Genomic approaches for mapping and predicting disease resistance in wheat (*Triticum aestivum L.*)

by

Cristiano Lemes Da Silva

B.S., Federal University of Technology of Parana, 2010
M.S., Federal University of Technology of Parana, 2013

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Abstract

Wheat diseases cause significant economic losses every year. To ensure global food security, newly released cultivars must possess increased levels of broadly-effective resistance against wheat pathogens, acceptable end-use quality, and high yield potential. Genetic host resistance stands out from other management strategies as the most viable option for controlling diseases. New genotyping platforms allow whole genome marker discovery at a relatively low cost, favoring the identification of novel loci underlying traits of interest. The work presented here describes genomic approaches for mapping and predicting the resistance to Fusarium head blight (FHB) and wheat rusts.

The first study used biparental mapping to identify quantitative trait loci (QTL) associated with Fusarium head blight (FHB) resistance. A doubled haploid population (DH) was originated from a cross of Everest and WB-Cedar, which are widely grown wheat cultivars in Kansas with moderately resistant and moderately susceptible reactions to FHB, respectively. We confirmed that neither of the parents carry known large-effect QTLs, suggesting that FHB resistance is native. Eight small-effect QTLs were identified as associated with multiple mechanisms of FHB resistance. All QTLs had additive effects, providing significant improvements in levels of resistance when they were found in combinations within DH lines.

In the second study, a genome-wide association mapping (GWAS) and genomic selection (GS) models were applied for FHB resistance in a panel of 962 elite lines from the K-State Wheat Breeding Program. Significant single nucleotide polymorphisms (SNPs) associated with the percentage of symptomatic spikelets were identified but not reproducible across breeding panels tested in each year. The Accuracy of predictions ranged from 0.25 to 0.51 depending on GS model, indicating that it can be a useful tool to increase levels of FHB resistance.

GWAS and GS approaches were also applied to a historical dataset to identify loci underlying resistance to leaf and stem rust at seedling stage in a panel of elite winter wheat lines. Infection types of multiple races of wheat rusts from the last sixteen years of the Southern Regional Performance Nursery (SRPN) were used in this study. A total of 533 elite lines originating from several breeding programs were tested in the SRPN during this period of time. GWAS identified significant SNP-trait associations for wheat rusts, confirming the effectiveness of already known genes and revealing potentially novel loci associated with resistance.

Genomic approaches for mapping and predicting disease resistance in wheat (*Triticum aestivum L.*)

by

Cristiano Lemes Da Silva

B.S., Federal University of Technology of Parana, 2010
M.S., Federal University of Technology of Parana, 2013

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Approved by:
Major Professor
Allan K. Fritz

Copyright

Abstract

Wheat diseases cause significant economic losses every year. To ensure global food security, newly released cultivars must possess increased levels of broadly-effective resistance against wheat pathogens, acceptable end-use quality, and high yield potential. Genetic host resistance stands out from other management strategies as the most viable option for controlling diseases. New genotyping platforms allow whole genome marker discovery at a relatively low cost, favoring the identification of novel loci underlying traits of interest. The work presented here describes genomic approaches for mapping and predicting the resistance to Fusarium head blight (FHB) and wheat rusts.

The first study used biparental mapping to identify quantitative trait loci (QTL) associated with Fusarium head blight (FHB) resistance. A doubled haploid population (DH) was originated from a cross of Everest and WB-Cedar, which are widely grown wheat cultivars in Kansas with moderately resistant and moderately susceptible reactions to FHB, respectively. We confirmed that neither of the parents carry known large-effect QTLs, suggesting that FHB resistance is native. Eight small-effect QTLs were identified as associated with multiple mechanisms of FHB resistance. All QTLs had additive effects, providing significant improvements in levels of resistance when they were found in combinations within DH lines.

In the second study, a genome-wide association mapping (GWAS) and genomic selection (GS) models were applied for FHB resistance in a panel of 962 elite lines from the K-State Wheat Breeding Program. Significant single nucleotide polymorphisms (SNPs) associated with the percentage of symptomatic spikelets were identified but not reproducible across breeding panels tested in each year. Accuracy of predictions ranged from 0.25 to 0.51 depending on GS model, indicating that it can be a useful tool to increase levels of FHB resistance.

GWAS and GS approaches were also applied to a historical dataset to identify loci underlying resistance to leaf and stem rust at seedling stage in a panel of elite winter wheat lines. Infection types of multiple races of wheat rusts from the last sixteen years of the Southern Regional Performance Nursery (SRPN) were used in this study. A total of 533 elite lines originating from several breeding programs were tested in the SRPN during this period of time. GWAS identified significant SNP-trait associations for wheat rusts, confirming the effectiveness of already known genes and revealing potentially novel loci associated with resistance.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First of all, I want to acknowledge the valuable guidance of my advisor Dr. Allan Fritz. I am glad to have had many opportunities of walking through segregating populations with him while discussing different aspects of wheat breeding. His advice made a big difference in my career and I will do my best to implement all the breeding lessons that I learned from him.

I also would like to thank all my committee members and co-advisors Drs. Jesse Poland, Robert Bowden, and John Fellers for the collaboration and dedication of time and effort to review this dissertation. I consider myself fortunate to be advised by such great mentors.

Thanks to my colleagues of the Kansas State Wheat Breeding Project: Anju Giri, Seth Filbert, Andy Auld, Shaun Winnie, and all undergraduate students for helping with planting and harvesting of experiments. Thanks to Angela Matthews and Shuangye Yu for assisting me with DNA extractions, preparation of GBS libraries, and overall procedures in the lab. Special thanks to Marshall Clinesmith for his friendship, suggestions, and productive discussions during long field trips and after hours in the lab. None of this would be possible without everyone's help.

I am also grateful to the former Ph.D. colleagues Sarah Battenfield and Robert Gaynor for introducing me to genomic data analysis in SAS and R environment. From the simplest analysis to complex genomic selection models, they were always there willing to share their knowledge with me in the beginning of my Ph.D. project.

To Dr. Floyd Dowell and Dr. Kamaranga Peiris from the USDA-ARS-CGAHR, thanks for providing me lab equipment and valuable suggestions regarding data analysis of FHB traits.

Finally, I must thank my wife Thais and my daughter Alice for their unconditional love and support during this journey. Thanks for always being by my side and understanding my absence while I was writing this dissertation.

Dedication

This dissertation is dedicated to all my family and friends who always offered me continuous support and encouragement throughout my long journey of education.

Chapter 1 - Literature Review

The genetic gain of cultivated crops has been around 1% per year during the last century which is on pace to feed the world population in the coming decades. Along with that, the rise of global temperatures will not only affect the adaptability of crops but can also favor the occurrence of certain crop diseases. Consequently, this represents the key challenge that a new generation of researchers will face to ensure global food security. Among all cultivated species, wheat stands out as one of the most widely grown crop in the world and it is considered a staple food in developing countries providing nearly 20% of the daily protein and food calories for almost half of the human population (FAO, 2017). The global wheat production and annual stocks increased respectively from to 233.4 and 82.8 million metric tons in 1960 to 753.8 and 267.5 million metric tons in 2017. It represents 3.2-fold increase whereas the global acreage remained relatively at the same level (USDA, 2017).

Wheat diseases cause substantial losses in grain yield and end-use quality. Fortunately, new advances in sequencing and genomic analysis has provided new tools to understand and assist disease resistance breeding. Sequencing-based genotyping platforms are generating abundant single nucleotide polymorphisms (SNPs) at relatively low cost while allowing simultaneous genotyping and discovery (Poland & Rife, 2012; Thomson, 2014). As a result, genotyping has become relatively more cost-effective than field phenotyping.

SNP-based genotyping has been extensively used for QTL mapping, genome-wide association studies, and genomic predictions and are being integrated into breeding programs and genetic studies to dissect the mechanisms underlying wheat agronomic traits in wheat such as: adaptability traits, yield components, abiotic stresses, and disease resistance. Moreover, breeding

for disease resistance indirectly impacts grain yield and leads to more stable wheat production over growing seasons.

1.0 Breeding for Disease Resistance in Wheat

Breeding exclusively by phenotypic selection limits the improvement of disease resistance because it relies on the evaluator's ability to rate wheat lines. Experimental errors such as uneven infection, hotspots, race variations, and genotype-by environment (G×E) interactions often lead to biased conclusions, while the genes controlling resistance remain unknown. Therefore, the use of genomic tools such as marker-assisted selection, biparental mapping, association studies, and genomic selection come into to play aiming to improve the genetic gain of agronomic traits by increasing the frequency of desirable alleles within breeding populations.

Disease resistance in plants has been long categorized into two distinct classes with different and confusing terms in the literature such as horizontal or vertical, major or minor genes, complete or partial resistance, etc. These categories refer to resistance conferred by one or a few genes and the resistance provided by multiple genes, respectively. Trying to overcome all these inadequate definitions Poland *et al*. (2009) advocated the use of the terms qualitative resistance when R-genes are controlling resistance following the gene-for-gene theory (Flor, 1942) and quantitative resistance when multiple quantitative resistant loci (QRL) and/or small-effect genes are governing resistance.

Each wheat disease has its own characteristics that affect the way we breed for resistance. For instance, so far significant race variation has not been identified amongst Fusarium head blight (FHB) isolates. As a result, durable resistance can be simply achieved with a single gene. In contrast, wheat rusts are characterized by the presence of multiple races that can rapidly

evolve and overcome deployed genetic resistance. Therefore, achieving durable rust resistance is substantiality more complex since the pathogen is constantly evolving and often overcoming the resistant genes present in the host.

In this review chapter, the importance and specific concerns of four wheat diseases are briefly discussed: Fusarium head blight, leaf rust, stem rust, and stripe rust. The following chapters contain separate studies where different strategies for increasing the host resistance of wheat to these diseases are discussed in detail.

1.1 Fusarium Head Blight

Fusarium head blight (FHB) is one of the most important wheat diseases in warm and humid environments. Weather conditions of high relative humidity and temperatures between 12 - 22 ^{o}C are ideal for the development of *Fusarium graminearum sensu lato*. This pathogen is classified as hemibiotrophic, since it behaves as a biotroph in the early stages of infection, then acts as a necrotroph while colonizing the spikes of the host around 72 hours after infection, then causing FHB (Trail, 2009). This fungus has asexual conidia as the main type of spores (Wegulo *et al*., 2015) and survives saprophytically overwinter in crop residue of wheat, barley, and corn in the form of chlamydospores or hyphae structures (Parry *et al*., 1990). In the spring, spores (mainly ascospores) are windblown or water-splashed onto wheat spikes (McMullen *et al*., 2012) causing the initial infections. FHB occurs primarily during anthesis and due to this short period ideal for the pathogen development, the disease is considered monocyclic. Symptoms appear in the spikes right after infection as small brown, water-soaked spots, that leads to a white-bleached color as the pathogen spreads throughout the spikes.

Another concern regarding FHB is the accumulation of mycotoxins in *Fusarium*-infected grains. The trichothecene deoxynivalenol (DON) is the most prevalent toxin produced by this pathogen, acting as a virulence factor facilitating disease spread within wheat spikes (Gunupuru *et al*., 2017; Shah *et al*., 2017). Ingestion of wheat grains or wheat-based products contaminated with DON may cause diarrhea, vomiting, (Moazami & Jinap, 2009), and even carcinogenic effects (Shephard, 2011). For these reasons, most countries have established limits of DON in food products derived from wheat. For instance, currently in the United States, the recommended limit of DON is 1 ppm in wheat products for human consumption (FDA, 2014).

Integrated management strategies are required to control of FHB and reduction of DON levels on wheat-derived products. Agronomic practices such as tillage, crop rotation, avoiding irrigation during anthesis, and preventive fungicide applications before anthesis can greatly contribute in controlling FHB epidemics (McMullen *et al*., 2012; Dweba *et al*., 2017). Nonetheless, genetic resistance seems to be one of the most effective forms to control this disease, since this resource is inexpensive and effortless for the wheat growers, and it can still be combined with other management practices.

FHB Resistance

According to Mesterhazy (1995), the genetic resistance to FHB can be classified into five types: initial infection/incidence (type I), spread within spike/severity (type II), DON accumulation in grain (type III), Fusarium kernel damage (type IV), and tolerance (type V). However, there are contradictions in the literature for what types III and IV are supposed to be. As a result, Sneller *et al*. (2012), decided to rename it as resistance to toxin accumulation (RTA) and resistance to kernel infection (RKI), respectively. FHB resistance can also be categorized based on its origin: native (present in adapted elite materials, i.e. Everest, Overland, Truman,

etc.,), exotic (varieties from other countries, i.e. Sumai #3 and Frontana), and alien (translocations introgressed from wild wheat relatives, i.e. *E. tsukushiensis, T. ponticum*).

FHB is a quantitative trait with relative complex inheritance (Buerstmayr *et al*., 2009) and gene effects are predominantly additive, which contributes to an effective gene pyramiding schemes. Due to the laborious phenotyping of FHB resistance, finding reliable and tightly linked markers to resistant loci has been the fundamental goal of the majority of studies. Mapping quantitative trait loci (QTL) in biparental populations has been the most popular method for identifying new genomic regions associated with FHB resistance. So far, hundreds of QTLs associated with FHB resistance have been reported in the literature (Liu *et al*., 2009) covering all wheat chromosomes with effects varying depending on the genetic background. Among all these mapped QTLs, seven with a larger effect on phenotype and validated across multiple studies have been formally named as: *Fhb1, Fhb2, Fhb3, Fhb4, Fhb5, Fhb6,* and *Fhb7*. Each one of these large-effect QTLs is briefly detailed in the following section.

Large-Effect QTLs and MAS for FHB Resistance

Fhb1 is the most well-known QTL and confers moderate levels of FHB resistance (especially type II) in different genetic backgrounds. It is located on the short arm of the chromosome 3B and can easily be detected by the tightly linked markers *unm10* (Liu *et al*., 2008) and/or *Xsnp3BS-8* (Bernardo *et al*., 2012). This QTL was first identified and mapped in the Chinese cultivar Sumai #3 and later also found in other Chinese materials. Currently, *Fhb1* has been transferred into several breeding programs but fewer commercial cultivars were released due to its association with yield penalties, especially in winter wheat backgrounds. So far, *Fhb1* is the only QTL associated with FHB resistance that has been cloned. A single gene called *pore-forming toxin-like* (*PFT*) was identified as the putative candidate for the QTL *Fhb1*.

This gene was expressed constitutively in resistant near-isogenic lines with the highest expression in pre-emerging spikes (Rawat *et al*., 2016). However, *PFT* did not significantly decrease DON levels, suggesting that an independent locus in the same genomic region may be controlling DON detoxification.

Another well-known QTL originated from Sumai #3 is *Fhb2*, which is associated with limited disease severity (type II) and low DON accumulation (type IV/RTA). This QTL is located on the short arm of 6B and is flanked by the microsatellites *gwm133* and *gwm644* in a 2 cM interval near to the centromere (Cuthbert *et al*., 2007). Several studies have identified a QTL in the same genomic region, suggesting that *Fhb2* is a true QTL and the most likely candidate located in this region of 6BS. Alone this QTL is capable of explaining up to 56% of the severity variation in the field, depending on the genetic background (Yang *et al*., 2003; Cuthbert *et al*., 2007). Recently, six putative genes were identified in the *Fhb2* interval revealing the underlying mechanisms of resistance, using integrated metabolon-transcriptomics (Dhokane *et al*., 2016). These genes were involved in cell wall reinforcement (decreasing the spread of pathogen within spike), and DON detoxification.

Fhb3 is an alien introgression from *L. racemosus* that was transferred onto the short arm of the chromosome 7A of wheat. The levels of FHB resistance provided by this translocation can be similar to Sumai #3 (Qi *et al*., 2008). Translocations, such as *Fhb3*, have the advantage of triggering a large effect on phenotype with a simple inheritance that facilitates its deployment in wheat breeding programs. Although combining multiple alien segments tend to cause deleterious effects on end-use quality. Using a maker assisted selection backcrossing approach, Brar *et al*. (2015) reported that pyramiding *Fhb1+Fhb2+Fhb3* reduced FHB severity by nearly 50% proving the additive nature of these QTL when they are combined together.

Fhb4 is another large-effect QTL that was first mapped in the Chinese landrace Wangshuibai (Lin *et al*., 2006) that confers type I resistance (barrier against initial infection). This QTL is located on the long arm of chromosome 4B in a 1.7 cM interval between the markers *Xhbg226* and *Xgwm149* (Xue *et al*., 2010). Using a biparental population from a cross between the winter cultivars Everest × Art, Clinesmith (2015) found a QTL from Art in the same region of *Fhb4* explaining 8.31- 17.80% of the percentage of symptomatic spikelets variation. Yet, in the same study, the microsatellite *Xgwm149* was located within the QTL interval suggesting that it could be in fact *Fhb4*. Several other studies have been repeatedly mapping QTL on *Fhb4* interval, indicating that this QTL is present in a relatively high frequency in wheat cultivars such as Ernie, Chockwang, Wuhan1, and Haiyanzhong (Cai *et al*., 2016).

Fhb5 is large-effect QTL found in Wangshuibai associated with type I resistance. It is located on the centromeric region of 5AS in a 0.3 cM interval flanked by the markers *Xgwm304* and *Xgwm415* (Xue *et al*., 2011). The same authors also verified that the genetic variation conferred by this QTL was significantly larger than the variation caused by G×E interactions. Steiner *et al*., (2004) also mapped a QTL from Frontana associated with FHB severity that may likely be *Fhb5*. Since pericentromeric regions usually have lower recombination, cloning this region near to the centromere might be challenging (Xue *et al*., 2011) as well as transferring it to a small genetic block from the donors to elite breeding materials.

Recently, *Fhb6* was transferred from *E. tsukushiensis* to the proximal part of chromosome 1AS via *ph1b*-induced homoeologous recombination and it can be followed with the Kompetitive Allele Specific PCR (KASP) marker *wg1S_snp* (Cainong *et al*., 2015). Yet in the same study, it was observed that plants homozygous for *Fhb6* presented on average 7 % of FHB severity in the field while null progenies averaged 35 % and the resistant check Everest was

rated at 27%. This segment confers resistance to type I and type II and it was released as a novel source of germplasm named KS14WGRC61 (Cainong *et al*., 2015). Currently, several combinations of *Fhb6* with other QTLs are being made aiming to improve FHB resistance.

Just as *Fhb3* and *Fhb6*, *Fhb7* it is also an alien introgression. *Fhb7* was transferred from *T. ponticum* to the chromosome 7D of wheat and it mapped closely linked to marker *Xcfa2240*, conferring resistant type II (Guo *et al*., 2015). Besides mapping and shortening the interval carrying *Fhb7*, the same study also investigated the effect of pyramiding *Fhb1+Fhb7*. However, none of the lines carrying both QTLs were significantly more resistant than the donor parent Ning 7840 (*Fhb1*), or the newly developed introgressions with *Fhb7*. Nevertheless, other QTL combinations should be made and tested targeting to improve FHB resistance.

So far, we have seen that significant progress of FHB resistance can be made through marker-assisted selection (MAS) of few QTLs as long as the effects are large, stable, with tightly linked markers. For instance, Eckard *et al*. (2015) were able to identify 15 QTLs of native FHB resistance while combining it with *Fhb1* into winter wheat cultivars, through identity-by-descent based linkage mapping in early generations. In the meantime, at the International Maize and Wheat Improvement Center (CIMMYT), resistant alleles from Chinese landraces have been introgressed since 1980 (Steiner *et al*., 2017) and currently, the QTLs *2DLc*, *Fhb4*, and *Fhb5* are the most frequently resistant loci found on their elite germplasm (Lu *et al*. 2013).

Minor-Effect QTLs Associated with FHB Resistance

Although marker-assisted selection (MAS) has been successfully applied for a few large-effect QTLs (Lu *et al*. 2013; Brar *et al*. 2015), more modest progress has been made in breeding programs where the FHB resistance is predominantly controlled by small-effect QTLs. In these

cases, resistance is being controlled primarily by small-effect genes, being relatively complex to breed for due to its quantitative nature and low heritability.

Increasing FHB resistance with minor-effect QTLs/genes has the advantage of avoiding introgression of large genetic blocks and/or alien segments from unadapted materials which often have yield penalties and not relying on biased marker associations. Besides, major well-known QTLs could still be easily combined with small-effect QTLs via MAS. Furthermore, resistance is rarely deployed by itself, but rather in combination with several other traits, such as yield potential and end-use quality, and finding the perfect balanced combination is the endless challenge of breeding.

Genome-wide association studies (GWAS) have been broadly used to identify small-effect loci underlying traits of interest in panels of non-pedigree related individuals. For instance, in a recent GWAS study, Wang *et al*. (2017) found six highly significant QTLs that were associated with FHB resistance in the Pacific Northwest Region of the United States and CIMMYT breeding program. The same authors also verified a QTL located on 5B was associated with low DON accumulation that could potentially be a novel gene. Another ten significant SNP-trait associations with FHB resistance were reported in an association study by Arruda *et al*. (2016a). Several minor-effect QTLs controlling FHB resistance were reported by Cai *et al*. (2016) while mapping and transferring it to American wheat background. Similar results have been also reported in Europe (Kollers *et al*., 2013).

In the second chapter of this dissertation, it was mapped several components of FHB resistance using a biparental doubled haploid population originating from a cross between Everest and WB-Cedar. These cultivars have been widely grown in Kansas in the last growing seasons and have a moderate level of native FHB resistance and are believed to carry native FHB

resistance. Likewise, the third chapter consists of an association study and genomic predictions of FHB resistance evaluated in a panel of winter wheat lines.

1.2 Wheat Rusts

The pathogens *Puccinia triticina* (*Pt*), *Puccinia graminis f. sp. tritici* (*Pgt*), and *Puccinia striiformis f. sp. tritici* (*Pst*) are known for causing leaf rust (LR), stem rust (SR), and stripe rust (YR) in wheat, respectively. These biotrophic fungi represent a major threat to wheat production worldwide. *Pt* is the least aggressive of all three species causing minor yield losses (Roelfs *et al*., 1992), while *Pgt* occurs predominantly in warm climate conditions with highly aggressive races, such as Ug99 (TTKSK), capable of destroying entire crops (Juliana *et al*., 2017). In contrast, *Pgt* epidemics happens mainly in temperate regions, such as the Great Plains of United States, causing smaller economic losses.

Cereal rusts produce five type of spores in a complete life cycle: spermatia, aeciospore, basidiospore, urediniospore, and teliospore. The first two spore types only occur during the sexual phase in alternate host species whereas urediniospores and teliospores are asexual structures formed by the pathogen while infecting wheat (Schumann & Leonard, 2000). Wheat rusts are polycyclic diseases since urediniospores are able to cause secondary infections. These spores are produced within uredinias and by the end of the wheat cycle, the uredinias also start to produce teliospores which are dikaryotic thick-walled spores capable of surviving overwinter in wheat residues (Leonard & Szabo, 2005). The symptoms of wheat rusts are characterized by round-shaped, orange, brown or black pustules caused by the uredinia structure that infects adaxial surface of leaves, stems, and even spikes in case of SR. Lesions caused by YR are

arranged in linear stripe patterns along leaves whereas LR pustules are relatively bigger and occur at random on wheat leaves.

Wheat rusts have relatively complex life cycles including asexual and sexual reproduction, multiple spore types, race variations, and off-season alternate hosts. For instance, barberry (*Barberis vulgaris*) was identified in mid-1910's as the alternative host for the sexual reproduction of *Pgt* in the United States. Since then collaborative efforts almost eradicated this secondary host species that consequently prevented the occurrence of the sexual phase of *Pgt* and *Pst* in North America (Roelfs & Bushnell, 1985). Although epidemics temporarily decreased, every year early infections in wheat fields of southern U.S. and northern Mexico produces urediniospores that are windblown to the north, going all the way to Canada. This process of spore migration was named as the '*Puccinia Pathway*' (Kolmer, 2001).

New races of wheat rusts are constantly emerging and evolving (Singh *et al*., 2015). Although, fungicides have been proven to be an effective form to control these races, it is not a viable option economically in developing countries. As a result, genetic resistance has been the preferred management strategy to prevent widespread epidemics. Therefore, it is required constant deployment of cultivars possessing multiple combinations of resistance genes to keep resistance moving ahead of new virulent races.

Genetics of Wheat Rust Resistance

Confusing terminologies are reported in the literature to classify genetic mechanisms of wheat rust resistance, which not always are synonyms. Regardless of having fewer genes controlling rust resistance than FHB resistance, achieving durable resistance is challenging due to the evolutionary nature of *Puccinia* pathogen species. Besides, the levels of resistance may vary depending plant developmental stages: seedling or adult plant resistance (APR). In addition,

gene interactions may occur; for instance, the APR gene *Lr34* is known to interact with other genes such as *Lr13,* resulting in lower infection types (Kolmer, 1996). Pleiotropic gene effects (or tight linkage) are also present in the wheat rust pathosystem, meaning that a single gene is able to control more than one phenotypic trait. For example, *Sr2* and *Lr34* not only confer APR resistance but also cause pseudo-black chaff (PBC) and leaf tip necrosis (LTN) onto wheat spikes and flag leaves, respectively. PBC and LTN are secondary traits controlled by the same genes that are associated with yield penalties and undesirable appearance (Juliana *et al*., 2015).

To date, 77 *Lr* genes, 78 *Yr* genes, and more than 50 *Sr* genes have been characterized (McIntosh *et al*., 2017), mapped, and classified into two main categories: resistance at seedling stage or adult plant resistance (discussed in details in following topics). These genes come from different sources including exotic accessions, landraces, and wild wheat relatives. However, reliable diagnostic markers are not available for all these genes and only a few have been cloned: *Lr10* (Feuillet *et al*., 1997), *Lr21* (Huang *et al*., 2003), *Lr34* (Krattinger *et al*., 2009), *Yr36* (Fu *et al*., 2009), *Sr33* (Periyannan *et al*., 2013), *Sr35* (Saintenac *et al*., 2013), *Lr67* (Moore *et al*., 2015), *Sr22* and *Sr45* (Steuernagel *et al*., 2016), and *Lr22a* (Thind *et al*., 2017).

Durable resistance is a recurring concept in the wheat rust pathosystem. It refers to resistance that remains effective while a cultivar possessing it is widely cultivated over time (Johnson, 1983). Several strategies have been proposed to extend the durability of resistance. The most relevant are: regional deployment throughout the pathogen path (i.e. The Puccinia Pathway), multi-lines, mixtures, gene rotation, releasing one gene at a time, and perhaps the most effective of all - gene pyramids (Mundt, 2014). Pyramiding refers to combining multiple genes within breeding lines to increase levels resistance, providing that the genes have additive effects.

From a breeding perspective, backcrossing followed by MAS and phenotypic selections have been a successful strategy for pyramiding genes associated with wheat rust resistance at CIMMYT (Singh *et al*., 2005). This strategy may be effective even when genes conferring resistance are unknown, as long as precise phenotyping is achieved in early generations. However, phenotypic selection can be confounded by the race structure presented in the target environment when selections were made. This confounding effect also might lead to mistaken genomic predictions if phenotypic data are from the training set that was only phenotyped for wheat rust resistance in a single or related environment.

Resistance at Seedling Stage

The seedling resistance of wheat rusts is predominantly monogenic, race-specific, and more effective on early developmental stages (Ellis *et al*., 2014), although some resistant genes (R-genes) may confer 'all-stage resistance' (Chen, 2005; Riaz *et al*., 2016). These R-genes segregate in a Mendelian fashion following the gene-for-gene theory, where for each R-gene in the host, there is a corresponding avirulence effector gene in the pathogen (Flor, 1942). Often these loci are also referred as 'major effect' genes. In general, the R-genes encode for immune receptors of the nucleotide-biding leucine-rich repeat class (NB-LRR), causing a hypersensitive reaction in the host (Van der Biezen & Jones, 1998; Marone *et al*., 2013).

Monogenic resistance deployed by itself creates a selection pressure in the fitness of certain races and favorable mutations and recombination events in the pathogens are naturally selected and resistance is defeated over time. For example, the race TTKSK (*Ug99*) became virulent to *Sr31* and after that, new variations of this race defeated several other resistant genes. In this case, genes were defeated mainly because they were deployed as a single race-specific gene within cultivars, rather than in pyramids. Currently, the most useful race-specific genes

against stem rusts are: *Sr22*, *Sr25*, *Sr26*, *Sr33*, *Sr35*, *Sr45*, and *Sr50* (Singh *et al*., 2015) and breeders are thoroughly advised to use them in combinations to avoid new gene breakdowns.

Seedling resistance is commonly assessed at early stages by inoculating pure single races or mixtures in a greenhouse, then rating infection types (IT) using the Stakman scale. This scale is semi-quantitative with a few possible phenotypic classes ranging from 0 to 4 which may be combined with the symbols ';', '+ or –', 'X', 'C', 'N' representing respectively: hypersensitive reaction, upper and lower limits of a given IT score, heterogeneous reaction, exceptionally pronounced chlorosis, and necrosis (Stakman, 1962). To simplify data analysis, Zhang *et al*. (2014) proposed a method to convert Stakman IT scores scale into a linear scale (0-9). Recently, a pipeline in Perl was created to automate this data conversion (Gao *et al*., 2016) facilitating data analysis of seedling resistance, especially for genome association studies.

Adult Plant Resistance

Adult plant resistance (APR) refers to genes that are effective in the host only at the adult plant stage. These APR genes are also referred as 'minor effect', partial resistance, and slow-rusting genes since it allows limited disease progress without causing significant damage in the host (Singh *et al*., 2005). Most of APR genes confer race-nonspecific resistance, meaning that they offer an acceptable level of resistance to a wide range of known races, rather than strong resistance against one or only a few specific races. Some of the well-known APR genes are co-localized, tightly linked, or pleiotropic to other genes that appear to convey race-nonspecific resistance, for example *Sr2/Yr30*, *Lr34/Yr18/Sr57*, *Lr46/Yr29/Sr58*, *Lr67/Yr46/Sr55* (Singh *et al*., 2016; Riaz *et al*., 2016). Furthermore, recent studies have reported several QTLs conferring minor APR effects that could potentially be novel APR genes (Yu *et al*., 2014; Gao *et al*., 2016).

It has been hypothesized that APR genes encode a more heterogeneous range of proteins that R-genes, which could help to explain its durability (Ellis *et al*., 2014). *Sr2* and *Lr34* are the most well-known APR genes and have been providing partial, but durable resistance in wheat varieties for more than 50 years (Ellis *et al*., 2014). Besides, combinations of race-specific and race-nonspecific such as *Lr13+Lr34* and *Lr16+Lr34* result in lower infection types (Kolmer, 1996, Kassa *et al*., 2017). Although gene combinations tend to make resistance more durable, underlying mechanisms of gene pyramids are somewhat unclear. However, it is speculated that the degree of host genotype×pathogen specificity may be lower for minor gene than major gene resistance (Mundt, 2014), meaning that APR genes are less likely to naturally selected more virulent races even when a cultivar containing it is widely grown. Besides, the probability of an asexual pathogen mutating to virulence against all resistance genes in a pyramid would be the product of all probabilities of each gene alone, thus making the probability of a new virulent race arising highly unlikely (Mundt, 2014).

In the fourth chapter, sixteen years of historical data from the Southern Regional Performance Nursery (SRPN) were compiled together to identify loci underlying wheat rust resistance across several races in elite breeding lines from several breeding programs in the US.

1.3 Genomic Tools for Improving Disease Resistance

During the last decade, advances in the next-generation of sequencing technologies enabled the use of sequencing-based genotyping platforms on a large scale at relatively low cost. For instance, the genotyping-by-sequencing (GBS) method developed by Poland *et al*. (2012) has been widely adopted in wheat genetic research. This approach uses two restriction enzymes

for targeted complexity reduction followed by multiplex sequencing which produces abundant SNP polymorphisms for discovery and genotype DNA samples, simultaneously.

GBS is a flexible genotyping method that allows to combine and re-call SNPs with new datasets compiled in databases (Poland & Rife, 2012). Yet, GBS reduces ascertainment bias when compared with array-based platforms such as SNPchip arrays (Thomson, 2014), being considered by many authors as an ultimate tool to accelerate breeding (He *et al*., 2014). As a result, this method of genotyping has been broadly applied in wheat genetic research, especially for biparental mapping, association studies, and genomic-based predictions.

Biparental Mapping

Biparental populations have been extensively used for mapping QTLs (quantitative trait loci) associated with disease resistance in wheat. This strategy provides information on the chromosome location and effect of a given loci in the genetic variation of a particular trait. There are three main statistical methods to perform QTL mapping analysis: standard interval mapping (SIM), composite interval mapping (CIM), and multiple QTL mapping (MQM).

SIM considers a single-QTL model at a time while accounting for missing data. However, it has limited ability to separate linked QTL and estimate possible interactions. To overcome some of SIM limitations, CIM enables to detect multiple loci of more modest effect and its interactions by using markers near to a putative QTL as a covariate. These covariate markers remove the effect of the major QTL allowing to identify others with smaller effects. Meanwhile, the MQM method is capable of fitting several QTLs into a regression model while accounting for QTL×QTL interactions, adjusting positions, estimating intervals and effects on phenotype, as well as, determining which parent is the donor of the desirable alleles at the QTL interval (Broman, 2009).

Nevertheless, the genetic variation exploited in these biparental mapping populations is often limited, because only alleles segregating between parents are captured in the analysis (Arruda *et al*., 2016a). As a result, mapping small-effect QTLs is challenging because these QTL often do not generate significant statistical differences. In addition, QTL locations and their effects on phenotype are specific to the particular population where they were mapped and often cannot be extrapolated to other non-related populations. Besides, QTL are mapped within relatively large confidence intervals that generally contain several candidate genes underlying a trait of interest (Kearsey, 2002; Broman, 2009). Furthermore, the phenotypic effect of genes/QTLs in common among parents cannot be estimated in the population, since it is not segregating.

Association Mapping

Genome-wide associations study (GWAS) appeared as an alternative to surpass the limitations of biparental mapping presented above. GWAS was initially developed for mapping studies with human data in early 2000's and was rapidly applied in plant genetics studies, as well (Thomas *et al*., 2005; Ikegawa, 2012). GWAS has the advantages of exploiting natural recombination events from diverse panels, without the upfront cost of funds, time, and effort associated with population development (Korte & Farlow, 2013).

Nonetheless, unbalanced populations and relatedness among individuals can lead to false marker-trait associations. Hence, it is reasonable to use statistical models that fit the population structure and the kinship matrix of genetic effects as covariates to reduce the false discovery rate of genetic markers (Tang *et al*., 2016). Several statistical methods have been developed and tested to minimize these confounding effects while optimizing computing speed. Among these methods, the compressed mixed linear model (CMLM) and its enriched version (ECMLM)

stands out with one of the highest statistical power (Lipka *et al*., 2012; Li *et al*., 2014) and can be easily implemented using the Genomic Association and Prediction Integrated Tool (GAPIT) in R environment with a minimal amount of code (Tang *et al*., 2016).

GWAS has been extensively applied in wheat genetics to identify significant marker-trait associations with quantitative resistance for FHB (Kollers *et al*., 2013; Wang *et al*., 2017), stem rust (Zhang *et al*., 2014; Juliana *et al*., 2015), stripe rust (Maccaferri *et al*., 2015; Liu *et al*., 2017), and leaf rust (Gao *et al*., 2016; Pasam *et al*., 2017). Regardless of its broad adoption, GWAS still has its own limitations such as spurious associations and failure to detect rare variants. Unfortunately for the GWAS aficionados, plant breeders are usually interested in discovering rare variations that could potentially make the next leading variety (Bernardo, 2016). However, another strategy to overcome these limitations, it is the use of genome-wide predictions which is detailed in the following section.

Genomic-based Predictions

In the early 2000's Meuwissen *et al*. (2001) proposed that with the increase of genomic marker density, it would be possible to estimate the genetic variance attributed to loci and predict phenotypic performance. It was the birth of genomic selection (GS). This analysis attempts to capture the total additive variance from all markers distributed across the whole genome rather than relying on performing significance tests at every single locus as done by GWAS models. Using genome-wide markers, every trait locus is likely to be in linkage disequilibrium (LD) with a minimum of one marker locus in the target population (Dreisigacker *et al*., 2016). As result, GS generally is more capable of accounting for small-effect loci associated with complex quantitative traits such as FHB and wheat rust resistance (Poland & Rutkoski, 2016).

GS uses a '*training population'* of individuals that have been genotyped and phenotyped for traits of interest in a target environment to predict the performance of a non-phenotyped but genetically related '*testing set*' that only have been genotyped (Jannik *et al*., 2010). This procedure produces genomic estimated breeding values (GEBV). These GBEVs are not identifying loci nor indicating genomic regions underlying genes/QTLs, but they summarize the presence of favorable alleles from a genome-wide markers analysis, that can be used a selection criterion. Several statistical models can be used to estimate GEBV. Here we focused on four of GS models: Ridge Regression Best Linear Unbiased Predictor (rrBLUP), Partial Least Squares Regression (PLSR), Elastic Net (ELNET), and Random Forest (RF).

rrBLUP is a mixed linear model developed by Elderman (2011) that considers random effects for markers where the marker variance is the quotient of total genetic variance divided by the number of markers, assuming that each locus contributes equally to the phenotype expression. The PLSR model (Mevik *et al*., 2013) takes into account principal components and multivariate regression on its predictions. ELNET is a generalized linear model (Friedman *et al*., 2010) that sets a mixing and tuning parameter to create a grid selection technique that allows markers to have variable and null effects at loci. The RF model (Breiman, 2011) is considered a machine-learning algorithm (Poland & Rutkoski, 2016) capable of capturing non-additive effects. Predictions from this model are based on a multiple decision trees that accounts for relatedness of individuals allowing the effect of markers depending on alleles present at each locus. It has been reported that RF generates higher prediction accuracies than other models when predicting quantitative traits that are controlled by multiple small-effect QTLs such as disease resistance (Rutkoski *et al*, 2012).

Recent studies are allowing to run multiple statistical models simultaneously, while estimating average prediction across different GS models (Gaynor, 2015), others included covariates to account for G×E interaction (Crossa *et al*., 2017), pedigree distance matrix (Juliana *et al*., 2017) and even crop growth modeling as covariates (Rincent *et al*., 2017) aiming to increase the accuracy of GS prediction. Others are reporting significant improvements made through the incorporation of fixed effect markers (Spindel *et al*., 2016). GS models that treat known QTLs as fixed effects may increase accuracy predictions by more than 30% when compared to conventional models with all SNPs treated as random effects (Arruda *et al*., 2016b).

