proceedings* 6(2):S10

Ohm, H.W., G. Shaner, J.E. Foster, F.L. Patterson, and G. Buechley. 1988. Registration of Clark wheat. *Crop Science* 28:1032-1032

Ohm, H.W., G. Shaner, R.H. Ratcliffe, D.M. Huber, H. Sharma, K.L. Perry, et al. 2000. Registration of Goldfield wheat. *Crop Sciences* 40:580-581

Oldenburg, E., J. Weinet, and G.A. Wolf. 2001. Effects of strobilurin containing fungicides on the deoxynivalenol content in winter wheat. *Mycotoxin Research* 17:10-14

Owens, B.F., A.E. Lipka, M. Magallanes-Lundback, T. Tiede, C.H. Diepenbrock, C.B.

Kandianis, et al. 2014. A foundation for provitamin a biofortification of maize: Genome-wide association and genomic prediction models of carotenoid levels. *Genetics* 198:1699-1716

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.). *Theoretical and Applied Genetics* 109:323-332

Paul, P.A., Lipps, P.E., Hershman, D.E., McMullen, M.P., Draper, M.A., and Madden, L.V. 2008. Efficacy of triazole-based fungicides for *Fusarium* head blight and deoxynivalenol control in wheat: a multivariate meta-analysis. *Phytopathology* 98:999-1011

Petersen, S., J.H. Lyerly, P.V. Maloney, R.A. Navarro, C. Cowger, G. Brown-Guedira, et al. 2014. *Fusarium* head blight resistance QTL in the NC-Neuse/AGS2000 recombinant inbred line population. Proceedings of the 2014 National *Fusarium* Head Blight Forum, St Louis, Missouri.

Parry, D.W., P. Jenkinson, and L. McLeod. 1995. *Fusarium* ear blight (scab) in small grain cereals: A review. *Plant Pathology* 44:207-238

Patel, D., M. Zander, J. Dalton-Morgan, and J. Batley. 2015. Advances in plant genotyping: where the future will take us. p. 1-11. In: J. Batley (ed). Plant Genotyping. Springer New York

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. *Molecular Breeding* 28:281-301

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. *The Plant Genome* 5:103-113

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

Ribichich, K.F., S.E. Lopez, and A.C. Vegetti. 2000. Histopathological spikelet changes produced by *Fusarium graminearum* in susceptible and resistant wheat cultivars. *Plant Disease* 84:794-802

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, et al. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics* 44:217-220

Riedelsheimer, C., J.B. Endelman, M. Stange, M.E. Sorrells, J.-L. Jannink, and A.E. Melchinger. 2013. Genomic predictability of interconnected biparental maize populations. *Genetics* 194: 493-503

Rutkoski, J.E., E.L. Heffner, and M.E. Sorrells. 2011. Genomic selection for durable stem rust resistance in wheat. *Euphytica* 179:161-173

Rutkoski, J.E., J. Benson, Y. Jia, G. Brown-Guedira, J.L. Jannink, and M.E. Sorrells. 2012. Evaluation of genomic prediction methods for *Fusarium* head blight resistance in wheat. *The Plant Genome* 5:51-61

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 Genes|Genomes|Genetics* 3:427-39

Rutkoski, J.E., J.A. Poland, R.P. Singh, J. Huerta-Espino, S. Bhavani, H. Barbier, et al. 2014. Genomic selection for quantitative adult plant stem rust resistance in wheat. *The Plant Genome*. 7:1-10

Salameh, A., M. Buerstmayr, B. Steiner, A. Neumayer, M. Lemmers, and H. Buerstmayr. 2011. Effects of introgression of two QTL for Fusarium head blight resistance from Asian spring wheat by marker-assisted backcrossing into European winter wheat on Fusarium head blight resistance, yield and quality traits. *Molecular Breeding* 28:485-494

Salgado, J.D., L.V. Madden, and P.A. Paul. 2014. Efficacy and economics of integrating in-field and harvesting strategies to manage Fusarium head blight of wheat. *Plant Disease* 98:1407-1421

SAS Institute. 2013. *The SAS System for Windows. Version 9.3.* SAS Institute Inc, Cary, NC.

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. 2005. Molecular mapping of Fusarium head blight resistance in the winter wheat population Dream/Lynx. *Theoretical and Applied Genetics* 111:747-756

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

Shapira, R. 2004. Control of mycotoxins in storage and techniques for their decontamination. pp.190-223. In N. Magan, and M. Olsen (ed). *Mycotoxins in food. Detection and control.* Cambridge: Woodhead Publishing Limited

- Shen, X., M. Zhou, W. Lu, and H.M. Ohm. 2003. Detection of *Fusarium* head blight resistance QTL in a wheat population using bulked segregant analysis. *Theoretical and Applied Genetics* 106:1041-1047
- Shen, X., M. Alam, F. Fikse, and L. Rönnegård. 2013. A novel generalized ridge regression method for quantitative genetics. *Genetics* 193:1255-1268
- Somers, D.J., R. Kirkpatrick, M. Moniwa, A. Walsh. 2003. Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. *Genome* 49:431-437
- Snijders, C.H.A. 1990. The inheritance of resistance to head blight caused by *Fusarium culmorum* in winter wheat. *Euphytica* 50:11-18
- Stack, R.W. 2003. History of *Fusarium* head blight with emphasis on North America. pp.1-34. In K.L. Leonard, and W.R. Bushnell (ed). *Fusarium head blight of wheat and barley*. St Paul: The American Phytopathological Society
- Stekhoven, D.J., and P. Bühlmann. 2011. MissForest - nonparametric missing value imputation for mixed-type data. *Bioinformatics* 28:112-118
- Strange, R.N., and H. Smith, 1971. A fungal growth stimulant in anthers which predisposes wheat to attack by *Fusarium graminearum*. *Physiological Plant Pathology* 1:141-150