Cross Predictions

Perhaps more important than using genome-wide marker data of make forward predictions of non-phenotyped elite lines in unknown environments, this resource could be also applied to predict which individuals would generate the most promising progeny when crossed. Consider a typical wheat breeding program that includes ~300 elite lines in the crossing block from where around a 1000 crosses are made every year. This number of crosses represent only 2.23% of the total number of the 44850 possible combinations, meaning that it would take nearly 40 years of work to generate all possible combinations from one season of the crossing block.

Cross-prediction refers to the ability to estimate and simulate cross combinations from a set of elite parents aiming to identify the ones that are more likely to generate a superior progeny. Although, crossing elite parents ensures the desired population mean, it does not guarantee that sufficient genetic variance will be created from which to select progeny (Bernardo *et al*., 2010). Several studies have recently proposed to predict the genetic variance of the progeny by simulating crosses where linkage structure and recombination is accounted for, and after that, the performance of the segregating population is simulated and predicted (Bernardo, 2014). Based

on these principles, Mohammadi *et al*. (2015) were pioneers in the development of the R package 'PopVar' which is capable of estimating the genetic variance in simulated populations based on phenotypic and genotypic data from a list of potential parents.

In general, most of the wheat breeding programs plan crosses exclusively based on the phenotypic data available from prior years and MAS data for a few genes. However, many crosses end up being discarded in early generations when superior progeny are not identified (Heslot *et al*., 2015). Cross-predictions have been applied to predict wheat crosses aiming superior grain yield and baking quality at CIMMYT and INIA (Lado *et al*., 2017), as well as FHB resistance traits in wheat and barley (Tiede *et al*., 2015; Mohammadi *et al*., 2015).

Although some progress has been reported in the literature, more cross validations for this approach are needed, as well as the integration of fixed effect markers that often explain a larger proportion of the genetic variance. Furthermore, a more complex cross design must be considered, since in wheat breeding programs the majority of progenies come from three-way crosses, instead of relying solely on biparental cross predictions, as currently done by 'PopVar'.

References

Arruda M. P., Brown P., Brown-Guedira G.B., Krill A.M., Thurber C., Merrill K.R., Foresman B.J., Kolb F.L. (2016a). Genome-Wide Association Mapping of Fusarium Head Blight Resistance in Wheat using Genotyping-by-Sequencing. The Plant Genome 9(1): 1-14.

Arruda M. P., Lipka A. E., Brown P. J., Krill A. M., Thurber C., Brown-Guedira G., Dong Y. Foresman B. J., Kolb F. L. (2016b). Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). Mol Breed. 36 (7):1-11.

Bernardo A.N. Ma H., Zhang D., Bai G. (2012). Single nucleotide polymorphism in wheat chromosome region harboring *Fhb1* for Fusarium head blight resistance. Mol Breed. 29(2): 477-488.

Bernardo, R. (2014). Genome-wide selection of parental inbreds: Classes of loci and virtual biparental populations. Crop Science. 54(1): 1-33.

Bernardo, R. (2016). Bandwagons I, too, have known. Theor Appl Genet. 129(12): 2323-2332.

Brar G. S., Pozniak C. J., Ruan Y. Thomas J., Kutcher H. R., Hucl P.J. (2015). Fusarium head blight resistance QTL *Fhb1*, *Fhb2*, and *Fhb3* reduce disease severity by up to 50% in near-isogenic wheat lines developed by marker-assisted selection. Crop Development Centre, Department of Plant Sciences, University of Saskatchewan, 51 Campus Drive, Saskatoon, SK S7N 5A8, Canada; (Y.R.) Semiarid Prairie Agricultural Research Centre-Agriculture and Agri-Food Canada, 1 Airport Road, P.O. Box 1030, Swift Current, SK R6M 1Y5, Canada; and (J.T.) 87, Pulberry Street, Winnipeg, MB R2M 2C4, Canada.

Breiman, L. (2001). Random Forests. Machine Learning 45: 5-32.

Broman K., & Sen S. (2009). A Guide to QTL Mapping with R/qtl: Springer.

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

Cai J., Wang S., Li T., Zhang G., Bai G. (2016). Multiple Minor QTL Are Responsible for Fusarium Head Blight Resistance in Chinese Wheat Landrace Haiyanzhong. PLoS One 11(9): e0163292.

Cai, J. (2016). Meta-analysis of QTL for Fusarium head blight resistance in Chinese wheat landraces using genotyping by sequencing. Ph.D. Dissertation, Kansas State University.

Cainong J.C., Bockus W.W., Feng Y., Chen P., Qi L., Sehgal S.K., Danilova T.V., Koo D.H., Friebe B., Gill B.S. (2015). Chromosome engineering, mapping, and transferring of resistance to Fusarium head blight disease from *Elymus tsukushiensis* into wheat. Theor Appl Genet. 128(6):1019-1027.

Chen X.M. (2005). Epidemiology and control of stripe rust [*Puccinia striiformis f. sp. tritici*] on wheat. Canadian Journal of Plant Pathology 27: 314-337.

Clinesmith M., (2016). Genetic mapping of QTL for Fusarium head blight resistance in winter wheat cultivars Art and Everest. Master Thesis. Kansas State University, Manhattan Kansas.

Crossa J., Pérez-Rodríguez P., Cuevas J., Montesinos-López O., Jarquín D, De Los Campos G., Burgueño J., González-Camacho J.M., Pérez-Elizalde S., Beyene Y., Dreisigacker S., Singh R. *et al*. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science 22(11): 961-975.

Cuthbert P. A., Somers D. J., Brulé-Babel A., (2007). Mapping of *Fhb2* on chromosome 6BS: a gene controlling Fusarium head blight field resistance in bread wheat (*Triticum aestivum* L.). Theor Appl Genet. 114(3):429-437.

Dhokane D., Karre S., Kushalappa A.C., McCartney C. (2016). Integrated Metabolo-Transcriptomics Reveals Fusarium Head Blight Candidate Resistance Genes in Wheat QTL-*Fhb2*. PLoS ONE 11(5): e0155851.

Dreisigacker S. *et al*. (2016). Molecular Marker-Based Selection Tools in Spring Bread Wheat Improvement: CIMMYT Experience and Prospects. In: Rajpal V., Rao S., Raina S. (eds) Mol Breed. for Sustainable Crop Improvement. Sustainable Development and Biodiversity, vol 11. Springer, Cham.

Dweba C.C., Figlan S., Shimelis H.A., Motaung T.E., Sydenham S., Mwadzingeni L., Tsilo T. J. (2017). Fusarium head blight of wheat: Pathogenesis and control strategies. Crop Protection 91: 114-122.

Eckard J., Gonzalez-Hernandez J., Caffe M., Berzonsky W., Bockus W., Marais G., & Baenziger P. (2015). Native Fusarium head blight resistance from winter wheat cultivars ‘Lyman,’ ‘Overland,’ ‘Ernie,’ and ‘Freedom’ mapped and pyramided onto ‘Wesley’- *Fhb1* backgrounds. Mol Breed. 35:6.

Ellis J.G., Lagudah E.S., Spielmeyer W., Dodds P.N. (2014). The past, present and future of breeding rust resistant wheat. Frontiers in Plant Science 5(641): 1-13.

Endelman J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome 4(3): 250-255.

Faostat F. (2017). Statistical Databases. Food and Agriculture Organization of the United Nations.

Feuillet C., Schachermayr G. and Keller B. (1997). Molecular cloning of a new receptor-like kinase gene encoded at the *Lr10* disease resistance locus of wheat. The Plant Journal 11: 45-52.

Flor H.H., (1942). Inheritance of pathogenicity races 22 and 24 of *Melampsora lini*. Phytopathology 32.5.

Friedman J., Hastie T., R. Tibshirani R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33(1): 1-22.

Fu D., Uauy C., Distelfeld A., Blechl A., Epstein, L., Chen X., Sela, H., Fahima T., Dubcovsky J. (2009). A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323: 1357-1360.

Gao L., Turner M.K., Chao S., Kolmer J., Anderson J.M. (2016). Genome Wide Association Study of Seedling and Adult Plant Leaf Rust Resistance in Elite Spring Wheat Breeding Lines. PLoS One 11(2): e0148671.

Gaynor R.C. (2015). Genomic Selection for Kansas Wheat. Ph.D. Dissertation. Kansas State University. Manhattan, Kansas.

Gunupuru L. R., Perochon A., Doohan F.M. (2017). Deoxynivalenol resistance as a component of FHB resistance. Tropical Plant Pathology 42(3): 175-183.

Guo J., Zhang X., Hou Y., Cai J., Shen X., Zhou T., Xu H., Ohm H.W., Wang H., Li A., Han F., Wang H., Kong L. (2015). High-density mapping of the major FHB resistance gene *Fhb7* derived from *Thinopyrum ponticum* and its pyramiding with *Fhb1* by marker-assisted selection. Theor Appl Genet. 128(11): 2301-2316.

He J., Zhao X., Laroche A., Liu H., Li Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science 5(484): 1-8.

Huang L., Brooks S. A., Li W., Fellers J. P., Trick H. N. and Gill B. S. (2003). Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164(2): 655-664.

Ikegawa S. (2012). A Short History of the Genome-Wide Association Study: Where We Were and Where We Are Going. Genomics Inform. 10(4): 220-225.

Johnson R. (1983). Genetic Background of Durable Resistance. In: Lamberti F., Waller J.M., Van der Graaff N.A. (eds) Durable Resistance in Crops. NATO Advanced Science Institutes Series (Series A: Life Sciences), vol. 55. Springer, Boston, MA.

Juliana P., Rutkoski J.E., Poland J.A., Singh R.P., Murugasamy S., Natesan S., Barbier H., Sorrells M.E. (2014). Genome-Wide Association Mapping for Leaf Tip Necrosis and Pseudo-Black Chaff in Relation to Durable Rust Resistance in Wheat. the Plant Genome 8(2): 1-12.

Juliana P., Singh R.P., Singh P.K., Crossa J., Huerta-Espino J., Lan C., Bhavani S., Rutkoski J.E., Poland J.A., Bergstrom G.C., Sorrells M.E. (2017). Genomic and pedigree-based

prediction for leaf, stem, and stripe rust resistance in wheat. Theor Appl Genet. 130: 1415-1430.

Kassa M. T., You F.M., Hiebert C.W., Pozniak C.J., Fobert P.R., Sharpe A.G., Menzies J.G., Humphreys G., Harrison N.R., Fellers J.P., McCallum B.D., McCartney B.C. (2017). Highly predictive SNP markers for efficient selection of the wheat leaf rust resistance gene *Lr16*. BMC Plant Biology 17(45): 2-9.

Kearsey M.J. (2002). QTL analysis: Problems and (possible) solutions. In MS Kang (ed.) Quantitative genetics, genomics, and plant breeding. CABI Pub., pp.45-58. New York.

Kolmer J.A. (1996). Genetics of resistance to wheat leaf rust. Annual Review of Phytopathology 34(1): 435-455.

Kolmer, J. A. (2001). Early Research on the Genetics of Puccinia graminis and Stem Rust Resistance in Wheat in Canada and the United States.

Leonard K. J., & Szabo L. J. (2005). Stem Rust of Small Grains and Grasses caused by Puccinia graminis. Molecular Plant Pathology 8(2):99-111.

Li M., Liu X., Bradbury P., Yu J., Zhang Y. M., Todhunter R.J. (2014). Enrichment of statistical power for genome-wide association studies. BMC Biol. 12(73): 2- 10.

Lin F., Xue S.L., Zhang Z.Z., Zhang C.Q., Kong Z.X., Yao G.Q., Tian D.G., Zhu H.L., Li C.J., Cao Y., Wei J.B., Luo Q.Y., Ma Z.Q. (2006). Mapping QTL associated with resistance to Fusarium head blight in the Nanda2419 × Wangshuibai population. II: Type I resistance. Theor Appl Genet. 112(3): 528-535.

Lipka A.E., Tian F., Wang Q., Peiffer J., Li M., Bradbury P.J. (2012). GAPIT: Genome association and prediction integrated tool. Bioinformatics. 28(18): 2397-2399.

Liu S., Hall M. D., Griffey C. A., McKendry A. L. (2009). Meta-analysis of QTL associated with Fusarium Head Blight in Wheat. Crop Science 49: 1955-1968.

Liu S., Pumphrey M. O., Gill B.S., Trick H. N. Zhang J. X., Dolezel J. Chalhoub B., Anderson J. A. (2008). Towards position cloning of *Fhb1*, a major QTL for Fusarium head blight resistance in wheat. 3th Int. FHB Symposium. Szeged, Hungry.

Liu W., Maccaferri M., Chen X., Laghetti G., Pignone D., Pumphrey M., Tuberosa, R. (2017). Genome-wide association mapping reveals a rich genetic architecture of stripe rust resistance loci in emmer wheat (*Triticum turgidum ssp. dicoccum*). Theor Appl Genet. 130(11): 2249-2270.

Lu Q.X., Lillemo M., Skinnes H., He X.Y., Shi J.R., Ji F., Dong Y.H., Bjornstad A. (2013). Anther extrusion and plant height are associated with Type I resistance to Fusarium head blight in bread wheat line 'Shanghai-3/ Catbird'. Theor Appl Genet. 126(2): 317-334.

Maccaferri M., Zhang J., Bulli P., Abate Z., Chao S., Cantu D., Bossolini E., Chen X., Pumphrey M., Dubcovsky J. (2015). A genome-wide association study of resistance to stripe rust (*Puccinia striiformis f. sp. tritici*) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). G3 Genes Genomes Genet. 5(3): 449-465.

Marone D., Russo M.A., Laido G., Leonardis A M., Mastrangelo A.M. (2013). Plant Nucleotide Binding Site–Leucine-Rich Repeat (NBS-LRR) Genes: Active Guardians in Host Defense Responses. Int. J. Mol. Sci. 14: 7302-7326.

McIntosh R.A., Dubcovsky J., Rogers W.J., Morris C., Xia X.C. (2017). Catalog of genes symbols for wheat: 2017 Supplement.

McMullen M., Bergstrom G., De Wolf E., Dill-Macky R., Hershman D., Shaner G., Van Sanford D. (2012). A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Disease. 96(12): 1712-1728.

Mesterhazy A. (1995). Type and components of resistance to Fusarium head blight of wheat. Plant Breed. 114(5): 377–386.

Meuwissen T.H.E., Hayes B.J., & Goddard M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4):1819–1829.

Mevik B.-H., Wehrens R., & Liland K.H. (2013). PLS: Partial Least Squares and Principal Component regression.

Moazami E.F., & Jinap S. (2009). Natural occurrence of deoxynivalenol (DON) in wheat based noodles consumed in Malaysia. Microchemical Journal. 93(1): 25-28.

Moore J. W., Herrera-Foessel S., Lan c., Schnippenkoetter W., Ayliffe M., Huerta-Espino J., Lillemo M., Viccars L., Milne R., Periyannan S., Kong X., Spielmeyer W., Talbot M., Bariana H., Patrick J.W., Dodds P., Singh R., Lagudah E. (2015). A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nature Genetics 47(12): 1494-1498.

Mundt, C.C. (2014). Durable resistance: A key to sustainable management of pathogens. Infection, Genetics and Evolution 27: 446-455.

Parry D.W. (1990). Diseases of small grain cereals. In: Plant Pathology in Agriculture. Cambridge University Press, Cambridge, p. 160-224.

Pasam R.J., Bansal U., Daetwyler H.D., Forrest K.L. Wong D., Petkowski J., Willey N., Randhawa M., Chhetri M., Miah H., Tibbits J., Bariana H., Hayden M.J. (2017). Detection and validation of genomic regions associated with resistance to rust diseases in a worldwide hexaploid wheat landrace collection using BayesR and mixed linear model approaches. Theor Appl Genet. 130(4): 777-793.

Periyannan S., Moore J., Ayliffe M., Bansal U., Wang X., Huang L., Deal K., Luo M., Kong X., Bariana H., Mago R., McIntosh R., Dodds P., Dvorak J., Lagudah E. (2013). The gene

Sr33, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. Science 341(6147): 786-788.

Poland J., Balint-Kurti J., Wisser R. J., Pratt R.C., Nelson R.J. (2009). Shades of gray: the world of quantitative disease resistance. Trends Plant Science 14(1): 21-29.

Poland J.A., Brown P.J., Sorrells M.E., Jannink J.L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. Plos One. 7(2): e32253.

Poland, J. A., & Rife T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. The Plant Genome 5(3):92-102.

Qi L.L., Pumphrey M.O., Friebe B., Chen P. D., Gill. B.S. (2008). Molecular cytogenetic characterization of alien introgressions with gene *Fhb3* for resistance to Fusarium head blight disease of wheat. Theor Appl Genet. 117(7):1155-1166.

Rawat N., Pumphrey M.O., Liu S., Zhang X., Tiwari V.K., Ando K., Trick H.N., Bockus W.W., Akhunov E., Anderson J.A., Gill B.S. (2016). Wheat *Fhb1* encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. Nature Genetics 48(12): 1576-1580.

Rincent R., Kuhn E., Monod H., Oury F.X., Rousset M., Allard V., Le Gouis, L. (2017). Optimization of multi-environment trials for genomic selection based on crop models. Theor Appl Genet.130(8): 1735-1752.

Roelfs A. P., & Bushnell W. (1985). The Cereal Rusts, Volume II.

Roelfs A.P., Singh R.P., Saari E.E. (1992). Rust Diseases of wheat: Concepts and methods of disease management. CIMMYT, Mexico, D.F.

Rutkoski J., Benson J., Jia Y., Brown-Guedira G., Jannink J.L., Sorrells M. (2012). Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. The Plant Genome 5 (2): 51-61.

Schumann, G.L., & Leonard K.J. (2000). Stem rust of wheat (black rust). The Plant Health Instructor.

Shah L., Ali A., Zhu Y., Wang S., Si H., Ma C. (2017). Wheat defense response to Fusarium head blight and possibilities of its improvement. Physiological and Molecular Plant Pathology 98: 9-17.

Shephard, G.S. (2011). Fusarium mycotoxins and human health: a review. Plant Breeding and Seeds Science 64: 113-121.

Singh R.P., Huerta-Espino J., Willian H.M. (2005). Genetics and Breeding for Durable resistance to leaf and stripe rusts in wheat. Turk. J. Agric. Fror. 29: 121-127.

Singh R.P., Singh P.K., Rutkoski J., Hodson D.P., He X., Jorgensen L.N., Hovmoller M.S., Huerta-Espino J. (2016). Disease impact on wheat yield potential and prospects of genetic control. Annu. Rev. Phytopathol. 54: 303-322.

Singh, R. P., Hodson, D. P., Jin, Y., Lagudah, E. S., Ayliffe, M. A., Bhavani, S., Rouse, M. N., Pretorius, Z. A., Szabo, L. J., Huerta-Espino, J., Basnet, B. R., Lan, C., and Hovmøller, M. S. (2015). Emergence and spread of new races of wheat stem rust fungus: Continued threat to food security and prospects of genetic control. Phytopathology 105(7): 872-884.

Sneller C., Guttuieri M., Paul P., Costa J., Jackwood R. (2012). Variation for Resistance to Kernel Infection and Toxin Accumulation in Winter Wheat Infected with Fusarium graminearum. Phytopathology 102(3): 306-314.

Stakman E.C., Stewart P.M., Loegering W. (1962). Identification of physiologic races of Puccinia graminis var. tritici. US Department of Agriculture, Agricultural Research Service. pp. Publ. E617.

Steiner B., Buerstmayr M., Michel S., Schweiger W., Lemmens M., Buerstmayr H. (2017). Breeding strategies and advances in line selection for Fusarium head blight resistance in wheat. Tropical Plant Pathology 42(3):165-174.

Steiner B., Lemmens M., Griesser M., Scholz U., Schondelmaier J., Buerstmayr H. (2004). Molecular mapping of resistance to Fusarium head blight in the spring wheat cultivar Frontana. Theor Appl Genet. 109(1): 215-224.

Steuernagel B., Periyannan S.K., Hernández-Pinzón I., Witek K., Rouse M.N., Yu G., Hatta A., Ayliffe M., Bariana H., Jones J.D.G., Lagudah E.S., Wulff B.BH. (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. Nature Biotechnology 34: 652-655.

Tang Y., Liu X., Wang J., Li M., Wang Q., Tian F., Su Z., Pan Y., Liu D., Lipka A. E., Buckler E. S., Zhang Z. (2016). GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. The Plant Genome 9(2): 1-9.

Thind A.K., Wicker T., Šimková H., Fossati D., Moullet O., Brabant C., Vrána J., Doležel J., Krattinger S.G. (2017). Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. Nat Biotechnol. 35(8):793-796.

Thomas DC, Haile RW, Duggan D. (2005). Recent developments in genome-wide association scans: a workshop summary and review. Am J Hum Genet. 77(3): 337-345.

Thomson M.J. (2014). High-Throughput SNP Genotyping to Accelerate Crop Improvement. Plant Breed. Biotech. 2(3): 195-212.

Trail F. (2009). For blighted waves of grain: Fusarium graminearum in the post-genomics era. Plant Physiol. 149:103–110.

U.S. Food and Drug Administration – FDA (2017). Guidance for Industry and FDA: Advisory Levels for Deoxynivalenol (DON) in Finished Wheat Products for Human Consumption and Grains and Grain By-Products used for Animal Feed. Available online on: https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ChemicalContaminantsMetalsNaturalToxinsPesticides/ucm120184.htm.

USDA - Foreign Agricultural Service (2017) Production, Supply, and Distribution Database.

Van Der Biezen E.A., & Jones J.D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. Trends Biochem. Sci. 23(12): 454-456.

Wang R., Chen J., Anderson J.A., Zhang J., Zhao W., Wheeler J., Klassen N., See D.R., Dong Y. (2017). Genome-Wide Association Mapping of Fusarium Head Blight Resistance in Spring Wheat Lines Developed in the Pacific Northwest and CIMMYT. Phytopathology 107(12): 1486-1495.

Wegulo, S. N., Baenziger P. S., Hernandez Nopsa J., Bockus W. W., & Hallen-Adams H. (2015). Management of Fusarium head blight of wheat and barley. Crop Protection 73: 100-107.

Xue S., Li G., Jia, H., Xu, F., Lin, F., Tang, M., Wang Y., An X., Xu H., Zhang L., Kong Z., Ma Z. (2010). Fine mapping *Fhb4*, a major QTL conditioning resistance to Fusarium infection in bread wheat (*Triticum aestivum* L.). Theor Appl Genet. 121: 147-156.

Xue S., Tang, M., Zhou Y., Li G., An X., Lin F., Xu H., Jia H., Zhang L., Kong Z., Ma Z. (2011). Precise mapping *Fhb5*, a major QTL conditioning resistance to Fusarium infection in bread wheat (*Triticum aestivum* L.). Theor Appl Genet. 123(6): 1055-1063.

Yang Z.P., Gilbert J., Somers D.J., Fedak G., Procunier J.D., McKenzie R. (2003). Marker-assisted selection of Fusarium head blight resistance genes in two doubled-haploid populations. of wheat. Mol Breed. 12:309–317.

Yu L.-X., Barbier H., Rouse M. N., Singh S., Singh R. P., Bhavani S., Huerta-Espino J., & Sorrells M. E. (2014). A consensus map for Ug99 stem rust resistance loci in wheat. Theor. Appl. Genet. 127(7): 1561-1581.

Zhang D., Bowden R.L., Yu J., Carver B. F., Bai G. (2014). Association Analysis of Stem Rust Resistance in U.S. Winter Wheat. PLoS One 9(7): e103747.

Chapter 2 - QTL Mapping of Fusarium Head Blight Resistance and Deoxynivalenol Accumulation in Kansas Wheat

Abstract

Fusarium head blight (FHB) is a wheat disease that reduces grain yield and accumulates mycotoxins in wheat-derived products. Deoxynivalenol (DON) is the most prevalent toxin and it has an advisory limit of 1 ppm in products for human consumption. The objective of this study was to map QTLs associated with components of native FHB resistance in Kansas wheat. A doubled haploid (DH) population was developed from a cross between Everest and WB-Cedar, which are moderately resistant and moderately susceptible to FHB, respectively. DH lines, parents, and checks were evaluated in the field for two years in a randomized complete block design with three replications. The evaluation of percentage of symptomatic spikelets (PSS) started 14 days after heading and repeated every three days. Genotypes were also evaluated for thousand kernel weight, DON accumulation, *Fusarium*-damaged kernels (FDK), and grain protein content. All DH lines and parents were genotyped using genotyping-by-sequencing. Standard, composite, and multiple QTL mapping was performed using Haley-Knott regression. Five QTLs from Everest were identified on 1BS, 3BS, 5AS, 5DS, and 6BS indicating that this cultivar is a source of native FHB resistance with multiple mechanisms of resistance. Another three QTLs from WB-Cedar were mapped on 1AS, 5BL, and 7AL. Additive QTL effects were confirmed by a contrasting analysis which showed that individuals containing all resistant alleles were significantly more resistant than lines with one or no favorable alleles at peak loci. A 67% reduction in DON content was observed in lines carrying all mapped QTLs for DON.

2.0 Introduction

Fusarium head blight (FHB), also known as scab, is a wheat disease caused by *Fusarium graminearum* that significantly reduces grain yield and produces mycotoxins that contaminate wheat grains and flour. Given the rise of global temperatures, major epidemics are likely to occur in the future, particularly in warm regions with high humidity (Dweba *et al*., 2017). Economic losses were ~$7.5 billion in the period from 1993 and 2001 in the United States alone due to severe FHB epidemics(Nganje *et al*., 2004). FHB also lead to accumulation of deoxynivalenol (DON) which is the most prevalent toxin produced by this pathogen being unsafe for human and animal health when consumed in concentrations higher than 1 ppm.

In Kansas, FHB occurs predominantly in the Eastern third of the state, where losses can be substantial, with occasional epidemics also occurring in central part of the state. Growers from these regions have access and are adopting, moderately resistant varieties, such as Everest, which has led the state in acreage for five consecutive years. In the South east district, 58.6% of the acreage was planted with Everest in the 2015-2017 growing seasons (Kansas Ag Statistics, 2017). As a result, an estimated ~$30 million are saved every year by the simple practice of using resistant varieties (Bockus *et al*., 2015).

QTL mapping of biparental populations has been the most popular method for identifying new genomic regions associated with FHB resistance. Hundreds of QTLs for FHB resistance have been reported covering all wheat chromosomes with varying effects (Liu *et al*., 2009). In addition, the quantitative nature and complex inheritance often complicate efficient mapping (Buerstmayr *et al*., 2009). FHB resistance has multiple components, of which the most important are: resistance to initial infection (Type I), disease spread after infection (Type II), resistance to

toxin accumulation (RTA or Type III), and resistance to kernel infection (RKI or Type IV) (Mesterhazy, 1995; Sneller *et al*., 2012).

In breeding programs where the FHB resistance is primarily based on native small-effect QTLs from phenotypic selection, mapping these loci and further selection by MAS is more challenging. Yet, breeding for FHB resistance with these type QTLs has its own advantages, such as avoiding introgression of large genetic blocks and/or alien segments from unadapted materials, which often have yield penalties and not relying on biased marker associations. Besides, major well-known QTLs could still easily be integrated with small-effect QTL in order to improve FHB resistance. Furthermore, it is important to point out that resistance is not deployed by itself, but rather in combination with other traits, such as yield potential and end-use quality, and finding the perfect balanced combination is the endless challenge of breeding.

In this study, a DH-biparental population was made with the goal of mapping the native FHB resistance of Everest in a single cross with WB-Cedar. Both parents are currently among of the most planted varieties in Kansas, representing nearly 20 % of the area grown with wheat the last three growing seasons (Kansas Ag Statistics, 2017). Therefore, the objective of this chapter was to: (1) map QTLs associated with FHB resistance types II, III, and IV; (2) perform a contrast grouping analysis to examine the effect of stacking QTLs within breeding lines; (3) and select the most resistant DH-lines to move forward in the K-State hard red winter wheat breeding Program. A better understanding of the genetic mechanisms underlying FHB resistance will allow breeders to build on the native resistance levels of Everest and assist in the development of even more resistant varieties for Kansas farmers in the near future.

2.1 Materials and Methods

Genetic Material

A doubled haploid (DH) population with 202 lines was developed from a cross between the cultivars Everest and WB-Cedar, which are moderately resistant and moderately susceptible to FHB, respectively (De Wolf *et al*., 2015). Everest was developed and released by Kansas State University in 2009, whereas WB-Cedar was developed by the company WestBred and released in 2011. The pedigree of Everest (PI 659807) is HBK1064-3/JAGGERW//X960103 whereas WB-Cedar (PI 661996) came from a cross of TAM-302/PIONEER-2180//EXP (NPGS, 2017). Both parents are hard red winter wheats and have the same alleles at the *Rht-B1*, *Rht-D1*, and *Ppd-D1* and *Ppd-B1* loci (unpublished data). As a result, this DH population is fairly homogenous in terms of plant height and maturity, which is a desirable characteristic for QTL mapping studies of FHB resistance since these traits may cause confounding effects.

Experimental Design

The study was conducted in the field at Rocky Ford FHB Nursery of the Department of Plant Pathology (Kansas State University) during the growing seasons of 2014/2015 and 2015/2016. The experiment was set up in a randomized complete block design with three replications, where each experimental unit was formed by a one-meter long single row plot. These were planted in sets of six rows where the susceptible and resistant checks (Overley and Everest, respectively) were planted as borders flanking sets of four genotypes. Full sets of checks (three consecutive rows) were also repeated three times within each block, with the goal of getting more accurate data for further comparative analysis.

Field Inoculations

The isolate GZ 3639 of *F. graminearum* which is native to Kansas was used as the primary inoculum source. The pathogen was isolated and purified in the lab. After that, corn kernels were autoclaved and colonized with this isolate to be further distributed in the field and cause disease. A homogeneous inoculation was achieved by scattering corn kernels infected with the pathogen in the field two weeks before heading. The field was then irrigated using misting sprinklers running three minutes per hour at night from the beginning of flowering to early dough stage, as described by Jin *et al*. (2013).

Phenotypic Evaluations of FHB

Heading date was recorded on plot basis when 50% of the spikes were exposed. Visual evaluation of the percentage of symptomatic spikelets (PSS) started 14 days after heading and was repeated every 3 days for a total of 5 evaluations in each year. This data was later used to estimate the spread intensity of the disease between different days of evaluation by calculating the area under the disease progress curve (AUDPC) as described by Madden *et al*. (2007).

Plant height (PH) was measured in centimeters from the soil surface to the top of spikes, excluding the awns, at maturity. Additionally, stripe rust notes were taken in the first year using a linear 1 to 9 scale where 1 is resistant and 9 is highly susceptible. Thousand kernel weight (TKW) was estimated by counting a sample of 300 kernels from each plot using an automated seed counter model 805-3 (International Marketing and Design, USA).

A representative sample of 100 grains from each harvested plot was collected to estimate DON and *Fusarium*-damaged kernel (FDK) using single kernel near-infrared spectroscopy (SKNIR). This automated system was developed by the USDA-ARS, CGAHR, Engineering and Wind Erosion Research and is capable of collecting and analyzing individual wheat kernels

using a spectrometer with an indium-gallium-arsenide detector that measures absorbance at 950–1650 nm (Dowell *et al*., 1999; Dowell *et al*., 2006). The SKNIR collects spectra values for each single wheat kernel and later different calibration curves are used to predict the trait of interest such as DON, FDK, protein content, and grain hardiness.

The DON calibration curve predicts the amount of mycotoxin in parts per million in each individual kernel. An average of all 100 kernels was calculated for each sample, as well as DON values for each sorted bin. FDK was calculated using the sorting feature, which sorts the kernels into different bins. Kernels with predicted values between -1.00 and 1.50 are classified as sound kernels whereas kernels with predicted values greater than 1.50 are considered FDK, which was expressed in percentage (FDK / FDK + sound) for each sample (Dowell *et al*., 2009; Peiris *et al*., 2010).

Genotyping and SNP Filtering

Leaf tissue from seedlings of all DH lines and the parents were collected at two-leaf stage. DNA was extracted and purified using the "BS96 DNA Plant" protocol for the Biosprint 96 workstation and the Biosprint 96 DNA plant kit (Qiagen, Hilden, Germany). The DNA samples were digested with the restriction enzymes *PstI* and *MspI* which are a rare-cuter and *a* common-cutter, respectively. After that, the samples were amplified and sequenced with an Illumina HiSeq equipment as described by Poland *et al*. (2012). A bioinformatics pipeline on TASSEL 4 was used to call and filter SNP markers. Only single nucleotide polymorphisms (SNPs) that were polymorphic between the parents and less than 50% of missing data were retained. A total of 13,752 SNPs were originally discovered.

The remnant SNP markers were filtered again using mapping construction functions of the R/qtl package (Broman, 2009) in R environment. Typically, it is expected that, on average,

50% of alleles come from each parent in a DH population originating from a single cross. Thus DH-lines with more than 90% matching markers were considered duplicates and removed using the function *comparegeno()*. Likewise, SNP markers were filtered for minor allele frequency greater than 20% while duplicate genotypes were removed by the function *findDupMarkers()*. Subsequently, markers that showed significant segregation distortion were also identified and discarded. The genotypic frequency of the parental alleles was assessed for each DH line and potentially switched alleles (estimated recombination fraction > 0.5) were discarded.

The DH population, parents, along with positive and negative controls were genotyped with three SSR markers: *gwm133* (Cuthbert *et al*., 2007), *Xgwm149* (Xue *et al*., 2010), and *Xgwm304* (Xue *et al*., 2011) which are respectively linked to *Fhb2*, *Fhb4*, and *Fhb5*. Based on results of prior genotyping (unpublished data) neither parent carries *Fhb1*, eliminating the need to screen the population for this gene.

Linkage Map Construction

The genotypic data was loaded into JoinMap 4.1 (Van Ooijen, 2006) to create a linkage map using the Kosambi mapping function, which estimates genetic distances between markers along with a maximum likelihood independence LOD for grouping. A total of 34 linkage groups were initially identified. When multiple linkage groups were assigned to the same chromosome, PopSeq positions were used to recalculate the genetic distances between markers for that chromosome (Chapman *et al*., 2015). This resource was also used to re-order the markers, from short arm to long arm, within each wheat chromosome.

Statistical Analysis of Phenotypic Traits

All the phenotypic data was analyzed using the software SAS version 9.4 (SAS Institute Inc., 2012). Initially, PROC UNIVARIATE was used to check the normality and distribution of

the phenotypic traits. The experimental factors were: genotype, block, year, block within year, and genotype by year (G×Y) interaction. The analysis of variance was conducted using the SAS procedure PROC GLM where the factors year and block were treated as random effects while genotype was considered fixed.

Plant height (PH) was included in the model as a covariate of fixed effect, since PH may be a confounding factor as shorter genotypes have greater exposure to the source of inoculum present in the ground. Ultimately, each year was analyzed separately due to the significant G×Y interaction for the majority of evaluated traits. The broad sense heritability (H^2) of all traits was calculated on an entry-mean basis using the following formula: $H^2 = \sigma_G^2/(\sigma_G^2 + \frac{\sigma_{G\times Y}^2}{n} + \frac{\sigma_E^2}{nr})$, where σ_G^2= genetic variance, $\sigma_{G\times Y}^2$= genotype-by-year interaction variance, σ_E^2 = error variance, n= number of experiments, and r= number of replications (Holland *et al*., 2002).

Mean comparisons between all DH lines and parents were calculated using the Tukey-Kramer test (Tukey, 1949) at 5% of probability of error. Pearson correlations and its graphic representations were estimated and drawn using the *scatterplotMatrix* function in R. DH lines were also sorted according to their QTL combinations, then the QTL effects on phenotypic traits were tested by a contrast grouping analysis in SAS with genotype groups considered as random. The adjust LSmeans of all traits were then used to perform further QTL mapping analysis.

QTL Mapping Analysis

The parental alleles of Everest and WB-Cedar were coded as ‘E’ and ‘C’ in the linkage map, respectively. Then, all QTL mapping analysis for the phenotypic traits was performed with the package ‘R/qtl’ (Broman *et al*., 2003; Broman *et al*., 2009) in Rstudio v. 0.98.1027 (Rstudio Team, 2017) using Haley-Knott regression. The significance of all mapped QTLs was

determined at $p<0.05$ probability of error using genome-wide LOD thresholds set using 1000 permutations.

Initially, standard interval mapping (SIM) analysis was used to identify the largest effect QTLs, which were then used as covariates to run composite interval mapping (CIM). Significant QTLs from CIM analyses were submitted to multiple QTL mapping (MQM) analysis to adjust positions, detect flanking and peak markers, determine confidence intervals, quantify individual and cumulative effects on regarding percent phenotypic variation explained, and identify parental origin of all mapped QTLs. Confidence intervals for each QTL were determined using the 95% Bayesian interval and genetic maps of chromosomes with the most significant QTLs and their flanking markers were drawn using MapChart version 2.1 (Voorrips, 2002).

2.2 Results and Discussion

Phenotypic Traits

The joint analysis of variance (Table 2-1) identified significant effects ($p < 0.01$) of genotype for all traits evaluated in this study. Similarly, the interaction of genotype by year (G×Y) was significant for all traits, except for protein content (PRO) in sound kernels. Since the effect of year was significant for all traits, except thousand kernel weight (TKW), initially each year was analyzed separately, and the two-year average was also considered.

Although this population does not segregate for the two major dwarfing genes (*Rht-B1*, *Rht-D1*), variation in plant height (PH) was observed. Therefore, PH was included as a covariate in the analysis variance since shorter plants are more exposed to the ground-based inoculum than taller plants. A significant effect of this covariate was observed for TKW, *Fusarium*-damaged kernel (FDK), and PRO in the joint analysis. A significant effect of these traits, as well as DON,

was associated with the PH in 2016. Coefficients of variation ranged from 4.83 to 20.91%, which can be considered acceptable for field experiments (Table 2-1). All six evaluated traits appeared to be normally distributed in both years (Figure2-1). Although the distribution of the final evaluation of symptomatic spikelets (PSS) and the area under the disease progress curve (AUDPC) was slightly skewed to the left, data transformation was not implemented.

The significant effect of the G×Y interaction shown in Table 2-1, and in the adjusted mean comparison across years presented in Table 2-2, demonstrates that a significantly more severe epidemic was achieved in 2015 than in 2016. Meanwhile, PRO was the only trait that did not show significant differences for the G×Y interaction, indicating that genotypes performed similarly across years for this trait. The broad sense heritability ranged from 0.25 to 0.69 with the lowest values for FDK and the highest for PSS. Graphical dispersion of Person correlations between FHB traits are shown in Figure 2-2 while correlations values for each individual year is presented in Table 2-3. Highly significant correlations with DON accumulation were observed for AUDPC (0.62^{***}), PSS (0.61^{***}) and FDK (0.53^{***}) in the two-year average.

Parents were statistically different from each other for PSS and AUDPC (Table 2-2), these results also confirmed that Everest is significantly more resistant to FHB than WB-Cedar. The average of the population fell close to the mid-parent mean for the majority of the traits, as expected. Transgressive segregation (presence of extreme phenotypes that significantly differ from the parents) was also observed. For instance, KS12DH0296-74 (DH074) and KS12DH0296-143 (DH143) presented significantly higher values of FDK, DON, and PSS, AUDPC indicating that these lines were significantly more susceptible than WB-Cedar (Table 2-2 and Figure 2-5), however, none of the DH lines were significantly more resistant than Everest. In contrast, KS12DH0296-208 (DH208) and KS12DH0296-147 (DH147) presented significantly

higher values of TKW and PRO than both parents, respectively (Table 2-2). These results suggest these both parents contain genes that influenced susceptibility and resistance.

The Genetic Map of the Everest/Cedar Population

After filtering the genotypic data and grouping makers within linkage groups based on marker distance and its segregation, our final linkage map contained 2,839 SNP markers spanning 3,040 cM and covering all the wheat chromosomes (Figure 2-3). The average marker spacing was 1.08 cM with the largest gap (43.7cM) being found chromosome 7D (Table 2-4). Strong linkage disequilibrium (LD) between markers within, but not across chromosomes was observed, indicating that the genetic map was well constructed (Figure 2-4).

The genetic map was compared against PopSeq positions (Chapman *et al*., 2015) to order markers from short to long arm within each chromosome and recalculate marker positions when more than one linkage group was assigned per chromosome. The majority of SNP markers were located on the A (42.9%) and B (47.1%) genomes, while only 10% of the polymorphism was detected on the D genome (Table 2-4). The low diversity on the D genome has been repeatedly reported in the literature and is mainly due to the fact that a limited number of ancestral genotypes of the D genome donor (*Aegilops tauschii*) contributed to the origin of hexaploid wheat (Wang *et al*., 2013; IWGSC, 2014; Jordan *et al*., 2015). This effect was obvious in this population as only 3 and 28 loci were identified as segregating between Everest and WB-Cedar on chromosomes 4D and 5D, respectively (Table 2-4). This also highlights the drawbacks of biparental mapping, since only loci segregating between the parents are taken into account.

Results from the SSR markers linked with the large- effect QTL *Fhb4* and *Fhb5* were monomorphic, indicating that these QTLs are not present in the current population. For *Fhb2*, the resistant allele consists of a 120-123 base pair (bp) product produced by the marker *gwm133*

(Cuthbert *et al*., 2007) and Sumai#3 was used as standard positive control. Everest and a few DH-lines yielded a 127 bp signal that was initially speculated to be associated with the presence of *Fhb2*, since this band size was just slightly larger than the positive control, Sumai#3. However, when data from this marker was included in the linkage map, it did not group within any linkage group. Calling the size of SSR markers is often challenging, and in these case, the appearance of an allele associated with *Fhb2* was the result of erroneous SSR peak calls. Moreover, based on pedigree, it is unlikely that any of the three known-QTLs are segregating within the DH population Everest/WB-Cedar.

QTL Mapping of FHB Resistance Components

Standard interval mapping (SIM) followed by composite interval mapping (CIM) (Figure 2-6) analyses were performed for all traits. Loci that were significant on prior analyses were then re-analyzed with a multiple QTL mapping analysis (MQM) to identify possible interactions and their combined effect on phenotypes. The MQM analysis releveled that all mapped loci have additive effects while no significant QTL×QTL interactions were identified. The predominance of additive effects is in agreement with the literature (Buerstmayr *et al*., 2009; Islam *et al*., 2016) and indicates that significant progress for FHB resistance can be made by stacking multiple QTLs within breeding lines. A total of 11 QTLs were mapped in this study, from which eight were associated with two or more resistance traits and/or both years of the experiment (Table 2-5). Everest was the donor for five of the eight reproducible loci. No significant QTLs were found for TKW and DON in the 2016 growing season which is likely due to the lower levels of disease infections observed in these second year of the experiment.

The last evaluation of PSS and the AUDPC measures the disease evolution within spikes over time after the initial infection, therefore these traits represent type II resistance. Three QTLs

(located on 1AS, 3BS, 7AL) were associated with disease spread within spikes were found in 2015 and together explained 31.76% of the phenotypic variation of PSS (Table 2-5). The allele associated with resistance on 3BS was from Everest and explained 15.32% of the PSS variation. This QTL also was remapped within the same interval for AUDPC15, AUDPC16, AUDPC15&16, FDK15 (type IV/RKI), DON15&16 (type III/RTA), and TKW15&16 (Figure 2-7). The association with TKW is likely an effect cofounded with FDK since lines with less damaged kernels tend to have higher TKW. Although, *Fhb1* is located on this same interval of 3BS (Cuthbert *et al*., 2006; Liu *et al*., 2009; Rawat *et al*., 2016), based on results from prior genotyping (unpublished data) neither parents carry resistant alleles at the *Fhb1* locus. Likewise, Islam *et al*. (2016) also reported a QTL in the same region of 3BS associated with FHB severity (Type II) in the cultivar Truman, a soft red winter wheat from the University of Missouri which is well-known for its native resistance. Everest contains several soft red winter wheat parents from the eastern US in its pedigree making it likely that this resistant locus in Everest has the same origin as the Truman QTL. The position of this QTL in relation to *Fhb1* will make pyramiding the two QTLs together more challenging since the will be likely linked in repulsion phase.

Two additional QTLs from Everest, located on 5DS and 6BS, were repeatedly mapped within the same confidence interval and were associated with significantly lower values for AUDPC15, AUDPC15&16, DON15, and DON15&16. The 5DS QTL explained from 4.15 to 11.65% of the phenotypic variation for these traits (Table 2-5). The QTL on 6BS mapped close to the interval of *Fhb2* (9 – 15cM) reported by Cuthbert *et al*. (2007). However, based on the results from the SSR marker associated with *Fhb2*, it is not present in our population. Thus the

genomic regions on 5DS and 6BS found in this study being associated with FHB resistance are likely novel native QTL present in North American winter wheat.

Another QTL on 5A from Everest was associated with increased TKW in 2015 and lower values for FDK15, FDK15&16, DON15, and DON15&16 (Figure 2-8). This QTL explained from 5.82 to 15.36% of the phenotypic variation for these traits, with the largest effect on reduced DON accumulation in 2015 and the average of DON over both years combined (Table 2-5). In a recent study, Cai *et al*. (2016a) found several minor-effect QTL from the Chinese landrace Haiyanzhong contributing to FHB resistance. The QTL with the largest effect in that study was mapped to 5A in an interval similar to the one identified in the present study. Everest and Haiyanzhong have little to no relatedness, by pedigree, making it less likely they would have a common QTL. These two cultivars have no pedigree connection in the history of modern breeding which makes the presence of a common QTL very unlikely.

The QTL on 1B from Everest was significantly associated with TKW15, TKW15&16, FDK15, FDK15&16, and DON15 (Table 2-5). It mapped within the interval 55.59 – 99.96 cM explaining from 3.50 to 6.96% of the phenotypic variation. In similar studies, Eckard *et al*. (2015) and Islam *et al*. (2016) reported a QTL on 1B associated with disease spread 21 days after the initial infection. These authors also verified that the QTL was present in the hard red winter cultivars Overland and Lyman, which increases the likelihood it is the same QTL that we detected on Everest in the current study.

Only three QTLs from WB-Cedar were associated with FHB resistance. The first one was located on the short arm of 1A and was significantly associated with lower values of PSS15, AUDPC15, FDK15, FDK15&16, and DON15&16 (Table 2-5). Other studies have also identified QTL on 1A associated with FHB resistance in the cultivars Lyman and Overland (Eckard *et al*.,

2015), Bess (Petersen, 2015), and Massey (Liu *et al*., 2013). All these cultivars were developed in the US and are well-known for having native resistance to FHB. Although the 1RS-1AL translocation is located on this genomic region, which is positively associated with FHB resistance (Costa *et al*., 2010), neither Everest nor WB-Cedar carries this alien segment.

Another locus from WB-Cedar was mapped on the long arm of 5B within the interval 113.51 – 180.29 cM (Table 2-5). The presence of this genomic region was associated with lower values of PSS and AUDPC in the average of both years of the experiment. Likewise, a QTL from WB-Cedar associated with PSS15, PSS15&16, AUDPC15, AUDPC15&16, and FDK16 was found on 7AL, explaining from 5.15 to 13.29% of the phenotypic variation (Table 2-5). This association with multiple traits in both years increases confidence in the validity of this QTL (Figure 2-9). Several QTLs associated with FHB resistance were identified on this genomic region of 7AL in the cultivars Wangshuibai and Spark (Liu *et al*., 2009), Huangfangzhu (Li *et al*., 2012) and NK93604 (Semagn *et al*., 2007), yet there is not pedigree relatedness between these materials and the Everest/Cedar population.

This study confirmed that FHB resistance is a complex trait with multiple loci contributing to resistance. Although resistance can be classified into different types, phenotypic traits are often highly correlated (Figure 2-2), leading to co-localization of QTLs. For example, QTLs for PSS and AUDPC frequently mapped to the same confidence interval. We also observed that the sum of all individual loci seldom explained more 40% of the phenotypic variation for the evaluated traits. The remaining non-explained variation is most likely due to experimental errors and other small-effect loci that the QTL mapping analysis does not have enough statistical power to capture. In addition, the small proportion of variation explained by mapped QTLs can yet be overestimated, as observed by Arruda *et al*. (2016).

Effect of QTL Combinations

A ‘genotype group’ refers to a set of alleles from different loci that are inherited together from a given parent and expressed in the progeny, thus individuals carrying most of the favorable alleles are expected to be significantly more resistant. In our study, DH lines were classified based on the parental allele present at the peak marker of all QTLs that were significantly associated with FHB resistance in the two-year average of the experiment (Table 2-5). Therefore, each QTL combination was considered as one genotype group and, since no more than one QTL was found per chromosome, loci were named after the chromosome on which they were found.

Three QTLs from Everest located on 3BS, 5DS, and 6BS, along with another two from WB-Cedar, located on 5BL and 7AL, were significantly associated with initial disease infection (last evaluation of PSS) in the average of both years. A total of 32 genotype groups were found in the population from these five loci, varying from 10 lines without any favorable allele to two lines (DH014, DH037) carrying all favorable alleles from both parents. Lines carrying all QTLs presented significantly lower PSS in 2015, 2016, and the two-year average than those lines without any of the mapped loci (Table 2-6). These results indicate the viability of stacking multiple QTLs together to increase Type II resistance of FHB.

AUDPC is estimated from multiple PSS evaluations over time, therefore it measures how fast the disease spreads after the initial infection and it is associated with type II resistance. Four loci located on 3BS, 5DS, 6BS, and 7AL mapped within the same confidence interval that was previously associated with PSS. In addition, another QTL on 1AL from WB-Cedar was also associated with lower AUDPC values. The contrast analysis for QTL combinations was statistically significant for the year 2015 and the across-experiment analysis, but not in 2016 (Table 2-7). Only one line (DH186) had all resistant alleles at the five QTLs associated with

AUDPC, presenting significantly lower values for this trait in 2015 and in the average of both years indicating that FHB spread slowly on this DH line.

Three QTLs explained 26.27% of FDK variation in the average of two years of the experiment (Table 2-5). A QTL on 1AL from WB-Cedar mapped in the same interval as the QTL associated with AUDPC suggesting that this genomic region is simultaneously associated with type II and type IV resistance. Another two QTLs from Everest, located on 1B and 5A, were also associated with low values of FDK. The genotype analysis identified 19 DH lines having all resistant alleles from both parents at these three loci. These lines had significantly lower values of FDK in 2015 and in the average of both years (Table 2-8). DH lines without any resistant allele at these QTLs presented 68.1% of FDK in the two-year average whereas 19 lines containing all resistant alleles averaged 51.9% which was statistically different from each other.

Four QTLs associated with low DON accumulation were mapped on 1A, 3B, 5A, 5D chromosomes. Together these loci explained 35.42% of the DON variation in the two-year average (Table 2-5). The QTL on 1A was the same one from WB-Cedar previously associated with AUDPC and FDK, while the loci on 3B and 5A from Everest were also mapped for PSS, AUDPC, and FDK, suggesting that these genomic regions affect FHB resistance through different physiological mechanisms. In contrast, the QTL on 5DS from Everest was exclusively associated with low DON values. The genotype grouping analysis identified five lines containing all resistant alleles at these four loci. These lines presented significantly lower values of DON in 2015 and in the average of both years, but not in 2016 (Table 2-9). Lines without any resistant allele at these QTLs averaged 6.5 ppm DON over the two years of the experiment while lines containing all resistant alleles averaged 4.37 ppm, representing a reduction of 67%.

Two QTLs from Everest were found on 1B and 3B explaining 15.62% of phenotypic variation for TKW (Table 2-4). However other QTL associated with PSS, AUDPC, FDK, DON mapped to these regions. Therefore, the effect on TKW is likely cofounded, since more susceptible lines tend to have lighter kernels than resistant lines due to increase damage from FHB. The group of lines containing favorable alleles at these two loci showed statistically higher values for TKW (Table 2-10). In contrast, the protein content could be estimated exclusively on sound kernels using the sorting features of the SKNIR system, avoiding the occurrence of confounding effects. As a result, four QTLs were significantly associated with PRO were found. Only one (5B) mapped in the same confidence interval as FHB resistant traits. The other loci were located on 2AL, 2BL and 7BL, and explained 42.3% of the grain protein variation. Only nine lines carried the favorable alleles for each of the QTLs and had significantly higher values for PRO than lines with zero or one favorable QTL (Table 2-11).

In summary, the presence of multiple QTLs within breeding lines significantly influenced the values of all phenotypic traits, confirming that additive effects are predominant in the genetic control of FHB resistance. Therefore, our results are in agreement with the majority of studies reported in the literature (Buerstmayr *et al*., 2009; Cai, 2016b). Substantial progress in FHB resistance can be achieved by having multiple significant QTLs stacked within breeding lines, as was also observed by Clinesmith (2016), suggesting that even when resistance is controlled by several minor-effect QTLs, pyramiding them together improved FHB resistance. Field notes of PSS and AUDPC relied on visual notes which led to larger error variance (Table 2-1) and generated only two or three groupings in the contrast analysis (Tables 2-6 and 2-7) whereas traits estimated by the SKNIR system presented smaller error variance and more statistically different

groups in the genotype group analysis. It highlights the need for automated phenotyping tools that would be able to distinguish minor phenotypic in further studies.

2.3 Conclusions and Future Prospects

Five QTLs from Everest were identified on 1BS, 3BS, 5AS, 5DS, and 6BS being repeatedly associated with FHB components indicating that this cultivar is a source of native FHB resistance with multiple mechanisms of resistance. Likewise, another three QTLs from WB-Cedar were discovered on 1AS, 5BL, and 7AL chromosomes. Transgressive segregation for resistance was not observed which indicates that these parents likely share genes associated with FHB resistance. Moreover, these eight mapped QTLs that were segregating in the progeny were not capable of generating any individual statistically more resistant than Everest.

The genotype grouping analysis confirmed the predominance of additive effects controlling FHB resistance traits. Although the percentage of phenotypic variation explained by small-effect QTLs were occasionally modest (less than 5%) when these loci were found in combinations within the population the level of resistance was improved. In general, DH lines containing all resistant alleles were significantly more resistant than lines with one or no favorable alleles at peak loci. It was verified an 67% reduction in DON content in those lines with all mapped QTLs compared to the ones without any QTL for DON. DH lines containing multiple QTL combinations were selected and included in the crossing block of the K-State Hard Red Winter Wheat Breeding Program.

The QTL on 3BS mapped in the same interval as *Fhb1* whereas the QTL on 6BS overlaps the interval of *Fhb2*, although results from tightly linked markers confirmed that they are likely not the same QTL. It also suggests that they could be linked in repulsion phase. Therefore,

combining these two QTL from Everest with the these well-known QTLs will require extra-large progenies and MAS to be able to identify crossovers events between these pairwise loci.

Several QTL were found in genomic regions where other loci associated with native FHB resistance were mapped in prior studies. It suggests that U.S. winter wheat cultivars likely share multiple genes of native FHB resistance indicating that limit genetic progress can be made by only breeding within this gene pool. Consequently, moving beyond the current levels of resistance will require pyramiding large-effect QTLs from exotic and alien sources with these minor-effect QTLs of native resistance. There are U.S. springs wheats that already carry such QTL and could be used as novel sources of FHB resistance, eliminating the need of crossing with unadapted sources that often carry undesirable agronomic traits.

These novel sources can be backcrossed into an elite winter wheat background such as Everest or its derivatives aiming to generate large progenies which could be then selected through MAS (since reliable markers are known) in early BC generations to increase the frequency of desirable alleles. After that, phenotypic selections could be performed under FHB inoculated conditions. Other breeding techniques may also be implemented such doubled haploids or single seed descent for fast generation advancement and whole-genome genotyping analysis to assist selection of recurrent genetic background, predict FHB resistance within and across families, as well as, identifying superior cross combinations among half-sib lines.

References

Arruda M. P., Brown P., Brown-Guedira G.B., Krill A.M., Thurber C., Merrill K.R., Foresman B.J., Kolb F.L. (2016). Genome-Wide Association Mapping of Fusarium Head Blight Resistance in Wheat using Genotyping-by-Sequencing. The Plant Genome 9(1): 1-14.

Bockus W.W., Appel J.A., De Wolf E.D., Todd T.C., Davis M.A., Fritz A.K. (2015). Impact of wheat cultivar Everest on yield loss in Kansas from Fusarium head blight during 2015. Proceedings of the 2015 National Fusarium Head Blight Forum. St Louis MO.

Broman K. W., Wu H., Sen S., & Churchill G. A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics 19(7): 889-890.

Broman K., & Sen S. (2009). A Guide to QTL Mapping with R/qtl: Springer.

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

Cai J., Wang S., Li T., Zhang G. Bai G. (2016a). Multiple Minor QTL Are Responsible for Fusarium Head Blight Resistance in Chinese Wheat Landrace Haiyanzhong. PLoS One 11(9): e0163292.

Cai, J. (2016b). Meta-analysis of QTL for Fusarium head blight resistance in Chinese wheat landraces using genotyping by sequencing. Ph.D. Dissertation, Kansas State University.

Chapman J. A., Mascher M., Buluç A., Barry K., Georganas E., Session A., Rokhsar D. S. (2015). A whole- genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biology 16(26): 1-17.

Clinesmith M., (2016). Genetic mapping of QTL for Fusarium head blight resistance in winter wheat cultivars Art and Everest. Master Thesis. Kansas State University, Manhattan Kansas.

Costa J.M., Bockelman H.E., Brown-Guedira G., Cambron S.E., Chen X., Cooper A. Cowger C., Dong Y., Grybauskas A., Jin Y., Kolmer J., Murphy J.P., Sneller C., Souza E. (2010). Registration of the Soft Red Winter Wheat Germplasm MD01W233–06–1 Resistant to Fusarium Head Blight. Journal of Plant Registrations 4(3): 255-260.

Cuthbert P. A., Somers D. J., Brulé-Babel A., (2007). Mapping of *Fhb2* on chromosome 6BS: a gene controlling Fusarium head blight field resistance in bread wheat (*Triticum aestivum* L.). Theor Appl Genet. 114(3): 429-437.

De Wolf, E. D., Bockus, W. W., & Whitworth, R. J. (2015). Wheat variety disease and insect rating 2015. Manhattan, KS: Kansas State University.

Dowell F. E., Maghirang E. B., & Baenziger P. S. (2009). Automated single-kernel sorting to select for quality traits in wheat breeding lines. Cereal Chem. 86(5): 527-533.

Dowell F. E., Maghirang E. B., Graybosch R. A., Baenziger P. S., Baltensperger D. D., & Hansen L. E. (2006). An automated near-infrared system for selecting individual kernels based on specific quality characteristics. Cereal Chem. 83(5): 537-543.

Dowell F. E., Ram M. S., & Seitz L. M. (1999). Predicting scab, vomitoxin, and ergosterol in single wheat kernels using near-infrared spectroscopy. Cereal Chem. 76(4): 573-576.

Dweba C.C., Figlan S., Shimelis H.A., Motaung T.E., Sydenham S., Mwadzingeni L., Tsilo T. J. (2017). Fusarium head blight of wheat: Pathogenesis and control strategies. Crop Protection 91: 114-122.

Eckard J., Gonzalez-Hernandez J., Caffe M., Berzonsky W., BockusW., Marais, G., & Baenziger P. (2015). Native Fusarium head blight resistance from winter wheat cultivars 'Lyman,' 'Overland,' 'Ernie,' and 'Freedom' mapped and pyramided onto 'Wesley'- *Fhb1* backgrounds. Mol Breeding 35:6.

Holland J.B., Nyquist W.E., & Cervantes-Martinez C.T. (2002). Estimating and interpreting heritability for plant breeding: An update. Plant Breed Rev. 22: 9-112.

Islam S., Brown-Guedira G., Van Sanford D., Ohm H., Dong, Y., Mckendry A.L. (2016). Novel QTL associated with the Fusarium head blight resistance in Truman soft red winter wheat. Euphytica 207(3): 571-592.

IWGSC - The International Wheat Genome Sequencing Consortium. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum* L.) genome. Science 345(6194): 1251788.

Jin F., Zhang D., Bockus W., Baenziger P.S., Carver B., Bai G. (2013). Fusarium Head Blight Resistance in U.S. Wheat Cultivars and Elite Breeding Lines. Crop Science 53: 2006-2013.

Jordan W. K., Wang S., Lun Y., Gardiner L-J., MacLachlan R., Hucl P., Wiebe K., Wong D., Forrest K.L., Sharpe A.G., Sidebottom C.H.D., Hall N., Toomajian C., Close T., Dubcovsky J., Akhunova A., Talbert L., Bansal U.K., Bariana H.S., Hayden M.J. Pozniak C., Jeddeloh J., Hall A., Akhunov E. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biology 16(48): 1-18.

Kansas Ag Statistics – USDA. (2017). Kansas Wheat Varieties. Online publication available at: https://www.nass.usda.gov/Statistics_by_State/Kansas/Publications/Crops/Wheat_Varieities/KS_whtvar17.pdf.

Li T., Bai G., Wu S., Gu S. (2012). Quantitative trait loci for resistance to Fusarium head blight in the Chinese wheat landrace Huangfangzhu. Euphytica 185(1): 93-102.

Liu S., Griffey C.A., Hall M D., McKendry A. L., Chen J., Brooks W. S. Brown-Guedira G., Van Sanford D., Schmale D.G. (2013). Molecular characterization of field resistance to

Fusarium head blight in two US soft red winter wheat cultivars. Theor Appl Genet. 126(10): 2485-2498.

Liu S., Hall M. D., Griffey C. A., McKendry A. L. (2009). Meta-analysis of QTL associated with Fusarium Head Blight in Wheat. Crop Science 49:1955-1968.

Madden L.V., Hughes G., Van den Bosch, F. (2007). The Study of Plant Disease Epidemics. The American Phytopathological Society, APS Press St. Paul, Minnesota.

Mesterhazy A. (1995). Type and components of resistance to Fusarium head blight of wheat. Plant Breed. 114(5): 377-386.

Nganje W.E., Kaitibie S., Wilson W.W., Leistritz F.L. & Bangsund D.A. (2004). Economic impacts of Fusarium head blight in wheat and barley: 1993–2001 (Agribusiness and Applied Economics Report No. 538) (North Dakota State University, 2004).

NPGS – U.S. National Plant Germplasm System. (2017). Available online at: https://npgsweb.ars-grin.gov/gringlobal/search.aspx.

Peiris K. H. S., Pumphrey M. O., Dong Y., Maghirang E. B., Berzonsky W., Dowell F.E. (2010). Near-Infrared Spectroscopic Method for Identification of Fusarium Head Blight Damage and Prediction of Deoxynivalenol in Single Wheat Kernels. Cereal Chem. 87(6): 511-517.

Petersen, S. (2015). Advancing marker-assisted selection for resistance to powdery mildew and Fusarium head blight in wheat (Doctor of Philosophy). Raleigh, North Carolina 2015.

Poland J.A., Brown P.J., Sorrells M.E., Jannink J.L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PLoS One 7(2): e32253.

Rawat N., Pumphrey M.O., Liu S., Zhang X., Tiwari V.K., Ando K., Trick H.N., Bockus W.W., Akhunov E., Anderson J.A., Gill B.S. (2016). Wheat *Fhb1* encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. Nature Genetics 48(12): 1576-1580.

RStudio Team (2017). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, URL http://www.rstudio.com/.

SAS Institute Inc. (2011) Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc.

Semagn K., Skinnes H., Bjørnstad Å., Marøy A.G., Tarkegne Y. (2007). Quantitative Trait Loci Controlling Fusarium Head Blight Resistance and Low Deoxynivalenol Content in Hexaploid Wheat Population from Arina and NK93604. Crop Science 47: 294-303.

Sneller C., Guttieri M., Paul P., Costa J., Jackwood R. (2012). Variation for Resistance to Kernel Infection and Toxin Accumulation in Winter Wheat Infected with *Fusarium graminearum*. Phytopathology 102(3): 306-314.

Tukey J. W. (1949). Comparing individual means in the analysis of variance. Biometrics 5(2): 99-114.

Van Ooijen, J. W. (2006). JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen.

Voorrips R.E. (2002) MapChart. Software for the graphical presentation of linkage maps and QTL. J Hered. 93: 77−78. Available at http://www.joinmap.nl.

Wang J., Luo M-C., Chen Z., You F.M., Wei Y., Zheng Y., Dvorak J. (2013). *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytologist 198(3): 925-937.

Xue S., Li G., Jia, H., Xu, F., Lin, F., Tang, M., Wang Y., An X., Xu H., Zhang L., Kong Z., Ma Z. (2010). Fine mapping *Fhb4*, a major QTL conditioning resistance to Fusarium infection in bread wheat (*Triticum aestivum* L.). Theor Appl Genet.121(1):147-156.

Xue S., u, F., Tang, M., Zhou Y., Li G., An X., Lin F., Xu H., Jia H., Zhang L., Kong Z., Ma Z. (2011). Precise mapping *Fhb5*, a major QTL conditioning resistance to Fusarium infection in bread wheat (*Triticum aestivum* L.). Theor Appl Genet. 123(6): 1055-1063.

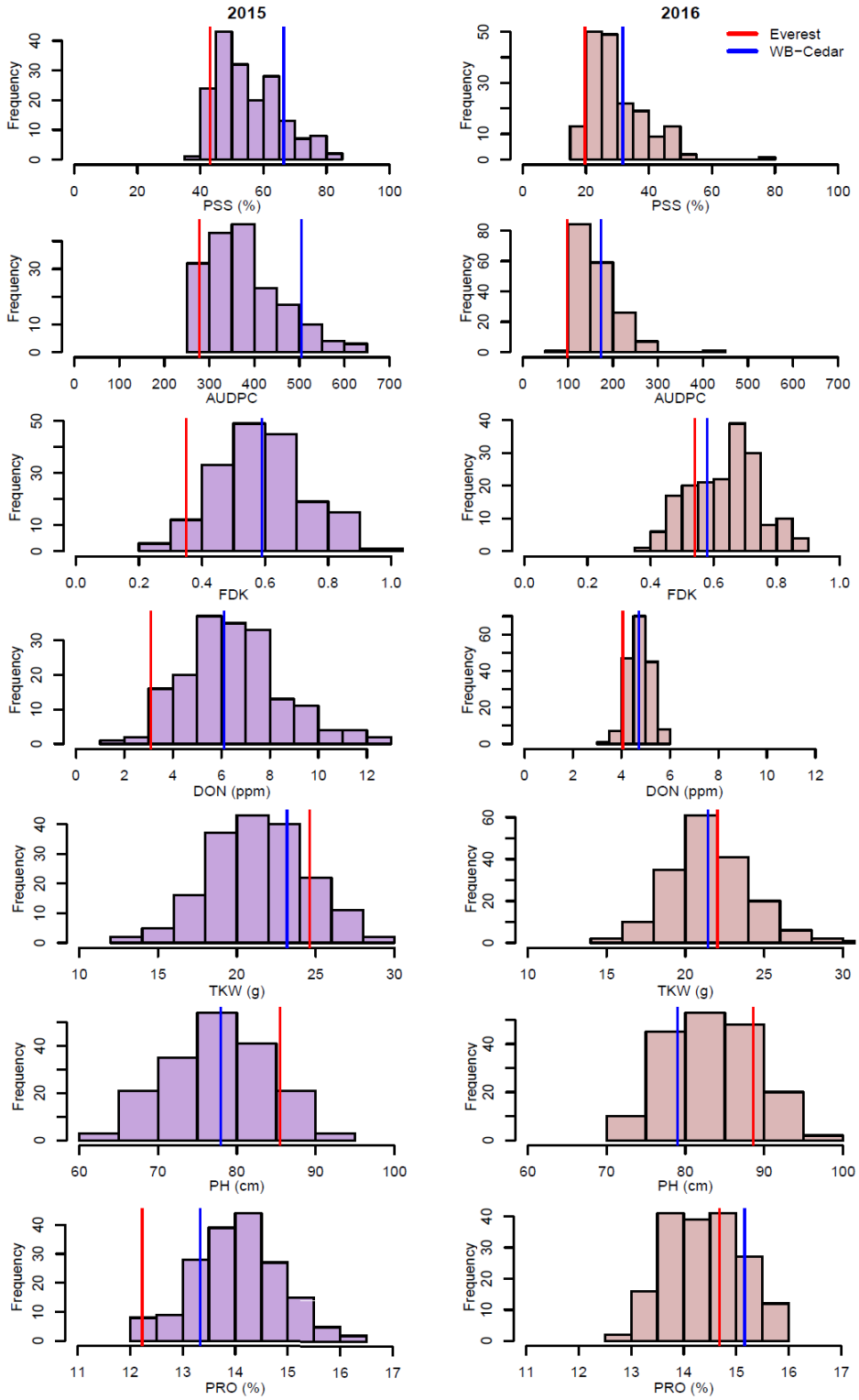

Figure 2-1. Normal distribution of all phenotypic traits evaluated in 2015 and 2016 growing seasons in the DH population Everest/Cedar. Red and blue vertical lines represent parent means.

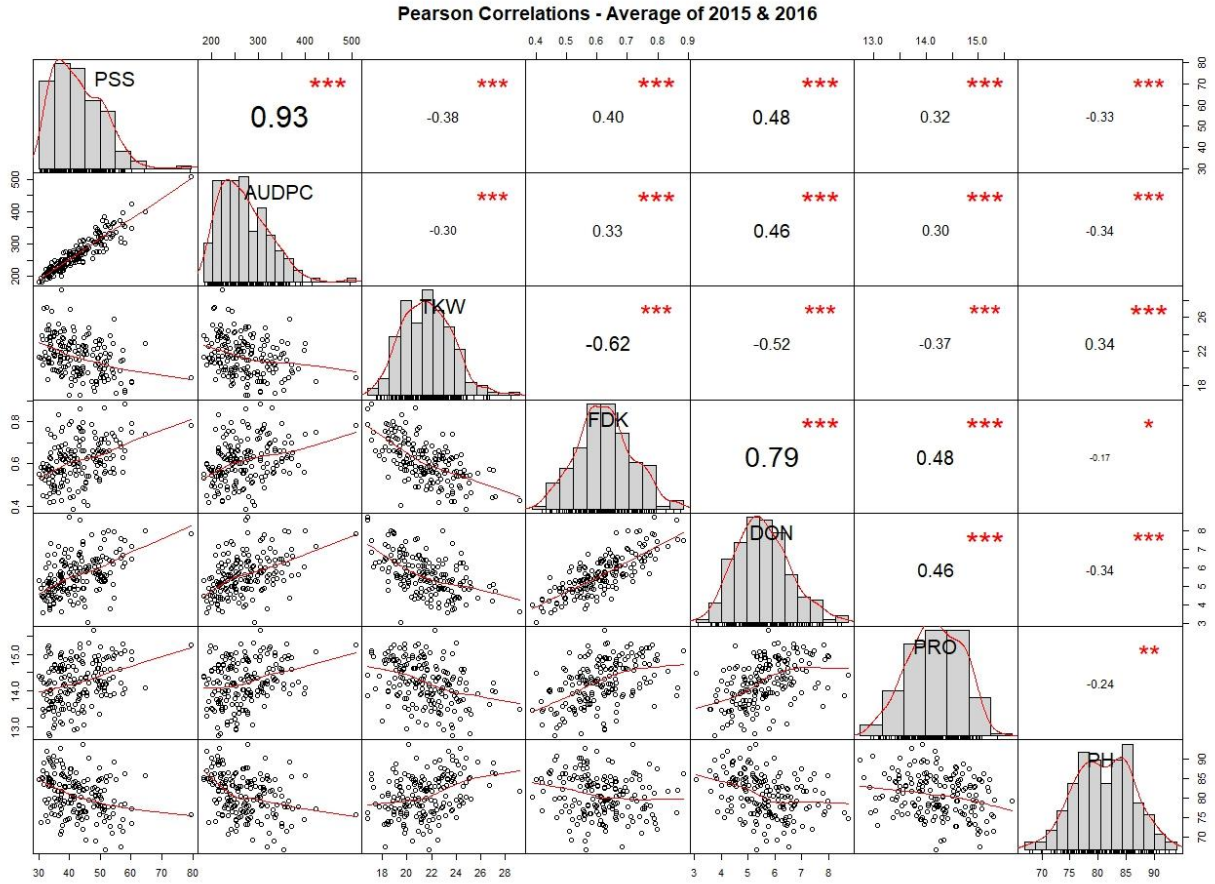

Figure 2-2. Graphical dispersion of Pearson correlations between phenotypic traits of the two-year average of the experiment and plotted using the R package 'performanceanalytics'.

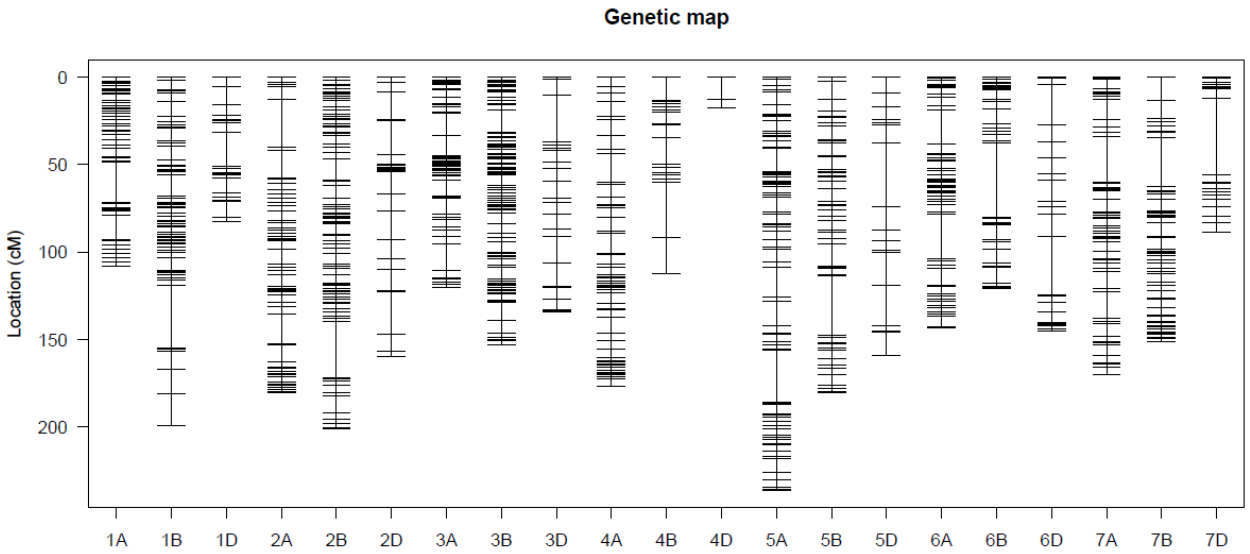

Figure 2-3. Genetic map of the Everest/Cedar DH population with 2839 GBS SNP markers distributed across the wheat genome.

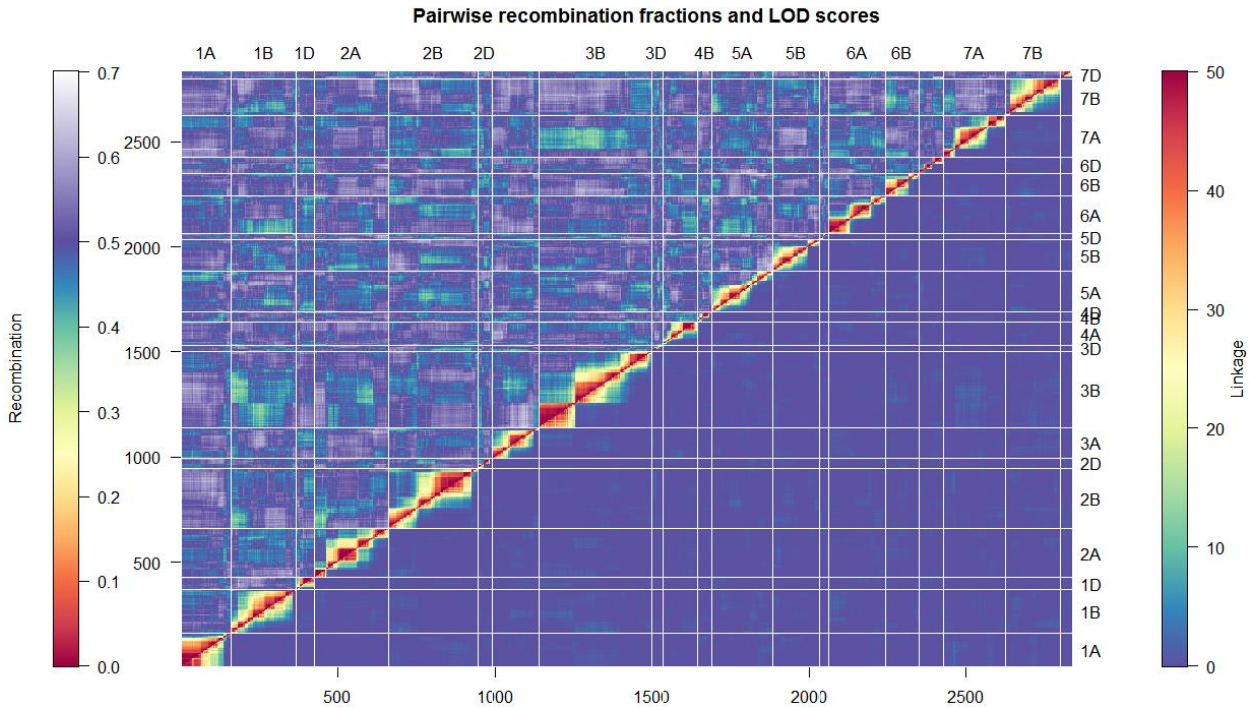

Figure 2-4. Heat map of pairwise recombination fractions between markers and their pairwise LOD scores drawn with the R package ASmap.