- Sutton, J.C. 1982. Epidemiology of wheat head blight and maize ear rot caused by *Fusarium graminearum*. Canadian Journal of Plant Pathology 4:195-209
- Suzuki, T., M. Sato, and T. Takeuchi. 2012. Evaluation of the effects of five QTL regions on *Fusarium* head blight resistance and agronomic traits in spring wheat (*Triticum aestivum* L.). Breeding Science 62:11-17
- Szabó-Hevér, A., S. Lehoczki-Krsjak, M. Varga, L. Purnhauser, J. Pauk, C. Lantos, et al. 2014. Differential influence of QTL linked to *Fusarium* head blight, *Fusarium*-damaged kernel, deoxynivalenol contents and associated morphological traits in a Frontana-derived wheat population. Euphytica 200:9-26
- Talukder, S.K., M.A. Babar, K. Vijayalakshmi, J. Poland, P.V.V. Prasad, R. Bowden, et al. 2014. Mapping QTL for the traits associated with heat tolerance in wheat (*Triticum aestivum* L.). BMC Genetics 15:97
- Thompson, C.J. 2010. Evaluation of an inoculation method and quantitative trait loci for *Fusarium* head blight resistance in wheat. Master of Science thesis, University of Illinois at Urbana-Champaign. Available at <http://hdl.handle.net/2142/18231>
- Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B 58:267-288

Tiemann, U., K.P. Brüssow, D. Dannenberg, et al. 2008. The effect of feeding a diet naturally contaminated with deoxynivalenol (DON) and zearalenone (ZON) on the spleen and liver of sow and fetus from day 35 to 70 of gestation. *Toxicology Letters* 179:113-117

Tomaszewski, J., R. Miturski, A. Semczuk, et al. 1998. Tissue zearalenone concentration in normal, hyperplastic and neoplastic human endometrium. *Ginekologia Polska* 69:363-366

Troyanskaya, O., M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, et al. 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics* 17:520-525

Tuite, J. 1969. *Plant pathological methods: Fungi and bacteria*. Minneapolis: Burgess Publishing Company.

USDA. 2015. National Agricultural Statistics Service. Available at <http://www.nass.usda.gov/> (accessed 20 Mar. 2015)

Van Ginkel, M., W. Van Der Schaar, Y. Zhuping, and S. Rajaram. 1996. Inheritance of resistance to scab in two wheat cultivars from Brazil and China. *Plant Disease* 80:863-867

Van Raden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, et al. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* 92: 16-24

Verges, V.L., D. Van Sanford, G. Brown-Guedira. 2006. Heritability estimates and response to selection for Fusarium head blight resistance in soft red winter wheat. *Crop Science* 46:1587-1594

Von der Ohe, C., E. Ebmeyer, V. Korzun, and T. Miedaner. 2010. Agronomic and quality performance of winter wheat backcross populations carrying non-adapted Fusarium head blight resistance QTL. *Crop Science* 50:2283-2290

Waldron, B.L., B. Moreno-Sevilla, J.A. Anderson, et al. 1999. RFLP mapping of QTL for Fusarium head blight resistance in wheat. *Crop Science* 39:805-811

Wang, Y., M.F. Mette, T. Miedaner, M. Gottwald, P. Wilde, J.C. Reif, et al. 2014. The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC Genomics* 15:556

Wegulo, S.N., W.W. Bockus, J.H. Nopsa, E.D.D. Wolf, K.M. Eskridge, K.H.S. Peiris. 2011. Effects of integrating cultivar resistance and fungicide application on Fusarium head blight and deoxynivalenol in winter wheat. *Plant Disease* 95:554-560

Wetterstrand, K.A. 2015. DNA sequencing costs: Data from the NHGRI large-scale genome sequencing program. National Human Genome Research Institute, Bethesda, MD.
<http://www.genome.gov/sequencingcosts> (accessed 22 Mar. 2015)

Windels, C. 2000. Economic and social impacts of Fusarium head blight: changing farms and rural communities in the Northern Great Plains. *Phytopathology* 90:17-21

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

Würschum, T., J.C. Reif, T. Kraft, G. Janssen, and Y. Zhao. 2013. Genomic selection in sugar beet breeding populations. *BMC Genetics* 14:85

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. *Nature Genetics* 38:203-208

Yu, J.B., G.H. Bai, W.C. Zhou, et al. 2008. Quantitative trait loci for Fusarium head blight resistance in a recombinant inbred population of Wangshuibai/Wheaton. *Phytopathology* 98:87-94

Zhang, Z., U. Ober, M. Erbe, H. Zhang, N. Gao, J. He, et al. 2014. Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS ONE* 9:e93017

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

Zhao, Y., M. Gowda, F.H. Longin, T. Würschum, N. Ranc, and J.C. Reif. 2012. Impact of selective genotyping in the training population on accuracy and bias of genomic selection.

Theoretical and Applied Genetics 125:707-713

Zhao, Y., M.F. Mette, M. Gowda, C.F.H. Longin, and J.C. Reif. 2014. Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat.

Heredity 112: 638-645

Zhong, S., J.C.M. Dekkers, R.L. Fernando, and J.L. Jannink. 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study.

Genetics 182:355-364

Zhou, W.C., F.L. Kolb, J.B. Yu, G. Bai, L.K. Boze, and L.L. Domier. 2004. Molecular characterization of Fusarium head blight resistance in Wangshuibai with simple sequence repeat and amplified fragment length polymorphism markers. Genome 47:1137-1143

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B 67:301-320