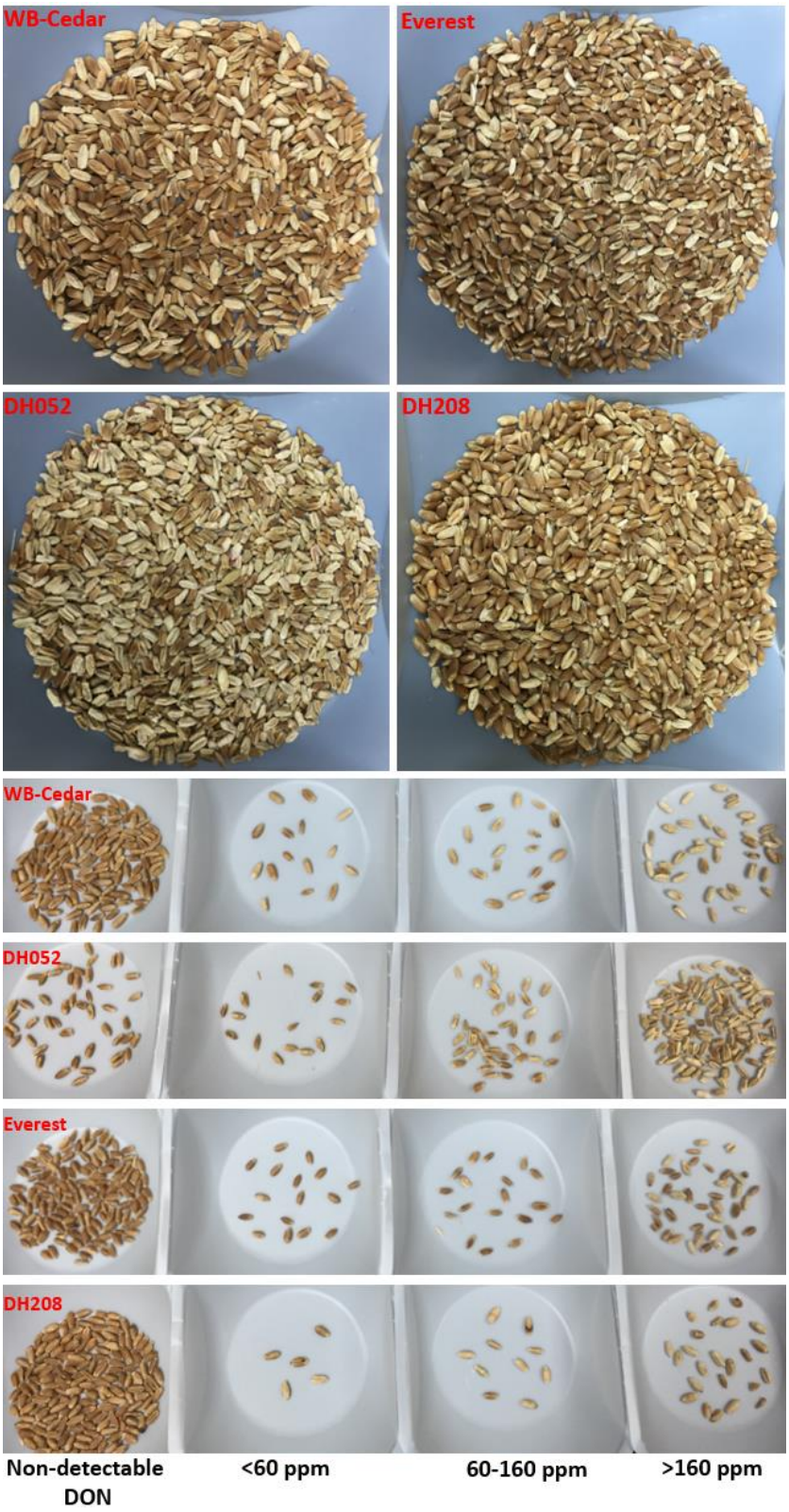

Figure 2-5. Visual aspects of samples evaluated by the SKNIR system. After evaluation samples are sorted in four bins based on the estimated DON content.

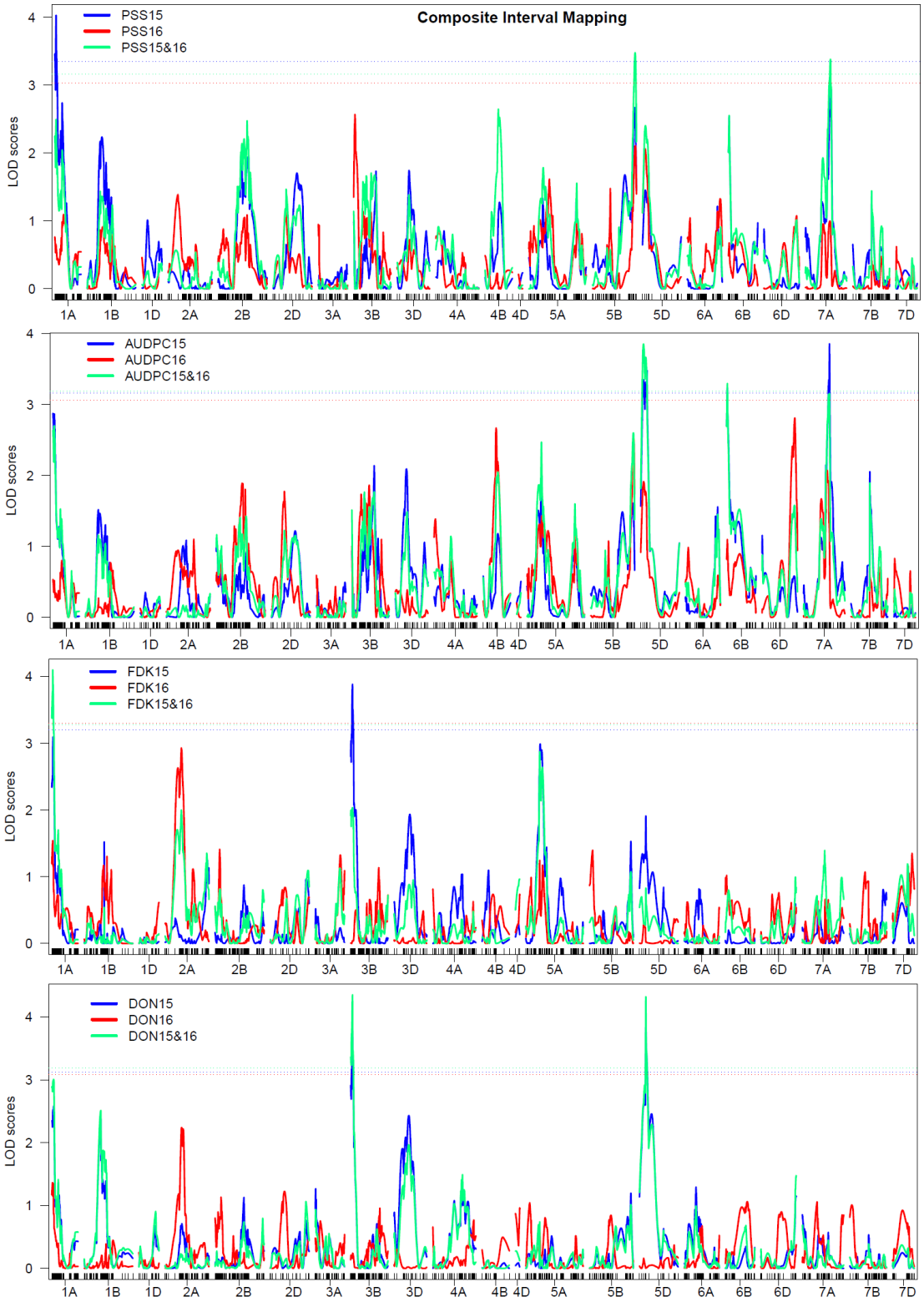

Figure 2-6. Genome-wide LOD scores from the Composite Interval Mapping analysis for six FHB components in the growing seasons of 2015, 2016, and the two-year average. Dotted horizontal lines represent p<0.05 probability of error using genome-wide LOD thresholds set using 1000 permutations for each trait.

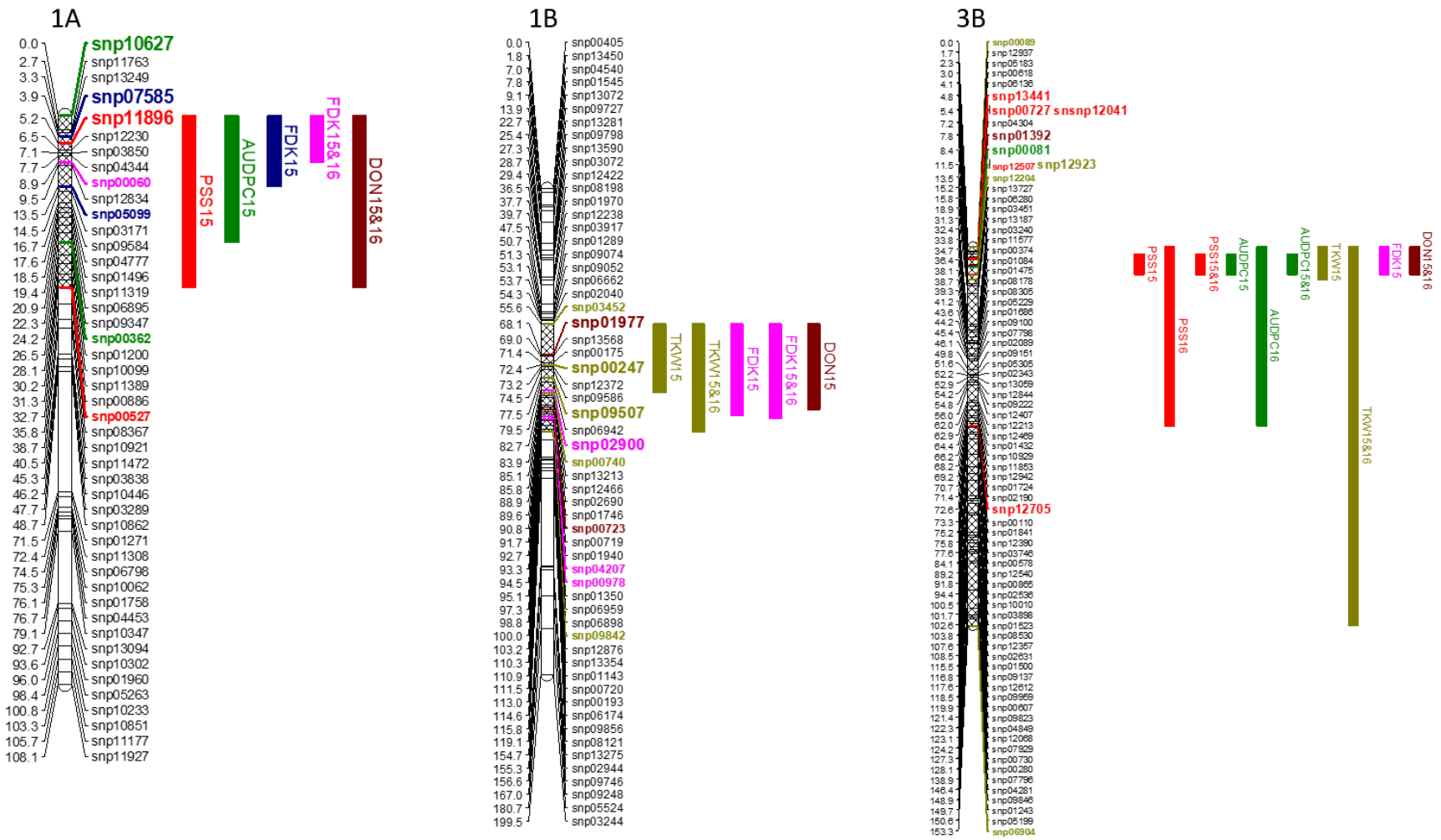

Figure 2-7. Chromosomal positions (in cM) of QTLs associated with percentage of symptomatic spikelets (PSS), area under de disease progress curve (AUDPC), thousand kernel weight (TKW), *Fusarium*-damaged kernels (FDK), deoxynivalenol accumulation (DON), and protein content (PRO). Bars to the right of the chromosome represent the 95% Bayesian interval of QTL with a different color for each trait. SNP marker flanking QTL are color coded in respective to each trait with larger font size for the marker at peak loci.

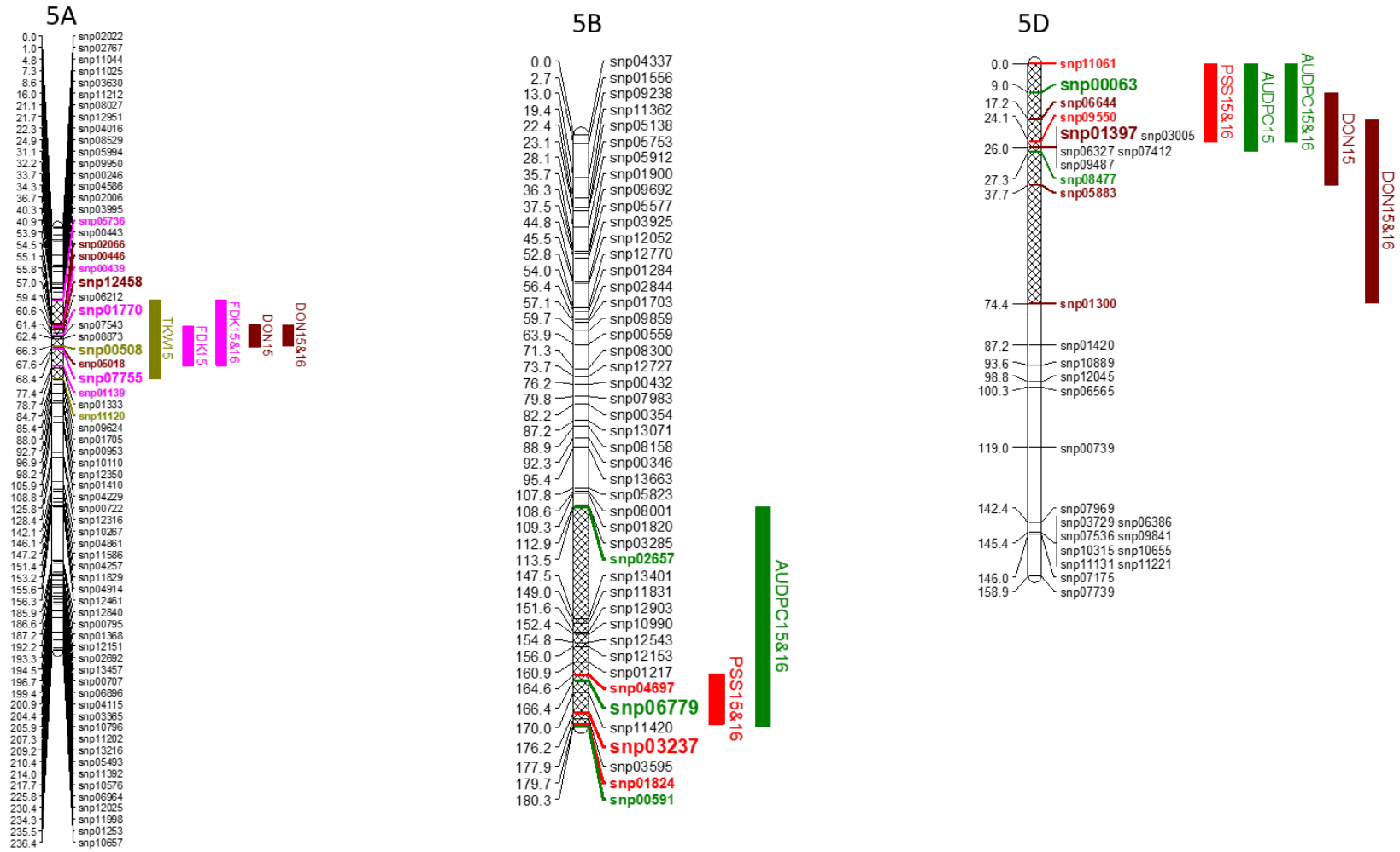

Figure 2-8. Chromosomal positions (in cM) of QTLs associated with percentage of symptomatic spikelets (PSS), area under the disease progress curve (AUDPC), thousand kernel weight (TKW), *Fusarium*-damaged kernels (FDK), deoxynivalenol accumulation (DON), and protein content (PRO). Bars to the right of the chromosome represent the 95% Bayesian interval of QTL with a different color for each trait. SNP marker flanking QTL are color coded in respective to each trait with larger font size for the marker at peak loci.

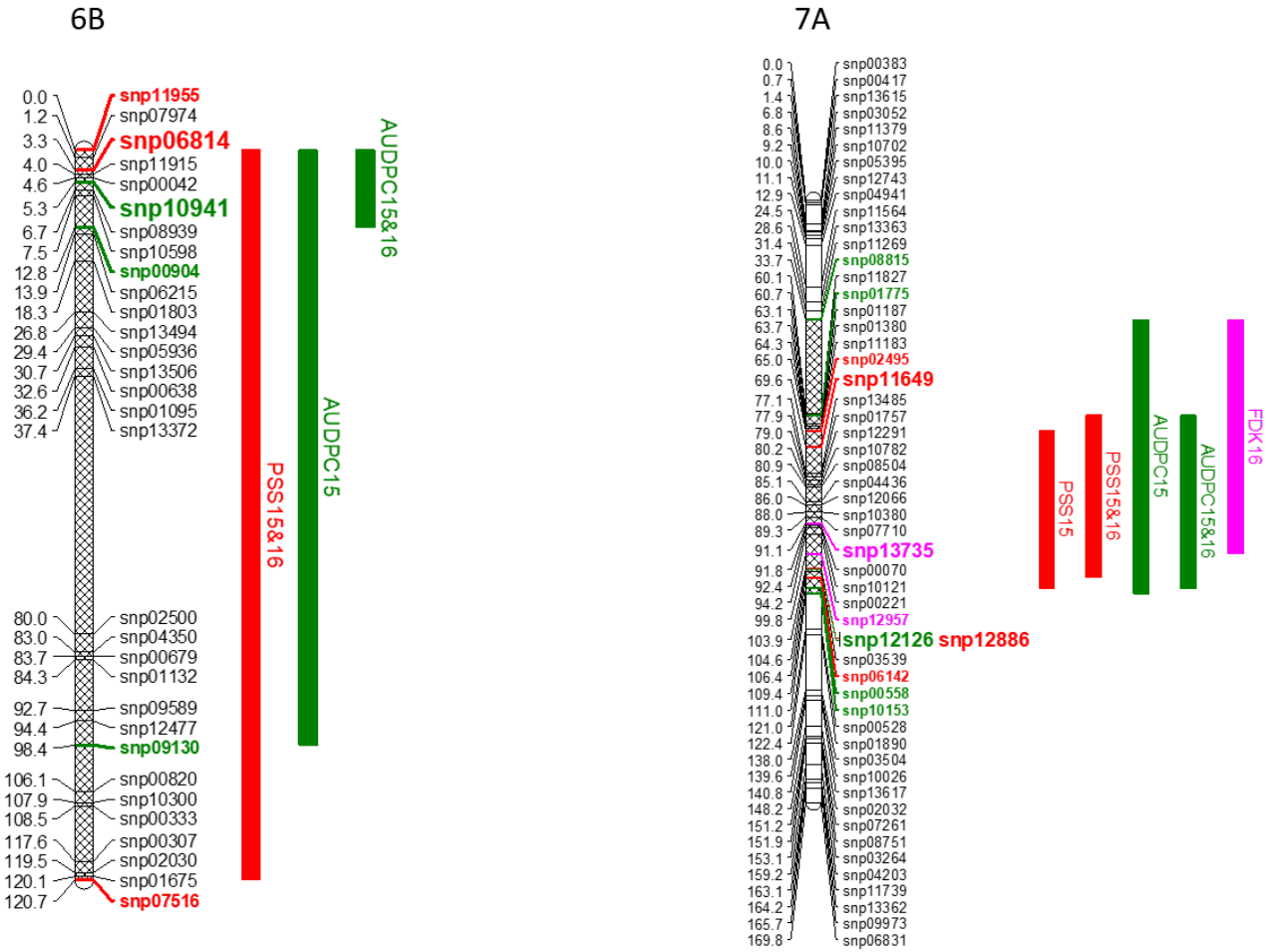

Figure 2-9. Chromosomal positions (in cM) of QTLs associated with percentage of symptomatic spikelets (PSS), area under the disease progress curve (AUDPC), thousand kernel weight (TKW), *Fusarium*-damaged kernels (FDK), deoxynivalenol accumulation (DON), and protein content (PRO). Bars to the right of the chromosome represent the 95% Bayesian interval of QTL with a different color for each trait. SNP marker flanking QTL are color coded in respective to each trait with larger font size for the marker at peak loci.

Table 2-1. Means squares from the joint and individual analysis of variance of a DH population and parents conducted in the field during the growing seasons of 2015 and 2016.

Year	Source of variation	d.f. †	PSS	AUDPC	TKW	FDK	DON	PRO
2015	Genotype	177	187.97**	12001.99**	16.37**	0.039**	7.40**	1.78**
	Block	1	256.19*	1341.46^{ns}	0.08^{ns}	0.015^{ns}	14.29**	3.75^{ns}
	Covariate PH	1	35.56^{ns}	2.78.39^{ns}	0.42^{ns}	0.001^{ns}	0.19^{ns}	0.08^{ns}
	Error	161	56.78	2126.6	3.14	0.097	1.81	0.42
	Mean	-	55.1	376.1	21.58	0.58	6.44	13.44
	C.V. (%)	-	13.68	12.26	8.5	16.8	20.91	4.83
2016	Genotype	177	265.17**	6836.12**	22.78**	0.033**	0.62**	1.42**
	Block	2	371.55**	1991.60*	0.64^{ns}	1.45**	22.36**	3.23*
	Covariate PH	1	30.11^{ns}	1804.23^{ns}	122.20**	0.15**	2.48*	4.18*
	Error	354	27.38	678.97	12.12	0.015	0.42	0.77
	Mean	-	30.13	163.37	21.48	0.63	4.74	14.41
	C.V. (%)	-	17.36	15.94	16.2	19.38	13.66	6.09
Combined	Genotype	177	317.30**	12663.27**	26.21**	0.04**	3.44**	2.12**
	Year	1	133400.11**	9525671.16**	0.88^{ns}	0.50**	600.32**	22.53**
	Genotype*Year	177	104.84**	5444.59*	13.03**	0.02**	4.52**	0.73^{ns}
	Block(Year)	3	341.26**	1781.43^{ns}	0.56^{ns}	0.96**	19.85**	73.87**
	Covariate PH	1	56.27^{ns}	955.55^{ns}	93.56**	0.13**	2.18^{ns}	5.30**
	Error	515	36.57	1136.73	9.46	0.013	0.85	0.75
	Mean	-	39.75	245.82	21.52	0.61	5.40	14.03
	C.V. (%)	-	15.21	13.71	14.29	18.82	17.08	6.20

† d.f: degrees of freedom, C.V.: coefficient of variation, last evaluation of percentage of symptomatic spikelets (PSS), area under de disease progress curve (AUDPC), thousand kernel weight (TKW), *Fusarium* damaged kernels (FDK), average deoxynivalenol content (DON), plant height (PH), and protein content in sound kernels (PRO). ** and *** represents respectively significance at $p<0.05$ and p<0.01 of probability of error, while ^{ns} indicates absence of statistical significance.

Table 2-2. Adjusted means, minimum and maximum values, 95% confidence limits, and broad-sense heritability (H^2) for last evaluation of percentage of symptomatic spikelets (PSS), area under de disease progress curve (AUDPC), thousand kernel weight (TKW), Fusarium-damaged kernels (FDK), average deoxynivalenol accumulation (DON), and protein content in sound kernels (PRO).

Trait	Year	Everest	WB-Cedar	DH pop.	Minimum	Maximum	H^2
PSS	2015	43.05±10.8†	66.47±10.5†	55.43	39.31±10.9 DH062	83.05±10.6 DH161	0.69
	2016	19.67±5.9†	31.69±6.0†	29.28	18.35±5.9 DH201	77.06±5.9 DH143*	
	Combined	31.12±5.47†	49.01±5.43†	42.67	30.26±5.5 DH056	79.71±5.47 DH143*	
AUDPC	2015	277.98±66.4†	504.82±64.4†	376.44	251.96±69.1 DH026	609.39±64.8 DH167	0.67
	2016	99.27±29.7†	173.94±29.7†	162.63	99.27±29.7 Everest	412.44±29.6 DH143*	
	Combined	190.68±30.5†	339.55±30.3†	269.27	181.88±34.9 DH056	505.21±30.5 DH143*	
TKW	2015	24.6±2.7	23.19±2.6	21.39	12.88±3.7 DH074*	29.77±2.6 DH183	0.38
	2016	22.03±3.9	21.42±3.9	21.49	15.86±3.9 DH136	30.92±4.1 DH208*	
	Combined	23.02±2.78	22.31±2.76	21.51	16.94±2.76DH209	29.10±2.76 DH183	
FDK	2015	0.35±0.14	0.59±0.13	0.59	0.25±0.14 DH188	0.94±0.05 DH074*	0.25
	2016	0.54±0.14	0.58±0.14	0.63	0.39±0.14 DH189	0.86±0.15 DH077	
	Combined	0.45±0.10	0.59±0.10	0.61	0.38±0.10 DH188	0.86±0.11 DH077	
DON	2015	3.08±1.94	6.12±1.88	6.53	1.15±1.92 DH009	12.8±2.7 DH074*	0.37
	2016	4.06±0.74	4.71±0.74	4.74	3.35±0.73 DH183	5.98±0.74 DH089	
	Combined	3.60±0.84	5.41±0.83	5.62	3.09±1.05 DH009	8.71±0.84DH209*	
PRO	2015	12.23±0.47	13.33±0.92	13.45	10.87±0.95DH135*	16.35±0.98DH147*	0.66
	2016	14.68±1.1	15.16±1.0	14.39	12.69±1.0DH163	15.92±1.0 DH026	
	Combined	13.88±0.81	14.43±0.81	14.09	12.63±0.81 DH094	15.71±0.80 DH028	

* Indicate lines that that were significantly different from the parents by a 95% confidence interval where maximum values were compared with susceptible parent and minimum values were compared with the resistant parent Everest.
† indicates traits where parents statistically differ from each other.

Table 2-3. Values of Person correlations between phenotypic traits evaluated in the DH population Everest/Cedar during the 2015 and 2016 growing seasons.

Traits	Year	AUDPC	TKW	FDK	DON	PH	PRO
PSS	2015	0.88 ***	-0.52 ***	0.60 ***	0.54 ***	-0.29 **	0.40 ***
	2016	0.98 ***	-0.09 ns	0.18 ns	-0.06 ns	-0.30 **	0.09 ns
AUDPC	2015	-	-0.38 **	0.48 ***	0.49 ***	-0.28 **	0.35 **
	2016	-	-0.08 ns	0.02 ns	-0.04 ns	0.11 ns	0.11 ns
TKW	2015	-	-	-0.78 ***	-0.61 ***	0.45 ***	-0.56 ***
	2016	-	-	-0.33 **	-0.37 **	-0.16 ns	-0.16 *
FDK	2015	-	-	-	0.86 ***	-0.33 **	0.67 ***
	2016	-	-	-	0.59 ***	0.28 **	0.28 **
DON	2015	-	-	-	-	-0.36 **	0.62 ***
	2016	-	-	-	-	0.39 **	0.39 ***
PH	2015						-0.42***
	2016						0.04 ns

ns represents non-significant correlations while ***, **, and * indicates significance at p>0.001, p>0.01, and p>0.05 respectively.

Table 2-4. Summary of genetic map of the Everest/Cedar DH population including number the markers, length, average spacing, and maximum spacing between markers within each wheat chromosome.

Wheat chromosomes	Number of Markers	Length (cM)	Ave. Spacing (cM)	Max. spacing (cM)
1A	159	108.12	0.68	22.85
1B	209	199.45	0.96	35.59
1D	58	82.38	1.45	19.61
2A	234	180.24	0.77	27.52
2B	287	201.18	0.7	32.11
2D	45	159.95	3.64	24.1
3A	148	120.42	0.82	15.0
3B	360	153.3	0.43	12.45
3D	33	134.21	4.19	26.65
4A	110	176.85	1.62	16.16
4B	45	112.17	2.55	31.6
4D	3	17.56	8.78	12.61
5A	192	236.4	1.24	29.61
5B	151	180.3	1.2	34.03
5D	28	158.93	5.89	36.7
6A	179	143.55	0.81	25.35
6B	109	120.75	1.12	42.6
6D	77	145.22	1.91	32.9
7A	196	169.77	0.87	26.35
7B	175	151.15	0.87	28.3
7D	41	88.69	2.22	43.7
Overall	2,839	3,040.59	1.08	43.7

cM: centimorgans.

Table 2-5. Significant QTL from the Multiple QTL Mapping (MQM) analysis associated with FHB resistance using adjusted phenotypic means calculates within and across years.

Trait	Chr.	Position	95% Bayesian interval	Peak marker	LOD[§]	%var.	Source	QTL effect at peak marker ± SE	
								EE	CC
PSS15	1AS	5.2	0.00 - 32.70	snp11896	4.43***	9.16	Cedar	58.95 ± 1.08	52.67 ± 0.99
	3BS	6.0	2.98 - 11.45	snp12041	7.11***	15.32	Everest	51.37 ± 1.08	58.81 ± 0.95
	7AL	103.9	64.96 - 109.44	snp12886	3.56***	7.28	Cedar	57.65 ± 1.04	53.28 ± 1.08
PSS16	3BS	5.4	0.00 - 72.59	snp00727	2.52**	6.19	Everest	27.19 ± 1.01	32.03 ± 0.89
	4BL	53	34.91 - 91.48	snp00386	3.12***	7.74	Everest	27.68 ± 0.87	32.9 ± 1.01
PSS15&16	3BS	6.0	2.98 - 11.45	snp12041	7.63***	14.31	Everest	39.23 ± 0.90	45.39 ± 0.79
	5BL	174	164.55 - 179.69	snp03237	4.81***	8.65	Cedar	44.89 ± 0.84	40.19 ± 0.91
	5DS	12	0.00 - 24.11	snp00063	4.17***	7.45	Everest	40.66 ± 0.85	45.20 ± 0.93
	6BS	3.3	0.00 - 120.74	snp06814	3.31***	5.84	Everest	40.88 ± 0.90	44.37 ± 0.86
	7AL	69.6	60.70 - 106.42	snp11649	3.46***	6.11	Cedar	44.17 ± 0.92	41.35 ± 0.87
AUDPC15	1AS	0.0	0.00 - 24.17	snp10627	3.40***	6.07	Cedar	404.21 ± 9.16	356.76 ± 8.09
	3BS	8.4	2.98 - 11.45	snp00081	6.18***	11.05	Everest	347.31 ± 8.94	401.71 ± 7.96
	5DS	11	0.00 - 27.25	snp00063	3.74***	6.72	Everest	354.43 ± 8.23	4.06.08 ± 9.05
	6BS	5.3	0.00 - 98.44	snp10941	3.50***	6.28	Everest	357.94 ± 8.92	395.31 ± 8.45
	7AL	103.9	33.74 - 110.98	snp12886	3.84***	6.92	Cedar	395.52 ± 8.50	358.18 ± 8.87
AUDPC16	3BS	4.8	0.00 - 72.59	snp13441	3.13***	8.37	Everest	147.80 ± 5.04	174.02 ± 4.48
AUDPC15&16	3BS	6.0	2.98 - 11.45	snp12041	7.41***	13.6	Everest	247.77 ± 6.03	286.90 ± 5.31
	5BL	166.4	113.51 - 180.29	snp06779	3.24***	5.6	Cedar	281.83 ± 5.98	258.52 ± 5.81
	5DS	11.0	0.00 - 24.11	snp00063	5.28***	9.41	Everest	253.54 ± 5.51	289.81 ± 6.06
	6BS	3.3	0.00 - 12.82	snp06814	4.27***	7.49	Everest	255.75 ± 5.97	282.75 ±5.72
	7AL	103	60.70 - 109.44	snp12126	2.99**	5.15	Cedar	281.90 ± 5.74	256.67 ± 5.99
TKW15	1B	77.5	55.59 - 83.86	snp09507	6.96***	12.52	Everest	22.80 ± 0.29	20.07 ± 0.29
	3BS	11.5	0.00 - 13.52	snp12923	5.02***	8.81	Everest	22.67 ± 0.33	20.47 ± 0.29
	3D	70	59.52 - 86.92	snp05382	3.96***	6.83	Everest	22.25 ± 0.32	20.60 ± 0.32
	5A	66.3	40.93 - 84.69	snp00508	3.40***	5.82	Everest	22.50 ± 0.34	20.65 ± 0.29
TKW15&16	1B	72	55.59 - 99.96	snp00247	4.72***	11.32	Everest	22.39 ± 0.22	20.74 ± 0.20
	3BS	5.4	0.00 - 153.30	snp00727	1.87**	4.3	Everest	22.22 ± 0.23	20.98 ± 0.21
FDK15	1AS	5.2	0.00 - 13.47	snp11896	3.63***	6.57	Cedar	0.63 ± 0.01	0.55 ± 0.01
	1B	82.7	55.59 - 93.27	snp02900	4.32***	7.88	Everest	0.53 ± 0.01	0.64 ± 0.01
	3BS	6.0	0.00 - 11.45	snp12041	5.19***	9.6	Everest	0.53 ± 0.01	0.63 ± 0.01
	5A	61	55.74 - 77.42	snp01770	4.61***	8.45	Everest	0.53 ± 0.01	0.63 ± 0.01
FDK16	2A	66.7	41.66 - 82.04	snp04743	2.93***	7.09	Cedar	0.66 ± 0.01	0.60 ± 0.01
	7AL	91.1	33.74 - 99.84	snp13735	3.55***	8.67	Everest	0.59 ± 0.01	0.66 ± 0.01
FDK15&16	1AS	3.9	0.00 - 8.91	snp07585	4.75***	10.15	Cedar	0.63 ± 0.01	0.58 ± 0.00
	1B	82.7	55.59 - 94.48	snp02900	4.59***	9.8	Everest	0.57 ± 0.00	0.64 ± 0.01
	5A	68.4	40.93 - 77.42	snp07755	3.21***	6.72	Everest	0.57 ± 0.01	0.63 ± 0.00
DON15	1B	68.1	55.59 - 90.82	snp01977	3.50***	6.98	Everest	6.00 ± 0.22	6.98 ± 0.20
	5A	58	54.48 - 67.58	snp12458	5.64***	11.59	Everest	5.63 ± 0.21	7.20 ± 0.19
	5DS	26	8.99 - 37.66	snp01397	5.67***	11.65	Everest	5.87 ± 0.20	7.27 ± 0.21
DON15&16	1AS	5.0	0.00 - 32.70	snp11896	3.08***	5.43	Cedar	5.93 ± 0.11	5.36 ± 0.10
	3BS	7.8	0.00 - 11.45	snp01392	4.04***	7.21	Everest	5.29 ± 0.11	5.88 ± 0.10
	5A	57	55.12 - 66.25	snp12458	8.12***	15.36	Everest	5.14 ± 0.11	5.99 ± 0.09
	5DS	26	17.19 - 74.26	snp01397	4.15***	7.42	Everest	5.30 ± 0.10	5.99 ± 0.11
PRO15	2AL	103	12.8 -108.86	snp02988	5.33***	8.28	Everest	13.65 ± 0.10	13.24 ± 0.09
	2BL	120	100.93 - 134.48	snp12925	3.32***	5.01	Cedar	13.27 ± 0.09	13.65 ± 0.11
	2DL	157	146.71 - 159.95	snp01069	6.06***	9.52	Cedar	13.11 ± 0.10	13.68 ± 0.09
	5AS	61	56.95 - 66.25	snp01770	9.97***	16.63	Cedar	12.98 ± 0.10	13.77 ± 0.08
	7BL	148.5	139.77 - 151.14	snp06733	8.18***	13.29	Everest	13.78 ± 0.10	13.12 ± 0.09
PRO16	2AL	71.4	12.80 - 82.04	snp01700	3.67***	7.95	Everest	14.63±0.07	14.24±0.07
	2BL	103	97.95 - 117.81	snp09736	4.46***	9.78	Cedar	14.20±0.06	14.68±0.07
	5AS	82	8.55 - 108.79	snp04340	2.96***	6.36	Cedar	14.23±0.08	14.56±0.06
PRO15&16	2AL	49	40.32 - 82.04	snp09392	6.36***	11.44	Everest	14.30±0.07	13.94±0.06
	2BL	106	100.93 - 117.81	snp07103	6.92***	12.55	Cedar	13.93±0.06	14.33±0.07
	5AS	57	40.93 - 77.42	snp12458	5.42***	9.61	Cedar	13.84±0.07	14.31±0.06
	7BL	149.7	131.66 - 151.14	snp03162	4.94***	8.7	Everest	14.30±0.07	13.94±0.06

QTL: quantitative trait loci, LOD: logarithm of odds, %var: percentage of variation explained by each QTL, SE: standard errors. EE and CC represent respectively the parental alleles of Everest and WB-Cedar at a given genomic location.

Table 2-6. Genotype groups containing QTL combinations associated with PSS. Adjusted means followed by different letters are statistically different by Tukey grouping test ($p<0.05$).

Genotypes	#lines	Line names	PSS15	PSS16	PSS1516
5B	4	DH023, DH167, DH189, DH206	64.05 abc	43.63 a	53.93 a
No QTL	10	DH031, DH036, DH050, DH108, DH132, DH137, DH143, DH160, DH219, DH223	68.98 a	36.81 ab	52.93 a
7A	8	DH059, DH066, DH077, DH119, DH131, DH157, DH161, DH190	65.61 ab	34.19 ab	49.83 a
3B	5	DH007, DH020, DH048, DH181, DH211	65.88ab	30.20 ab	48.12 ab
6B+7A	11	DH013, DH052, DH058, DH082, DH091, DH117, DH123, DH149, DH155, DH159, DH214	61.30 abc	33.16 ab	47.21 ab
5B+5D+7A	3	DH001, DH177, DH210	62.43 abc	30.66 ab	46.47 ab
5B+7A	9	DH035, DH076, DH086, DH088, DH098, DH129, DH135, DH174, DH215	59.04 abc	32.75 ab	45.82 ab
6B	6	DH019, DH021, DH028, DH080, DH093, DH162	55.63 abc	35.24 ab	45.44 ab
5D+6B	6	DH106, DH136, DH151, DH178, DH180, DH201	59.60 abc	30.29 ab	44.97 ab
5D	5	DH043, DH147, DH176, DH188, DH191	53.38 abc	36.54 ab	44.88 ab
5B+5D	6	DH049, DH168, DH171, DH175, DH198, DH204	62.03 abc	27.52 ab	44.82 ab
5B+6B	4	DH010, DH068, DH092, DH164	54.09 abc	35.05 ab	44.78 ab
3B+5D	9	DH047, DH072, DH094, DH115, DH138, DH146, DH185, DH187, DH194	56.38 abc	30.41 ab	43.37 ab
5B+5D+6B	4	DH002, DH009, DH101, DH220	55.11 abc	31.05 ab	42.97 ab
3B+6B	5	DH030, DH103, DH107, DH113, DH195	52.37 abc	30.57 ab	41.41 ab
5D+6B+7A	10	DH033, DH054, DH060, DH065, DH078, DH096, DH114, DH116, DH197, DH203	53.69 abc	27.40 ab	40.48 ab
3B+6B+7A	4	DH041, DH090, DH109, DH224	50.26 abc	28.75 ab	39.35 ab
3B+7A	6	DH042, DH051, DH121, DH141, DH173, DH200	47.12 bc	29.97 ab	38.48 ab
5D+7A	3	DH011, DH017, DH148	48.78 bc	28.30 ab	38.48 ab
5B+6B+7A	6	DH045, DH069, DH079, DH179, DH196, DH208	51.69 abc	24.78 ab	38.13 ab
3B+5B+6B+7A	4	DH008, DH046, DH140, DH207	50.42 abc	25.25 ab	37.82 ab
3B+5B	2	DH122, DH127	49.21 abc	25.44 ab	37.42 ab
3B+5D+7A	5	DH029, DH073, DH139, DH153, DH183	49.23 abc	25.84 ab	37.28 ab
3B+5D+6B	5	DH056, DH067, DH172, DH199, DH209	50.69 abc	22.8 ab	36.80 ab
3B+5B+6B	4	DH038, DH156, DH165, DH184	46.10 bc	25.60 ab	35.77 ab
3B+5B+7A	4	DH057, DH095, DH102, DH202	47.88 bc	23.75 ab	35.72 ab
3B+5B+5D	4	DH062, DH089, DH120, DH205	46.09 bc	24.67 ab	35.38 ab
3B+5D+6B+7A	2	DH158, DH186	48.79 abc	21.90 ab	35.10 ab
5B+5D+6B+7A	3	DH110, DH118, DH163	47.52 bc	22.64 ab	35.01 ab
3B+5B+5D+7A	3	DH026, DH087, DH097	45.27 bc	24.50 ab	34.65 ab
3B+5B+5D+6B	3	DH084, DH130, DH212	44.99 c	21.35 bc	33.09 b
3B+5B+5D+6B+7A	2	DH014, DH037	46.08 bc	19.44 c	32.64 b

Table 2-7. Genotype groups containing QTL combinations associated with AUDPC. Adjusted means followed by different letters are statistically different by Tukey grouping test (p<0.05).

Genotypes	#lines	Line names	AUDPC15	AUDPC16	AUD1516
No QTL	7	DH066, DH077, DH129, DH143, DH190, DH219, DH223	476.41 a	206.90 a	341.85 a
6B	5	DH013, DH021, DH086, DH093, DH123	466.28 a	189.35 a	327.43 a
1A	10	DH023, DH031, DH050, DH119, DH132, DH135, DH160, DH167, DH189, DH206	458.39 a	191.98 a	323.50 a
1A+7A	3	DH036, DH076, DH157	431.75 ab	199.06 a	316.73 ab
5D	3	DH171, DH175, DH191	447.07 ab	180.89 a	313.93 ab
7A	8	DH035, DH059, DH088, DH098, DH108, DH131, DH161, DH174	442.91 ab	182.15 a	313.07 ab
3B	7	DH007, DH020, DH048, DH141, DH181, DH211, DH215	445.29 ab	173.57 a	307.66 ab
5D+6B	2	DH078, DH114	428.93 ab	179.28 a	303.45 ab
5D+6B+7A	5	DH065, DH116, DH136, DH151, DH220	390.65 ab	164.15 a	278.09 ab
1A+3B	5	DH051, DH122, DH137, DH164, DH173	374.58 ab	185.21 a	277.52 ab
1A+6B	8	DH010, DH058, DH068, DH080, DH082, DH091, DH155, DH159	377.80 ab	177.51 a	275.87 ab
3B+5D	6	DH029, DH047, DH073, DH089, DH187, DH194	380.47 ab	161.70 a	270.98 ab
6B+7A	10	DH019, DH028, DH045, DH052, DH092, DH149, DH179, DH207, DH208, DH214	360.11 ab	170.98 a	265.70 ab
5D+7A	6	DH001, DH011, DH049, DH138, DH148, DH176	365.25 ab	160.08 a	263.52 ab
1A+3B+5D	10	DH062, DH072, DH094, DH115, DH139, DH146, DH147, DH183, DH185, DH205	357.79 ab	159.23 a	257.83 ab
3B+6B	2	DH113, DH184	368.82 ab	141.95 a	257.07 ab
1A+5D+7A	6	DH017, DH043, DH096, DH188, DH198, DH210	363.76 ab	148.25 a	255.80 ab
3B+6B+7A	5	DH008, DH090, DH109, DH140, DH156	346.30 ab	160.88 a	253.08 ab
1A+3B+6B	4	DH038, DH103, DH165, DH195	344.09 ab	161.65 a	251.70 ab
3B+5D+6B	5	DH037, DH067, DH101, DH172, DH209, DH110	361.31 ab	134.75 a	247.56 ab
1A+5D+6B	6	DH002, DH009, DH033, DH178, DH180, DH201	339.05 ab	154.47 a	245.98 ab
1A+5D+6B+7A	8	DH014, DH054, DH060, DH106, DH118, DH163, DH197, DH203	347.56 ab	142.93 a	245.22 ab
1A+5D	4	DH158, DH168, DH177, DH204	340.32 ab	146.96 a	243.65 ab
1A+6B+7A	4	DH069, DH079, DH117, DH196	340.51 ab	142.90 a	241.76 ab
3B+7A	3	DH030, DH057, DH202	337.93 ab	143.33 a	241.25 ab
1A+3B+6B+7A	4	DH041, DH046, DH107, DH162	331.56 ab	128.74 a	231.66 ab
1A+3B+7A	7	DH042, DH095, DH102, DH121, DH127, DH200, DH224	309.53 ab	133.35 a	221.98 ab
1A+3B+5D+7A	5	DH026, DH087, DH097, DH120, DH153	292.90 b	133.61 a	215.49 ab
3B+5D+7A	1	DH110	303.27 ab	126.11 a	214.71 ab
1A+3B+5D+6B	5	DH056, DH084, DH130, DH199, DH212	293.07 b	119.08 a	206.25 b
1A+3B+5D+6B+7A	1	DH186	281.46 c	119.77 a	201.29 b

Table 2-8. Genotype groups containing QTL combinations associated with FDK. Adjusted means followed by different letters are statistically different by Tukey grouping test ($p<0.05$).

Genotypes	#lines	Line names	FDK15	FDK16	FDK1516
No QTL	24	DH001, DH020, DH037, DH048, DH052, DH066, DH073, DH082, DH086, DH088, DH093, DH103, DH109, DH110, DH123, DH129, DH136, DH141, DH149, DH171, DH174, DH175, DH207, DH209, DH215	0.683 a	0.684 a	0.681 a
5A	12	DH011, DH047, DH065, DH067, DH077, DH078, DH098, DH108, DH114, DH181, DH191, DH202	0.663 ab	0.686 a	0.675 ab
1A	32	DH002, DH010, DH023, DH031, DH033, DH042, DH043, DH050, DH051, DH054, DH060, DH068, DH079, DH084, DH091, DH096, DH106, DH115, DH118, DH121, DH146, DH147, DH159, DH167, DH173, DH178, DH186, DH189, DH195, DH198, DH199, DH206	0.649 ab	0.630 ab	0.634 abc
1B	22	DH007, DH013, DH019, DH029, DH030, DH045, DH049, DH059, DH089, DH090, DH097, DH113, DH116, DH137, DH140, DH143, DH156, DH187, DH194, DH203, DH214, DH223	0.595 ab	0.649 ab	0.620 abc
1B+5A	21	DH008, DH021, DH035, DH057, DH062, DH092, DH101, DH131, DH148, DH151, DH158, DH161, DH172, DH176, DH179, DH184, DH190, DH208, DH211, DH219, DH220	0.572 ab	0.603 ab	0.587 bcd
1A+1B	20	DH014, DH028, DH038, DH072, DH087, DH102, DH117, DH130, DH155, DH160, DH163, DH165, DH177, DH185, DH196, DH197, DH200, DH201, DH212, DH224	0.538 bc	0.608 ab	0.572 cd
1A+5A	14	DH009, DH036, DH041, DH046, DH069, DH080, DH127, DH132, DH157, DH164, DH168, DH180, DH183, DH205	0.553 abc	0.555 b	0.551 cd
1A+1B+5A	19	DH017, DH026, DH056, DH058, DH076, DH094, DH095, DH107, DH119, DH120, DH122, DH135, DH138, DH139, DH153, DH162, DH188, DH204, DH210	0.431 c	0.603 ab	0.519 d

Table 2-9. Genotype groups containing QTL combinations associated with DON. Adjusted means followed by different letters are statistically different by Tukey grouping test ($p<0.05$).

Genotypes	#lines	Line names	DON15	DON16	DON1516
No QTL	18	DH019, DH031, DH059, DH066, DH088, DH091, DH108, DH123, DH129, DH143, DH149, DH196, DH198, DH207, DH214, DH219	8.121 a	4.933 a	6.522 a
1A	9	DH002, DH010, DH023, DH028, DH033, DH079, DH160, DH167, DH206	8.066 a	4.983 a	6.489 a
3B	15	DH008, DH029, DH030, DH048, DH084, DH090, DH102, DH109, DH113, DH141, DH156, DH162, DH202, DH209, DH215	7.153 abc	4.893 a	6.0196 ab
5D	18	DH001, DH049, DH052, DH082, DH086, DH093, DH116, DH117, DH136, DH140, DH148, DH155, DH159, DH163, DH171, DH174, DH175, DH187	7.357 ab	4.605 a	5.965 ab
1A+5D	9	DH043, DH054, DH060, DH068, DH096, DH106, DH177, DH178, DH203	6.956 abc	5.001 a	5.953 abc
5A	11	DH013, DH021, DH045, DH077, DH092, DH098, DH131, DH157, DH161, DH190, DH208	6.993 abc	4.736 a	5.862 abc
3B+5D	8	DH020, DH037, DH067, DH073, DH089, DH110, DH137, DH224	6.446 abc	4.950 a	5.669 abc
1A+3B	9	DH007, DH038, DH046, DH120, DH121, DH153, DH185, DH195, DH200	6.666 abc	4.508 a	5.584 abc
1A+5A	9	DH036, DH058, DH076, DH119, DH132, DH135, DH164, DH189, DH197	6.588 abc	4.579 a	5.555 abc
5A+5D	15	DH009, DH011, DH035, DH065, DH069, DH078, DH114, DH118, DH130, DH151, DH176, DH179, DH191, DH210, DH220	5.630 abc	4.754 a	5.200 abc
1A+3B+5D	11	DH014, DH051, DH072, DH087, DH097, DH115, DH146, DH147, DH173, DH199, DH212	5.565 abc	4.714 a	5.126 abc
1A+3B+5A	6	DH095, DH101, DH103, DH122, DH165, DH186	5.712 abc	4.322 a	5.021 abc
3B+5A	3	DH138, DH184, DH211	5.544 abc	4.358 a	4.960 abc
1A+5A+5D	7	DH017, DH080, DH168, DH180, DH188, DH201, DH204	4.892 bc	4.648 a	4.759 bc
3B+5A+5D	12	DH026, DH041, DH047, DH057, DH062, DH094, DH107, DH127, DH158, DH172, DH181, DH194	4.678 c	4.790 a	4.735 c
1A+3B+5A+5D	5	DH042, DH056, DH139, DH183, DH205	4.201 c	4.541 a	4.379 c

Table 2-10. Genotype groups containing QTL combinations associated with TKW. Adjusted means followed by different letters are statistically different by Tukey grouping test (p<0.05).

Genotypes	#lines	Line names	TKW15	TKW16	TKW1516
No QTL	61	DH001, DH002, DH009, DH011, DH019, DH023, DH031, DH033, DH035, DH036, DH043, DH049, DH050, DH052, DH054, DH059, DH060, DH065, DH066, DH068, DH069, DH077, DH078, DH079, DH080, DH082, DH086, DH088, DH091, DH093, DH096, DH098, DH106, DH108, DH114, DH117, DH118, DH123, DH132, DH135, DH136, DH149, DH157, DH159, DH164, DH167, DH168, DH171, DH174, DH175, DH176, DH177, DH178, DH179, DH180, DH189, DH191, DH198, DH204, DH206, DH220	19.86 c	21.10 a	20.58 c
3B	31	DH020, DH037, DH041, DH047, DH048, DH056, DH067, DH072, DH073, DH095, DH103, DH109, DH110, DH115, DH121, DH127, DH141, DH146, DH147, DH173, DH181, DH183, DH194, DH195, DH199, DH202, DH205, DH207, DH209, DH212, DH215	20.98 bc	21.44 a	21.32 bc
1B	35	DH010, DH013, DH017, DH021, DH028, DH030, DH045, DH051, DH058, DH076, DH092, DH101, DH116, DH119, DH129, DH131, DH138, DH143, DH148, DH151, DH155, DH160, DH161, DH163, DH188, DH190, DH196, DH197, DH201, DH203, DH208, DH210, DH214, DH219, DH223	22.04 b	21.65 a	21.87 ab
1B+3B	38	DH007, DH008, DH014, DH026, DH029, DH038, DH042, DH046, DH057, DH062, DH084, DH087, DH089, DH090, DH094, DH097, DH102, DH107, DH113, DH120, DH122, DH130, DH137, DH139, DH140, DH153, DH156, DH158, DH162, DH165, DH172, DH184, DH185, DH186, DH187, DH200, DH211, DH224	23.74 a	21.99 a	22.85 a

Table 2-11. Genotype groups containing QTL combinations associated with PRO. Adjusted means followed by different letters are statistically different by Tukey grouping test ($p<0.05$).

Genotypes	#lines	Line names	PRO15	PRO16	PRO1516
No QTL	11	DH076, DH080, DH114, DH119, DH130, DH135, DH138, DH148, DH180, DH188, DH205	12.36 e	13.72 c	13.28 e
2A	12	DH042, DH047, DH058, DH095, DH103, DH131, DH132, DH139, DH168, DH190, DH201, DH210	12.49 de	14.03 bc	13.56 de
5A	11	DH029, DH037, DH048, DH079, DH096, DH110, DH120, DH163, DH177, DH209, DH212	13.10 bcde	14.04 bc	13.67 cde
7B	11	DH062, DH117, DH118, DH122, DH127, DH151, DH162, DH172, DH184, DH186, DH191	12.98 cde	14.02 bc	13.78 bcde
2B	13	DH008, DH041, DH051, DH065, DH078, DH092, DH094, DH121, DH175, DH183, DH189, DH211, DH220	12.96 cde	14.35 abc	13.90 bcde
2A+2B+7B	3	DH179, DH181, DH208	13.56 abcde	14.40 abc	14.11 abcde
2A+5A	11	DH014, DH073, DH084, DH086, DH089, DH136, DH156, DH160, DH173, DH178, DH206	13.57 abcde	14.36 abc	14.13 abcd
2B+7B	7	DH017, DH045, DH069, DH161, DH165, DH194, DH197	13.61 abcde	14.50 abc	14.21 abcd
2A+7B	13	DH011, DH013, DH026, DH035, DH057, DH060, DH098, DH101, DH107, DH157, DH158, DH176, DH204	13.44 bcde	14.56 abc	14.24 abcd
5A+7B	15	DH002, DH019, DH038, DH043, DH049, DH116, DH123, DH153, DH155, DH159, DH167, DH171, DH174, DH202, DH203	13.82 abc	14.35 abc	14.24 abcd
2B+5A	14	DH001, DH030, DH033, DH072, DH088, DH091, DH113, DH129, DH140, DH143, DH207, DH215, DH223, DH224	13.42 bcde	14.71 ab	14.26 abcd
2A+2B	7	DH009, DH021, DH036, DH056, DH077, DH097, DH164	13.98 abc	14.55 abc	14.40 abcd
2B+5A+7B	9	DH020, DH023, DH031, DH054, DH090, DH137, DH146, DH187, DH214	14.14 abc	14.74 ab	14.43 abc
2A+5A+7B	11	DH010, DH066, DH068, DH082, DH102, DH106, DH108, DH141, DH147, DH196, DH219	14.21 ab	14.63 abc	14.43 abc
2A+2B+5A	8	DH007, DH046, DH087, DH115, DH149, DH195, DH198, DH200	13.67 abcd	14.96 ab	14.58 ab
2A+2B+5A+7B	9	DH028, DH050, DH052, DH059, DH067, DH093, DH109, DH185, DH199	14.66 a	15.16 a	14.98 a

Chapter 3 - A Genome-Wide Association Study and Genomic Selection for FHB Resistance in Winter Wheat

Abstract

Fusarium head blight (FHB) is one of the most important wheat diseases and its resistance is essentially controlled by quantitative small-effect loci. The objective of this chapter was to perform a genome-wide association study (GWAS) and test genomic selection (GS) models to map and predict FHB resistance. For three years, a total of 962 breeding lines from the K-State Wheat Breeding Program were phenotyped for FHB in a non-replicated design. Genotyping-By-Sequencing (GBS) was used to identify 23,157 single nucleotide polymorphisms (SNPs) which spanned more than 85% of the physical reference genome. Lines were evaluated for percentage of symptomatic spikelets, starting 14 days after heading. Grain samples were collected to estimate levels of deoxynivalenol (DON) and *Fusarium*-damaged kernels (FDK). Significant marker-associations were identified for FHB in each breeding panel tested within year but not across panels. This lack of consistency across years is likely due to variability in the frequency of resistance alleles based on changing parental germplasm from year to year. Although no significant differences were observed among GS models and training population sizes, the accuracy of predictions was relatively high (>0.45) when 80% of the data was assigned to the training set. Our results suggest that GS can be successfully implemented in wheat breeding programs to improve the levels of FHB resistance.

3.0 Introduction

Fusarium head bight is a wheat disease caused by *Fusarium graminearum*, that occurs under conditions of high humidity and warm temperatures during anthesis, leading to substantial economic losses in epidemic years. In Kansas alone, 7.1 million bushels of wheat were lost in 2008 due to FHB (McMullen *et al*., 2012) and more recently an estimated of 11.6 million bushels of wheat were lost to FHB in Kansas in 2015 (Bockus *et al*., 2015). Genetic resistance, used in conjunction with management practices, has been shown to be the most effective strategy to control FHB. As result, wheat breeders and pathologists are constantly working to discover, map, and introgress new sources of resistance into elite cultivars.

Although biparental populations have been broadly used for QTL mapping of FHB resistance, this approach exploits limited genetic variation and produces linkage maps with low resolution that consequently has low statistical power to detect minor-effect loci. Genome-wide association studies (GWAS) are an alternative that can overcome these limitations. GWAS uses all natural recombination events from diverse panels, without the upfront cost, time, and effort associated with population development (Korte & Farlow, 2013). Additionally, the relatedness of individuals, population structure, and covariates can be taken into account in the analysis.

During the last decade, markers availability and cost were the main restrictions for GWAS studies. Consequently, a limited number of association studies aiming to identify genomic regions associated with components of FHB resistance was found in the literature. For instance, Miedaner *et al*. (2011) used only 115 SSR markers for an association study while in a similar study Kollers *et al*. (2013) used 732 SSRs to perform a GWAS of FHB resistance in European wheat. Yet, both studies were able to identify significant marker-trait associations. For a short period of time fixed genotyping arrays such as the 90K SNPchip array (Wang *et al*.,

2014) became relevant to obtain a large number of markers. However due to its expensive design, ascertainment bias, and failure to detecting rare and unique variants (Thomson, 2014) fixed genotyping platforms were not widely adopted for breeding. Then the development of the genotyping-by-sequencing (GBS) protocol (Poland *et al*., 2012) has resulted in an unprecedented capability to discovering and utilize genome-wide SNP markers at a relatively low cost. The majority of recent mapping studies in wheat have adopted GBS for genotyping in wheat genetic research (He *et al*., 2014). Likewise, the number of markers in the genetic maps and the number of QTLs associated with disease resistance increased substantially as a result.

Significant marker-trait associations with FHB resistance were reported by Arruda *et al*. (2016b) where eight QTLs were identified in a diverse panel of U.S. cultivars. Similarity, Wang *et al*. (2017) found another six highly significant QTLs associated with multiple components of FHB resistance in the Pacific Northwest Region of the United States and CIMMYT breeding program. Yet in the same study, it was found a QTL on 5B which was speculated to potentially be a novel locus of FHB resistance and explained a large proportion of DON accumulation.

Nevertheless, GWAS has its own limitations as it often fails to detect rare variants and spurious associations may occur (Bernardo, 2016). Thus, another approach to assist breeding disease resistance is the use of genome-wide markers to calculate genomic estimated breeding values (GEBVs). This approach, known as genomic selection (GS), attempts to capture the total additive variance from markers distributed throughout the entire genome, rather than relying on performing significance tests at every single locus. GS uses genome-wide marker data from related materials to predict the performance of another set of individuals that have not yet been phenotyped. Therefore, GS has more power to account for small-effect loci associated with complex traits such as FHB resistance (Poland & Rutkoski, 2016), although genomic regions

associated with the trait of interest remain unknown. Rutkoski *et al*. (2012) were pioneers in the application of GS for FHB resistance in wheat, by testing several GS models. Since then, other studies have investigated alternative ways to increase the prediction accuracies of GS for FHB resistance in wheat breeding (Arruda *et al*., 2016; Jiang *et al*., 2017).

The literature is lacking studies that simultaneously use multiple genomic approaches to improve disease resistance. Here in this chapter, we use an integrated strategy to combine results from biparental QTL mapping and GWAS to genomic selection. For this purpose, three large panels of elite wheat lines from the Kansas State Hard Red Winter Wheat Program that were tested in an FHB nursery from 2015-2017 were used.

3.1 Materials and Methods

Panel of Breeding Lines

The K-State Hard Red Winter Wheat Breeding Program is based on a modified bulk selection method where single plants are selected within populations up until F_4 or F_5. After selection and harvest of derived lines, DNA is collected for genotyping. Derived lines are tested in Individual Plant Short Rows (IPSRs) where selections are made and the superior candidates advanced to yield trails. At this stage, all breeding lines are also tested in short head-row plots in disease nurseries.

Breeding for FHB resistance is primarily based on the phenotypic selection of minor genes associated with native resistance in the K-State Breeding Program. Every year, preliminary and advanced lines are tested at Rocky Ford FHB Nursery of the Department of Plant Pathology. Panels of 377, 349, and 163 breeding lines were evaluated in 2015, 2016 and 2017, respectively. A total of 21 lines initially evaluated in 2015 were re-tested in 2016 whereas another 42 lines

evaluated in 2016 were advanced to 2017, and only 10 lines from 2015 were included in 2017 panel. Susceptible (Overley) and moderately resistant (Everest) checks were also included in every set of six individual rows in the FHB nursery to facilitate rating comparisons in the non-replicated design. Each experimental unit consisted of one breeding line planted in a one-meter long single row plot.

Evaluations of Phenotypic Traits

Details of FHB inoculation and nursery conduction were performed according to the methodology described by Jian *et al*. (2013) previously described in Chapter 2. Heading date was recorded as when 50% of the spikes emerged from the boot. The evaluation of the percentage of symptomatic spikelets (PSS) started 14 days after heading and was repeated at 21 and 28 days after heading, depending on the year. Additionally, in 2015 a stripe rust epidemic occurred and notes were taken using a 1-9 linear scale where 1 represents resistance and 9 refers to a highly susceptible reaction.

After the last PSS evaluation, a total of 38, 45, and 112 lines were selected, based on lower PSS ratings, to be evaluated for secondary components of FHB resistance in 2015, 2016, and 2017, respectively. For this purpose, a random sample of 100 grains was taken from each breeding line to estimate *Fusarium*-damaged kernels (FDK) and accumulation (DON) using the single kernel basis using the SKNIR system which was described in Chapter 2 (Dowell *et al*., 1999; Peiris *et al*., 2010).

Genotypic Data

Leaf tissue of all breeding lines and the biparental mapping population (Chapter 2) were collected from 3-5 plants and pooled together at the seedling stage for DNA extraction. The complete panel consisted of 986 unique individuals. The “BS96 DNA Plant” protocol was used

for DNA extraction (Qiagen, Hilden, Germany). All DNA samples were normalized and genotyping-by-sequencing (GBS) libraries were prepared using the protocol described by Poland *et al*. (2012). GBS libraries were then sequenced on an Illumina HiSeq equipment.

A bioinformatic pipeline on TASSEL 5.0 (Trait Analysis by Association, Evolution and Linkage Software) was used to call and filter single nucleotide polymorphism (SNPs). From the 141,193 SNPs initially discovered, only those with minor allele frequency (MAF) greater than 1% and less than 25% missing genotypes were retained. Markers that yielded multi-allelic calls and/or heterozygosity higher than 10% were discarded. Breeding lines with more than 50% missing makers were also removed from the dataset. Unanchored SNPs were retained in the genotypic data and assigned to an unknown chromosome (UN). Physical positions of SNP markers were corrected based on the reference genome and ordered from the telomere region of the short arm to the long arm of each chromosome using the *161010_Chinese_Spring_v1.0* pseudomolecule reference (IWGSC, 2017). After filtering, a total of 23,157 SNP markers and 962 lines were retained in the final data set.

Genome-Wide Association Analysis

No marker imputation was performed for the genome-wide association study (GWAS). Association analysis was performed in R using the Genome Association and Prediction Integrated Tool (GAPIT) (Tang *et al*., 2016) with an enhanced compression of the mixed linear model (ECMLM) (Li *et al*., 2014). A Bayesian information criterion (BIC) was estimated by setting the parameter Model.selection as 'TRUE' in order to determine the optimal number of principal components (PCs) that should be included in the association analysis. An additional GWAS analysis was run with the rrBLUP package (Endelman, 2013) using the EMMA model (Kang *et al*., 2008) by setting the population parameter '*P3D*' as false to avoid overestimation of

marker significance. The kinship matrix and the first three principal components were included in the GAPIT and rrBLUP analysis as covariates to account for the population structure.

Significant trait-marker associations were detected using multiple test correction with three different thresholds of significance: 5% and 10% false discovery rate (FDR 5% and FDR 10%) and Bonferroni correction. FDR values were obtained from the GWAS analysis with rrBLUP whereas the default option of GAPIT determined significant associations by the Bonferroni correction test. This last method is highly conservative, limiting the identification of significant marker-trait associations as reported by several authors (Sham & Purcell, 2014; Gao *et al*., 2016), therefore a less stringent test (FDR) was also implemented in our study.

A pairwise linkage disequilibrium analysis (LD) between all SNP markers was calculated using the R package 'Genetics' for each wheat chromosome. The extent of LD was determined to decay when the $r^2 < 0.2$. Subsets of wheat lines from each year were analyzed separately for the GWAS analysis. GWAS results from GAPIT and rrBLUP were used to reconstruct Manhattan plots using the 'qqman' R package.

Genomic Selection Models

SNP markers from prior analysis were numerically coded as 1, -1, 0, and NA respectively for major allele, minor allele, heterozygosity, and missing data. Loci with missing data were imputed using the EM method with the *A.mat* function of the 'rrBLUP' package in R (Endelman,2011). Four genomic selection models (GS) were estimated using the R package 'GSwGBS' (Gaynor, 2015). The first GS model was the ridge regression best linear unbiased predictor (rrBLUP) obtained using the package 'rrBLUP' while partial least squares regression (PLSR), elastic net (ELNET), and random forest (RF) were calculated using the R packages 'pls' (Mevik & Wehrens, 2007), 'glmnet' (Friedman *et al*., 2009), and 'randomForest' (Liaw &

Wiener, 2002), respectively. The average prediction (AVE) across these four prediction models was calculated using standardized values to avoid overly weighting the average towards any single prediction model as described in detail by Battenfield (2015).

Cross-validation of accuracy predictions was conducted within years considering four training population sizes, where 20%, 40%, 60%, 80% of the data were randomly masked to predict the remnant data, and replicated 100 times. The accuracies of predictions were obtained by dividing the correlation coefficient by the square root of broad-sense heritability in the average of 100 replications, as described by Battenfield *et al*. (2016). Forward predictions were conducted using data from prior years to predict the following year, (i.e., 2015 predicts 2016 and 2015 and 2016 predict 2017, etc.). In these case, the relationship between predicted and observed values was assessed by Pearson correlation. Additionally, these data were also used as a 'training set' to predict FHB resistance in $F_{5:6}$ lines (IPSRs) in the K-State Wheat Breeding Program.

3.2 Results and Discussion

Distribution of SNPs on the Physical Map

The filtered genotypic dataset consisted of 23,157 SNP markers on 787 breeding lines and 175 DH lines from the biparental DH population studied in Chapter 2. The majority of the marker polymorphism was found on the A (37.1%) and B (40.8%) genomes while only 19.4% of the markers were located on the D genome (Table 3-1). SNPs present on unanchored contigs represented only 2.7% of the total number of markers and were kept apart in the genotypic dataset as 'UN'.

The SNP markers spanned 14.5 gigabases (Gb) of the physical map representing a complete coverage of the wheat reference genome. The average marker spacing was 0.63

megabases (Mb) with the largest gap (55.9 Mb) being found in the centromeric region of chromosome 3D (Table 3-1). SNPs were more densely distributed towards the telomeres (Figure 3-2), which confirms the trend reported by several studies of higher recombination rates occurring in those regions (IWGSC, 2014). Likewise, Avni *et al*. (2017) reported that the density of genes was up to 14-fold higher in the distal, compared to pericentromeric, regions of chromosome arms of wild emmer (*T. turgidum ssp. dicoccoides*). Moreover, since the restriction enzymes used in the GBS protocol are methylation sensitive, it discriminates against highly methylated regions such near to centromeres.

Identifying Significant SNP-trait Associations

The percentage of symptomatic spikelets (PSS) was normally distributed in all three breeding panels tested within each year (Figure 3-2). Spurious associations between markers and traits may occur when the population structure is not accounted for in the GWAS analysis. To minimize this issue, the level of stratification was assessed via principal components analysis (PCA) using all 23,157 SNPs. A moderate level of population structure was detected with the clear formation of three clusters within each panel, and also when all panels were analyzed together (Figure 3-3). The first three PCA explained the majority of the genetic variance in the wheat panels and they were further included in the GWAS analysis as covariates along with the kinship matrix.

The association analysis for the Everest/Cedar population identified nine SNPs on 3BS 2DL, 6BS, 7AL, 7BL that were significantly associated with the percentage of symptomatic spikelets (PSS1516) and area under the disease progress curve (AUDPC1516) using the average over the two years of the experiment (Figure 3-4). The SNPs S7A_PART2_259768092 and S7B_PART2_286166114 showed the highest significance of p-values for PSS1516 and

AUDPC1516, simultaneously (Table 3-2). Apart from one SNP on 2DL, all other SNP-trait associations mapped in the same genomic regions where QTL were found in the biparental mapping analysis of Chapter 2. Several studies have reported the presence of a QTL associated with FHB resistance (typically type II) on the long arm of 2D from Wuhan1, Wangshuibai, Sumai #3 (Liu *et al*., 2009) and CJ9306 (Jiang *et al*., 2007). Likewise, Clinesmith (2015) identified a QTL from Art in the same interval of 2DL, explaining more than 10% of the PSS, FDK, and DON variation, indicating that it could be the same QTL previously mapped.

Only two SNPs (S4A_PART2_208236032 and S1B_PART2_200695538) were associated with the first and second evaluation of PSS in the panel of 377 breeding lines phenotyped in the 2015 growing season (Figure 3-5). These loci are located on long arm of 4A and 1B and together explained 7.3% of the phenotypic variation of PSS (Table 3-2). Other studies have reported the presence of QTL associated with FHB resistance in these genomic regions. For example, Arina and Pirat are European cultivars known for their high level of resistance and both carry QTL associated with FHB resistance type II in these chromosomal regions (Holzapfel *et al*., 2008; Liu *et al*., 2009).

A severe stripe rust epidemic occurred in 2015 and field notes were taken for all breeding lines. Based on the GWAS analysis, several SNPs on the short arm of 2A were significantly associated with stripe rust resistance (Table 3-2 and Figure 3-5). The translocation 2NS·2AS from *Ae. ventricosa* is located in this genomic region and is associated with disease resistance against multiple wheat pathogens (Mondal *et al*., 2016), The gene cluster *Yr17*/*Lr37*/*Sr38* resides in this region (Helguera *et al*., 2003). Therefore, *Yr17* is likely explaining the majority of the stripe rust resistance in the breeding panel evaluated in 2015, since the 2NS·2AS translocation is present at a relatively high frequency within the K-State Wheat Breeding Program (MAF>0.40).

In the panel of 349 breeding lines phenotyped in 2016, significant SNP associations with PSS1 were detected on 1DS, 2AL, 3DS, 4DS, and 5DS whereas only one SNP located on 1AL was significantly associated with PSS2. The presence of a QTL on 4DS associated with FHB resistance is often reported in the literature, since this region harbors the dwarfing gene *Rht-D1* where the short allele (*Rht-D1b*) is associated with susceptibility to FHB (Type I) and low anther extrusion (Buerstmayr *et al*., 2016; Steiner *et al*., 2017). The underlying mechanisms of *Rht* genes on FHB are speculated to be a tight linkage, pleiotropy, or disease escape (He *et al*., 2016a). The QTL on 5DS mapped within the same interval as the one found in Chapter 2, thus it is likely the same locus. The other significant loci are in regions where other QTL have been reported in European material (Holzapfel *et al*., 2008), North American germplasm (Liu *et al*., 2009) and synthetic accessions from CIMMYT (He *et al*., 2016b).

In the panel of 163 breeding lines phenotyped in 2017, two SNPs on 2BL and 7DL were associated with lower values of PSS2 and PSS3, respectively. An SNP located on 6AL fell just below the FDR>10% threshold, suggesting a weak association with PSS2 and PSS3. A QTL on a similar region of 6A associated with FHB resistance has been repeatedly reported in the literature, with resistant alleles present in the European varieties Apache (Holzapfel *et al*., 2008) and 'Dream' (Kollers *et al*., 2013). Likewise, a QTL on 2BL associated with type II resistance has been recently identified in the cultivar Truman (Islam *et al*., 2016), which was used as a source of native FHB resistance in the K-State Wheat Breeding program. Truman is the potential donor of the QTL found in this study. Additionally, 109 lines were harvested and tested for FDK and DON accumulation using the SKNIR system in the 2017 growing season. Data collected in 2015 and 2016 were not included in the GWAS analysis due to the small number of samples that

were collected in those years. Only two SNPs, located on 3BL, were associated with FDK and no significant associations were found for DON accumulation.

Although a less conservative test of significance (FDR 5% and FDR 10%) was chosen to declare significant trait to genotype associations, no significant markers associated with FHB resistance were found across breeding panels tested within each year. Similar results are reported in the literature where Arruda *et al*. (2016) and Wang *et al*. (2017) also used FDR<10% and p<0.001 to declare significance of markers. It suggests that the highly quantitative nature of FHB resistance makes it difficult to find strong SNP-trait associations. Another explanation for these findings is based on the structure of the breeding program. Relatively few lines are advanced from one year to the next and are not exclusively selected for FHB resistance. Therefore, the frequency and origin of alleles associated with FHB resistance can change drastically from year to year. Moreover, rare variants, are often associated with traits of interest, especially in less diverse panels such as elite breeding lines and are difficult to detect in GWAS (Bernardo, 2016). FHB resistance has been shown to be a quantitative trait with many minor genes controlling expression on the phenotype. Therefore, even if a set of SNPs were positively associated with lower FHB severity, their modest effect may not be able to generate a –log10 p-value high enough to meet the threshold of significance.

These results suggest the allele frequency of resistance alleles varied from one panel to the other as parental germplasm used in the breeding program changes from year to year. Furthermore, SNP positions were obtained by direct alignment against the *Chinese Spring v1.0 pseudomolecule* however, this variety used as a reference does not represent well the genetic diversity present in modern cultivars as recently reported by (Montenegro *et al*., 2017). These authors observed that more than 12,000 genes absence un the Chinese Spring reference were

present in all 18 modern wheat cultivars sequenced in their study. Moreover, recombination events and linkage among markers were not taken into account in the association analysis since SNPs were exclusively ordered within chromosomes according to their physical positions and it may have affected the GWAS results.

Recently an innovative GWAS procedure was proposed to overcome some of these limitations by accounting for historical linkage disequilibrium (LD) blocks that have been accumulated in breeding lines throughout multiple cycles of selection (He *et al*., 2017). This approach initially groups tightly linked sequential SNPs into LD blocks to form markers with multi-allelic haplotypes, then the markers are preselected as single-loci followed by multi-locus, multi-allele model stepwise regression during the association analysis. This procedure tends to generate a larger number of significant associations, suggesting that it could also be implemented in wheat to identify minor-effect loci underlying FHB resistance.

Applying Genomic Prediction Models

Five genomic selection models (rrBLUP, RF, PLSR, ELNET, and their average prediction AVE) were tested for each year of data considering four training population (TP) sizes (20, 40, 60, and 60%). Only phenotypic data from the last evaluation of PSS of each panel was used to test these models. The cross-validation was performed within years using the prior proportions that were randomly selected to predict the FHB resistance in the remnant individuals with 100 replications. Broad-sense heritability was calculated for each GS model and across all four TP sizes and it is presented in Table 3-2.

Accuracy predictions ranged from 0.37 to 0.51 in 2015, 0.34 to 0.47 in 2016, and 0.25 to 0.51 in 2017. However, no significant differences between GS models were observed in any breeding panel (Figure 3-8) considering 100 iterations. The large 95%-confidence error bars

verified for all GS models are likely due to the unbalanced nature of the breeding panels. In contrast to the majority of studies reported in the literature, our study used data from an actual wheat breeding program. Therefore, when less genetically related individuals were randomly assigned to the training and testing set, lower prediction accuracies were obtained due to the population structure previously discussed (Figure 3-3). The necessity of accounting for genetic relatedness between training and testing sets to optimize GS accuracy predictions has been reported in the literature. (Akdemir *et al*., 2015; Akdemir, 2017) and should be further implemented in the K-State Wheat Breeding Program.

Overall rrBLUP, PLSR, and AVE tended to give numerically higher prediction accuracy values across all three years (Figure3-8), although none of the models significantly differed from each other. Model averaging (AVE) has been considered ideal when performing forward predictions (Raftery, *et al*., 2010). Additionally, higher prediction accuracies with rrBLUP and AVE were observed previously in the K-State Wheat Breeding Program for grain yield (Gaynor, 2015) and end-use quality (Battenfield *et al*., 2016). Further research is needed in order to increase the accuracy of predictions for FHB. Specifically, obtaining more precise phenotypic assessment such as aerial image analysis, testing genetic algorithms to account for the relatedness of training and target sets, replicated trials of disease nurseries, and including covariate variables in the GS models should be explored.

3.3 Conclusions

Significant SNP associations were verified for all three breeding panels with two, six and our QTLs found respectively for the panels tested in 2015, 2016, and 2017. However, no significant marker associations were observed across panels, suggesting the allele frequency of resistance may have changed from one year to the other as there is a limited number of favorable loci controlling FHB resistance in the K-State Wheat Breeding program and parental germplasm used in the breeding program changes from year to year. Although significant loci have been reported in similar genomic regions as other studies, we were not able to confirm whether these QTLs were the same.

The genomic selection models resulted in moderately high values of prediction accuracies, especially when only 20% of the data was masked in the training set. This indicates that levels of FHB resistance can be improved using genomic-based predictions, regardless of the GS method chosen. The absence of statistical differences for the accuracy of predictions between GS models and training population sizes was greatly influenced by the level of population structure present in the wheat breeding panels.

References

Akdemir D., Sanchez J. I., & Jannink J.L. (2015). Optimization of genomic selection training populations with a genetic algorithm. Genetics Selection Evolution 47(1) 38: 1-10.

Akdemir D. (2017). STPGA: Selection of training populations with a genetic algorithm. bioRxiv 111989.

Arruda M. P., Brown P., Brown-Guedira G.B., Krill A.M., Thurber C., Merrill K.R., Foresman B.J., Kolb F.L. (2016a). Genome-Wide Association Mapping of Fusarium Head Blight Resistance in Wheat using Genotyping-by-Sequencing. The Plant Genome 9(1): 1-14.

Arruda M. P., Lipka A. E., Brown P. J., Krill A. M., Thurber C., Brown-Guedira G., Dong Y. Foresman B. J., Kolb F. L. (2016b). Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). Mol Breeding 36 (7): 1-11.

Avni R., Nave M., Barad O., Baruch K., Twardziok S.O., Gundlach H., Mascher M., Spannagl M., Wiebe K., Jordan K.W., Golan G., Deek J., Ben-Zvi B., Ben-Zvi G., Himmelbach A., MacLachlan R.P., Sharpe A.G., Fritz A.K., Ben-David R., Budak H., Fahima T., Korol A., Faris J.D. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357 (6346): 93-97.

Battenfield, S. D., C. Guzmán, R. C. Gaynor, R. P. Singh, R. J. Peña, S. Dreisigacker, A. K. Fritz, and J. A. Poland. (2016). Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. The Plant Genome 9(2): 1-12.

Bernardo R. (2014). Genome-wide selection of parental inbreds: Classes of loci and virtual biparental populations. Crop Sci. 54(1):1–33.

Bernardo R. (2016). Bandwagons I, too, have known. Theor Appl Genet. 129(12): 2323-2332.

Bockus W.W., Appel J.A., De Wolf E.D., Todd T.C., Davis M.A., Fritz A.K. (2015). Impact of wheat cultivar Everest on yield loss in Kansas from Fusarium head blight during 2015. Proceedings of the 2015 National Fusarium Head Blight Forum. St Louis MO.

Buerstmayr M., Buerstmayr H. (2016). The semi-dwarfing alleles *Rht-D1b* and *Rht-B1b* show marked differences in their association with anther retention in wheat heads and with Fusarium head blight susceptibility. Phytopathology 106(12): 1544-1552.

Endelman J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4(3): 250-255.

Friedman J., Hastie T., & Tibshirani R. (2009). Glmnet: Lasso and elastic net regularized generalized linear models. R package version 1.

Gaynor R.C. (2015). GSwGBS: An R package genomic selection with genotyping-by-sequencing. Genomic selection for Kansas wheat. K-State Research Exchange, Manhattan, KS.

He J., Meng S., Zhao T., Xing G., Yang S., Li Y., Guan R., Lu J., Wang Y., Xia Q., Yang B., Gai J. (2017). An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. Theor Appl Genet. 130(11): 2327-2343.

He J., Zhao X., Laroche A., Liu H., Li Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science 5(484): 1-8.

He X., Lillemo M., Shi J., Wu J., Bjørnstad Å., Belova T., Dreisigacker S., Duveiller E., Singh P. (2016b). QTL Characterization of Fusarium Head Blight Resistance in CIMMYT Bread Wheat Line Soru#1. PLoS One 11(6): e0158052.

He X., Singh P.K., Dreisigacker S., Singh S., Lillemo M., Duveiller E. (2016a). Dwarfing Genes *RhtB1b* and *Rht-D1b* Are Associated with Both Type I FHB Susceptibility and Low Anther Extrusion in Two Bread Wheat Populations. PLoS One 11(9): e0162499.

Helguera M., Khan I.A., Kolmer J., Lijavetzky D., Zhong-qi L., Dubcovsky. J. (2003). PCR assays for the *Lr37-Yr17-Sr38* cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. Crop Science 43:1839-1847.

Heslot N., J.-L. Jannink & Sorrells M.E. (2015). Perspectives for genomic selection applications and research in plants. Crop Science 55(1): 1-12.

Holzapfel J., Voss HH., Miedaner T., Korzun V., Häberle J., Schweizer G., Mohler V., Zimmermann G., Hartl L. (2008). Inheritance of resistance to Fusarium head blight in three European winter wheat populations. Theor Appl Genet. 117(7): 1119-1128.

Islam S., Brown-Guedira G., Van Sanford D., Ohm H., Dong, Y., Mckendry A.L. (2016). Novel QTL associated with the Fusarium head blight resistance in Truman soft red winter wheat. Euphytica 207(3): 571-592.

IWGSC -The International Wheat Genome Sequencing Consortium. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345(6194): 1251788.

Jiang G.L., Shi J.R., Ward R.W. (2007). QTL analysis of resistance to Fusarium head blight in the novel wheat germplasm CJ 9306. I. Resistance to fungal spread. Theor Appl Genet. 116(1): 3-13.

Jiang Y., Schulthess A.W., Rodemann B. *et al*. (2017). Validating the prediction accuracies of marker-assisted and genomic selection of Fusarium head blight resistance in wheat using an independent sample. Theor Appl Genet. 130(3): 471-482.

Jiang Y., Schulthess A.W., Rodemann B., Ling J., Plieske J., Kollers S., Ebmeyer E., Korzun V., Argillier O., Stiewe G., Ganal M. W., ROder M.S., Reif J.C. (2017). Validating the prediction accuracies of marker-assisted and genomic selection of Fusarium head blight resistance in wheat using an independent sample. Theor Appl Genet. 130(3): 471–482.

Kang H.M., Zaitlen N.A., Wade C.M., Kirby A., Heckerman D., Daly M.J., Eskin E. (2008). Efficient control of population structure in model organism association mapping. Genetics 178(3): 1709-1723.

Kollers S., Rodemann B., Ling J., Korzun V., Ebmeyer E., Argillier O., Hinze M., Plieske J., Kulosa D., Ganal M.W. Roder M.S. (2013). Whole Genome Association Mapping of Fusarium Head Blight Resistance in European Winter Wheat (*Triticum aestivum* L.). PLoS One 8(2): e57500.

Korte A. & Farlow A. (2013). The advantages and limitations of trait analysis with GWAS: A review. Plant Methods 9(29): 1-9.

Li M., Bradburry P., Yu J., Zhang Y.M., Todhunter R. J., Buckler E., Zhang Z. (2014). Enrichment of statistical power for genome-wide association studies. BMC Biology 12(73): 1-10.

Liaw A., & M. Wiener. (2002). Classification and regression by Random Forest. R News 2(3): 18-22.

Liu S., Hall M. D., Griffey C. A., McKendry A. L. (2009). Meta-analysis of QTL associated with Fusarium Head Blight in Wheat. Crop Science 49: 1955-1968.

McMullen M., Bergstrom G., De Wolf E., Dill-Macky R., Hershman D., Shaner G., Van Sanford D. (2012). A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Disease 96(12): 1712-1728.

Mevik B.H., & Wehrens R. (2007). The PLS package: Principal component and partial least squares regression in R. J. Stat. Softw. 18(2): 1-24.

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

Mohammadi M., Tiede T., & Smith. K. P. (2015). PopVar: A Genome-Wide Procedure for Predicting Genetic Variance and Correlated Response in Biparental Breeding Populations. Crop Science 55(5): 2068-2077.

Montenegro J.D., Golicz A.A., Bayer P.E., Hurgobin B., Lee H., Chan C.K., Visendi P., Lai K., Doležel J., Batley J., Edwards D. (2017). The pangenome of hexaploid bread wheat. The Plant Journal 90(5): 1007-1013.

Peiris K. H. S., Pumphrey M. O., Dong Y., Maghirang E. B., Berzonsky W., Dowell F.E. (2010). Near-Infrared Spectroscopic Method for Identification of Fusarium Head Blight Damage

and Prediction of Deoxynivalenol in Single Wheat Kernels. Cereal Chem. 87(6): 511-517.

Poland J. & Rutkoski J. (2016). Advances and Challenges in Genomic Selection for Disease Resistance. Annu. Rev. Phytopathol. 54(): 79-98.

Poland J. A., & Rife T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. The Plant Genome 5(3): 92-102.

Poland J.A., Brown P.J., Sorrells M.E., Jannink J.L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PloS One 7(2): e32253.

R Development Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Raftery A.E., Kárný M., & Ettler P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics 52(1): 52-66.

Rutkoski J., Benson J., Jia Y., Brown-Guedira G., Jannink J., Sorrells M. (2012). Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. The Plant Genome 5(2): 51-61.

Sham P.C., & Purcell S.M. (2014). Statistical power and significance testing in large-scale genetic studies. Nature Genetics 15(5): 335-346.

Spindel J.E., Begum H., Akdemir D., Collard B., Redoña E., Jannink L.L., McCouch S. (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity 116(4): 395-408.

Steiner B., Buerstmayr M., Michel S., Schweiger W., Lemmens M., Buerstmayr H. (2017). Breeding strategies and advances in line selection for Fusarium head blight resistance in wheat. Tropical Plant Pathology 42(3): 165-174.

Tang Y., Liu X., Wang J., Li M., Wang Q., Tian F., Su Z., Pan Y., Liu D., Lipka A. E., Buckler E. S., Zhang Z. (2016). GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. The Plant Genome 9(2): 1-9.

Thomson M.J. (2014). High-Throughput SNP Genotyping to Accelerate Crop Improvement. Plant Breed. Biotech. 2(3): 195-212.

Wang R., Chen J., Anderson J.A., Zhang J., Zhao W., Wheeler J., Klassen N., See D.R., Dong Y. (2017). Genome-Wide Association Mapping of Fusarium Head Blight Resistance in Spring Wheat Lines Developed in the Pacific Northwest and CIMMYT. Phytopathology 107(2): 1486-1495.

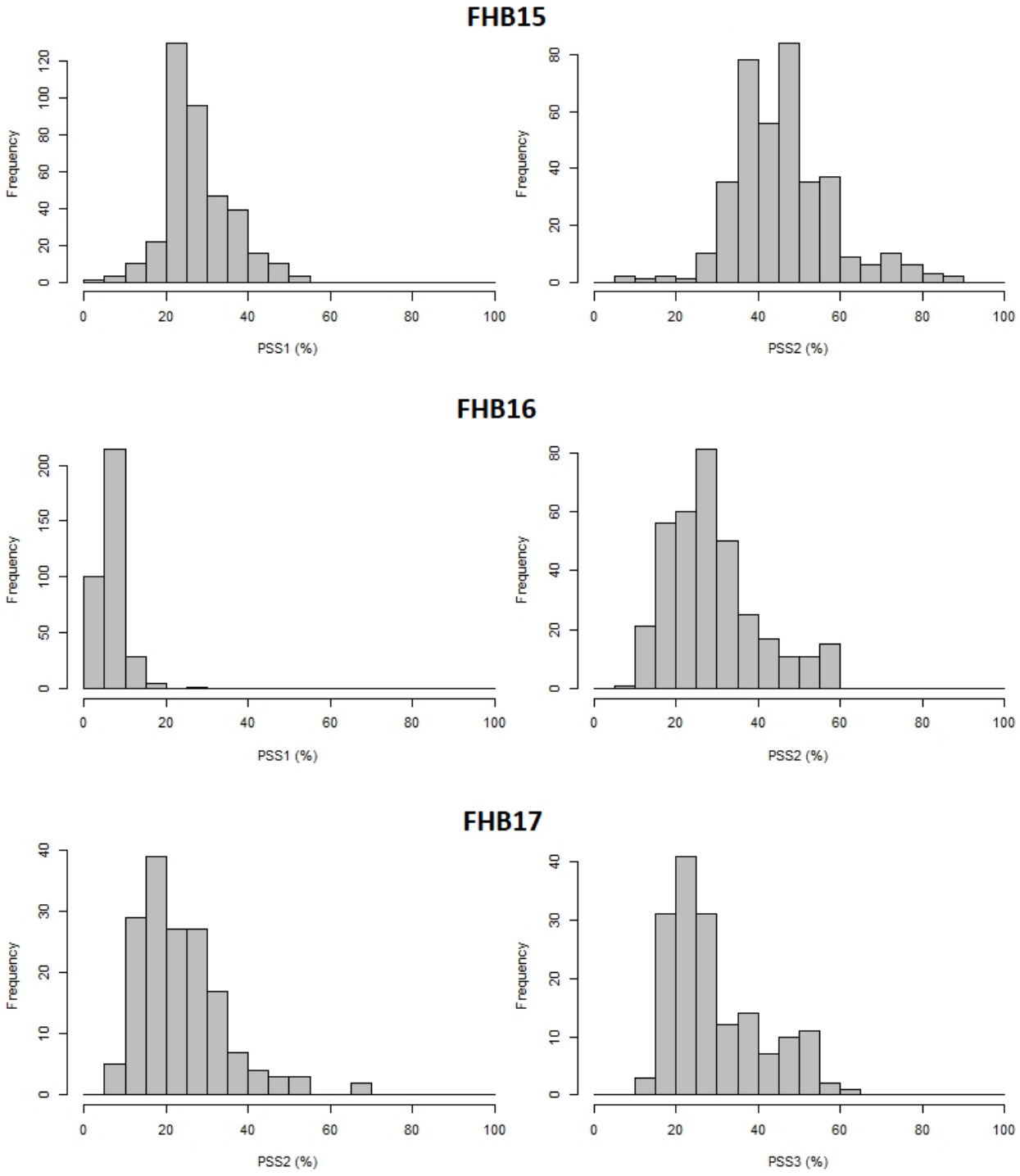

Figure 3-1. Normal distribution for the first, second, and third evaluation of percentage of symptomatic spikelets (PSS1, PSS2, PSS3) evaluated in the breeding panels phenotyped in 2015, 2016, and 2017.

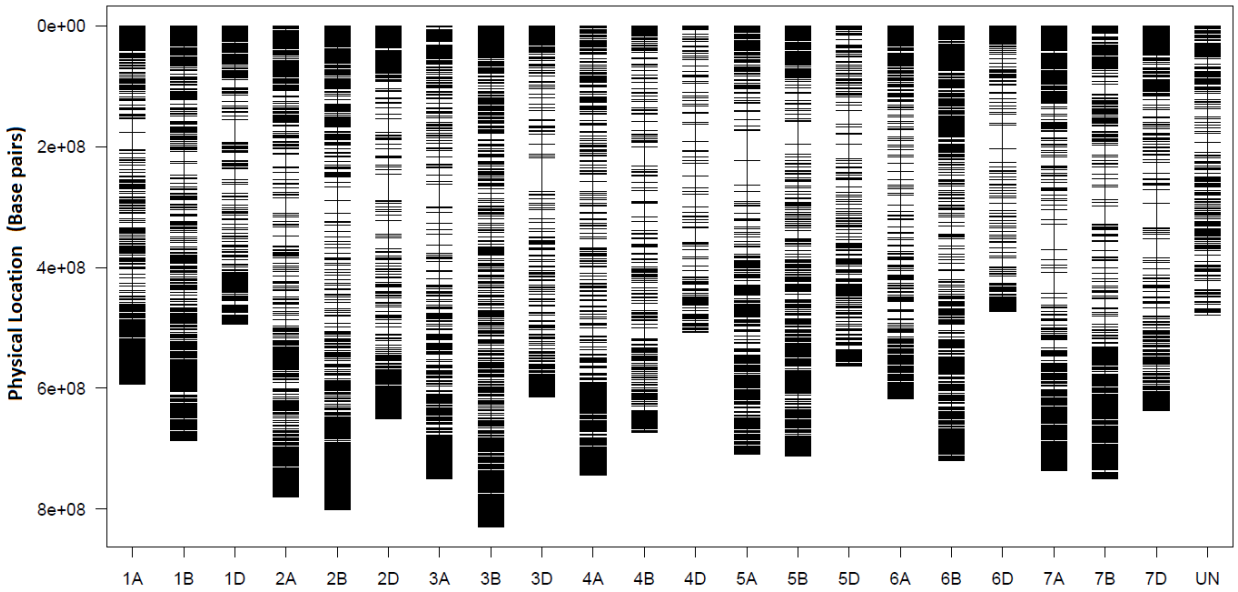

Figure 3-2. Physical map positions of 23,157 SNP markers identified in panel of 962 wheat lines. Map was drawn using the 'plotMap' function from the R/qtl package.

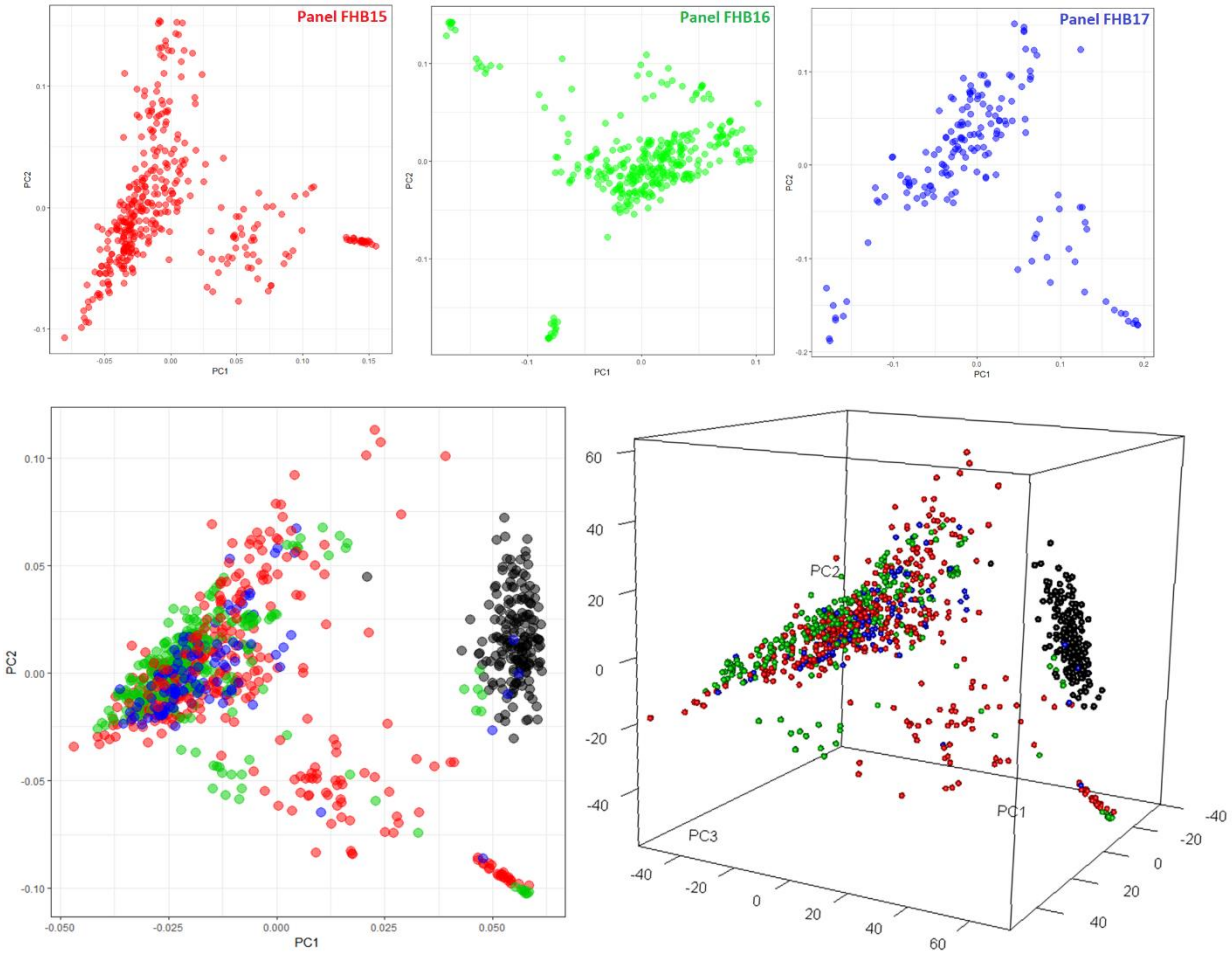

Figure 3-3. Principal components analysis of a wheat association panel based on 23,157 SNP markers drawn using the '*autoplot*' function from the ggfortify R package. Red, green, blue and gray dots represent breeding lines tested in 2015, 2016, 2017, and the Everest/Cedar population, respectively.

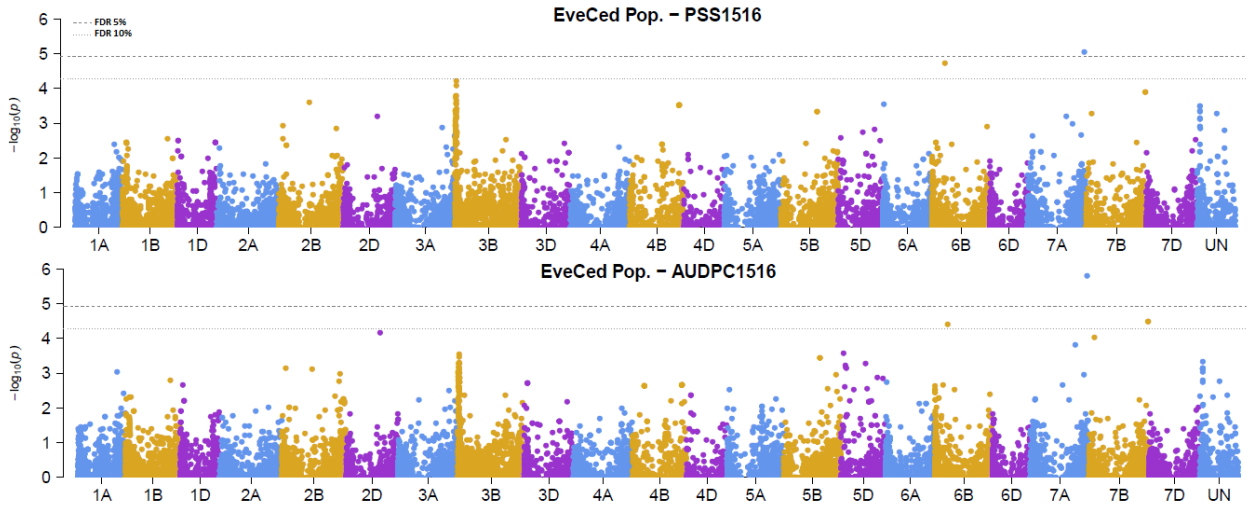

Figure 3-4. Manhattan plots showing association results for the population Everest/Cedar based on 23,157 common SNPs for AUDPC and PSS. The x-axis represents physical positions of the SNPs in the wheat genome and the y-axis represents the –log$_{10}$ of P-values.

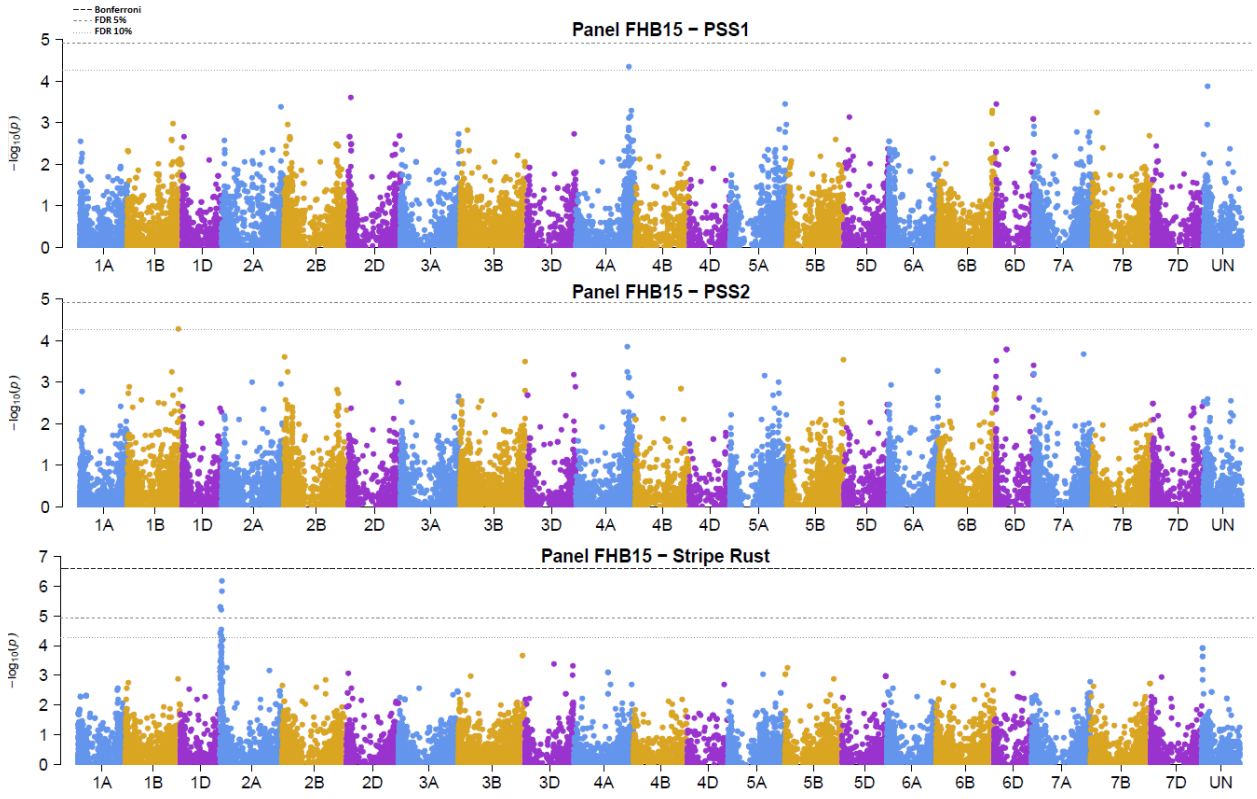

Figure 3-5. Manhattan plots showing association results for breeding panel phenotyped in 2015 based on 23,157 common SNPs for AUDPC and PSS. The x-axis represents physical positions of the SNPs in the wheat genome and the y-axis represents the –log$_{10}$ of P-values.

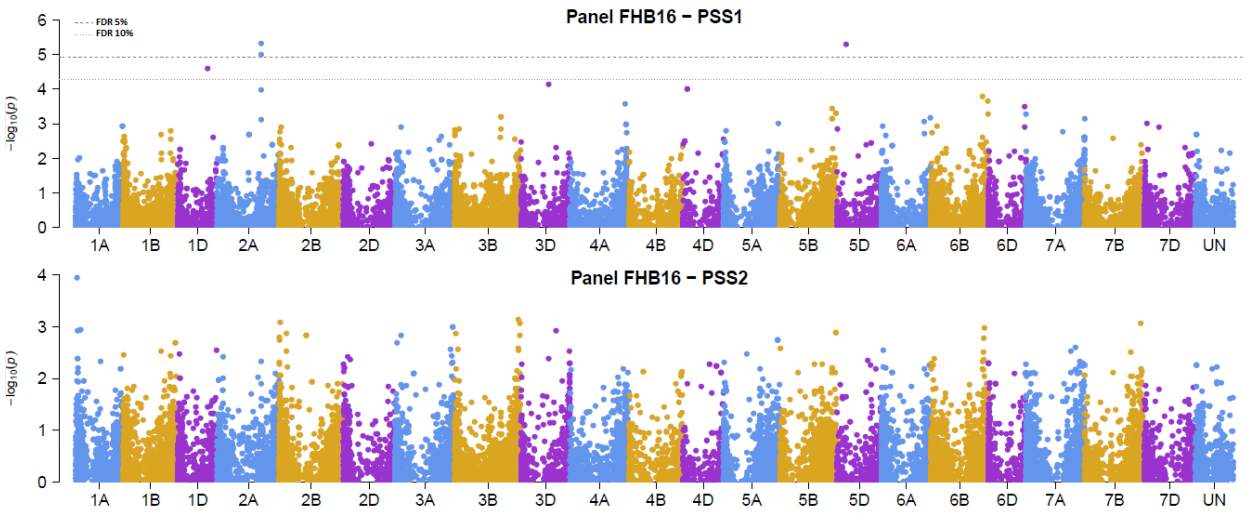

Figure 3-6. Manhattan plots showing association results for breeding panel phenotyped in 2016 based on 23,157 common SNPs for the first and second evolution of percentage of symptomatic spikelets (PSS1 and PSS2). The x-axis represents physical positions of the SNPs in the wheat genome and the y-axis represents the $-\log_{10}$ of P-values.

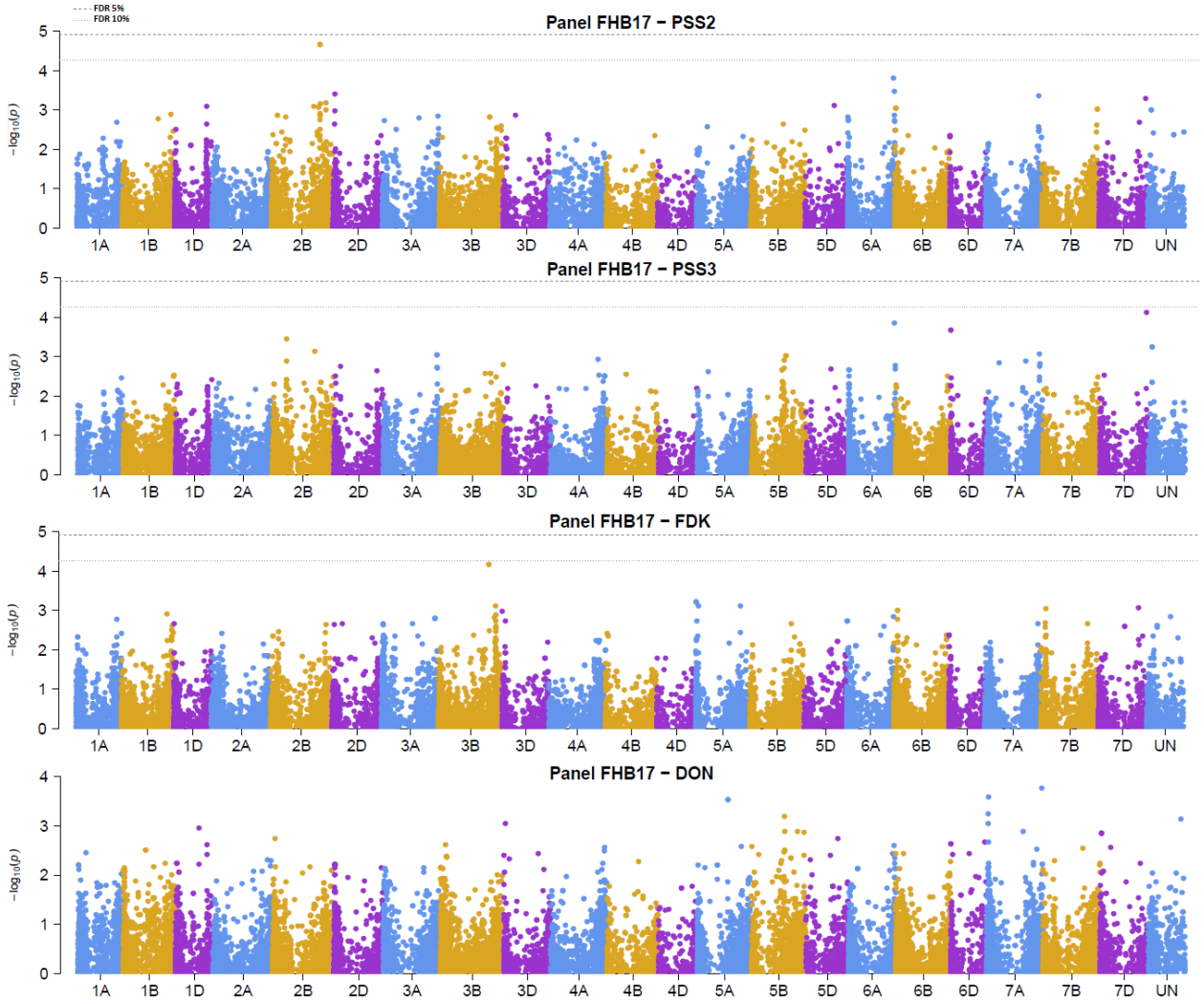

Figure 3-7. Manhattan plots showing association results for breeding panel phenotyped in 2017 based on 23,157 common SNPs for the first and second evolution of percentage of symptomatic spikelets (PSS1 and PSS2), Fusarium-damaged kernels, and deoxynivalenol content in ppm (DON). The x-axis represents physical positions of the SNPs in the wheat genome and the y-axis represents the $-\log_{10}$ of P-values.

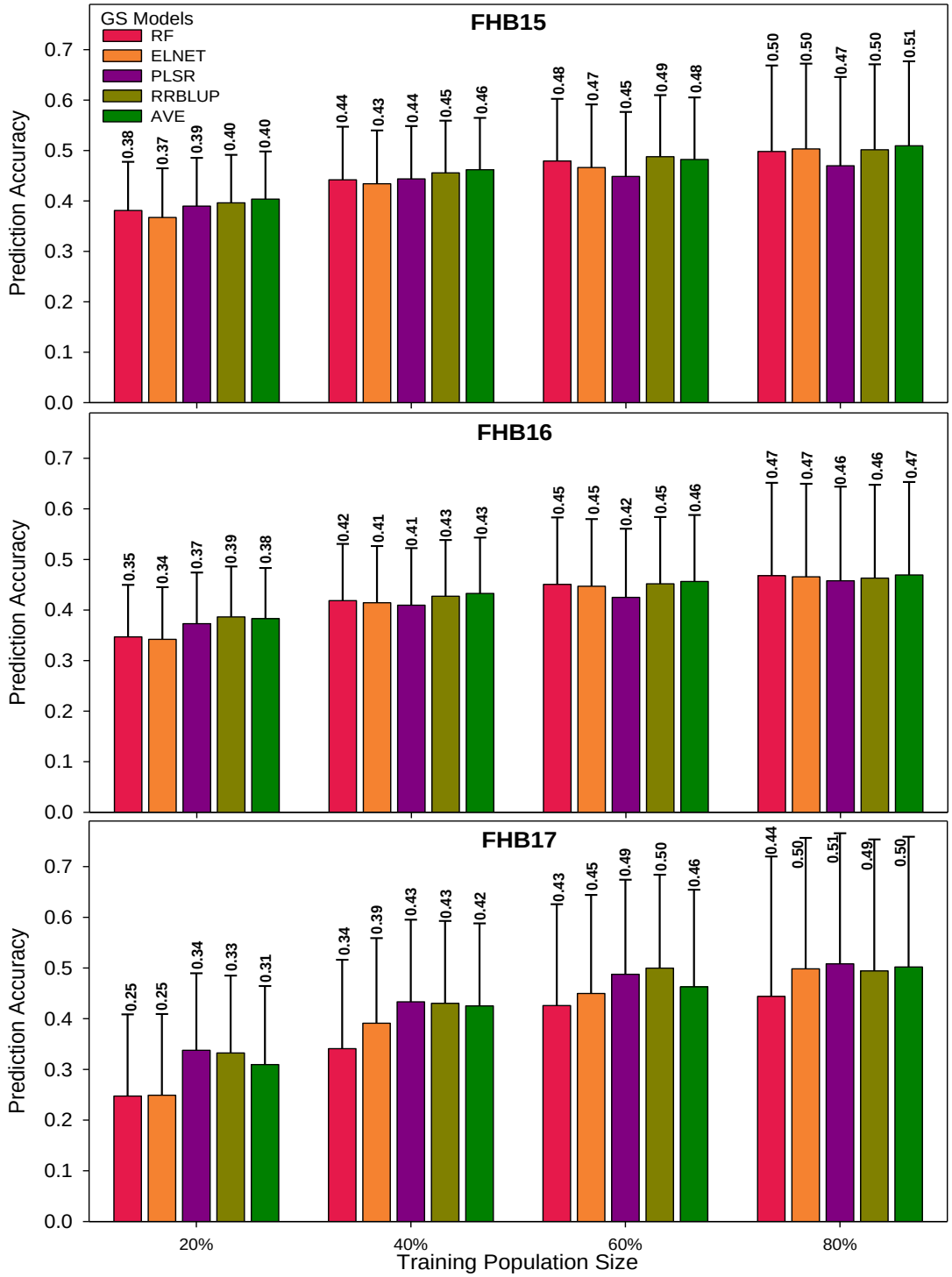

Figure 3-8. Prediction accuracies of four GS models assessed in elite breeding lines. RRBLUP: ridge regression best linear unbiased predictor; RF: random forest; PLSR: partial least squares regression; ELNET: elastic net; and AVE: average prediction across all four GS methods.

Table 3-1. Summary of the physical map of 23,157 SNP markers found in a panel of 962 wheat lines. Distances are measured in base pairs (bp).

Wheat chromosomes	Number of Markers	Length (bp)	Ave. Spacing (bp)	Max. spacing (bp)
1A	1,261	592,644,648	470,352.9	27,957,004
1B	1,428	687,224,147	481,586.6	21,024,379
1D	769	493,966,030	643,184.9	37,660,377
2A	1,310	779,984,717	595,863.0	24,793,520
2B	1,689	800,987,430	474,518.6	22,454,577
2D	1,046	650,089,090	622,094.8	43,165,540
3A	1,264	750,402,957	594,143.3	37,950,092
3B	1,724	829,321,630	481,324.2	9,263,962
3D	732	613,718,297	839,559.9	55,925,809
4A	1,113	743,640,867	668,741.8	17,679,614
4B	514	672,893,047	1,311,682.4	32,447,334
4D	229	506,896,734	2,223,231.3	27,857,968
5A	1,035	709,619,497	686,285.8	48,768,904
5B	1,229	712,388,397	580,120.8	37,170,534
5D	498	563,584,484	1,133,972.8	23,710,188
6A	1,025	617,227,849	602,761.6	17,013,863
6B	1,486	720,602,263	485,254.0	15,409,092
6D	520	472,863,651	911,105.3	40,233,947
7A	1,573	736,384,301	468,437.9	41,407,438
7B	1,374	749,497,273	545,882.9	30,390,060
7D	710	636,711,965	898,042.3	40,831,073
UN	628	479,274,591	764,393.3	16,029,401
Overall	23,157	14,519,923,865	627,617.2	55,925,809

Table 3-2. Values of broad-sense heritability (H^2) of percentage of symptomatic spikelets (PSS2) evaluated in three panels of elite winter wheat breeding lines.

GS Models	Panels	Training Population Sizes			
		20%	40%	60%	80%
RF	2015	0.76	0.72	0.67	0.66
	2016	0.63	0.57	0.55	0.52
	2017	0.67	0.90	0.95	0.97
ELNET	2015	0.74	0.74	0.71	0.66
	2016	0.65	0.60	0.55	0.52
	2017	0.64	0.89	0.92	0.96
PLSR	2015	0.76	0.76	0.71	0.66
	2016	0.69	0.60	0.55	0.52
	2017	0.64	0.88	0.95	0.96
rrBLUP	2015	0.74	0.73	0.68	0.66
	2016	0.67	0.60	0.55	0.52
	2017	0.81	0.85	0.93	0.96
AVE	2015	0.74	0.73	0.69	0.65
	2016	0.61	0.60	0.54	0.52
	2017	0.70	0.90	0.93	0.97

Table 3-3. Details of single nucleotide polymorphisms (SNPs) significantly associated PSS, AUDPC and stripe rust detected by the enhanced compression of the mixed-linear model.

Panel	Trait	SNP	Chr.	Position	p-values	MAF	R^2
EveCed	AUDPC1516	S7A_PART2_259768092	7A	709815078	1.51E-06	0.006	0.107
		S7B_PART2_286166108	7B	739988745	3.17E-05	0.020	0.078
		S7B_PART2_286166114	7B	739988751	3.17E-05	0.020	0.078
		S6B_PART1_159139004	6B	159139004	3.72E-05	0.023	0.077
		S2D_PART1_426927369	2D	426927369	6.52E-05	0.186	0.072
	PSS1516	S7A_PART2_259768092	7A	709815078	8.69E-06	0.057	0.093
		S6B_PART1_159139004	6B	159139004	1.77E-05	0.023	0.086
		S3B_PART1_17625393	3B	17625393	6.04E-05	0.429	0.075
		S3B_PART1_17994913	3B	17994913	8.19E-05	0.434	0.072
Panel FHB15	PSS1	S4A_PART2_208236032	4A	660791124	4.41E-05	0.126	0.037
	PSS2	S1B_PART2_200695538	1B	639415692	5.07E-05	0.023	0.036
	SR	S2A_PART1_24002751	2A	24002751	6.35E-07	0.459	0.042
		S2A_PART1_24002749	2A	24002749	1.39E-06	0.462	0.039
		S2A_PART1_2800603	2A	2800603	4.85E-06	0.432	0.035
		S2A_PART1_2800562	2A	2800562	4.85E-06	0.432	0.035
		S2A_PART1_2800596	2A	2800596	4.85E-06	0.432	0.035
		S2A_PART1_15449240	2A	15449240	6.28E-06	0.473	0.034
Panel FHB16	PSS1	S2A_PART2_89346135	2A	551722308	4.63E-06	0.072	0.050
		S5D_PART1_116655089	5D	116655089	4.86E-06	0.074	0.050
		S2A_PART2_90900748	2A	553276921	9.51E-06	0.063	0.047
		S1D_PART1_372880513	1D	372880513	2.50E-05	0.103	0.042
		S3D_PART1_350667570	3D	350667570	7.17E-05	0.139	0.037
		S4D_PART1_57896297	4D	57896297	9.99E-05	0.153	0.036
		S4D_PART1_57896305	4D	57896305	9.99E-05	0.153	0.036
	PSS2	S1A_PART1_15325346	1A	15325346	1.14E-04	0.493	0.035
Panel FHB17	PSS2	S2B_PART2_182562323	2B	635781247	2.11E-05	0.276	0.082
	PSS3	S7D_PART2_158473918	7D	612286186	7.34E-05	0.064	0.082
	FDK	S3B_PART2_210340155	3B	658495424	6.68E-05	0.261	0.145
		S3B_PART2_210340154	3B	658495423	6.68E-05	0.261	0.145

MAF: minor allele frequency; R^2: proportion of phenotypic variance explained by SNPs; FDR: false discovery rate.

Chapter 4 - A Genome-Wide Association Study of Stem and Leaf Rust Resistance in a Historical Dataset of Elite Breeding Lines

Abstract

Stem rust (SR) and leaf rust (LR) are wheat diseases that cause substantial economic losses to global wheat production. New races of these of pathogens are constantly emerging which makes breeding for resistance more challenging. Therefore, the objective of this study was to identify genomic regions associated with resistance to wheat rusts at the seedling stage. For this purpose, we used historical data of advanced breeding lines from the Southern Regional Preliminary Nursery (SRPN). This nursery tests dozens of advanced wheat lines from public and private breeding programs every year. Genotyping-By-Sequencing (GBS) was used to identify 35,467 single nucleotide polymorphisms (SNPs) in 533 unique breeding lines were tested in the SRPN nursery from 2000 to 2015. A total of 51 LR races and 34 races SR races were separately inoculated under controlled conditions during this period of time. A QTL on 2AS was repeatedly found associated with low infection types for multiple races of stem and leaf rust. After comparing SNPs at this QTL peak with the SSR marker *ventriup-ln2*, we confirmed that this locus corresponds to the 2NS·2AS translocation from *Ae. ventricosa.* Two loci located on 2DS and 4AL were highly associated with resistance to the stem rust races TPMKC and TTTTF, respectively. Another four QTLs on 1BS, 2DS, 3AL, and 3BL were also associated with broad-spectrum resistance to stem and leaf rust. Another three QTLs located on 4AL, 4BL, and 7AS are possibly novel loci associated with stem and leaf rust resistance at seedling stage.

4.0 Introduction

Wheat rusts are among the most destructive wheat diseases causing significant economic losses globally. In addition, the recurrent emergence of new races complicates deployment of durable resistance. Stem (SR), leaf (LR), and stripe rust (YR) are caused by the pathogen species *Puccinia graminis f. sp. tritici* (*Pgt*), *P. triticina* (*Pt*), and *P. striiformis f. sp tritici* (*Pst*) respectively. There are two main strategies to control cereal rusts: genetic resistance and use of fungicides. Host plant resistance has economic and environmental advantages (Ellis *et al*., 2014) and, for these reasons, it is widely used as a management practice to control wheat rusts.

Stem rust (SR) is undeniably the most aggressive of all three rusts, with the emergence of highly aggressive races. For instance, the race TTKSK (Ug99) became virulent against the gene *Sr31* and since then variants of this race defeated the resistance provided by several genes deployed in Africa and Asia, (Singh *et al*., 2015), representing a threat to global wheat production and food security. Extensive efforts have been made to find and map new sources resistance for these highly virulent races, as a result, several minor effect-loci have been recently identified (Mohammadi *et al*., 2013; Singh *et al*., 2014).

Deployment of cultivars containing multiple resistance genes has proven to be the most effective strategy to prevent widespread epidemics of wheat rusts (Singh *et al*., 2016). Hence, wheat breeders and pathologists are constantly developing and testing new sources of germplasm and elite lines to find resistance genes that are broadly effective against all rust species and their race variations. Other breeding and research approaches have included backcrossing, combining genes of race-nonspecific and race-specific resistance (Singh *et al*., 2005), association studies with elite lines and global collections (Zhang *et al*., 2014; Gao *et al*., 2017), and genomic

prediction of resistance to wheat rusts (Rutkoski *et al*., 2011; Rutkoski *et al*., 2015; Juliana *et al*., 2017).

The Southern Regional Performance Nursery (SRPN) is a public nursery established by the USDA-ARS in 1932 to characterize the performance of advanced winter wheat lines from multiple breeding programs in the Central Plains of the United States (Reitz and Salmon, 1959). Every year, between 35 and 50 new entries are tested in this nursery for grain yield, end-use-quality and wheat diseases with phenotypic data being collected across more than 20 locations. In addition, all lines are inoculated for multiple races of wheat rusts in greenhouse conditions at the USDA Cereal Disease Lab in Saint Paul, MN. In the meantime, these breeding lines were also screened for important known genes at the USDA Small Grain Genotyping Lab in Manhattan, KS. The majority of samples submitted to the nursery since 2000 have been genotyped via GBS (Rife, 2016). Therefore, the SRPN can be considered one of the most valuable resources to study the genetics of winter wheat cultivars and elite lines in the United States.

In this chapter, sixteen years of historical data from the SRPN (2000-2015) were used to perform a genome-wide association study of seedling resistance for multiple races of *Pgt* and *Pt*. This effort should effectively identify genomic regions associated with race-specific and broadly effective resistance in early stage of development while the effect of adult plant resistance (APR) genes such as *Lr34*, *Lr46*, *Lr67*, and *Lr68* will likely not be detected. Afterward, identifying QTLs/genes broadly effective at the seedling stage will assist breeding for rust resistance since these loci may be used in combinations with the well-known APR genes.

4.1 Materials and Methods

The SRPN Historical Dataset

A total of 687 unique winter wheat breeding lines from multiple breeding programs were submitted to the (SRPN) from 2000 to 2015. All the historical data is available online at https://www.ars.usda.gov/Research/docs.htm?docid=16642. For this study, we used a subset of 533 breeding lines that were inoculated for wheat rusts. A few lines were submitted to the SRPN for more than one year and, in some cases, tested for more than one race of wheat rust. Infection types were averaged across years in these cases.

Inoculation and Infection Types

The inoculation of multiple races of wheat rusts was conducted under controlled conditions at the USDA Cereal Disease Laboratory in St. Paul, MN. A total of 51 races of *Pt* and 34 races of *Pgt* were inoculated onto breeding lines submitted to the SPRN in the time period under consideration (2000-2015) (Table 4-1). Only field notes of stripe rusts from unknown races were available, thus it was not included in the association analysis. Each race of *Pt* and *Pgt* was separately inoculated onto wheat lines in the greenhouse. Then, two weeks after inoculation, the infection types (ITs) were scored using the Stakman scale (Stakman, 1962). Screening with the *Pgt* race TTKSK (Ug99) began in 2008.

The IT scores can be converted into a linear scale using the algorithm proposed by Zhang *et al*., (2014). A more recent study developed a fully automated pipeline in Perl (Gao *et al*., 2016) to facilitate the conversion of IT scores using the prior algorithm with slight modifications. This more recent algorithm was implemented in our study. Subsequently, IT scores were averaged across races for each pathogen, with the exception of TTKSK which was analyzed separately.

Genotyping-by-Sequencing

All breeding lines tested in the SRPN were genotyped using the genotyping-by-sequencing (GBS) protocol described by Poland *et al*. (2012). Single nucleotide polymorphisms (SNPs) were called using an automated pipeline in Tassel 5. Only those SNP markers with minor allele frequency (MAF) greater than 1%, heterozygosity lower than 15%, and less than 50% missing values across genotypes were retained. Markers that yielded multi-allelic calls were discarded. A total of 35,467 SNPs and 687 wheat lines remained in the genotypic dataset after filtering. The physical positions of SNP markers were corrected for each chromosome using the *161010_Chinese_Spring_v1.0* pseudomolecule reference (IWGSC, 2017). SNP markers were ordered from the distal region of short arm to the distal part of the long arm within each chromosome.

Genome-Wide Association Analysis

The association mapping analysis (GWAS) was conducted in R using the Genome Association and Prediction Integrated Tool (GAPIT) (Tang *et al*., 2016). We used an enhanced compression of the mixed linear model (ECMLM) (Li *et al*., 2014). The Bayesian information criterion (BIC) was estimated by setting the parameter '*Model.selection=TRUE'* to determine the optimal number of principal components (PCs) for the association analysis. The kinship matrix and the first two principal components were included in the model as covariates to account for the population structure.

GWAS results from GAPIT were reloaded into R using the package 'qqman' to generate Manhattan plots. To declare significant SNP-trait associations, we use a multiple correction test with three different thresholds of significance: false discovery rate (FDR) at 5% and 10%, and the Bonferroni correction. FDR values were obtained from the GWAS analysis run with rrBLUP

whereas values for the Bonferroni correction was obtained with the default output from GAPIT. This last method is highly conservative, limiting the identification of significant marker-trait associations, as reported by several authors (Sham & Purcell, 2014; Gao *et al*., 2016) which justified the adoption of a less stringent test such as the FDR.

4.2 Results and Discussion

Genotypic Data of SRPN

The genotypic data consisted of 35,467 SNP markers and 687 breeding lines from multiple breeding programs that were tested in the Southern Regional Performance Nursery (SRPN) from 2000 to 2015. A larger proportion of single nucleotide polymorphisms (SNPs) was identified on the A genome (37.5%) and B (44%) while only 18.5% of SNPs were assigned to the D genome (Table 4-2). The average distance between markers was 0.39 megabases (Mb) with the largest gap (34.9 Mb) found near to the centromere of chromosome 2B (Table 4-2). SNPs were more densely distributed towards the telomeres (Figure 4-2) which confirms the trend reported by several studies of higher recombination rates and gene density being more concentrated in distal regions (IWGSC, 2014; Avni *et al*., 2017). Although the average of infection types (ITs) for all races within rust species tended to be normally distributed, ITs of individual races were observed to fit a bimodal distribution with a large frequency of lines falling into the susceptible (9) and resistant (1) classes (Figure 4-1). Phenotypic data were not transformed prior association mapping analysis. The first two principal components explained the majority of genetic variation and a strong population structure was not observed (Figure 4-3).

GWAS Results for Stem Rust

A total of 34 unique races of *Pgt* were separately inoculated on the entries submitted to the SRPN between 2000 and 2015. TTTTF, TPMKC, and TTKSK were the most commonly inoculated races of *Pgt* and were analyzed separately (Table 4-1). The IT scores of all races, except TTKSK, were also averaged (bulk) with the goal of identifying SNPs associated with broad-spectrum seedling resistance. The number of QTLs identified varies according to the multiple correction tests used to declare the significance of SNP-trait associations (Table 4-4). For the Bonferroni correction threshold which is the most conservative, we only identified two QTLs (2AS, 2DS) significantly associated with the average IT scores of 33 races of *Pgt*, whereas another nine QTLs were identified when using a false discovery rate of 5% (Figure 4-4 and Table 4-4).

The minor alleles of two SNPs at the QTL peak on 2AS (S2A_PART1_2800562 and S2A_PART1_2336941) matched with the results of the marker *ventriup-ln2* for the translocation 2NS·2AS, confirming that the QTL on 2AS corresponds to this alien segment. Additionally, the physical position of these SNPs (at ~23 and ~28Mb away from the distal end of 2AS) indicates that they are in fact within the 2NS segment. The MAF for these two SNPs ranged from 0.36 to 0.40, showing that this translocation is present in a high frequency in elite winter wheat lines across several breeding programs in the United States. This translocation carries the *Yr17/Lr37/Sr38* gene cluster and is widely known for conferring resistance to multiple wheat pathogens at seedling and adult stages, including stripe rust (Helguera *et al*., 2003), leaf rust (Kolmer, 2017), stem rust (Mohammadi *et al*., 2013), wheat blast (Cruz *et al*., 2016), and even nematode resistance (Williamson *et al*., 2013). Jagger, a highly successful hard winter wheat variety that is prominent in the pedigree of many U.S. hard winter wheat lines developed in the

Central and Southern Plains, was the primary source of the 2NS·2AS translocation in the SRPN materials (Table 4-1). The QTL on 2DS mapped in the interval where the genes *Lr2a* and *Sr6* are located. These two genes are well-known for conferring resistance to leaf and stem rust at seedling stages (Tsilo *et al*., 2014) and are commonly present in U.S. winter wheat lines (Kolmer *et al*., 2007; Zhang *et al*., 2014). Therefore, these genes are the most likely candidates for the QTL associated with broad stem rust resistance (average of all races) at seedling stage found in this study on the distal region of 2DS.

TTTTF is one of the most widely virulent *Pgt* races to be identified in the United States and produces high infection types on all stem rust differential lines (Roelfs & Martens, 1988) whereas TPMKC was the predominant race in all regions of the United States during the late 1990's (McVey *et al*., 2002). In contrast, TTKSK (Ug99) emerged in Uganda and has currently spread to several countries within Africa and Asia but is not yet present in North nor South America. A highly significant QTL was identified on the distal end of 4AL, which was particularly associated with low infection types of the race TTTTF. *Sr7a* and *SrND643* were previously reported in this genomic region (Zhang *et al*., 2014; Tsilo *et al*., 2014). The gene in this study is most likely *Sr7a* since it was recently confirmed to confer resistance to SR in the cultivar Jagger (Turner *et al*., 2016). As noted previously, Jagger was extensively used as a parent in the hard winter wheat region of the U.S (Table 4-4). Also, this region has been recently reported to confer resistance to TTKSK (Basnet *et al*., 2015; Yu *et al*., 2017). These results were verified in our study (Figure 4-4) when a less stringent threshold of significance (FDR 10%) was considered.

A strong association was identified on 4AL for resistance to race TPMKC. As previously noted, this region contains the genes *Sr7a* and *SrND643* (Figure 4-4). At least two independent

dominant genes were reported to confer resistance to TPMKC, but these genes are not effective against TTTTF (Oliveira *et al*., 2008). These results were partially confirmed in our study, since the QTL detected on 2DS was significantly associated with resistance to both races. Considering the results from the avirulence/virulence formula (Roelfs & Martens, 1988), the gene *Sr6* the most likely candidate for this locus (Table 4-3). Four QTLs associated with resistance to TTTTF (1AS, 1BS, 2AL, 4AL) were identified using FDR 5% as the significance threshold whereas only two QTLs (1AS, 2DS) were associated with resistance to TPMKC.

Regarding TTKSK, the most significant association was verified on the distal end of 3AL (Figure 4-4). This genomic region contains the genes *Sr27* and *Sr35*. However, *Sr27* is present in a rye translocation that is not present in this panel of wheat lines. Similarly, *Sr35*, which has been recently cloned, and it is known for conferring near immunity to Ug99 and related races (Saintenac *et al*., 2013) but is an improbable candidate as it has not been widely deployed in the hard winter wheat region of the U.S. Therefore, the significant locus found in our study on this genomic region is likely novel. Another five QTLs associated with TTKSK were found on 1AL, 1BS, 1DS, 3DL and 7BL which can potentially correspond to $1RS^{Amigo}$, *Sr31*, *Sr33/Sr45*, *Lr24*/*Sr24* and *Lr68*, respectively. The linked genes *Lr24/Sr24* came from translocation from *Ag. elongatum* to the long arm of 3D and are known for providing resistance to multiple races of *Pgt* and *Pt*, including TTKSK (Smith *et al*., 1968; Mago *et al*., 2005; Imbaby *et al*., 2014). However, comparing the genotypes results of the SNPs at peak locus found in our study with marker Sr24#12 which is linked to *Lr24*/*Sr24*, it was not observed a clear trend, suggesting the occurrence of marker failure to detect this alien segment, or spurious associations since the gene *Sr24* was considered effective against the races TTTTF, TPMKC, and TTKSK according to the avirulence/virulence postulation (Roelfs & Martens, 1988) presented in Table 4-3.

Lr68 is an APR race-nonspecific gene known for conferring durable resistance to all three rusts and frequently found in spring wheats originated from CIMMYT (Herrera-Fosessel *et al*., 2012). Although further evidence is required to confirm whether the QTL we identified on 7BL actually is *Lr68.* It would be the first time that this gene was found associated with broad spectrum stem rust resistance in the U.S winter wheat. Several of these genes, especially those effective against TTKSK, were previously found in a relatively high frequency in U.S. wheat cultivars (Zhang *et al*., 2014). Similar SNP- trait associations for Ug99 races in spring wheat (Yu *et al*., 2017) corroborate with our findings. Considering the SNP associations for the bulk of *Pgt* races (average of all IT scores) at FDR<10%, another four QTLs were identified on 1BL, 2BL, 5DL, and 6DL. Three of these loci associated with broad-spectrum stem rust resistance at the seedling stage are potentially novel since the locus on 1BL is likely the 1BL·1RS translocation. A locus on 2BL associated stem rust resistance have been recently identified in a global spring wheat germplasm collection (Gao *et al*., 2017) and it is speculated to be novel.

GWAS Results for Leaf Rust

A total of 51 unique races of *Pt* were separately inoculated on the entries submitted to the SRPN between 2000 and 2015. The *Pt* races THBJ, TNRJ, and KFBJ were repeatedly inoculated during multiple years, therefore they were analyzed separately. The average of IT scores of all races (bulk) were used in the GWAS analysis to identify SNPs associated with resistance against multiple races of leaf rusts (Figure 4-5 and Table 4-4). Only one QTL on 2DS was found associated with resistance to multiple races of *Pgt* using the Bonferroni multiple correction test. As discussed in the stem rust section, *Sr6* have been previously mapped to this region. Additionally, another three seedling-resistant *Lr* genes (*Lr2a*, *Lr15*, and *Lr39*) and one APR gene (Lr22a) have been detected on the short arm of 2D (Lan *et al*., 2017). According to the

avirulence/virulence formula (Long & Kolmer, 1989), all these three *Pt* races are virulent to *Lr2a* (Table 4-3). Therefore, the genes *Lr15* and/or *Lr39* are the most likely candidate(s) found in our study on this genomic region.

Using the SNP significance of FDR<5% allowed identification of another five QTLs located on 2AS, 3BS, 7AS, 7DS, 4BL (Figure 4-5 and Table 4-4). The genes *Lr37* (located on the 2NS·2AS segment), *Lr27/Sr2, Lr47,* and *Lr34* are the most likely candidates for the first four QTLs, whereas the QTL on 4BL is potentially a novel locus for leaf rust resistance. Gao *et al*. (2016) also reported a QTL located in a similar region of 4BL which was associated with APR but not seedling resistance against *Pt* races. The locus on 7AS may also be novel as *Lr47* was introgressed from *Ae tauschii* and is unlikely to be common in U.S wheat cultivars. Similar findings were also recently reported in a spring wheat core collection (Turner *et al*., 2017).

No significant marker associations were observed for the *Pt* races THBJ, TNRJ, and KFBJ when using the multiple test correction of Bonferroni and FDR at 5% as the threshold for significance (Figure 4-5). THBJ was commonly found in the Great Plains during early 2000's (Kolmer *et al*., 2004) and is virulent against *Lr16*, *Lr9*, and *Lr24* (Oelke & Kolmer, 2004). Five QTLs associated with low infection types for the race THBJ (Figure 4-5) and another eleven QTLs associated with resistance to race TNRJ were identified in this work. TNRJ is virulent to *Lr9*, *Lr10*, *Lr11*, *Lr24*, and *Lr41* and occurs primarily found in the southern Great Plains (Kolmer *et al*., 2007) while KFBJ is virulent to *Lr26* and it is considered one of most virulent races of *Pt* in the United States (German & Kolmer, 2012; Bruce *et al*., 2014). A total of five QTL were identified for KFBJ in our study. Although several of these loci mapped within the interval of known rust resistance genes, a larger proportion of this loci are speculated to be novel genes.

All IT scores were averaged across 34 races of *Pgt* and 51 races of *Pt* to identify genomic regions associated with race-nonspecific resistance at seedling stage (Figure 4-6). The GWAS results revealed a highly significant (Bonferroni 5%), single locus located on 2AS which was associated with the broad spectrum resistance of stem and leaf rust at seedling stage. As previously discussed, we confirmed that this locus corresponds to the 2NS·2AS translocation, which is widely known for providing resistance to multiple wheat diseases (Helguera *et al*., 2003; Mondal *et al*., 2016). Another four QTLs on 1BS, 2DS, 3AL, and 3BL were also associated with average IT over races when a significant threshold of FDR 5% was used. So far, there is no of evidence in the literature whether seedling resistance genes can be expected to provide durable resistance. However, finding the ones that are effective against multiple races has its importance reported by other studies (Mago *et al*., 2005) in terms of designing and deploying combinations with APR genes that can lead to more durable resistance.

4.3 Conclusions

The number of QTLs detected in the GWAS analysis varied from 1 to 15 according to the strictness of the multiple correction test applied. Several loci were identified in the interval of previously characterized genes, confirming the effectiveness of these prior reported genes against multiple races of *Pgt* and *Pt* pathogen species. Occasionally, significant loci were mapped near newly introgressed genes which are unlikely to be commonly found in winter wheat elite lines.

A locus on 2AS was repeatedly found associated with low infection types for multiple races of stem and leaf rust. After comparing genotypes from SNPs at this locus with the SSR marker *ventriup-ln2*, we confirmed that this locus corresponds to the 2NS·2AS translocation from *Ae. ventricosa*. This segment is universally known among wheat breeders and pathologists

for conferring resistance against multiple wheat diseases. So far, no yield or quality penalties have been reported in the literature due to the presence of 2NS·2AS, demonstrating that this alien segment brings multiple benefits breeding programs.

Another two loci located on 2DS and 4AL were highly associated with resistance to Pgt races TPMKC and TTTTF, respectively. These loci were also significantly associated with low infection types for the bulk of races *Pgt* and *Pt*. There are several *Lr* and *Sr* genes located in these genomic regions. However, due to the absence of marker data, it was not possible to identify the actual genes conferring resistance at these loci. Moreover, several other potentially novel loci associated with rust resistance at seedling resistance were identified in this study. Further research is needed for validation of these loci. From a breeding perspective, identifying novel resistant loci in elite winter lines is advantageous in that sources of resistance are available in adapted backgrounds without deleterious effects on yield and end-use quality.

References

Basnet B.R., Singh S., Lopez-Vera E.E, Huerta-Espino J., Bhavani S., Jin Y., Rouse M. N., Singh R.P. (2015). Molecular Mapping and Validation of *SrND643*: A New Wheat Gene for Resistance to the Stem Rust Pathogen Ug99 Race Group. Phytopathology 105(4): 470-476.

Bruce M., Neugebauer K.A., Joly D.L., Migeon P., Cuomo C.A., Wang S., Akhunov E., Bakkeren G., Kolmer J.A. and Fellers J.P. (2014). Using transcription of six *Puccinia triticina* races to identify the effective secretome during infection of wheat. Front. Plant Sci. 4(520): 1-7.

Cruz, C. D., Peterson, G. L., Bockus, W. W., Kankanala, P., Dubcovsky, J., Jordan, K. W., Valent, B. (2016). The 2NS Translocation from *Aegilops ventricosa* Confers Resistance to the Triticum Pathotype of *Magnaporthe oryzae*. Crop Science 56(3): 990–1000.

Ellis J.G., Lagudah E.S., Spielmeyer W., Dodds P.N. (2014). The past, present and future of breeding rust resistant wheat. Front Plant Sci. 5(641): 1-13.

Gao L., Rouse M. N., Mihalyov P.D., Bulli P., Pumphrey M.O., Anderson J.A. (2017). Genetic Characterization of Stem Rust Resistance in a Global Spring Wheat Germplasm Collection. Crop Science 57: 2575-2589.

Gao L., Turner M.K., Chao S., Kolmer J., Anderson J.M. (2016). Genome Wide Association Study of Seedling and Adult Plant Leaf Rust Resistance in Elite Spring Wheat Breeding Lines. PLoS One 11(2): e0148671.

Gao X., Starmer J., Martin E.R. (2008). A Multiple Testing Correction Method for Genetic Association Studies Using Correlated Single Nucleotide Polymorphisms. Genetic Epidemiology 32(4): 361-369.

German S.E., Kolmer J.A. (2012). Leaf Rust Resistance in Selected Uruguayan Common Wheat Cultivars with Early Maturity. Crop Science 52: 601-608.

Helguera, M., Khan I.A., Kolmer J., Lijavetzky D., Zhongqi L., and Dubcovsky J. (2003). PCR assays for the *Lr37-Yr17- Sr38* cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. Crop Science 43(5): 1839-1847.

Herrera-Foessel, S.A., Singh, R.P., Huerta-Espino, J. *et al*. (2012). *Lr68*: a new gene conferring slow rusting resistance to leaf rust in wheat. Theor Appl Genet. 124(8): 1475-1486.

Imbaby, I. A., Mahmoud, M. A., Hassan M. E. M., Abd-El-Aziz A. R. M. (2014). Identification of Leaf Rust Resistance Genes in Selected Egyptian Wheat Cultivars by Molecular Markers. The Scientific World Journal 2014: 574285.

Juliana P., Rutkoski J.E., Poland J.A., Singh R.P., Murugasamy S., Natesan S., Barbier H., Sorrells M.E. (2014). Genome-Wide Association Mapping for Leaf Tip Necrosis and

Pseudo-black Chaff in Relation to Durable Rust Resistance in Wheat. The Plant Genome 8(2):1-12.

Juliana P., Singh R.P., Singh P.K. Crossa J., Huerta-Espino J., Lan C., Bhavani S., Rutkoski J.E., Poland J.A., Bergstrom G.C., Sorrells M.E. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. Theor Appl Genet. 130(7): 1415-1430.

Kolmer J.A. (2017). Genetics of Leaf Rust Resistance in the Hard Red Winter Wheat Cultivars Santa Fe and Duster. Crop Science 57(5):2500-2505.

Kolmer J.A. Jim Y., Long L. (2007). Wheat leaf rusts and stem rust in the United States. Australian Journal of Agricultural Research 58: 631-638.

Kolmer J.A., Long D.L., and Hughes M.E. (2004). Physiologic specialization of *Puccinia triticina* on wheat in the United States in 2002. Plant Disease 88(10): 1079-1084.

Lan C., Hale I. L., Herrera-Foessel S. A., Basnet B. R., Randhawa M. S., Huerta-Espino J., Singh R. P. (2017). Characterization and Mapping of Leaf Rust and Stripe Rust Resistance Loci in Hexaploid Wheat Lines UC1110 and PI610750 under Mexican Environments. Frontiers in Plant Science 8(1450): 1-11.

Long D.R, & Kolmer J.A. (1989). A North American System of Nomenclature for *Puccinia triticina*. Phytopathology 79:525-529.

Mago R., Bariana H. S., Dundas I. S., Spielmeyer W., Lawrence G. J., Pryor A. J., Ellis J. G. (2005). Development of PCR markers for the selection of wheat stem rust resistance genes *Sr24* and *Sr26* in diverse wheat germplasm. Theor Appl Genet. 111(3): 496-504.

McVey D. V., Long D. L., and Roberts J. J. (2002). Races of *Puccinia graminis* in the United States during 1997 and 1998. Plant Disease 86():568- 572.

Mohammadi M., Torkamaneh D., Patpour M. (2013). Seedling stage resistance of Iranian bread wheat germplasm to race Ug99 of *Puccinia graminis f. sp. tritici*. Plant Disease 97(3): 387-392.

Mondal S., Rutkoski J. E., Velu G., Singh P. K., Crespo-Herrera L. A., Guzmán C., Singh R. P. (2016). Harnessing Diversity in Wheat to Enhance Grain Yield, Climate Resilience, Disease and Insect Pest Resistance and Nutrition Through Conventional and Modern Breeding Approaches. Frontiers in Plant Science 7(991): 1-15.

Mundt C.C. (2014). Durable resistance: A key to sustainable management of pathogens. Infection. Genetics and Evolution 27: 446-455.

Oelke L.M., & Kolmer J.A. (2004). Characterization of leaf rust resistance in hard red spring wheat cultivars. Plant Disease 88(10):1127-1133.

Oliveira P. D., Millet E., Anikster Y., and Steffenson B. J. (2008). Genetics of resistance to wheat leaf rust, stem rust, and powdery mildew in *Aegilops sharonensis*. Phytopathology 98(3): 353-358.

Poland J.A., Balint-Kurt P.J., Wisser R.J., Pratt R.C., Nelson R.J. (2009). Shades of gray: the world of quantitative disease resistance. Trends Plant Science 14(1): 21-29.

Poland J.A., Brown P.J., Sorrells M.E., Jannink J.L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PloS One 7(2): e32253.

Reitz L.P., & Salmon S.C. (1959). Hard red winter wheat improvement in the plains: A 20-year summary. Tech. Bull. 1192. USDA. U.S. Gov. Print. Office, Washington, DC.

Roelfs A. P., & Martens J. W. (1988). An international system of nomenclature for *Puccinia graminis f. sp. tritici*. Phytopathology 78:526- 533.

Rutkoski J. E., Poland J. A., Singh, J. Huerta-Espino R. P., Bhavani S., Barbier H., Rouse M. N., Jannink J., Sorrells M. E. (2014). Genomic Selection for Quantitative Adult Plant Stem Rust Resistance in Wheat. The Plant Genome 7(3): 1-10.

Rutkoski J., Singh R.P., Huerta-Espino J., Bhavani S., Poland J., Jannink J.L., Sorrells M.E. (2015). Genetic Gain from Phenotypic and Genomic Selection for Quantitative Resistance to Stem Rust of Wheat. The Plant Genome 8(2): 1-10

Rutkoski J.E., Heffner E. L., Sorrells M.E. (2011). Genomic selection for durable stem rust resistance in wheat. Euphytica 179(1): 161-173.

Saintenac C., Zhang W., Salcedo A., Rouse M.N., Trick H.N., Akhunov E., Dubcovsky J. (2013). Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. Science 341(6147): 783-786.

Singh R.P., Herrera-Foessel S., Huerta-Espino J. Singh S., Bhavani S., Lan C., Basnet B.R. (2014). Progress Towards Genetics and Breeding for Minor Genes Based Resistance to Ug99 and Other Rusts in CIMMYT High-Yielding Spring Wheat. Journal of Integrative Agriculture 13(2): 255-261.

Singh R.P., Huerta-Espino J., Willian H.M. (2005). Genetics and Breeding for Durable Resistance to Leaf and Stripe Rusts in Wheat. Turk J Agric For. 29:121-127.

Singh R.P., Singh P.K., Rutkoski J., Hodson D. P., He X., Jørgensen L.N., Hovmøller M.S., Huerta-Espino J. (2016). Disease Impact on Wheat Yield Potential and Prospects of Genetic Control. Annu Rev Phytopathol. 54: 303-322.

Smith E.L., Schlehuber A.M., Young H.C. Jr., Edwards L.H. (2005). Registration of agent wheat. In: Crop Science 8:511-512.

Stakman E.C., Stewart P.M., Loegering W. (1962). Identification of physiologic races of Puccinia graminis var. tritici. US Department of Agriculture, Agricultural Research Service. pp. Publ. E617.

Turner M.K., Jin Y., Rouse M.M., Anderson J.A. (2016). Stem Rust Resistance in 'Jagger' Winter Wheat. Crop Science 56: 1-7.

Turner M.K., Kolmer, J.A., Pumphrey, M.O. (2017). Association mapping of leaf rust resistance loci in a spring wheat core collection. Theor Appl Genet. 130(2): 345-361.

Williamson V.M, Thomas V., Ferris H., Dubcovsky J. (2013). An *Aegilops ventricosa* Translocation Confers Resistance Against Root-knot Nematodes to Common Wheat. Crop Science 53(4): 1412-1418.

Yu L.X., Chao S., Singh R.P., Sorrells M.E. (2017). Identification and validation of single nucleotide polymorphic markers linked to Ug99 stem rust resistance in spring wheat. PLoS One 12 (2): e0171963.

Zhang D., Bowden R.L., Yu J., Carver B. F., Bai G. (2014). Association Analysis of Stem Rust Resistance in U.S. Winter Wheat. PloS One 9(7): e103747.

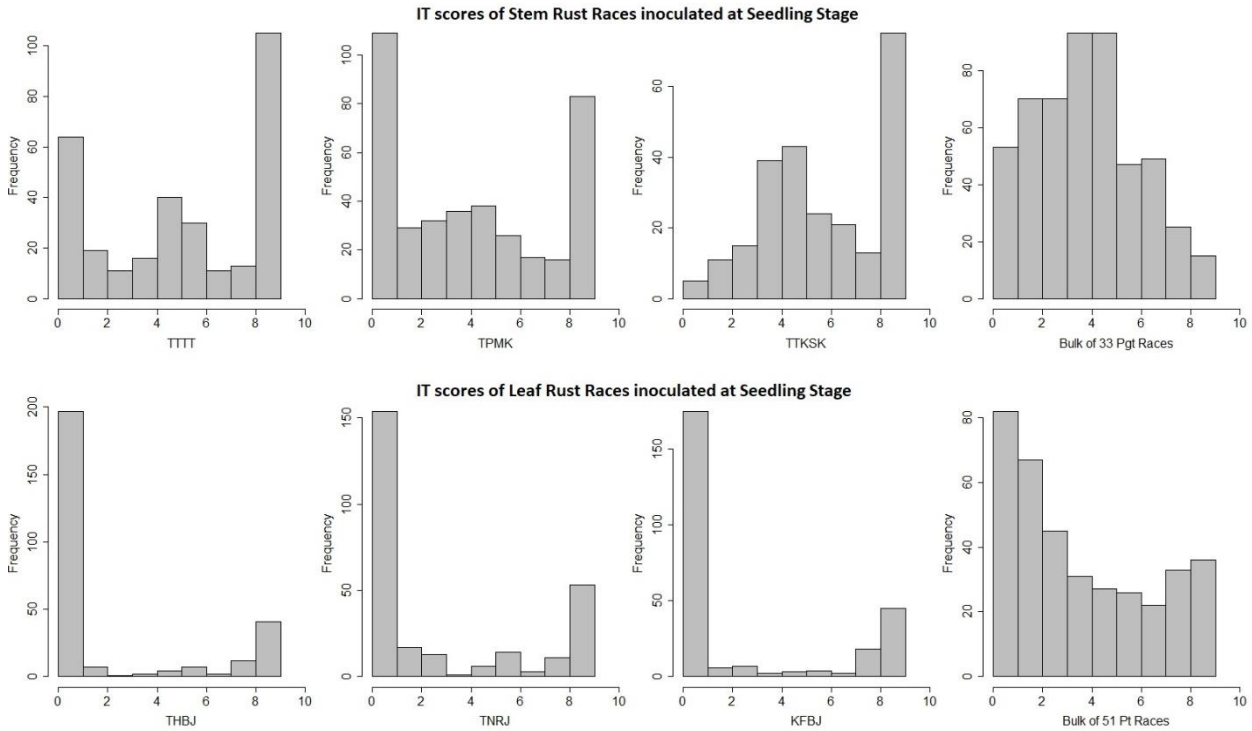

Figure 4-1. Normal distribution infection types of the three most commonly inoculated races of stem rust and leaf rust, and the average of all races inoculated in the SRPN from 2000 to 2015.

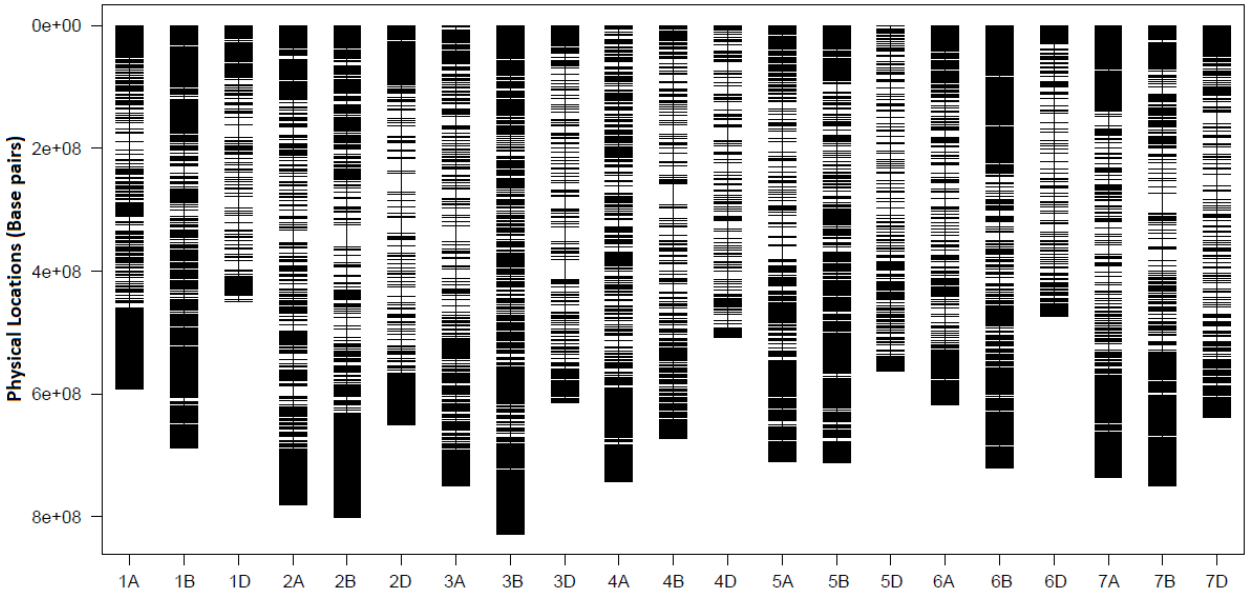

Figure 4-2. Physical map positions of 35,467 SNP markers identified in panel of 687 wheat lines tested in the SRPN from 2000 to 2015. Map was drawn using the '*plotMap*' function from the R/qtl package.

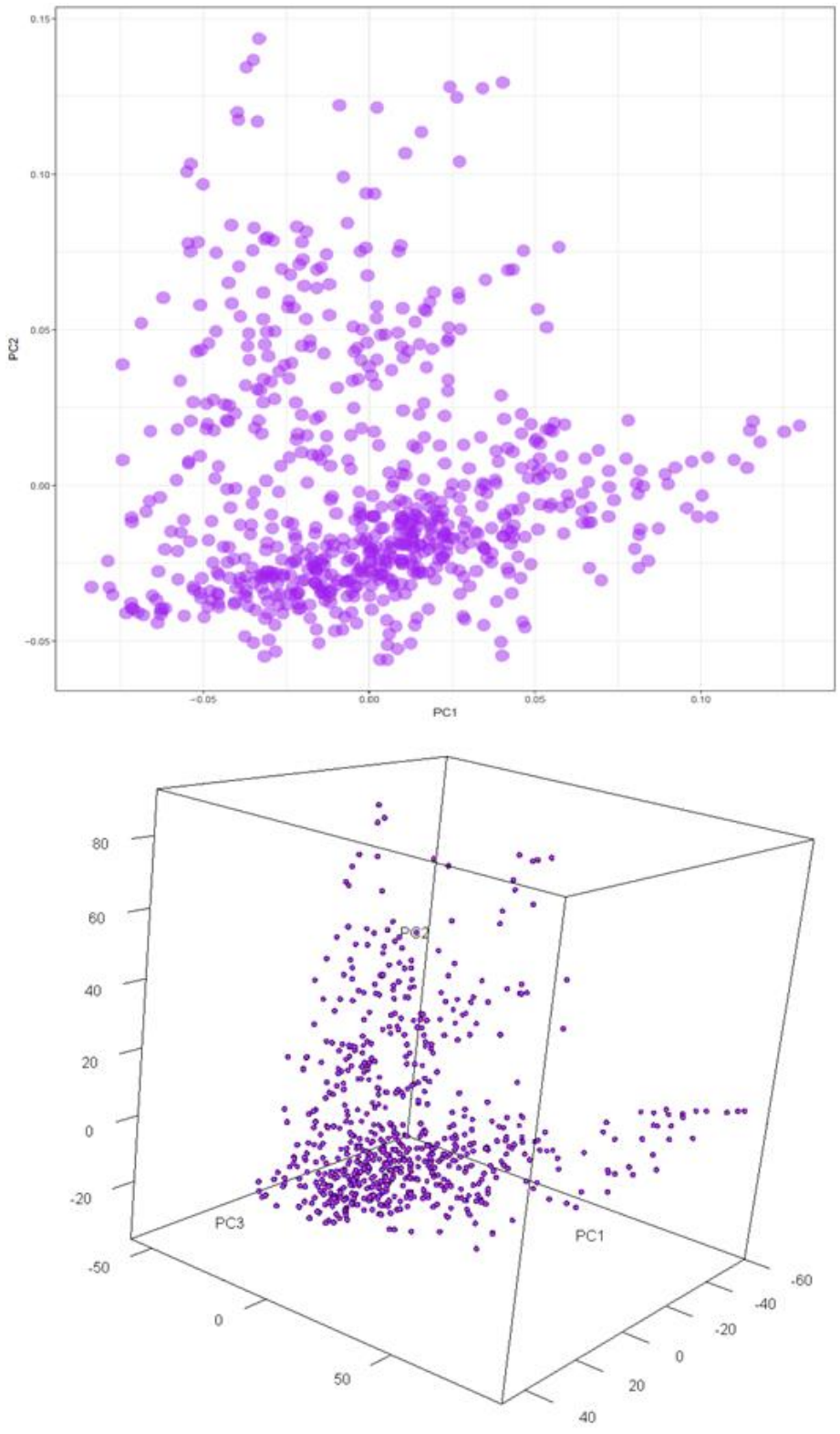

Figure 4-3. Principal components analysis of 687 wheat lines constructed with 35,467 SNP markers.

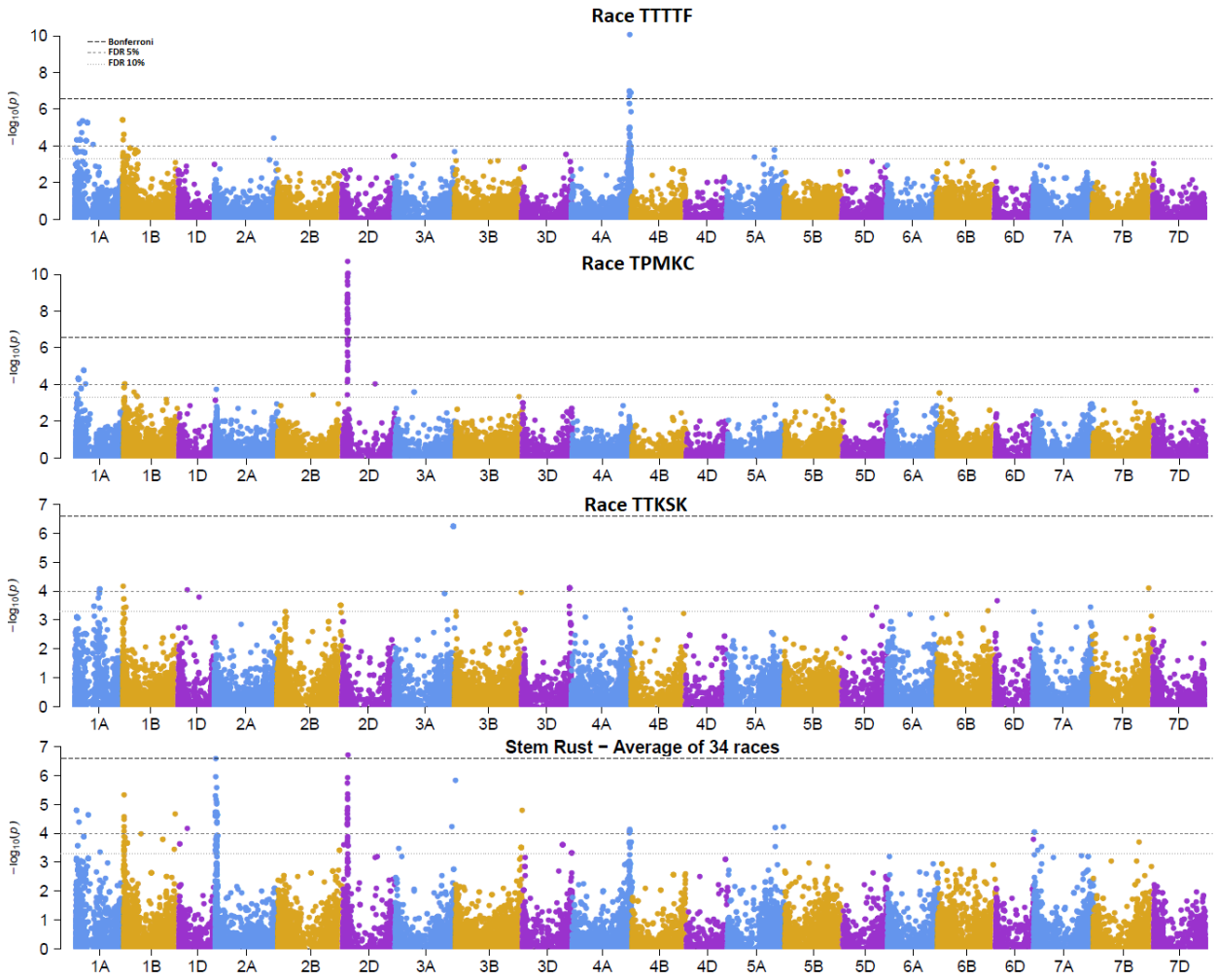

Figure 4-4. Manhattan plots showing association results for infection types of stem rust races based on 35,467 common SNPs. The x-axis represents physical positions of the SNPs in the wheat genome and the y-axis represents the $-\log_{10}$ of p-values.

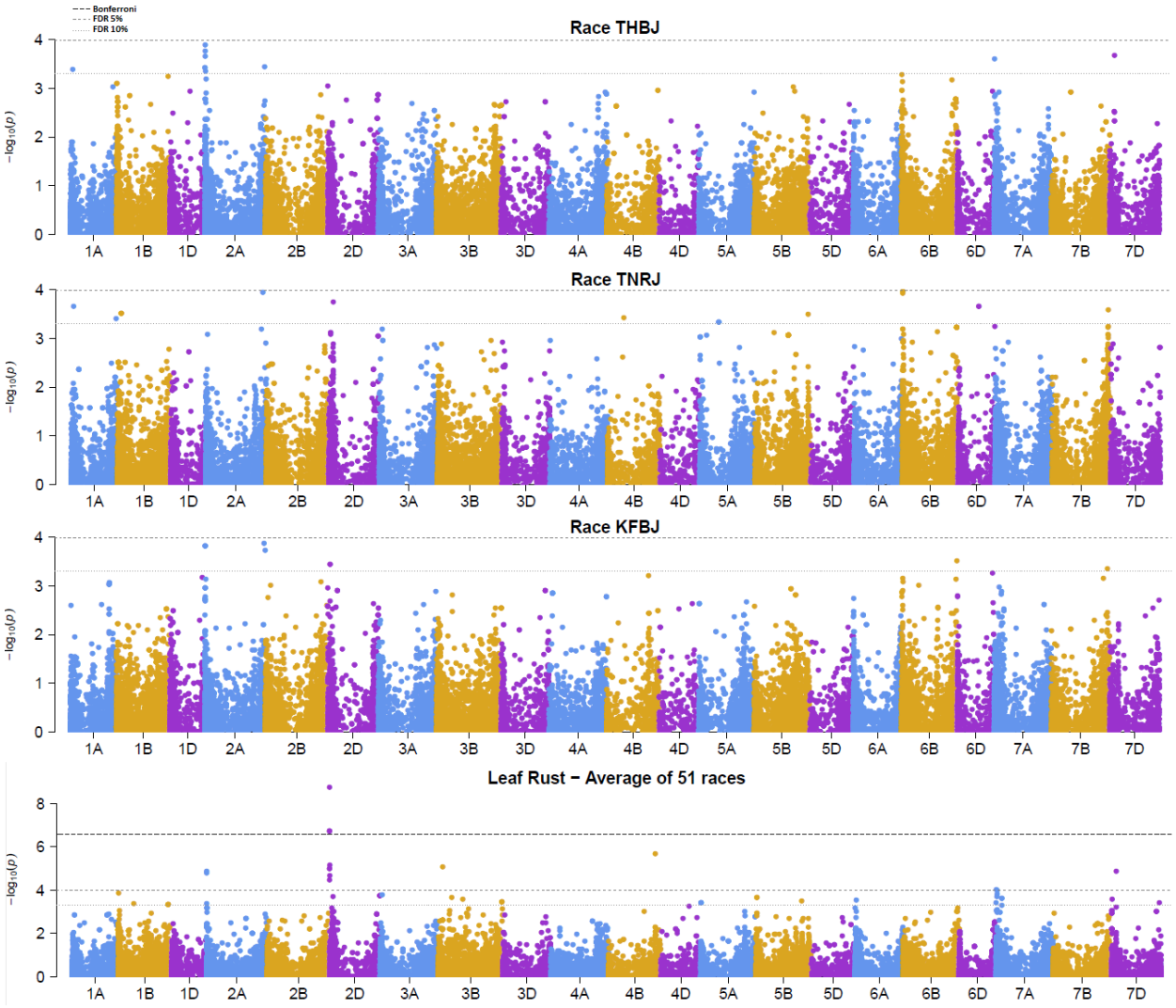

Figure 4-5. Manhattan plots showing association results for infection types of leaf rust races based on 35,467 common SNPs. The x-axis represents physical positions of the SNPs in the wheat genome and the y-axis represents the $-\log_{10}$ of p-values.

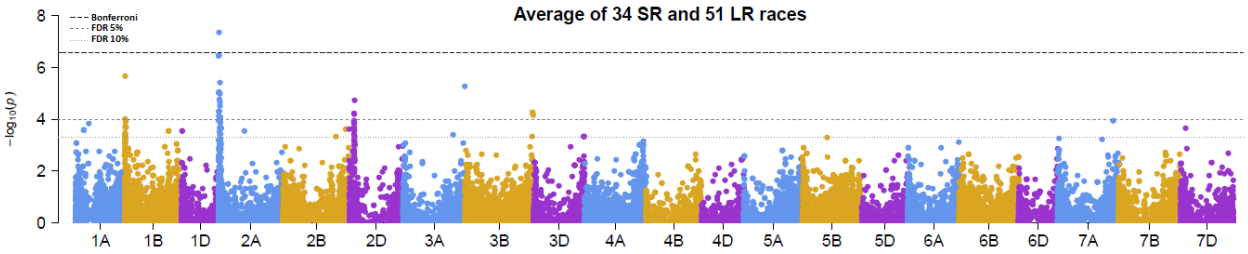

Figure 4-6. Manhattan plots showing association results for the average of infection types of 34 races of stem rust and 51 races of leaf rust based on 35,467 common SNPs. The x-axis represents physical positions of the SNPs in the wheat genome and the y-axis represents the $-\log_{10}$ of p-values.

Table 4-2. Summary of entries tested in the Southern Regional Preliminary Nursery (SRPN) from 2000 to 2015 and inoculate for wheat rust at seedling stage in greenhouse.

Year	Number of Entries	Races of Wheat Rusts	
		Leaf Rust (LR)	Stem Rust (ST)
2000	45	(8) KBBM, MLPR, MFDM, TFPM, TDDM, TBQM, TFGM, KCBM	(7) QTHJ, PTHS, TPMKC, TTRS, RKRQ, RTQQ, RCRS
2001	45	(6) MLRT, MFBP, TKBP, TDGT, MFBP, KBQT	(5) RTQQ, QTHJ, TTRS, RTHJ, TPMKC
2002	46	-	(6) TPMKC, RTQQ, RRTS, QTHJ, PTHS, TTTTF
2003	46	(4) THBJ, MCDS, TNRJ, KFBJ	(4) TTTTF, TPMKC, RTQQ, QTHJ.
2004	50	-	(5) TPMKC, QTHJ, TTTTF, RCRS, QFCS
2005	48	(7) KDBG, MCDS, TCTD, MFBJ, THBJ, MJBJ, TNRJ	(5) TPMKC, QFCS, TTTTF, RCRS, RKQQ
2006	50	(7) MCDS, KFBJ, THBJ, TNRJ, KDBG, TLGF, MJBJ	(7) QFCS, MCCF, RKQQ, TPMKC, QTHJ, TTTTF, TTKS
2007	50	(9) MCRK, THBJ, MJBJ, TGBG, MHDS, KFBJ, TNRJ, MFPSC, MLDSB	(7) QFCS, QTHJ, RCRS, RKQQ, TPMKC, TTTTF, TTKS
2008	50	(8) MLDS, THBJ, MJBJ, MFPS, TDBJ, TDBG, MHDS, KFBJ	(11) QFCS, QTHJ, RCRS, RKQQ, TPMKC, TTTTF, TTKSK, TTKS, TTKST, TTTSK, TRTT
2009	46	(8) MFPS, MHDS, TNRJ, MLDS, THBJ, KFBJ, TDBG, TMGJ	(13) QFCS, QTHJ, MCCF, RCRS, RKQQ, TPMKC, TTTTF, TTKSK, QCCSM, TTKST, TTTSK, TRTT, RFCS
2010	48	(8) TMGJ, MFPS, TNRJ, TDBG, HDS, KFBJ, MLDS, THBJ	(13) QFCS, QTHJ, MCCF, RCRS, RKQQ, TPMKC, TTTTF, RFCS, TTKSK, TTKST, TTTSK, TRTT, SCCSC
2011	38	(8) TMGJ, TDBG, MFPS, MHDS, MLDS, TNRJ, TFBJ, KFBJ	(12) QFCSC, QTHJC, MCCFC, RCRSC, RKQQC, TPMKC, TTTTF, TTKSK, TTKST, TTTSK, TRTTF, SCCSC
2012	44	(10) TDBGG, TBBGJ, TCRKG, MBDSD, TNRJJ, TFBJQ, MHDSB, KFBJG, TGBGG, MLDSD	(13) QFCSC, QTHJC, MCCFC, RCRSC, RKQQC, TPMKC, TTTTF, TRTTF, RRTTF, TTKSK, TTKST, SCCSC, TTTSK
2013	43	(9) TDBGG, TBBGJ, MBDSD, TFBJQ, MHDSB, KFBJG, MLDSD, TCRKG, TNRJJ	(12) QFCSC, QTHJC, MCCFC, RCRSC, RKQQC, TPMKC, TTTTF, TTKSK, TTKST, TTTSK, TRTTF, SCCSC
2014	40	(9) TNBGJ, MCTNB, MFPSB, KFBJG, MBDSD, TFBJQ, MHDSB, TCRKG, PBLRG	(15) QFCSC, QTHJC, MCCFC, RCRSC, RKQQC, TPMKC, TTTTF, SCCSC, TTKSK, TTKST, TTTSK, GFMNC, TRTTF, RRTTF, TKTTF
2015	42	(11) TNBGJ, TNRJ, MCTNB, TBBGJ, KFBJG, MBDSD, TFBJQ, MJBJG, TCRKG, PLBRG, TBBGS	(15) QFCSC, QTHJC, MCCFC, RCRSC, RKQQC, TPMKC, TTTTF, TKTTF, TRTTF, TKKTP, TTKSK, QCCSM, TTKST, TTTSK, TTKTT
Total	731*	51 races	34 races

*Out of 731 breeding lines 687 were unique and 533 had disease data available.

Table 4-3. Summary of the physical map of 35,467 SNP markers identified in panel of 687 wheat lines tested in the SRPN from 200 to 2015. Distances are measured in base pairs (bp).

Wheat chromosomes	Number of Markers	Length (bp)	Ave. Spacing (bp)	Max. spacing (bp)
1A	1,956	592,038,631	302,833.06	13,618,692
1B	2,429	687,265,770	283,058.39	11,056,157
1D	1,028	449,197,921	437,388.43	20,902,583
2A	2,050	780,357,174	380,847.82	19,535,754
2B	2,621	800,782,797	305,642.29	34,942,402
2D	1,650	651,112,066	394,852.68	31,491,912
3A	1,841	750,100,719	407,663.43	25,039,334
3B	2,785	828,074,948	297,440.71	7,576,284
3D	951	614,934,427	647,299.40	32,491,856
4A	1,682	743,133,265	442,078.09	12,665,164
4B	833	672,808,430	808,663.98	33,157,972
4D	299	507,803,682	1,704,039.20	21,820,594
5A	1,591	709,647,995	446,319.49	21,216,708
5B	2,150	712,221,813	331,420.11	13,259,725
5D	652	562,634,450	864,261.83	12,914,193
6A	1,710	617,367,234	361,244.72	14,841,500
6B	2,546	720,118,785	282,954.34	12,898,215
6D	966	473,049,027	490,206.25	15,680,336
7A	2,483	735,492,876	296,330.73	16,524,583
7B	2,233	750,094,085	336,063.66	32,311,225
7D	1,011	637,809,028	631,494.09	19,415,041
Overall	35,467	13,996,045,123	394,855.41	34,942,402

Table 4-4. Results from the avirulence/virulence formula for gene postulation based on the inoculation of differential sets as described by Roelfs & Martens (1988) and Long & Kolmer (1989) for *Pgt* and *Pt* races, respectively.

Rust	Races	Avirulent On Genes	Virulent On Genes
Leaf (*Pt*)	THBJ	*Lr9, Lr3ka, Lr11, Lr17, Lr18, Lr24, Lr30, LrB*	*Lr1, Lr2a, Lr2c, Lr3,Lr10, Lr14a, Lr16, Lr26*
	TNRJ	*Lr16, Lr17, Lr18, Lr26, LrB*	*Lr1, Lr2a, Lr2c, Lr3, Lr3ka, Lr9, Lr11, Lr10, Lr14a, Lr24, Lr30*
	KFBJ	*Lr1, Lr3ka, Lr9, Lr11, Lr16, Lr17, Lr18, Lr30, LrB*	*Lr2a, Lr2c, Lr3, Lr10, Lr14a, Lr24, Lr26*
Stem (*Pgt*)	TTTTF	*Sr24, Sr31, 1A·1R*	*Sr5, Sr6, Sr7b, sr8a, Sr9a, Sr9b, Sr9d, Sr2, Sr9g, Sr10, Sr11, Sr17, Sr21, Sr30, Sr36, Sr38, SrMcN*
	TPMKC	*Sr6, Sr9a, Sr9b, Sr24, Sr30, Sr31, Sr38, 1A·1R*	*Sr5, Sr7b, Sr8a, Sr9a, Sr9d, Sr9e, Sr9g, Sr10, Sr11, Sr17, Sr21, Sr36, SrTmp, SrMcN*
	TTKSK	*Sr24, Sr36, SrTmp, 1A·1R*	*Sr5, Sr6, Sr7b, Sr8a, Sr9b, Sr9d, Sr9e, Sr9g, Sr10, Sr11, Sr17, Sr21, Sr30, Sr31, Sr38, SrMcN*

Table 4-5. Number Significant SNP- trait associations and quantitative trait loci (QTL) identified in the GWAS analysis according three different multiple correction tests.

Rusts	Races	Bonferroni correction<5%		False Discovery Rate (FDR) <5%		False Discovery Rate (FDR) <10%	
		# SNPs	# QTL	# SNPs	# QTL	# SNPs	# QTL
Stem Rust (*Pgt*)	TTTTF	7	1 (4AL)	48	4 (1AS, 1BS, 2AL, 4AL)	70	7 (1AS, 1BS, 2AL, 2DL, 3AL, 3DL, 4AL, 5AL)
	TPMK	39	1 (2DS)	28	2 (1AS, 2DS)	51	9 (1AS, 1BS, 2AS, 2BL, 2DL, 3AS, 5BL, 6BS, 7DL)
	TTKSK	-	-	22	6 (1AL, 1BS,1DS, 3AL, 3DL, 7BL)	36	16 (1AL, 1BS, 1DS, 2BS, 2BL, 3AL 3BL, 3DL, 4AL, 4BL, 5DL, 6BL, 6DS, 7AS, 7AL, 7BL)
	Bulk of 33 *Pgt* races	3	2 (2AS, 2DS)	117	11 (1AS, 1BS, 1BL, 1DL, 2AS, 2DS, 3AL, 3BL, 4AL, 5AL, 7AS)	77	15 (1AS, 1BS, 1BL, 1DS, 2AS, 2BL, 2DS, 3AS, 3BL, 3DL, 4AL, 5AL, 6DL, 7AL, 7BL)
Leaf Rust (*Pt*)	THBJ	-	-	-	-	10	5 (1AS, 2AS, 2AL, 7AS, 7DS)
	TNRJ	-	-	-	-	18	11 (1AS, 1AL, 1BS, 2AL, 2DS, 4BS, 5AS, 5BL, 6BS, 6DL, 7BL)
	KFBJ	-	-	-	-	10	5 (2AS, 2AL, 2DS, 6BL, 7BL)
	Bulk of 51 *Pt* races	2	1 (2DS)	51	6 (2AS, 2DS, 3BS, 4BL, 7AS, 7DS)	43	15 (1BS, 1BL, 2AS, 2DS, 2DL, 3AS, 3BS, 3BL, 5AS, 5BS, 5BL, 6AS, 7AS, 7DS, 7DL)
SR+LR	Bulk of all races of *Pgt* and *Pt*	1	1 (2AS)	29	5 (1BS, 2AS, 2DS, 3AL, 3BL)	66	12 (1AS, 1BS, 1BL, 1DS, 2ASa, 2ASb, 2BL, 2DS, 3AL, 3BL, 7AL, 7DS)

FDR: False Discovery Rate.

Table 4-6. Details of the five resistant and ten susceptible entries submitted to SRPN from 2000 to 2015 and inoculated for stem and leaf rust. The five selected SNPs are the ones located at peak loci associated with resistance to multiple races of *Pgt* and *Prt*.

SRPN ID	Entry	Pedigree	SNPs found at peak loci					SRPN Marker Data				
			A/G	A/T	C/G	T/C	G/C	2NS-2AS	Sr24/Lr24 (3DL)	1RS trasnloc.	Postulated gene(s) based on IT reactions	Average of all IT scores of *Pgt* and *Pt*
			S2A_PART 1_2800562	S1B_PART1_ 5234750	S3A_PART2_ 292817024	S2D_PART1_6 1759932	S3B_PART2_ 361378376	ventriup-ln2	Sr24#12	TSM0120		
2004SRPN022	KS950811-5-1	Ogallala/KS95WGRC33//Jagger	G	T	G	T	C	-	-	non-1RS	-	0.00
2005SRPN032	HV9W99-558	Freedom/Tomahawk//Jagger	G	-	-	T	-	2NS+	-	non-1RS	-	0.00
2005SRPN045	AP02T4342	Coronado//1174-27-46/X960210	G	A	-	T	G	2NS+	-	non-1RS	-	0.00
2009SRPN032	TX05A001822	2145/X940786-6-7	G	T	G	T	C	2NS+	Sr24+	non-1RS	Lr24/Sr24 Lr41	0.82
2008SRPN021	HV9W02-942R	53/3/ABL/1113//K92/4/JAG/5/ KS89180B	G	T	G	T	C	2NS+	Sr24+	1BL.1RS	Lr34	0.94
2009SRPN013	KS0603A-58-1	Overley*3/Amadina	G	-	C	T	G	2NS+	non-Sr24	1BL.1RS	-	0.05
2009SRPN044	HV9W04-1594R	KS89180B-2-1-1/ CMBW91M02959T//JGR	G	-	-	T	G	2NS+	non-Sr24	1BL.1RS	Lr41	0.13
2011SRPN013	KS020638-5-1	KS940786-17-2/Jagalene//Trego	G	T	G	T	C	2NS+	Sr24+	non-1RS	Lr24/Sr24	0.69
2012SRPN020	HV9W07-1942	JAGALENE//W99-331/ X940786-6-4	G	-	-	T	-	2NS+	non-Sr24	1BL.1RS	-	0.15
2012SRPN031	NE09517	W96x1080-21 =(Jagger/Thunderbolt)/Jagalene	G	T	G	T	C	2NS+	Sr24+	non-1RS	Lr24/Sr24	1.39
2014SRPN022	LCH10-187	B88/2180//T81-1	A	A	C	T	G	non-2NS	Sr24+	1RS:1AL	-	7.02
2011SRPN011	OK06336	Magvars/2174//Enhancer F4:12	A	A	C	T	G	non-2NS	Sr24+	non-1RS	-	7.38
2007SRPN027	KS970093-8-9-#1	HBK1064-3/KS84063-9-39-3-4W//X960103	A	A	C	T	G	non-2NS	non-Sr24	non-1RS	Lr1, Lr14a	7.64
2005SRPN026	CO00796	Transvaal/Arlin/2/CO910424/Halt	A	A	C	T	G	non-2NS	non-Sr24	1AL.1RS	14a	7.69
2007SRPN031	CO02W280	98HW521(93HW91/93HW255)/ 98HW165(ARL/WGRC15)	A	A	C	T	G	non-2NS	non-Sr24	non-1RS		7.71
2007SRPN033	CO03W239	KS01-5539/CO99W165	A	A	C	T	G	2NS+	non-Sr24	non-1RS	Lr14a	8.16
2007SRPN034	CO03W269	KS01-5539/CO99W191	A	A	C	T	G	2NS+	non-Sr24	non-1RS	-	8.00
2006SRPN013	T150	T81/T201	A	A	C	T	G	non-2NS	non-Sr24	non-1RS	-	8.21
2015SRPN010	CO11D446	CO050270/Byrd	A	A	C	T	G	non-2NS	non-Sr24	non-1RS	-	8.36
2014SRPN010	CO11D174	TAM 112/Byrd	A	A	C	T	G	non-2NS	non-Sr24	non-1RS	-	8.83

Chapter 5 - Final Remarks and Future Prospects

5.1 Overall Conclusions

The biparental mapping revealed eight small-effect QTLs located on 1AS, 1BS, 3BS, 5AS, 5BL, 5DS, 6BS and 7AL, which were associated with multiple mechanisms of FHB resistance. All these mapped QTLs presented additive effects. A significant reduction of 67% in DON content was observed in lines carrying all mapped QTLs for DON in comparison to lines without any of the mapped loci. We found that three DH lines (DH014, DH026, and DH130) carried seven of the eight alleles for resistance, whereas only one line (DH108) possessed all susceptible alleles for the mapped QTLs. Since all these loci explained a relatively small proportion of phenotypic variation, converting GBS sequences flanking QTLs into markers to assist wheat breeding via MAS is not viable. Instead, we recommend the implementation of genomic selection strategies, as shown in Chapter 3, which can lead to more significant progress of FHB resistance in wheat breeding programs.

The GWAS identified significant SNP associations with the percentage of symptomatic spikelets in all three breeding panels but the results were not reproducible across years. This likely indicates the frequency of resistant alleles associated FHB resistance is low, especially since selections are not exclusively based on FHB resistance in the K-State Wheat Breeding program. Allele frequencies change from one year to the other based on parents used in a given crossing cycles and this changing set of germplasm likely contributed to the inability to consistently identify regions associated with resistance. Although significant loci are reported in similar genomic regions as other studies, based on the pedigree information, they are unlikely the same as the ones described in the literature. Accuracy predictions of genomic selection (GS)

models were relatively high (>0.45) when only 20% of the data was masked in the training set. However, GS models did not statistically differ from each other, indicating that, in this case, improvements in FHB resistance can be made regardless of the model adopted. Upcoming studies must investigate alternatives to increase prediction accuracies. Examples include the development of tools to obtain more precise PSS ratings, testing of genetic algorithms to account for the genetic relatedness between training and target sets, and the inclusion of covariates such as heading date, plant height and anther extrusion in the GS models. Furthermore, GWAS analysis could be performed in training populations to identify SNPs associated with the trait of interest, such as FHB. These SNPs could then be further considered as fixed effect markers in GS models to increase the accuracy of GS predictions.

In chapter 4, the association mapping with historical data from the SRPN revealed multiple loci conferring stem and leaf rust resistance at seedling stages. A highly significant locus was repeatedly detected on 2AS, which was associated with resistance to multiple races and the overall bulk average of *Pgt* and *Pt*. We confirmed that this loci, in fact, corresponds to the 2NS·2AS translocation and nearly 40% of the lines submitted to the SRPN from 2000-2015 carries this alien segment. Another two highly significant loci were found on 2DS and 4AL. There are multiple candidate genes on these genomic regions, therefore further research will be needed to confirm which genes are putative candidates. Evaluation of wheat rusts in natural conditions is often confounded with the predominant race in a given environment. Here, we have shown the viability of using data from individual race inoculations and their combinations to map loci associated with resistance to rusts at seedling stage. Future studies may use GWAS to identify the most significant loci associated with traits of interest in breeding panels that could be used as training sets in GS schemes that consider significant QTLs as fixed effect factors, aiming

to increase prediction accuracies. Finally, GS models must be considered as a tool to assist breeding, not necessary a replacement for phenotypic selections. Genomic tools discussed in in this dissertation may be integrated into wheat breeding programs aiming to increase genetic gains for disease resistance breeding.

5.2 The Future of Disease Resistance Breeding

The cost of genotyping has significantly decreased in the past few years and it is no longer a limiting factor for wheat genetic research. Hereafter, the availability of a complete annotation reference (IWGSC, 2017) and the pangenome of hexaploid wheat (Montenegro *et al*., 2017) will enable more extensive studies of genetic mechanisms underlying traits of interest. For instance, the tools now exist to investigate how selection pressure operates across homoeologous genes at different ploidy levels, facilitate gene cloning (especially the ones involved in host resistance) and understand gene expression (Uauy, 2017). Multiplex targeted sequencing (Rife *et al*., 2015) research must be extended to other important genes as a fast, low-cost strategy allowing the assay of the whole-genome profile of elite lines while following specific genes. Moreover, other genotyping approaches, such as exome capture and RNA-sequencing can be used for high-resolution mapping in early generations (Liu *et al*., 2012; Ramirez-Gonzalez *et al*., 2015) whereas reverse genetics resources such as TILLING populations may be used to identify gene functions (Krasileva *et al*., 2017). Furthermore, recent advances in genome-editing, such as CRISPR/Cas9 (Cong *et al*., 2013), have been speculated as a promising tool not only for switching on or turning off genes, but also for guiding recombination events which could potentially double the gain from selection in the near future (Bernardo, 2017).

Regarding field data collection, the key challenge for genetic studies in the coming years will be the development of precise phenotyping tools. Exclusive reliance on visual notes that are evaluator-dependent and often inaccurate is a large limitation. Progress needs to be made towards the development of high-throughput image-sensing platforms able to measure disease-infected areas on leaves and spikes at plant and plot levels. It would allow us to distinguish minor phenotypic variations that even a trained evaluator cannot differentiate and collect more data-points per unit of area in a reasonable time. Fully automated pipelines will be required for rapid, real-time data analysis to assist breeder's decision-making prior to phenotypic selection and/or harvest. Likewise, more effective mapping results, genome-wide associations, and genomic predictions will be achieved with high quality of phenotypic data available.

Training the next-generation of "genomics-enabled researchers/breeders" is also crucial. It will ensure that advances in wheat genomics will be translated into higher genetic gains and delivered to farmers' fields as higher-yielding and disease-resistant varieties (Uauy, 2017). Having these multidisciplinary skills also will be fundamental for the new scientists, including automated data collection, genotyping, programming, and big data analysis. Undoubtedly, professionals with these skills could be the bridge between classical breeders and bioinformaticians, helping to develop/integrate complex genomic approaches to surpass the current levels of disease resistance, end-use quality, and grain yield in wheat breeding.

References

Bernardo R. (2017). Prospective Targeted Recombination and Genetic Gains for Quantitative Traits in Maize. The Plant Genome 10(2):1-9.

Cong L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W., Marraffini L.A. *et al*. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339: 819-23.

IWGSC RefSeq v1.0. (2017). International Wheat Genome Sequencing Consortium. Available online at: https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies.

Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R. H., Wang, X., Borrill, P. *et al*. (2017). Uncovering hidden variation in polyploid wheat. Proc. Natl. Acad. Sci. 114: E913-E921.

Liu S., Yeh C.T., Tang H.M., Nettleton D., Schnable P.S. (2012). Gene Mapping via Bulked Segregant RNA-Seq (BSR-Seq). PLoS ONE 7(5): e36406.

Montenegro J.D., Golicz A.A., Bayer P.E., Hurgobin B., Lee H., Chan C.K., Visendi P., Lai K., Doležel J., Batley J., Edwards D. (2017). The pangenome of hexaploid bread wheat. The Plant Journal 90(5): 1007-1013.

Ramirez-Gonzalez R.H., Segovia V., Bird N., Fenwick P., Holdgate S., Berry S., Jack P., Caccamo M., Uauy C. (2015). RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant Biotechnology Journal 13(5): 613-624.

Rife T.W., Wu S., Bowden R.L.; Poland J.A. (2015). Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. BMC Genomics 16(1):248.

Uauy C. (2017). Wheat genomics comes of age. Current Opinion in Plant Biology 36: 142-148.